Phonological variation in German Learner English

Lukas Sönning

Lukas Sönning

Phonological variation in German Learner English

Sönning, Lukas. 2020. *Phonological variation in German Learner English*. University of Bamberg dissertation.

Bibliographische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliographie; detaillierte bibliographische Informationen sind im Internet über http://dnb.d-nb.de/ abrufbar.

Diese Arbeit hat der Fakultät Geistes- und Kulturwissenschaften der Otto-Friedrich-Universität Bamberg als Dissertation vorgelegen.

Gutachter: Prof. Dr. Manfred Krug
Gutachterin: Prof. Dr. Ulrike Gut

Tag der mündlichen Prüfung: 1. Dezember 2017

This title can be downloaded at:
https://fis.uni-bamberg.de/handle/uniba/49135

Typesetting: Lukas Sönning
Typesetting software: This book was typeset with LuaLaTeX using *Texmaker* (Brachet 2018) and the *R* package *bookdown* (Xie 2016); the design is based on the Tufte-LaTeX document classes (http://code.google.com/p/tufte-latex/).

URN: urn:nbn:de:bvb:473-irb-491352
DOI: http://dx.doi.org/10.20378/irb-49135

Open Science Framework (OSF) repository: https://osf.io/2pdh6/
Source code available from https://osf.io/gnzu4/

Acknowledgements

This book grew out of a dissertation project at the University of Bamberg. I would like to thank my supervisor, Manfred Krug, for his guidance and patience, and for giving me the freedom to develop my ideas. I am also grateful for being part of a team of colleagues who turned my work at the department of English linguistics into a delightful and rich experience.

My gratitude also extends to Ulrike Gut, who co-supervised the present thesis and offered helpful and constructive feedback throughout. Special thanks are due to Robert Fuchs for reading an earlier version of this book, and for many thoughtful comments and suggestions.

I would also like to thank my colleagues, past and present, for the warm and productive atmosphere and the extraordinary team spirit: Katharina Beuter, Urška Grum, and Michaela Hilbert for generous support at the coal-face of teaching; Fabian Vetter for the computational infrastructure; Julia Schlüter, Ole Schützler, and Valentin Werner for setting academic examples and providing challenging and stimulating input.

I am grateful to Graham Pascoe, Isabell Rank, and Christina Wunder for allowing me to (re)analyze and share their data. Thanks also to Gunter Narr Verlag Tübingen, for permission to adapt and use images which first appeared in Wieden & Nemser (1991).

The present work relies heavily on open source software tools, most notably *R* (R Core Team 2020) and *RStudio* (RStudio Team 2020). To draw the many graphs that appear in this book, I relied almost exclusively on the *lattice* package developed by Deepayan Sarkar (2008). Paul Bürkner's (2017) fabulous *brms* package eased the transition to the *Stan* programming language (Stan Development Team 2018). The *R* packages *bookdown* (Xie 2016) and *kableExtra* (Zhu 2020) provided invaluable assistance with typesetting the book in Tufte-LaTeX, a package to imitate the layout of Edward Tufte's books (created by Kevin Godby, Bill Kleb, and Bill Wood).

Finally, I would like to thank my family and friends for their support; above all my wife Katharina, for embracing the ups and downs of this project. For full disclosure, I should mention that we met in the *Introduction to English Linguistics*, which was taught by Manfred Krug, who then hired both of us as student assistants and set the ball rolling.

List of abbreviations

AmE	American English
AOL	Age at onset of learning
AS	Articulatory Settings (Honikman 1964)
Bk	Bark
BrE	British English
CAH	Contrastive Analysis Hypothesis (Lado 1957)
D	Euclidean distance
D_{Bk}	Euclidean distance between vowels in Bark units
D_{zBk}	Euclidean distance between vowels in Lobanov-normalized Bark units (z-scores)
Δ	Absolute difference on the proportion scale
Δ_{Bk}	Difference between formants in Bark units
Δ_{Hz}	Difference between formants in Hertz units
DH	Desensitization Hypothesis (Bohn 1995)
DST	Dynamic Systems Theory (De Bot et al. 2007)
F_1	First formant
F_2	Second formant
FAR	Foreign accent rating
FCM	Feature Competition Model (Hancin-Bhatt 1994)
FM	Functional Model of Phonological Acquisition (Boersma 1998)
FVO	Final voiced obstruent
GDM	Gradual Diffusion Model (Gatbonton 1978)
GLE	German Learner English
Hz	Hertz
IL	Interlanguage
IPA	International Phonetic Alphabet
L1	First language
L2	Second language
LTD	Linguistic Theory of L2 Phonological Development (James 1988)
M_{Age}	Average age
MDH	Markedness Differential Hypothesis (Eckman 1977)
Mdn	Median
MSA	Model of Segmental Acquisition (Colantoni & Steele 2008)
MSV	Model of Sociolinguistic Variation (Fasold & Preston 2007)
n	Number of units/sample size
NDH	Naturalness Differential Hypothesis (Schmid 1997)
NM	Natural Model of L2 Phonological Acquisition (Dziubalska-Kołaczyk 1990b)
OPM	Ontogeny Phylogeny Model of Second Language Acquisition (Major 2001)
OT	Optimality Theory (Prince & Smolensky 1993)
PAM	Perceptual Assimilation Model (Best 1995)
PIM	Phonological Interference Model (Brown 1998)
SCH	Structural Conformity Hypothesis (Eckman 1991)
SHG	Standard High German
SD	Standard deviation
SDRH	Similarity Differential Rate Hypothesis (Major & Kim 1996)
SLA	Second Language Acquisition
SLM	Speech Learning Model (Flege 1995)
SpIL	Spoken Interlanguage corpus (Pascoe 1987)
TL	Target language
U	Universal
UG	Universal Grammar
zBk	Bark measurements converted to z-scores (using the method proposed by Lobanov 1971)

Contents

1
Introduction

The study of second language (L2) speech is concerned with the description and explanation of the sound systems acquired by non-native speakers. Over the past 60 years, the field has profited from insights gained across a range of disciplines including linguistics, the cognitive sciences and education. As a result, L2 speech is studied from a wide range of perspectives, which is reflected in the state of theoretical knowledge about the acquisition of an L2 sound system.

Current approaches to L2 phonology view learner speech as an autonomous system in the sense that its organization is to a certain extent independent of the target language (TL) and the learner's native language (L1). This assumption is captured by the term *interlanguage*[1] (IL), which refers to the L2 system internalized by a learner. It is the objective of the field of L2 phonological acquisition to advance our understanding of the factors that constrain and systematically shape IL variation.

[1] The term *interlanguage* was introduced independently by Corder (1967), Nemser (1971a), and Selinker (1972).

A broad distinction can be made between learner-related and linguistic constraints. The former, which are often summarized under the label *individual differences*, are operative through the personality and experience of the learner and include attributes such as language aptitude, age at onset of learning (AOL), and TL experience.[2] The second class of constraints turns to linguistic constructs to account for systematic patterns of variation in learner speech. These mainly revolve around two notions: (i) cross-linguistic influence, which describes the way in which learner speech is constrained by the native language, and (ii) language universals, where the focus is on influences that are universal in the sense that they operate independently of the L1 and TL.

[2] See Piske et al. (2001) for an overview.

This study concentrates on linguistic constraints and pursues two general aims. The first is to survey theoretical work on L2 phonology from an empirical perspective. Theoretical contributions are reviewed and applied to a diverse set of segmental structures in German Learner English (GLE). The second aim is to offer empirical insights into phonological variation in GLE in the domain of vowels, sonorants, and obstruents. Accordingly, this book is structured as follows.

Chapter 2 surveys existing theoretical work on L2 speech learning. In total, twenty contributions are evaluated with regard to their tenets, predictive scope, and the type of IL variation they account for. It will be argued that the latter aspect offers a new and fruitful perspective on the theoretical landscape and is of relevance for the interaction between theory and empirical research.

Chapter 3 gives a general outline of the methodology and data used. This includes a description of materials and procedures, data sources, and the strategies used for statistical inference.

In the chapters that follow, the predictive scope and adequacy of theoretical contributions is examined in the light of different segmental structures in German Learner English. Each chapter is structured in the same way. Following a contrastive analysis and an exploration of linguistic constraints, a set of hypotheses is constructed to map the expectations based on the accounts surveyed in Chapter 2. This is followed by a review of existing empirical work and a quantitative analysis of the focal structure.

Chapter 4 addresses German learners' acquisition of a new vowel contrast that has received considerable attention in the literature: the difference between front open TRAP (*had*, *bad*) and front mid DRESS (*head*, *bed*). TRAP is a new vowel for German learners. The relationship between length and quality features, which has been largely neglected in the literature on GLE, is assessed at various proficiency levels by means of instrumental techniques. This offers insights into the relative weights of these cues at different proficiency levels. The quantitative analysis is substantiated by an extensive survey of acoustic studies on monophthongs in British English, American English, and German.

Chapter 5 deals with the acquisition of English lateral allophones by German learners whose L1 lacks a velarized ("dark") variant. With instrumental studies on this feature of GLE being scarce, the present study aims to offer an acoustic account of laterals produced by German learners. To this end, the acoustic degree of /l/-velarization in prevocalic and non-prevocalic position is assessed across proficiency levels and compared to native speaker productions. Context for the interpretation of findings is provided by a comprehensive review of the instrumental literature on laterals in English and German.

The focus of Chapter 6 is on the acquisition of English /r/-allophones by German learners. Building on previous empirical and theoretical work, this study employs an auditory analysis to examine the realization of prevocalic and postvocalic /r/ across different levels of L2 pronunciation ability. Insights are provided into the sensitivity of /r/-production to various constraints, including complexity (singleton vs. cluster), linking context, structural strength, and speaking style.

Chapter 7 is concerned with the acquisition of the labio-velar glide /w/, a novel structure for German learners. Based on an auditory analysis, the present investigation complements earlier work by assessing the accuracy rate across proficiency levels and charts new territory by studying phonetic context effects. Motivated by theoretical considerations and anecdotal evidence in the literature, we examine /w/-production conditional on properties of the phonetic context.

Chapter 8 investigates the production of the voiced labiodental fricative /v/ in onset position. This study expands upon previous work and examines onset /v/-production with a focus on three variants that are reflective of different types of constraints on IL. An auditory analysis offers a systematic account of an overgeneralization error (/v/ produced as [w]) that has been referred to as *hypercorrection* and *over-correction* in the literature.

Chapter 9 turns to dental fricatives and presents an auditory analysis of German learners' production of onset /ð/. Apart from the distribution of variants across proficiency levels, attention is given to phonetic context effects, which have not been investigated in previous work on GLE.

Chapter 10 offers a description of final voiced obstruent (FVO) production by German learners. Unlike English, German regularly devoices obstruents in coda position strongly, thus neutralizing the voicing opposition in pairs like *Rat* vs. *Rad* [ʁaːt]. By means of an auditory analysis, this study expands upon earlier work by documenting the realization of FVOs across proficiency levels and its sensitivity to properties of the obstruent and the phonetic context. An acoustic analysis provides further insights into the degree to which variation in the preceding vowel duration is employed as an extrinsic cue to final obstruent voicing.

Chapter 11 closes with a discussion of some broader issues that emerged when communicating between L2 phonological theory and data. These may be of relevance to the field of L2 phonology more generally.

2 L2 phonological acquisition: Theoretical work

This chapter reviews theoretical work on L2 phonology. For expository clarity, contributions will be grouped by three major themes. Following some preliminary remarks, §2.2 discusses work that is primarily concerned with the influence of the learner's native language. Advances that put forward more general, language-universal constraints are covered in §2.3. Models with a focus on IL development and variation are dealt with in §2.4. §2.5 closes with a summary and comparative overview.

2.1 Preliminary remarks

L2 phonological theories mainly differ in two regards: (i) the constraints they stipulate to account for observable patterns and (ii) the type of variation they aim to predict or explain. From an empirical point of view, much is gained from differentiating two broad dimensions of IL variation: that of the learner and that of the structure. Put simply, there is variation among learners, as some are more proficient than others, and there is variation among structures, as some are more difficult than others. We may refer to these types of variability as *between*-learner and *between*-structure variation. Furthermore, a learner or structure may also show internal variation. For instance, a learner's accuracy rate may vary systematically with speaking style and a given structure may show variation that is sensitive to the phonetic context. Thus, there is also *within*-learner and *within*-structure variation.

The following overview will attempt to describe theoretical accounts in terms of their scope and empirical application. Since the focus of this study is on the role of linguistic constraints over the course of L2 phonological acquisition, the only type of between-learner variation that is of interest is L2 pronunciation ability. Consequently, the following treatment will distinguish between within-structure, between-structure, and developmental variation. This distinction is made to clarify the type of variation a particular theory aims to account for. As for terminology, the term *structure* will serve as a general label for any type of phonetic, phonological or prosodic unit. *Prosodic*, in turn, is understood in a broad sense – that is, as referring to structure(s) above the level of the segment. The labels *prosodic* and *suprasegmental* will be used interchangeably.

2.2 *Cross-linguistic influence*

The central theme of theoretical approaches discussed in this section is the influence of the learner's native language. §2.2.1 deals with Contrastive Analysis, which relies on the construct of transfer to account for systematic errors in L2 speech production. Subsequent sections discuss contributions that consider L1 transfer to operate primarily on the level of articulatory routines (§2.2.2) and speech perception (§2.2.3).

2.2.1 *Contrastive analysis*

Lado's (1957) Contrastive Analysis Hypothesis (CAH) marks the first formal account of cross-linguistic influence in L2 acquisition. The CAH relies on the assumption that language learning involves the formation of a set of habits that are transferred to the L2. Correct results of habit-shifts were termed *positive transfer*; errors were attributed to *negative transfer* (also called *interference*). Lado (1957) proposed that, based on a contrastive analysis, it should be possible to identify difficult structures, the fundamental claim being that similar structures should be simple and different structures difficult. The greatest difficulty was predicted to occur in the learning of allophonic splits, where L1 allophones realize different TL phonemes.[1] The association between difficulty and functional status is presumably rooted in the correspondence between form and meaning. Communicative relevance makes phonemic contrasts – and by implication distinctive features – more salient, which facilitates learning.

The formulation of the CAH triggered a wave of empirical work with a strong focus on phonology. In essence, studies noted that the role of the L1 was not as straightforward as posited by the CAH: Certain different sounds appeared to be acquired with ease and certain similar structures posed enormous difficulties. Empirical findings also pointed to learner errors that could not be explained by reference to L1-TL contrasts (e.g. Briere 1966; Tarone 1978; Wode 1978, 1977; Garnica & Herbert 1979). Nevertheless, the fundamental tenets of the CAH continue to form a basis for most research into L2 phonological acquisition today (Major 2008; Zampini 2008; Broselow & Kang 2013).

Contrastive analysis can be employed in two ways to predict the relative ease of structures. First, in the original version of CAH, differences and similarities correspond to difficulty and ease of acquisition, respectively. Second, the functional status of a particular structure or contrast in the L1 and TL can be taken into account to elaborate on the degree of difficulty involved. In particular, meaning-distinguishing features are predicted to be more noticeable, making phonemic contrasts easier to acquire than allophonic ones. This allows us to predict between-structure variation on the basis of well-defined criteria.

[1] More refined approaches to Contrastive Analysis formulated hierarchies of difficulty and elaborated on the effect of the functional status of sounds or contrasts in the L1 and TL – that is, whether they are phonemic, allophonic or absent. Stockwell & Bowen (1965), for instance, establish a hierarchy for L1 English-TL Spanish and ascribe the greatest difficulty to TL allophones and phonemes that are absent from the L1. This arrangement was partially supported by Hammerly (1982), a large-scale study on Spanish learners of English, where inaccuracies dominated at the allophonic level. Other studies have also reported asymmetries between allophonic and phonemic contrasts (Hardy 1993; Elliott 1995; Elliott 1997; Eckman et al. 2003).

2.2.2 *Articulatory settings*

One answer to the question of what exactly is transferred from the native to the target language is offered by Articulatory Settings (AS) theory

(Honikman 1964). AS theory holds that cross-linguistic influence surfaces in the transfer of L1 articulatory settings, which are speech-motor routines of the active articulators. Automatized language-specific settings not only allow for motor economy but also conspire to give each language a "specific timbre" (ibid.: 73). A distinction is made between external settings, which are visible in face-to-face communication (e.g. lip gestures and jaw movement), and internal settings (e.g. tongue gestures). Established routines are shaped by the most frequently occurring segments and their phonotactic distribution. English, for example, can be characterized by generally neutral lip rounding. German and French, on the other hand, exhibit more energetic lip rounding. Further, a small degree of jaw opening is characteristic of English, where differences in vowel quality are predominantly accomplished by tongue gestures.

L1-TL contrasts in articulatory settings necessitate the learning of novel routines. As pointed out by Colantoni et al. (2015: 58), most studies invoking AS theory have relied on acoustic measurements and thus only provide indirect evidence for cross-linguistic influence in gestural routines (see Wilson & Gick 2014 for an exception). The main caveat of AS theory is the fact that the investigation of gestural parameters requires specialized instruments. Nevertheless, reference to gestural routines may help to explain certain speech motor biases in non-native productions. For instance, L1-TL differences in lip rounding and jaw opening may explain the prevalence of certain substitutes for TL structures. The strength of AS theory lies in its explanatory adequacy. It considers motor behavior to be the result of automatization, a domain-general learning mechanism that is well-attested in L1 and L2 acquisition.

2.2.3 Speech perception

In the 1980s, a prominent research program emerged, which ascribes inaccuracies in speech production to perceptual distortions. In this line of research, the perception of L2 speech sounds is assumed to be constrained by the sound categories established in L1 acquisition. In adults, L1-adapted perception is largely automated and robust (Strange & Shafer 2008). As a result, reliable L1 cues are foregrounded and sensitivity to non-native contrasts declines (e.g. Werker & Tees 1999; Kuhl et al. 1992; Polka & Werker 1994). Desensitized cues may therefore pose perceptual difficulties for learners.[2]

A re-evaluation of the role of transfer in the light of similarity vs. dissimilarity sparked a wave of research that culminated in several perceptual models of L2 acquisition. Two aspects that received attention in research on L2 speech perception were the degree of difference between L1-TL phonological entities and the learner's perception of such differences.[3]

Speech Learning Model (SLM)

One of the most influential contributions is Flege's (1995) Speech Learning Model (SLM). L2 learners are assumed to have access to the same learning mechanisms that operate in L1 acquisition, but their perceptual sensitivity

[2] This was stated early by Trubetzkoy (1939), who noted that TL sounds have to pass through the "sieve" of the L1 phonological system.

[3] Speech perception research also focused on learner-specific factors that could account for between-learner variation in perception accuracy. Thus, perceptual sensitivity has been observed to vary as a function of several factors, including age at onset of learning and TL experience.

changes with age and as a result of L1 experience. In the view taken by Flege (1995), perceptual learning as a result of TL experience yields the modification of established L1 categories and the creation of new TL phonological representations. L1 and TL representations coexist in the same phonological space.

The crucial step in phonological learning is the formation of new categories. Whether a learner succeeds in establishing a new category depends on their ability to perceptually distinguish L1 and TL phonetic segments. The SLM hypothesizes that learners perceptually relate TL sounds to the closest L1 sound. The unit of perceptual analysis is the position-sensitive allophone. Perception is thus held to operate on a level less abstract than the phoneme and more abstract than the surface phone.

The decisive factor is the *perceived phonetic dissimilarity* between the TL sound and the closest L1 sound. The more similar the two sounds are perceived to be, the less likely it is that a new category will be formed. For a change in the IL grammar to occur, at least some of the phonetic differences must be discerned. If the sounds are perceived as dissimilar, the TL sound is parsed as a new sound and a new category will be established. This new category may differ from the TL category in feature weights. If, on the other hand, the learner detects no difference, category formation will be blocked by what Flege (1995) terms *equivalence classification*. The perceptually linked sounds will then be processed using a single phonetic category and will be equivalent in production. A critical factor is age, as perceptual capacity is assumed to vary with AOL: Younger learners are more sensitive to acoustic differences that become backgrounded as a result of L1 experience.[4]

SLM-based hypotheses about IL variation hinge on two premises: (i) the identification of the L1 sound that is closest to the TL segment and (ii) knowledge about the degree of perceived similarity between the two. A quantification of the degree of perceived similarity then translates into predictions about the likelihood of equivalence classification and the relative difficulty of a given TL sound. Tests of the SLM therefore rely on a valid assessment of similarity. Bohn (2002) argues that direct perceptual assessments provide the most valid measurement of phonetic similarity and should therefore be preferred to indirect acoustic comparisons. The type of variation the SLM is concerned with is that between segmental structures. Invoking the notion of perceived phonetic similarity, it resorts to inherent characteristics of position-sensitive allophones to explain and predict why certain segments are more difficult to learn than others. Essentially, it thereby arrives at relative statements about perception-based difficulty, allowing the researcher to formulate specific hypotheses on L2 speech production. Similar to the CAH, however, its scope is limited to segmental structures and between-structure variation.[5]

[4] Flege's (1995) SLM was formulated on the basis of a large body of empirical work investigating segmental perception and production by learners from different L1s. Bohn & Flege (1992), for instance, found that experienced L1 German learners of English showed authentic production of the dissimilar sound /æ/ but not the similar sounds /iː ɪ e/. Similar findings were reported by Flege (1987) for advanced L1 English learners of French, who succeeded in learning the new sound /y/ but not the similar sound /uː/.

[5] Since the SLM focusses on position-sensitive allophones, however, the model does account for within-structure variation if the term *structure* refers to a phonemic category.

Similarity Differential Rate Hypothesis (SDRH)

While the relationship between perceived phonetic dissimilarity and relative difficulty has been substantiated by empirical research, a number of studies have produced conflicting findings. Thus, it has been observed that the accuracy of similar (vs. dissimilar) structures may decline in more experienced (vs. less experienced) learners (Major 1997, 1987; Major & Kim 1996; Flege et al. 1997). In light of this evidence, Major & Kim (1996) argued that similarity may be better described as affecting *rate* rather than *ease* of acquisition. Replacing the notion of *difficulty* with *rate*, the Similarity Differential Rate Hypothesis (SDRH) states that "[a]n L2 phenomenon that is dissimilar to an L1 phenomenon is acquired faster than an L2 phenomenon that is similar to this same L1 phenomenon" (ibid.: 474). In longitudinal or cross-sectional studies, this is reflected in the amount of change in accuracy observed over time for two structures. The SDRH also allows negative slopes – that is, deterioration as a result of IL development.

The SDRH's notion of rate of acquisition adds a developmental dimension to between-structure comparisons. As it is conventional to think of change over time graphically, differences in rate translate into different slopes. The SDRH expects steeper slopes for dissimilar phenomena. This permits the inclusion of time as an additional variable in hypothesis-formulation. Upon closer inspection, however, two points appear to complicate the empirical evaluation of the SDRH:

- First, a consequential methodological decision is the operationalization of production accuracy. Assessments using acoustic measurements, for instance, will yield higher accuracy rates for similar phenomena, since these are (at least initially) expected to be more similar to the TL category. Different results may be obtained if production accuracy is assessed categorically (e.g. a binary classification into *target* vs. *non-target*). The operationalization of production accuracy therefore affects estimates of the rate of acquisition.

- Second, the improvement of structures that are initially more target-like is constrained by a ceiling effect. As a result, the slope is more constrained for similar structures in terms of absolute value. It is therefore almost a logical necessity that dissimilar structures will exhibit steeper slopes.

While the SDRH does not stipulate novel constraints to explain IL phenomena, its shift of focus from *difficulty* to *rate* adds a developmental perspective to the central tenets of the SLM. This opens up new avenues for between-structure comparisons, but also raises new questions about the comparability between structures and measurements of production accuracy. These issues encourage careful reflection on methodology. Finally, like the CAH and the SLM, the SDRH is limited in scope to segmental structures and between-structure variation.

Perceptual Assimilation Model (PAM-L2)

Embedded in the framework of Browman & Goldstein's (1992) Articulatory Phonology, Best's (1995) Perceptual Assimilation Model (PAM) assumes that a speech sound is assimilated to the *articulatorily closest* L1 sound. This assimilation may lead to the sound being *categorized* – that is, perceived as a more or less prototypical exemplar of an L1 phonological segment

– or *uncategorized* – that is, perceived as unlike any L1 segment. The prototypicality of an exemplar can be assessed empirically via goodness-of-exemplar ratings, where subjects are asked to rate on a continuous scale how well a token represents a certain category.

PAM-L2 extends this model to L2 perceptual learning (Best & Tyler 2007). It differs from the SLM in two respects: (i) It assumes (cross-linguistic) perception to operate in terms of articulatory (rather than acoustic-phonetic) similarity, and (ii) its primary interest lies in contrasts between TL sounds. While the SLM attends to contrasts between a TL sound and the closest L1 sound, PAM-L2 goes one step further. While the perceptual assimilation patterns of TL to L1 sounds also form the basis of cross-linguistic influence in speech perception, the model is ultimately interested in how L1-TL interference bears on the perception of contrasts in the TL. Thus, while the SLM restricts its focus to a single TL category, a PAM-L2-informed approach deals with pairs of TL sounds.

The model focuses on the perception of TL minimal contrasts and stipulates two relevant factors: (i) the type of perceptual assimilation (*categorized* vs. *uncategorized*) and (ii) the prototypicality of *categorized* sounds as exemplars of the respective L1 category. PAM-L2 assimilation patterns are shown in Table 2.1.

Pattern	Assimilation	Discrimination
Two Category	Assimilated to different L1 categories	Excellent
Single Category	To same L1 category, equal goodness-of-exemplar ratings	Poor
Category Goodness	To same L1 category, different goodness-of-exemplar ratings	Intermediate
Uncategorized-Categorized	Only one phone is perceived as a speech sound and assimilated	Very good
Uncategorized-Uncategorized	Neither phone is perceived as a speech sound and assimilated	Poor to excellent

Table 2.1: Assimilation patterns and discrimination performance posited in PAM-L2.

Depending on which assimilation pattern applies, perceptual learning – that is, new category formation – is more or less likely to occur. If the two focal TL phones are assimilated to different L1 categories (Two Category), learners will be able to detect the contrast, but further perceptual learning is unlikely for either phone. If both phones are assimilated to the same L1 category and both are equally good or bad representatives of this category (Single Category), perceptual learning will be difficult. If both phones are assimilated to the same L1 category with different goodness-of-exemplar ratings (Category Goodness), category formation is more likely to occur for the more peripheral exemplar.

Facilitating factors for perceptual learning are frequency and functional load of the contrasting phones in the TL. If only one of the phones is assimilated to an L1 category, perceptual learning for this phone is unlikely, but expected for the other one. Finally, if both phones are uncategorized, their discriminability determines perceptual learning (cf. Best & Tyler

2007: 28). The predictions of PAM-L2 have found empirical support from a number of studies, which mainly focused on Single Category, Two Category, and Category Goodness contrasts (cf. Colantoni et al. 2015: 41).

There are clear parallels between the SLM and PAM-L2: Uncategorized sounds correspond to new sounds, and category goodness is largely analogous to the degree of perceived dissimilarity. In considering TL contrasts, however, the model goes one step further than the SLM, since perceptual learning is not only determined by L1-TL comparisons (in terms of gestural similarity) but also by comparisons within the TL sound system. Arguably, this provides a more realistic account of L2 perceptual learning.

Since PAM-L2 is restricted to perceptual learning, it does not explicitly address speech production. Extrapolating its tenets, however, we would expect learner productions to reflect different underlying assimilation patterns. PAM-L2-derived hypotheses about L2 speech production require an experimental assessment of the assimilation patterns and goodness-of-exemplar ratings of the focal TL sounds and relevant L1 exemplars. In the absence of such evidence, any hypothesis invoking the notions of PAM-L2 must be considered at best suggestive. The incorporation of psycholinguistically informed concepts such as frequency and prototypicality effects is a strength of the model. Similar to the CAH, SLM, and SDRH, PAM-L2 is restricted to segmental structures and between-structure variation.

Desensitization Hypothesis (DH)

The SLM and PAM-L2 account for perception and production difficulties in terms of cross-linguistic influence. Regarding non-native vowels, however, Bohn (1995) argues that interference alone may not be sufficient for explaining L2 perceptual categorization. Learners are assumed to be biased towards durational cues irrespective of whether or not they operate in L1. Duration is regarded as more salient and thus easier to access relative to other signals even if the learner has had no prior experience with this cue. Bohn's (1995) Desensitization Hypothesis (DH) holds that if learners are desensitized to spectral differences in a particular region of the vowel space, they will use duration to distinguish between vowels. Empirical support was reported in a number of studies (Bohn & Flege 1990; Escudero & Boersma 2004; Cebrian 2006).[6]

The tenets of the DH can complement SLM-informed reasoning about new category formation by formulating testable hypotheses about the way in which newly formed IL categories and TL categories may differ in feature weighting. Specifically, it predicts (initial) bias toward length as a distinctive feature. Since the DH only applies to new vowels, its scope is restricted to this subclass of TL segments.

What the perception-based models discussed so far have in common is their phonetic approach to speech perception. With a focus on acoustic and articulatory levels of description, the SLM and PAM-L2 advance physically observable surface properties of speech sounds as the primary units of analysis. Cross-linguistic influence on speech perception has also been addressed from a phonological perspective, where abstract underlying representations rather than surface properties are held to determine

[6] However, it has been questioned whether duration is in fact universally accessible. No evidence for the DH was found by McAllister et al. (2002), who investigated the acquisition of Swedish quantity distinctions by L1 American English, Latin American Spanish and Estonian speakers.

sensitivity to non-native contrasts. Specifically, L1-TL differences in the inventory of phonological features are put forward to explain the relative difficulty of TL segments. L2 perception research in this domain relies on phonological theory to provide an abstraction of the mental organization of features. In what follows, two such models will be discussed.

Phonological Interference Model (PIM)

Brown's (1998) Phonological Interference Model (PIM) aims to account for perceptual difficulty by reference to contrasts between the L1 and TL inventory of phonological features. The underlying assumption is that learners are able to redeploy L1 features to the perception of TL contrasts. According to the PIM, contrasts involving at least one feature that also operates in L1 can be discerned by the learner, as L1 features can be rearranged to form new representations. The decisive factor for perceptual sensitivity to a TL contrast is therefore presence/absence of (a) relevant feature(s) in the L1. Considering the formal description of the underlying grammar, Brown (1998) adopts a version of Feature Geometry proposed by Rice & Avery (1995). In general, Feature Geometry holds that phonological features are hierarchically structured, with co-occurrence and dependency relationships between them having an articulatory basis. In the characteristic tree-like representations, each node is assumed to have a preferred or unmarked instantiation. For example, if a language exhibits peripheral consonants, it will have labial (unmarked) and possibly also dorsal consonants – but not vice versa. If a language has only one (the default) instantiation of a node, it is predictable and therefore not represented in the speaker's grammar. For a non-native contrast to be detected, at least one relevant feature must be represented in the listener's L1 feature tree; otherwise it will (at least initially) go unnoticed.

The fundamental claim of the PIM is straightforward: If a TL contrast is signaled by features that are absent in the L1, the learner will struggle to perceive this contrast. Predictions about the relative difficulty of a structure hinge on the type of phonological grammar adopted for analysis, which is not amenable to empirical observation. In some settings, Feature Geometry may not provide an adequate representation of the type of contrast to be investigated.[7] In such cases, PIM-based predictions may resort to different types of phonological feature inventories. We will return to this point shortly.

[7] In particular, due to the disputed status of the features [lateral] and [sibilant] in the feature tree as well as the unsettled debate about the representation of complex segments with a double articulation such as [w] and affricates (Hall 2011: 199–207), Feature Geometry provides an unreliable basis for predictions about a number of relevant segmental features in GLE.

Feature Competition Model (FCM)

While the PIM relies on presence/absence of features in the L1, the Feature Competition Model (FCM) proposed by Hancin-Bhatt (1994) introduces the gradient notion of feature prominence. Languages are assumed to differ in the weight and salience of their phonological features, which can be determined on the basis of their functional load. In particular, features that are involved in fewer contrasts in the L1 are assumed to be less prominent and thus less likely to trigger cross-linguistic influence. The type of phonological model employed by the FCM is Underspecification

Theory, which thinks of phonological representations in the form of feature hierarchies. Similar to Feature Geometry, these inventories are structured in an economic fashion, with redundant or predictable features and feature specifications not being represented in the grammar. In general, listeners are assumed to be biased toward certain features as a result of L1 experience. Hancin-Bhatt (1994) proposed a metric for determining perceptual salience, which is simply the ratio of the number of phonemes that contain the feature (assuming underspecification) to the total number of phonemes.[8]

In contrast to the PIM, which posits phonological interference to surface in an all-or-nothing fashion, the FCM views perceptual bias as gradient. Similar to the PIM, the FCM adopts a particular model of phonological representation. In the context of the present study, Underspecification Theory appears to be problematic given the lack of agreement on the underspecified inventory for German.[9] Nevertheless, the tenets of the PIM and FCM are autonomous in the sense that hypotheses can be formulated against the backdrop of any particular phonological theory that invokes the notion of features.

In this study, an inventory of distinctive features closely corresponding to that of Chomsky & Halle (1968) will form the basis of the application of both models. The feature matrices established for English and German consonants and vowels[10] are given in Tables 2.2, 2.3, and 2.4. These were derived from Wiese (1996: 152, 164), Hall (2011: 131–132) and Giegerich (1992: 110, 128). Following Clements (1999), affricates are classified as [–continuant, +strident, –anterior].

From the viewpoint of the PIM, a contrastive analysis based on these matrices shows identical sets of features for vowels, and one difference in the inventory for consonants: German lacks the feature [strident], as it has no meaning-distinguishing function (see Hall 2011: 132). This severely limits the scope of the PIM for German Learner English. A quantification of feature prominence in the spirit of the FCM is given in Table 2.5 for German vocalic and consonantal features. Features are ordered by prominence, which is indicated by the prominence ratio proposed in Hancin-Bhatt (1994).

What the perception-based models of L2 phonological acquisition have in common is their limited focus on segmental structures. In principle, the tenets of these contributions can be extended to the prosodic level of analysis. What remains unclear, however, is whether (non-native) listeners can consciously attend to structures above the level of the segment and whether they are able to perceive them contrastively. To my knowledge, the SLM, PAM-L2, PIM, and FCM have not been discussed in the light of prosodic phenomena. In the present study, their scope will therefore be considered as exclusively segmental.

2.3 *Language universals*

While contrastive analysis dominated early L2 speech research, the 1970s marked a shift in SLA theorizing. Influenced by the cognitive current of

[8] Hancin-Bhatt & Govindjee (1999) extend the prominence metric to include a further factor, the frequency of specific contrasts in the lexicon.

[9] Thus, Hancin-Bhatt's (1994) matrix for German obstruents, which is based on Benware (1986), differs considerably from the one proposed by Wiese (1996: 165), both in number and choice of features.

[10] We follow the literature and disregard diphthongs in our survey of distinctive vowel features.

Table 2.2: Distinctive features for German and English vowels.

	German																English											
Feature	iː	ɪ	yː	ʏ	eː	ɛː	ɛ	øː	œ	aː	a	oː	ɔ	uː	ʊ	ə	iː	ɪ	e	æ	ʌ	ɑː	ɒ	ɔː	uː	ʊ	ɜː	ə
[Back]	−	−	−	−	−	−	−	−	−	+	+	+	+	+	+	+	−	−	−	−	+	+	+	+	+	+	+	+
[High]	+	+	+	+	−	−	−	−	−	−	−	−	−	+	+	−	+	+	−	−	−	−	−	−	+	+	−	−
[Low]	−	−	−	−	−	−	−	−	−	+	+	−	−	−	−	−	−	−	−	+	−	+	+	+	−	−	−	−
[Round]	−	−	+	+	−	−	−	+	+	−	−	+	+	+	+	−	−	−	−	−	−	−	+	+	+	+	−	−
[Tense]	+	−	+	−	+	−	−	+	−	−	−	+	−	+	−	−	+	−	−	−	−	+	−	+	+	−	−	−
[Long]	+	−	+	−	+	+	−	+	−	+	−	+	−	+	−	−	+	−	−	−	−	+	−	+	+	−	−	−

Table 2.3: Distinctive features for English consonants.

	Obstruents																Sonorants							
Feature	p	b	t	d	k	g	f	v	s	z	ʃ	ʒ	tʃ	dʒ	θ	ð	m	n	ŋ	l	j	w	r	h
[Consonantal]	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	−	+	+
[Sonorant]	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	+	+	+	+	+	+	−
[Voice]	−	+	−	+	−	+	−	+	−	+	−	+	−	+	−	+	+	+	+	+	+	+	+	−
[Continuant]	−	−	−	−	−	−	+	+	+	+	+	+	−	−	+	+	−	−	−	+	+	+	+	+
[Nasal]	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	+	+	−	−	−	−	−
[Strident]	−	−	−	−	−	−	−	−	+	+	+	+	+	+	−	−	−	−	−	−	−	−	−	−
[LABIAL]	+	+	−	−	−	−	+	+	−	−	−	−	−	−	−	−	+	−	−	−	−	+	−	−
[CORONAL]	−	−	+	+	−	−	−	−	+	+	+	+	+	+	+	+	−	+	−	+	+	−	+	−
[Anterior]	+	+	+	+	−	−	+	+	+	+	−	−	−	−	+	+	+	+	−	+	−	−	−	−
[DORSAL]	−	−	−	−	+	+	−	−	−	−	−	−	−	−	−	−	−	−	+	−	−	+	−	−

Table 2.4: Distinctive features for German consonants.

	Obstruents														Sonorants						
Feature	p	b	t	d	k	g	f	v	s	z	ʃ	ʒ	tʃ	x	m	n	ŋ	l	j	ʁ	h
[Consonantal]	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	+	+
[Sonorant]	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	+	+	+	+	+	−
[Voice]	−	+	−	+	−	+	−	+	−	+	−	+	−	+	+	+	+	+	+	+	−
[Continuant]	−	−	−	−	−	−	+	+	+	+	+	+	−	−	−	−	−	+	+	+	+
[Nasal]	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	+	+	−	−	−	−
[LABIAL]	+	+	−	−	−	−	+	+	−	−	−	−	−	−	+	−	−	−	−	−	−
[CORONAL]	−	−	+	+	−	−	−	−	+	+	+	+	+	+	−	+	−	+	+	+	−
[Anterior]	+	+	+	+	−	−	+	+	+	+	−	−	−	−	+	+	−	+	−	−	−
[DORSAL]	−	−	−	−	+	+	−	−	−	−	−	−	−	−	−	−	+	−	−	−	−

Consonants			Vowels		
Feature	N	Prominence ratio	Feature	N	Prominence ratio
[Consonantal]	21	.95	[Round]	8	.50
[Voice]	12	.55	[Long]	8	.50
[Continuant]	12	.55	[Back]	7	.44
[Anterior]	11	.50	[High]	6	.38
[CORONAL]	10	.45	[Tense]	6	.38
[Sonorant]	6	.27	[Low]	2	.13
[DORSAL]	5	.23			
[LABIAL]	5	.23			
[Nasal]	3	.14			
[Strident]	0				

Table 2.5: Prominence of consonantal and vocalic features in German.

the time, the concept of interlanguage was introduced independently by several scholars and learner language came to be viewed as a system in its own right (Corder 1967; Nemser 1971a; Selinker 1972). With researchers observing parallels to universal tendencies identified in typological comparisons of natural languages and L1 acquisition, the focus shifted from L1-TL contrasts to more general linguistic constraints on non-native speech. Language universals emerged as a further explanatory paradigm for IL phenomena. This section deals with theories that invoke such concepts to account for the patterns observed in non-native speech.

Broadly speaking, the term *universals* refers to generalizations about natural languages; depending on the particular school of thought, however, it may denote different ideas. The two major streams in the study of language universals are the generative and the typological school. Apart from differences in methodology and abstractness of analysis, the two strands put forward different explanations for what they consider as universal (Comrie 1989). The generative grammar approach is characterized by a highly formalized and abstract description of linguistic structures and is usually based on a fine-grained analysis of a small number of languages. This perspective is most closely associated with the work of Chomsky (1959, 1965). In order to account for cross-linguistic regularities and ease of L1 acquisition, generativists assume that human beings are biologically endowed with a language faculty or universal grammar (UG), a genetic predisposition for learning language. The ongoing endeavor is to seek a description of this universal grammar, which can be thought of as the building plan for human language.

The typological approach to language universals originated in the work of Greenberg (1976). Cross-linguistic generalizations are usually stated on a more concrete level of analysis and are derived from the comparison of a large number of natural languages. Such generalizations can be differentiated along two dimensions: (i) implicational vs. non-implicational universals and (ii) absolute universals vs. universal tendencies. The latter binary concerns universality as such: While absolute universals are exceptionless, tendencies reflect statistical trends. The notion of implication, on the other hand, distinguishes conditional and unconditional statements about the presence of structures. Non-implicational universals are autonomous statements in the sense that the existence of a particular phenomenon does not depend on the presence/absence of other structures. One example is that all languages have vowels. Implicational universals, on the other hand, are conditional generalizations. Typically, these are distributional statements of the form "if B then A" – the presence of structure B implies the presence of structure A. In other words, if we observe B in a language, the language must also have A. Regarding onset complexity, for example, a language that permits clusters in syllable onsets (e.g. *spy* /spaɪ/) also has syllable-initial singletons (e.g. *pie* /paɪ/).

Both approaches to language universals have contributed to the study of second language phonology. The implementation of typological universals in SLA is most notably associated with the work of Eckman (1977, 1991), whose contributions will be discussed in §2.3.1. While generative accounts have predominantly dealt with (morpho)syntax, the apparatus of Chomsky's (1981) Government and Binding Theory has been brought to bear on L2 phonology in the work of Archibald (1994); his UG-based model of L2 stress acquisition is presented in §2.3.2.

2.3.1 Typological universals

The notion of markedness, which originated in the Prague School of Linguistics (Trubetzkoy 1939; Jakobson 1968), holds that binary oppositions between certain language structures are asymmetric: One member of the pair is more basic, natural and frequent – in markedness terminology: *unmarked*. Marked structures generally occur in fewer languages, are learned later in L1 acquisition and are historically unstable (Kager 1999; Hall 2011).

Markedness Differential Hypothesis (MDH)

Typological markedness was introduced into SLA through Eckman's (1977) Markedness Differential Hypothesis (MDH), which builds on Greenberg's (1976) notion of implicational relationships between structural binaries. As noted above, the presence of a marked structure B implies and thus depends on the presence of the corresponding unmarked structure A. According to the MDH, learners will experience difficulty with new L2 structures that are marked, the degree of difficulty being proportional to the degree of markedness. A new structure that is unmarked, on the other hand, will cause no difficulty. Empirical support for the MDH was primarily reported for the acquisition of onset and coda clusters as well as syllable-final consonants (e.g. Anderson 1987; Eckman 1981; Carlisle 1991), owing to the fact that typological generalizations about these structures are relatively robust (cf. Eckman 2008: 102).

Markedness has emerged as an important concept not only in L2 phonological research, but in the disciplines of phonetics and phonology (and other linguistic fields) more generally. Its appeal as an explanatory principle stems from the fact that it establishes a connection to linguistic constraints and tendencies that have been identified in other fields such as historical linguistics, typology, and L1 acquisition. This allows SLA research to draw on a diverse knowledge base for formulating hypotheses and understanding IL phenomena. Findings from L1 acquisition and language typology are particularly valuable since they may inform predictions about IL patterns and guide the interpretation of data.

However, the employment of markedness in L2 research has also met with criticism mainly due to the concept itself. More generally, the term has been attacked because of its vagueness and status as "an almost theory-neutral everyday term in linguistics" (Haspelmath 2006: 27). The added value of the concept has been challenged, since what it refers to or aims to describe may be captured by more substantive explanations. This is especially true for phonology, where its relevance and explanatory power have been questioned, not only from a general theoretic perspective (e.g. Hume 2004; Blevins 2004), but also in SLA theorizing. Archibald (1998: 150), for example, evaluates the contribution of the functional-typological approach exemplified by the work of Eckman (1991, 1977) as "an interesting description of the phenomena to be explained".

As pointed out by Haspelmath (2006), phonological markedness can be pinned down to frequency of use and phonetic factors such as articulatory and acoustic properties. Frequency and phonetic difficulty are related and in fact reinforce each other, which strengthens their association: Highly frequent segments may undergo articulatory simplification, and due to articulatory complexity, segments may be rare. These psycholinguistic and physiological attributes decompose the abstract term into concrete observables and thereby provide a more concrete basis for incorporating markedness in applied research.[11] Further, the concept gains explanatory adequacy and reveals parallels to other fields in linguistics in which equivalent constraints have been identified.[12]

An attractive feature of markedness is the scope of its applicability. It is not restricted to comparisons between isolated structures but allows an assessment of other linguistic factors such as phonetic or prosodic context. For example, syllable onsets are considered to be less marked relative to codas. In terms of more concrete characteristics, onsets are perceptually stronger (Strange 1992), they license a larger number of consonants (Colantoni et al. 2015: 186), and they are cross-linguistically more frequent (Maddieson 2013). Such markedness relationships can be employed to explain or predict within-structure variability. Since markedness considerations also apply to prosodic phenomena, the MDH is not restricted to segmental units. It therefore exceeds most of the models discussed so far in terms of scope.

Structural Conformity Hypothesis (SCH)

Since the MDH was originally proposed as a refinement of the CAH, strictly speaking its scope is limited to areas of L1-TL difference. This led Eckman (1991: 24) to postulate the more general Structural Conformity Hypothesis (SCH), which states that "[t]he universal generalizations that hold for the primary languages hold also for ILs". The SCH abstains from the CAH stance of L1-TL contrasts and does not make explicit statements

[11] The criticism markedness has attracted should be considered an opportunity for L2 phonology researchers rather than a disqualification of the concept. In fact, it may be considered a useful cover term for a number of explanatory factors rooted in human cognition and the physiology of articulation. As these attributes can be assessed based on the empirical literature, SLA researchers are able to assess the *markedness* of structures for which no markedness relationships have been posited.

[12] What appears to be difficult to operationalize, however, is the MDH's notion of relative degree of markedness, which translates into relative degree of difficulty. While it is conceivable to compare structures using attribute lists including (and possibly weighting) several factors such as frequency, articulatory complexity and acoustic salience, it is not clear how to establish comparability between different types of structures such as vowels and consonants.

about learning difficulty. ILs are simply posited to be natural languages which obey the same universal principles that hold for primary languages, one such universal being markedness.[13]

[13] One type of evidence for the SCH are IL structures that occur independently of the L1 and TL. If this particular type of structure is attested in other languages and resembles the outcome of what may be considered a universal generalization or natural tendency, this would be consistent with the hypothesis.

The fact that Eckman's (1991) SCH is stated very broadly may be seen both as a strength and a weakness. The investigation of a particular structure in learner speech thus presupposes an understanding of how this phenomenon behaves in natural languages synchronically, diachronically, as well as in language acquisition. Identification of parallel mechanisms is attractive for several reasons. First, it adds to the credibility of SLA findings as it identifies a familiar phenomenon in a new population of speakers. Such parallels are also desirable from a broader viewpoint as they further underline the existence of universal constraints on language. Second, as pointed out above, the researcher has at her disposal a larger body of literature for an explanation of observations and hypothesis formulation. Third, if a parallel tendency has been identified but its cause remains subject to speculation, the burden of advancing an explanation rests on many shoulders. The SCH can thus also be understood as an impulse for interdisciplinary work striving to find common ground.

Its vague statement on the equivalence of ILs and natural languages, however, invites subjective interpretation. The results of a study may to a large part be dependent on the researcher's judgement of whether a particular constraint qualifies as a universal. The SCH's fuzziness is not inevitably a weakness, but it does leave room for selective quotation to account for patterns in the data. It is therefore essential to lay out the assumptions on which the application of the SCH is based. Indeed, the premises or universals a researcher relies on may be the most consequential part of an empirical study within this framework.

An advantage of Eckman's (1991) SCH is its broad scope. Surely, universals are to be found at many if not all levels of phonetic and phonological structure. The hypothesis thus reaches out to segmental as well as suprasegmental phenomena and is capable of accounting for variation between and within structures.[14] In principle, the scope of the SCH is only limited by our knowledge about language universals.

[14] In fact, it may also apply to speech perception. As was mentioned above, universal dispositions in the perception of vowels have been put forward by Bohn's (1995) Desensitization Hypothesis (see §2.2.3).

2.3.2 Universal Grammar

The theory of universal grammar (UG) originated in the 1960s in the early works of Chomsky (1959, 1965). The aim of the generative enterprise is to describe the distinctive characteristics of human language, including the abstract representation of language in the mind. Language and the ability to learn it are seen as genetically endowed. Prominent arguments in favor of this claim are the independence of L1 mastery from cognitive abilities and the input, with certain structures being acquired despite limited positive evidence.[15] In the UG model established in Chomsky's (1981) Government and Binding Theory, the universal footing of language consists of a set of principles and parameters, which constrain language acquisition. Principles are invariant – they are true for all languages. Parameters, on the other hand, allow for variation between languages in that they comprise a limited number of settings. Human beings are thus

[15] This is referred to as the *poverty-of-the-stimulus argument* or the *logical problem of language learning*.

assumed to be genetically equipped with a fixed set of principles and a limited choice of parameter settings. This is called the initial state, the starting point for language acquisition. Based on this universal grammar, children construct their L1 grammar by inferring the relevant settings from the input. From the viewpoint of SLA, the crucial – and controversial – question is whether an L2 learner still has access to UG, and by implication whether the IL system developed by learners conforms to the same UG principles as primary languages (cf. Mitchell et al. 2012: 90ff.).

UG-based model of L2 stress acquisition

A UG-based model of the L2 acquisition of lexical stress was proposed by Archibald (1994). Embedded in Chomsky's (1981) principles-and-parameters framework, it rests on the assumption that cross-linguistic variation in stress assignment can be modeled in terms of differential parameter settings. It is posited that ILs abide by metrical universals and learners can access parametric options to arrive at a TL-appropriate representation. The model includes a learning theory to account for the process by which a learner can tap into UG. An important assumption is that learners have access to indirect negative evidence, which means they can infer the ungrammaticality of a structure based on its absence in the input.[16] Drawing on work by Saleemi (1992), Archibald (1994) proposes such inductive learning to occur after a certain frequency or time threshold is crossed. Up to this point, the learner's performance may show variation. A further component of the learning theory is the notion of cue appropriateness (Dresher & Kaye 1990). Upon noticing a mismatch between input and internal representation, a learner is able to find the correct parameter to change. Finally, lexical dependency explains the propagation of a change in the grammar, with new information being generalized to lexemes within the scope of the new parameter setting via feature-copying mechanisms.

[16] In contrast, UG accounts of L1 acquisition consider positive evidence to be sufficient.

Archibald's (1994) UG-based model of lexical stress acquisition exemplifies the application of UG thinking to L2 phonological theory. It rests on a number of theoretical premises that are idiosyncratic for the school of generative grammar, most notably the idea of innate principles and parameters.[17] Archibald's (1994) model crucially relies on the presence of metrical universals and a fixed set of parametric options specifying lexical stress placement. More generally, the model only applies to lexical stress acquisition and is thus limited in scope.

[17] The postulation of a genetic blueprint for language has been questioned, most prominently by the current stream of functional-cognitive approaches to language acquisition, which argue that language can indeed be learned through reliance on domain-general skills (e.g. Tomasello 2003).

The following sections are concerned with theoretical approaches to L2 phonology that also invoke universal constraints on IL phonology. These include contributions nested in the schools of Natural Phonology (§2.3.3) and Functional Phonology (§2.3.4), and frameworks stipulating articulatory constraints (§2.3.5).

2.3.3 Natural Phonology

Natural approaches to L2 speech hold that the innate natural phonological processes that frame L1 acquisition also shape L2 acquisition. The theory of

Natural Phonology (Stampe 1979) stresses the role of universal articulatory and perceptual constraints, positing that sound patterns of languages are governed by a language-universal set of natural processes. What makes these processes natural is the fact that they have an articulatory basis – that is, they are explainable by reference to the physiology and movement of the speech organs. An example for a natural process is assimilation. In order to save a tongue tip gesture, *unpopular* is usually articulated as [ʌmˈpɒpjələ]. The realization of /n/ anticipates the feature LABIAL of the following consonant, the phonetic motivation being ease of articulation.

To account for the sound patterns of languages, Natural Phonology distinguishes between two opposing functional principles: the speaker-oriented principle of ease of articulation and the hearer-oriented principle of ease of perception. Accordingly, there are two types of natural processes: fortitions and lenitions. Perception-directed fortitions aim at maximizing intelligibility by foregrounding distinctive features. Such strengthening processes apply, for example, to vowels in stressed syllables or, more generally, in situations where perceptibility is critical. If uttered insistently, for instance, the directive *Get out* likely reflects two fortitions: (i) aspiration of both alveolar stops and (ii) onset strengthening in *out* through glottal stop insertion – [ˈgetʰ ˈʔaʊtʰ].[18] Lenitions, on the other hand, serve to minimize articulatory effort by reducing complexity in the speech stream. These weakening processes apply, say, in unaccented syllables and in casual speech. The utterance *Can't find them*, for instance, my exhibit several lenitions: (i) deletion of [t d ð], (ii) anticipatory place assimilation of [n] in *Can't*, and (iii) reduction of vowel quality in *them* – [kæɱ̩ˈfaɪnəm].[19]

[18] Other examples of fortitions are dissimilation, epenthesis and diphthongization.

[19] Further examples of lenition processes are monophthongization and desyllabification.

Natural Phonology takes the view that infants are equipped with a latent set of innate universal processes. Language acquisition involves the suppression, limitation and ordering of these processes. Through perception, the child constructs the L1 phonological system, which is "largely the residue of an innate system of phonological processes, revised in certain ways by linguistic experience" (Stampe 1979: vii).

Natural Model of L2 Phonological Acquisition (NM)

A model of L2 phonological acquisition within the framework of Natural Phonology was proposed by Dziubalska-Kołaczyk (1990a,b). The Natural Model of L2 Phonological Acquisition (NM) posits that learners construct an abstract L2 phonological representation by accessing the same set of natural processes that apply in L1 acquisition. Perception plays a central role and is assumed to operate on the level of abstract phonemic categories ("sound intentions") rather than phonetic surface realizations.

Initially, learners relate L2 phones to L1 phonemic categories ("sound intentions"). The first step in L2 phonological acquisition is for the learner to perceive L2 surface realizations in their own right without reference to L1 phonemes. Perceptual learning is facilitated by input frequency, positive attitudes toward the TL and explicit instruction.

Once TL phones are perceived independently of the L1, L2-specific sound intentions (phonological categories) can be formed. Having established a mental representation of an L2 sound intention, a learner can access universal processes. In order to arrive at the appropriate L2 phonological grammar, processes may have to be suppressed, unsuppressed, or reordered if their organization is inconsistent with the L2-specific sound intention. Consistent perception in terms of L2 phonological categories requires alternations at the level of universal processes. The NM also takes into account the notion of markedness, which is re-expressed in terms of the concepts advanced by Natural Phonology. Structures are unmarked if they are the result of a universal process, and marked if they require the suppression of a natural process.

Dziubalska-Kołaczyk (1990b) offers a comprehensive model of L2 phonological acquisition, which blends a number of recurrent themes in L2 theorizing. Similar to several accounts discussed above, perception is assumed to play a critical role in L2 phonological acquisition. In contrast to models such as the SLM and the PAM-L2, however, the NM is less specific on the mechanisms underlying the categorization of TL sounds by non-native listeners. Similar to the PIM and the FCM, perception is assumed to operate on the level of abstract representations. In contrast to most theoretical approaches discussed so far, the NM provides a formal description of the learning process, abstracting L2 phonological acquisition into three steps. Facilitating psycholinguistic factors such as frequency in the input and learner variables such as language attitude and aptitude are taken into account as additional influences on progress in IL development.

The application of the model to the empirical study of non-native speech requires the researcher to relate surface structures to underlying natural processes. As noted by Blevins (2006), this is in many cases far from straightforward. Natural sound patterns are commonly defined as being grounded in articulatory and perceptual properties of speech, which need not reverberate in attributes such as cross-linguistic frequency or diachronic stability. In any case, the types of natural processes that are assumed to be relevant to the acquisition of a particular structure must be stated explicitly so they are open to critical scrutiny.

Naturalness Differential Hypothesis (NDH)

As pointed out by Schmid (1997), naturalness can be considered a logical counterpart to markedness. He proposed the Naturalness Differential Hypothesis (NDH), a transposition of Eckman's (1977) MDH into the Natural framework, replacing the notion of markedness with naturalness. In line with Dziubalska-Kołaczyk's (1990b) model, phonological learning is assumed to require access to and modification of processes (suppression, limitation, reordering). A contribution of the NDH is its prediction of relative ease of acquisition based on the status of a given process in the L1 and TL grammar – that is, whether it is latent, active or suppressed. The constellations in L2 acquisition are listed in Table 2.6. According to the NDH, natural processes that are active in the L1 but suppressed in the TL are difficult to suppress. Those that are suppressed or latent in the native language are easy to activate. Finally, processes that are latent in the NL and suppressed in the target language will surface in IL. While Schmid's (1997) NDH does not bring forward new concepts, his "natural" translation of the MDH allows the researcher to formulate testable hypotheses. The application of the NDH in empirical work is subject to the same caveats outlined above and thus hinges on the type(s) of natural process(es) that are assumed to underlie the acquisition of a TL structure.

Status of process		
L1	TL	Effect
Suppressed	Active	Activation of process easy
Latent	Active	Activation of process easy
Active	Suppressed	Suppression of process difficult
Latent	Suppressed	Process will appear in the IL

Table 2.6: NDH predictions about learning difficulty.

2.3.4 Functional Phonology and Optimality Theory

Optimality Theory (OT) was proposed by Prince & Smolensky (1993) as a universal grammar scheme for phonology. The basic assumption is that a grammar consists of a finite set of constraints, which pose demands on the well-formedness of the output. These are assumed to be genetically endowed – that is, part of UG – and violable, with conflicts being resolved through a rank order. Phonological grammars select the surface realization that is most harmonious with the underlying constraints.

There are two major classes of constraints, which place conflicting demands on the output: faithfulness and markedness constraints. Faithfulness constraints require the output to resemble the underlying representation and thus aim to preserve lexical contrast. The specification IDENT-IO(voice), for example, demands voicing agreement between input and output. Markedness constraints, on the other hand, promote ease of articulation by favoring unmarked structures or feature values. An example of a markedness constraint is *VOICEDCODA (obstruents must not be voiced in coda position).

Depending on the rank order of IDENT-IO(voice) and *VOICEDCODA, the input /bɪd/ may be parsed in different ways. In English, the faithfulness constraint IDENT-IO(voice) dominates the markedness constraint *VOICEDCODA. As a result, the OT grammar of an English native speaker selects [bɪd] as the optimal candidate and the input feature [+voice] is preserved. In German, this ranking is reversed. Since *VOICEDCODA outranks IDENT-IO(voice), the optimal candidate is [bɪt].

Within the descriptive apparatus of OT, cross-linguistic influence can be formalized as the transfer of constraint rankings. Consequently, the task in L2 phonological acquisition is to reorder constraints to establish an IL grammar that conforms to the TL hierarchy. What triggers a modification in the system is the perception of a mismatch between the optimal output parsed by the speaker's grammar and the underlying input. The mechanism by which the constraint hierarchy is modified is referred to as a learning algorithm.

Functional Phonology, a model of phonological acquisition that relies on OT's computational framework, was proposed by Boersma (1998). In contrast to Prince & Smolensky's (1993) OT, however, Boersma (1998) takes an emergentist view of constraints, maintaining that they are learned through speech input rather than being innate. Though couched within the same framework, Boersma's (1998) functional approach thus differs from OT in this fundamental assumption.

Functional Model of Phonological Acquisiton (FM)

Boersma's (1998) Functional Model of Phonological Acquisition (FM) postulates separate production and perception grammars and distinguishes between articulatory and perceptual features. Its production grammar operates by OT principles and comprises faithfulness (FAITH) and articulatory (ART) constraints. These reflect functional principles of perception (listener-oriented minimization of perceptual confusion) and speech production (speaker-oriented minimization of articulatory effort); markedness constraints do not exist. Newly acquired ART constraints enter the grammar as undominated constraints at the top of the constraint hierarchy while FAITH constraints enter at the bottom.

The FM does not distinguish between L1 and L2 acquisition and assumes that both proceed according to the same principles. Its formalization of phonological acquisition consists of six stages. For illustration, consider a German learner's acquisition of the dental fricative /θ/.

- Initially, the learner's grammar is empty with respect to the new category /θ/. If an acoustic input [θɪk] is perceptually categorized as /sɪk/, it is identical to the learner's perceived own output – no mismatch will be detected and no learning will occur. A new perceptual category emerges in the learner's grammar through a stochastic construction process. Once [θɪk] is perceived as /θɪk/, a new faithfulness constraint is created: *DELETE(dental) (Do not delete an apico-dental gesture). This constraint enters at the bottom of the hierarchy.

- At stage 2, the newly created faithfulness constraint is violated as the learner will perceive their own output [sɪk] as /sɪk/. Mismatch with the model perception /θɪk/ leads to further learning. Next, the sensorimotor skills for the new gesture are learned. At this point *GESTURE(dental) enters the learner's grammar as an undominated ART constraint. In order for the production grammar to render a target-like output, ART needs to be demoted and FAITH promoted.[20]

- At stage 3, the output is sensitive to speaking style. Focus on form produces renditions that the comparison module perceives as a match. Due to processing demands on working memory, however, mismatches permeate unmonitored speech.

- Stage 4 marks the successful learning of the correct relative ordering of the ART and FAITH constraints. The learner's production grammar is able to faithfully produce the model utterance as [θɪk], and this stage is characterized by overly faithful production. Next, the learner proceeds to the post-lexical level, which includes connected speech phenomena

[20] The mobility of constraints depends on brain plasticity, with younger learners requiring fewer learning steps. The model thereby accounts for age effects in L2 acquisition.

such as assimilation or deletion. Consider the perceptual specification /mʌnθs/.[21] An overly faithful production grammar will yield [mʌnθs], while connected speech processes usually result in place assimilation [mʌnts] or deletion [mʌns].

- The learner's grammar at stage 5 allows for context-dependent promotion of ART constraints. Ultimately, the learner identifies patterns of stylistic and functional morphophonological alternations. As a result, abstract underlying forms emerge. The input to the learner's grammar now shifts from the model utterance to this newly created representation and the production grammar's output no longer depends on the input.
- The final stage 6 reflects a grammar with interacting ART and FAITH constraints. According to the two functional principles outlined above, FAITH constraints that demand the preservation of perceptual features that are relevant for communication rank high. For ART constraints, components blocking the production of difficult gestures rank high.

[21] Note that we use simplified notation. Boersma (1998) uses slashes to denote a perceptual representation, that is, the categorization of an acoustic event by the perception grammar. The input to the production grammar, which is stored in the mental lexicon, is shown using pipes. The correct notation here would therefore be |mʌnθs|.

The FM thus posits different forms of variation across developmental stages. Figure 2.1 illustrates the FM's predictions about within-structure variation. While the initial stages are characterized by little within-structure variation, the learner is subsequently able to produce the structure in a target-like fashion given sufficient monitoring and working memory capacities. Variation at stage 3 is thus sensitive to focus on form. At stage 4, overly faithful production is expected to yield little variation. Increasingly native-speaker-like stylistic variation is expected to emerge at stages 5 and 6, when the learner's speech production is governed by functional principles.

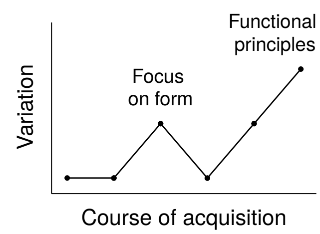


Figure 2.1: Schematic representation of the type and extent of variation predicted by the FM over the course of L2 acquisition.

OT has contributed to the study of SLA by offering a descriptive scheme that is able to incorporate various theoretical insights into L2 phonological acquisition. Transfer can be formalized as the influence of L1 constraint rankings and markedness constraints represent typological universals. The FM breaks with OT in two important respects, however. First, it does not posit constraints to be innate but views them as emergent phenomena that are learned. Second, it replaces the notion of markedness with articulatory difficulty and substitutes markedness constraints with articulatory constraints.

Boersma (1998) proposes a comprehensive model of L2 phonological acquisition, which integrates a number of recurrent theoretical notions. By replacing the concept of markedness with articulatory contraints operating within the speaker, the model relies on phonetic attributes whose validity holds across languages. The FM thereby encompasses markedness considerations if these are rooted in the physiology of articulation. Speech perception also plays a central role, with a perceived mismatch between (own) perceptual output and (model) input being the prerequisite for stochastic learning. While identical learning mechanisms are assumed for L1 and L2 acquisition, the FM allows for differences – that is, transfer-induced delays in perceptual and sensorimotor learning. L2-specific phenomena may thus be attributable to speech perception, a view that is in accordance with several influential models of L2 phonological acquisition (see §2.2.3). Further, cross-linguistic differences may constrain sensorimotor learning, which may be biased toward entrenched L1 articulatory settings (Honikman 1964).

A further attractive feature of the model is its reliance on general, implicit learning mechanisms that have been documented in other areas of L1 and L2 acquisition as well as other disciplines such as the cognitive sciences. With learning being assumed to proceed in an associative fashion, the FM incorporates frequency as a factor. In particular, input frequency drives the stochastic learning of new faithfulness constraints and output frequency affects the mastering of sensorimotor skills involved in learning new gestures. By incorporating stylistic variation, the model is also able to account for within-structure variability.

While the FM primarily deals with segmental phenomena, it does assume post-lexical phenomena to emerge after acquisition at the lexical stage has been completed. Other things being equal, connected speech phenomena would thus be expected to emerge relatively late.

2.3.5 Articulatory constraints

While physiological factors are a recurrent theme in the study of IL phonology, the status they have been assigned differs across theoretical strands. Honikman's (1964) AS theory, for example, considers cross-linguistic influence to surface in the transfer of entrenched L1 gestural routines. Articulatory constraints were thus not studied in their own right, but considered the outcome of L1 influence. In a similar fashion, Best & Tyler's (2007) PAM-L2 does not view gestures as an independent source of variation, but as a mediator of L1 influence on non-native speech perception. In AS theory and PAM-L2, articulatory constraints are thus attributed a secondary role.

Other models consider gestural constraints as a primary, independent factor rather than a vehicle for L1 transfer. Articulatory constraints thus play a critical role in Boersma's (1998) FM, where they impose restrictions on the output by promoting motor economy. In this regard, these constraints can be considered language-universal, since they shape natural languages in similar ways. Their universality derives from anatomical similarities between human beings. More generally, the investigation of speech-motor constraints on L2 speech production can build on the findings put forward by Experimental and Laboratory Phonology (Ohala 1993, 1971), whose primary aim is to forward phonetic explanations for recurrent sound patterns and sound change.

Model of Segmental Acquisition (MSA)

The relevance of Laboratory Phonology for L2 phonological theorizing was stated explicitly by Colantoni & Steele (2008). Based on their research on the acquisition of French [ʁ] and Spanish [ɾ] by US native speakers, they identify a number of shortcomings of current models of L2 phonological acquisition. Predictions formulated on the basis of the MDH, SLM, and PIM were found to be too general since they do not account for the observed, prosodically determined variability. The authors advocate an integration of insights from Laboratory Phonology into existing models of L2 speech learning and propose the Model of Segmental Acquisition (MSA). The MSA encompasses perception and production constraints. Perceptual processing is assumed to operate according to the mechanisms stipulated by the SLM. While the SLM somewhat vaguely posits that production will eventually resemble (newly established) mental categories, Colantoni & Steele (2008) elaborate on the correspondence between mental representation and output.

More specifically, universal properties of the speech mechanism are assumed to constrain learner output. Learners then compare their output to the perceived TL input and – upon detection of a mismatch – may further modify articulatory gestures. Fossilization results from a failure to perceive an input-output mismatch or from lack of articulatory control. Universal constraints related to the physics of speech production thus play a central role and may delay, impede, or cause variation in the correspondence between underlying representation and output. Such constraints have been identified for a number of segmental categories, most notably consonants.[22]

[22] For instance, it has been argued that, due to aerodynamic factors, voicing is easier to maintain in sonorants than in obstruents (Ohala 1993; Hoole 1999). Voicing preferences have also been related to place of articulation: the further back a consonant is articulated, the more difficult it is to sustain vocal fold vibration (Ohala 1993). Accordingly, typological comparisons have revealed that velar stops are least compatible with voicing (Greenberg 1970; Sherman 1975).

The MSA seeks to refine well-established models of L2 speech perception by adding a production channel that is regulated by physiological constraints operating within the talker. A strength of the model is its straightforward outline. This partly stems from the fact that it builds on Flege's (1995) SLM, an influential and widely familiar framework. The existence of production constraints is not only intuitively appealing but also adds substantive explanations to the observed variation in non-native speech. By reference to articulatory constraints the MSA further captures within-structure variation linked to the phonetic context.

2.4 Development and variation

One of the most notable characteristics of non-native speech is its high degree of variability, which is observable both within and between learners (Colantoni et al. 2015: 18). Corder (1981) introduced the distinction between *vertical* and *horizontal* variation. While horizontal variation is observable at a single point in time, vertical variation refers to change over time – that is, the course of IL development. This section is concerned with theoretical approaches that concentrate on IL variation. These can be assigned to two broad classes. Perspectives in the tradition of the variationist sociolinguistic study of language are discussed in §2.4.1. Theoretical contributions that are characterized by a developmental focus are presented in §2.4.2.

2.4.1 Variationist perspectives

The variationist approach to L2 phonology aims to uncover the regularities that condition variable learner productions. Preston (1996) notes that two variationist paradigms have influenced the study of SLA. The dynamic paradigm focusses on the spread of linguistic variants over time and sees variation as a transitory stage rather than part of language competence. The Labovian school, on the other hand, considers variation an inherent property of language and aims to describe the factors that account for the observed variability in speech. Its approach is typically strongly quantitative and characterized by efforts to disentangle the influence of different factors on alterations in language use. Both schools have found their way into SLA theorizing.

Gradual Diffusion Model (GDM)

Ideas from the dynamic paradigm were introduced into SLA research through Gatbonton's (1978) Gradual Diffusion Model (GDM), which builds on insights gained in studies on language variation and change in creole communities (Bickerton 1975, 1971) and L1 speaker groups (Bailey 1973; Chen & Wang 1975). Its basic tenet is that the target-like production of a segment develops gradually and conditional on phonetic context. This sequence is claimed to be predictable on the basis of linguistic constraints; precisely, the sonority of the immediate phonetic environment. Sonority refers to relative differences between speech sounds in (perceived) loudness and (acoustic) intensity. A sonority scale for the English phonemes is shown in Figure 2.2 (cf. Giegerich 1992: 133), where sonority increases from left (voiceless stops) to right (low vowels).

The GDM makes precise statements on the context-dependent progression of segmental accuracy. Two phases are postulated: an acquisition and a replacement phase. In the acquisition phase, the new sound begins to surface in certain preferred (sonorous) environments, where it alternates with non-target forms. The emergence of the newly acquired sound then progresses systematically to less sonorous contexts. At the end of the acquisition phase, the output shows variable patterns in all contexts. In the replacement phase, non-target forms begin to disappear, again starting in preferred environments. Acquisition and replacement thus advance in the same sequence of environments and occur in succession: Replacement begins after acquisition has been completed in all environments.

The GDM contributes to the study of L2 phonology by highlighting phonetic context effects in learner productions. With its introduction of sonority as a linguistic factor conditioning production accuracy, it was one of the first formal accounts of within-structure variation in non-native speech. Further, its notion of a context-dependent progression of segmental accuracy offers a framework for the systematic study of IL variation. While the GDM was proposed based on the factor sonority, the postulated two-stage development is independent of the specific explanatory factor employed for determining a sequence of preferred environments.[23] A

[23] As Trovimovich et al. (2007) demonstrate, other predictors can be identified and applied to the same data to compare their relative descriptive adequacy. A sonority-based analysis yielded results that were generally in agreement with those of Gatbonton (1978). In a second analysis of the data, the authors invoked two different explanatory principles: cross-language similarity and lexical frequency. While the predictions on the basis of these factors were more accurate, the developmental stages were found to closely reflect the stages and phases posited by the GDM.

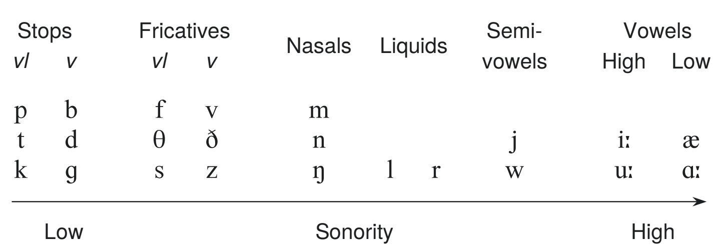


Figure 2.2: Sonority scale for English speech sounds (after Giegerich 1992: 133).

strength of the model is the fact that it generates predictions about the distribution of target and non-target forms in learner speech both at a single point in time and across time. The notion of favorable environments and the two-stage implicational nature of IL development can therefore be applied to account for within-structure and developmental variation.

Model of Sociolinguistic Variation (MSV)

Research in the Labovian tradition typically takes a multifactorial approach to IL sound patterns, where the distribution of discrete choices of forms is determined as a function of different predictors.[24] Usually, internal and external sources of variability are distinguished, which broadly refer to linguistic and social factors, respectively.

[24] Historically, the primary analytical tool is variable rule analysis (VARBRUL), a procedure that jointly considers a range of factors and gauges their association with the variation observed in the structure of interest. The notion of a variable rule, which was first introduced by Labov (1969), captures the systematic nature of language variation in the form of conditional probabilities for a set of alternative forms.

The Model of Sociolinguistic Variation (MSV) for native and non-native speech was proposed by Fasold & Preston (2007). The MSV holds that the choice among variants is probabilistically related to factors at three levels of variation: (i) social context, (ii) linguistic context, and (iii) time.

The social context includes characteristics of the communicative situation, gender of the speaker, extent of L1/L2 use and social identity. Stylistic variation induced by focus on form and formality is also located at this level, with speech styles that allow a speaker to monitor their speech yielding more target-like utterances. Among the linguistic factors in IL phonology are phonetic and prosodic context effects. Factors such as the voicing of the adjacent segment(s) or prosodic position may be systematically related to the preference for a specific variant.[25] The third level of variation, time, refers to the relative time of acquisition of a particular variant. Structures that are acquired early are more entrenched and automatized, and can thus be accessed more easily relative to variants that are acquired at a later stage or, from the perspective of the learner, have been acquired more recently. The relative time of acquisition is thus also assumed to exert influence on the probabilistic weight of variants.

[25] Studies on final /t d/ deletion in Chinese learners of English, for instance, have noted similar phonetic constraints in learner and native speech (Wolfram 1985; Bayley 1996; Hansen Edwards 2001).

As our discussion of theoretical approaches to L2 phonology has shown, different constraints have been postulated to account for non-native speech patterns. These include L1 transfer of articulatory and perceptual routines, functional status, perceived degree of (dis)similarity, markedness, articulatory constraints, speaking style, frequency, language universals, natural processes, and the sonority of adjacent sounds. A strength of the Labovian perspective is its traditionally multifactorial approach to (learner) speech. This has two advantages. With non-native speech being the result of a variety of linguistic and non-linguistic influences, a multifactorial approach does justice to the complex ensemble of processes underlying learner productions. Importantly, this allows the researcher to compare the relative weight of factors, and thus to compare, in a statistical sense, the explanatory power of constraints. Further, the connection to the field of sociolinguistics discloses a rich inventory of systematic sources of variability that are documented for native speech and which may also be operative in learner speech. The sociolinguistic approach thus not only contributes methodologically but also conceptually, as it puts variation at center stage and seeks to uncover its systematic components.

2.4.2 *IL development*

What Corder (1981) labeled vertical variation refers to developmental change in learners over time. Theories with a developmental focus aim to account for the systematic patterns that are observable in this dimension of IL variation. Three contributions will be discussed in this section.

Linguistic theory of L2 phonological development

James' (1988) Linguistic Theory of L2 Phonological Development (LTD) focuses on the interplay of different levels of phonological representation in the course of L2 development. The LTD is explicit about the assumed representation of phonological structure in the learner's grammar. Three levels of phonological representation – lexical, prosodic, and rhythmic – are posited to interact systematically over the course of L2 phonological development. A non-linear framework similar to metrical phonology is employed for the representation above the segmental level (see James 1986 for details). The prosodic level comprises seven hierarchical layers (cf. Figure 2.3). Through *s*/*w*-marking, binary strength values (strong vs. weak) are assigned to constituent nodes at each layer. These add up to determine the structural weight, which in turn conditions the degree to which contrastive features at the lexical level are implemented.

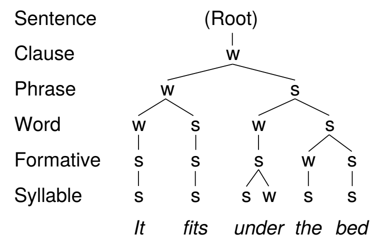


Figure 2.3: Supra-syllabic *s*/*w*-marking at five hierarchical levels. The *s*-marks add up to determine the structural strength of a particular structure (James 1986, 1988).

Sub-syllabic strength marking determines the realization of lexical phonological properties and specifies the scope of syntagmatic segmental influence, with *s*-marked units dominating adjacent *w*-marked units. Thus, *w*-marked segments are more likely to be the targets rather than the triggers of connected speech processes such as coarticulation or assimilation. Figure 2.4 illustrates sub-syllabic strength marking, which specifies strength asymmetries within syllables (onset > nucleus, coda) and clusters (initial > non-initial). Table 2.7 summarizes strength oppositions at the level of the syllable, set and segment.

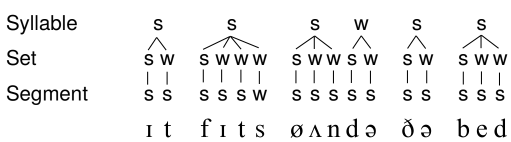


Figure 2.4: Sub-syllabic *s*/*w*-marking (James 1986, 1988).

The acquisition of L2 phonological representation is assumed to follow a universal sequence from the lexical to the prosodic to the rhythmic level, with the lower levels providing substance for the acquisition of higher-level structure. The notion of structural strength is employed to make predictions about within-structure variation, which is assumed to be

Level	Strong (*s*-marked)		Weak (*w*-marked)
Syllable	Stressed	>	Unstressed
Set	Onset	>	Coda
Segment	Initial cluster element	>	Non-initial cluster element

Table 2.7: Sub-syllabic strength marking in the LTD.

sensitive to prosodic representation. Assuming a learner has established prosodic *s*/*w*-marking in sufficient depth, James (1986) hypothesizes a tendency toward strengthened variants in *s*-marked and weakened variants in *w*-marked positions. The likelihood of strengthening/weakening is assumed to depend on the relative degree of dominance, with slots that are maximally *s*-marked being most likely to trigger strengthening and vice versa. With this type of segmental variation is also sensitive to prosodic embedding, learners at initial stages are not expected to show such systematic alternations.

At the segmental level, James (1988, 1990) makes a (somewhat idiosyncratic) distinction between *variability* and *variation*. *Variation* describes the deviation of IL phones from the target and is assumed to decrease steadily over the course of acquisition (see Figure 2.5). In other words, production accuracy increases over time. *Variability*, on the other hand, describes the extent to which different IL phones are used as realizations of a single TL category. Initial stages are assumed to show little variability – that is, stable L1-derived substitutes. As further variants enrich the learner's repertoire, variability increases and then decreases again toward the final stages. Crucially, variability grows increasingly systematic over the course of acquisition. At later stages, IL phones are posited to show sensitivity to phonetic and prosodic context.

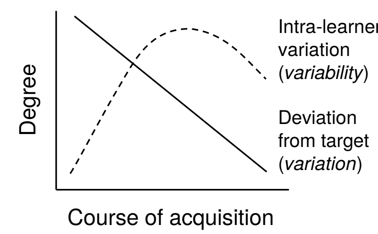


Figure 2.5: The extent of "variability" and "variation" of phones during the course of L2 acquisition; schematic illustration of the patterns predicted by the LTD (James 1988, 1990).

A unique feature of the LTD is its focus on the interdependence between different levels of phonological representation. It thereby offers a framework for the study of IL as a coherent system rather than a collection of independently acquired structures and advances factors that are operative across levels. In particular, its notion of structural strength identifies a constraint for the explanation and prediction of within-structure variation. To generate predictions at the segmental level of analysis, strength values can be determined according to the asymmetries shown in Table 2.7. Further, the bottom-up advancement of L2 phonological acquisition generates predictions about sound patterns at subsequent developmental stages. In general, strength and rate asymmetries grow during IL development. With structural strength being cumulative, an increasing *s*/*w*-differentiation is expected to yield a gradual increase in strength effects in syllables and segments.

An important strength of the theory is its predictive adequacy, as it generates quantitative hypotheses about IL variation.[26] While James (1988) invokes a metrical model of phonological representation, the fundamental assumptions of the LTD are in fact independent of any particular scheme.[27]

[26] See Sönning (to appear) for an application to the L2 acquisition of speech rhythm.

[27] For instance, its claims are also testable using simplified templates of prosodic organization such as a two-way distinction between lexical and syntactic stress.

Ontogeny Phylogeny Model (OPM)

Another model primarily concerned with IL development is the Ontogeny Phylogeny Model (OPM) proposed by Major (2001). The OPM merges theoretical insights into transfer, similarity, and typological markedness into a single coherent framework, which outlines the dynamic nature of IL development. The OPM rests on two basic assumptions: (i) a learner's IL consists of three types of structural components: L1, L2 and U; and (ii) the relationship between these components changes systematically over time.

Thus, every structure found in learner speech is attributable to one of three sources:

- L1: It has been transferred from the native language.
- L2: It is a target language structure.
- U: It is a universal structure that is not part of L1 or L2.

Major (2001: 83) takes a broad view of universals; U is defined as "the universal set of properties of the human language capacity and the resulting universal characteristics of languages. In addition to abstract linguistic constructs, U includes anatomical, functional and processing properties of the human mind". The OPM states that the interplay of L1, L2 and U during IL development depends on the (i) type of structure that is acquired and (ii) speaking style. The basic chronological assumption is that, over the course of acquisition, the presence of transferred L1 structures decreases, while L2 (TL) structures increase. The influence of universal language properties first increases and then decreases. This is shown graphically in Figure 2.6a, where the vertical scale denotes degree of influence of the respective component.

The OPM further states that normal, similar and marked structures show different trajectories; these are shown in Figure 2.6a, b and c. Normal structures are acquired relatively early – hence, there are only 5 stages. While transfer dominates initially, its influence decreases quickly. Target language structures, which are absent at stage 1, increase. At the hypothetical final stage of SLA, the IL resembles the TL. The role of U increases, reaches a maximum at stage 3, and then decreases again.

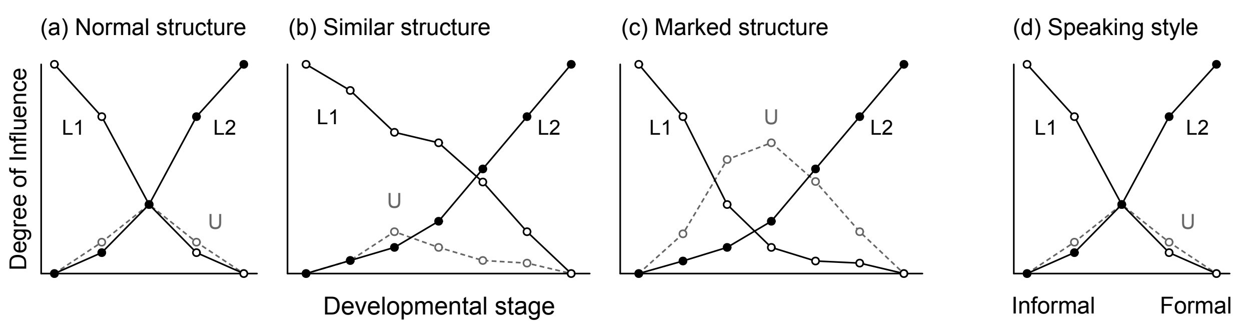


Figure 2.6: Graphical illustration of the corollaries of Major's (2001) OPM: The relative weight of L1, L2, and U interlanguage components in the course of acquisition of (a) normal, (b) similar, and (c) marked structures. Panel (d) shows variation according to the formality of the situation.

Compared to normal structures, similar ones are acquired at a slower rate, which is reflected in the number of stages (7 vs. 5). Likewise, marked structures are acquired at a slower pace. While equivalent L2 trajectories are posited for similar and marked structures, the relative influence of L1 and U differs. In the acquisition of similar structures, L1 transfer is persistent; U, on the other hand, exerts no notable influence. For marked structures, on the other hand, transfer is assumed to decrease rapidly, while U rises to exert considerable influence.

The OPM also takes into account stylistic variation (see Figure 2.6d). In casual speech with little focus on form, L1 influence is more likely to surface. In formal situations, on the other hand, transfer is least influential and L2 structures dominate. In between, U increases and decreases, reaching a maximum in stylistically neutral situations.

The OPM differs from previously discussed models in several respects. Its scope is more comprehensive since it incorporates various findings of L2 phonology research: transfer, universals, similarity, and markedness. The OPM thus brings together well-documented linguistic constraints and hypothesizes about how they conspire to shape non-native speech. The effect of transfer and universals is harmonized by assigning different weights to these factors, conditional on developmental stage. Whether traces of L1 influence or language-universal biases are observable thus depends on the proficiency level of the learner.[28]

The developmental arrangement of L1, L2, and U is further assumed to vary as a function of markedness and similarity attributes of the TL structure. Speaking style is included as a source of variation and posited to alter the relative weight of the structural components at every stage of IL development. A strength of the model is the fact that it offers a comprehensive and coherent account of IL sound patterns. While its empirical foundation is predominantly concerned with segmental phenomena, the OPM's predictions are posited to hold for other types of phonological structure and indeed other areas of (second) language acquisition and language change. The model also captures different types of IL variation, including within-structure, between-structure, and developmental variation. Extrapolating its claims to the notion of variable productions in learner speech, the developmental and stylistic patterns in Figure 2.6 can be translated into probabilistic predictions about the distribution of L1-, L2-, and U-derived variants at a single point in time. The model's broad definition of U leaves it to the researcher to determine in which way U might influence the acquisition of a particular structure.[29]

We now turn to the last framework in our survey of the theoretical literature: Dynamic Systems Theory, which is rooted in the Complexity Theory paradigm.

[28] See Sönning (2014, to appear) for an application to the acquisition of vowel reduction and speech rhythm by German learners of English.

[29] It should be noted that the model has also met with criticism due to its vague definition of universals and lack of specificity regarding the precise interplay of L1, L2, and U over time (e.g. Gut 2009: 27).

Dynamic Systems Theory (DST)

Complexity theory originated in the natural sciences and was introduced into SLA theorizing by Larsen-Freeman (1997). ILs exemplify complex systems in that they comprise a large number of interacting components, with the behavior of the whole emerging from this interplay. The isolated study of these components is considered inadequate and it is argued that a dynamic view of language acquisition may be more appropriate as it embraces the complex nature of ILs and discourages reductionist explanations. This stream of SLA theorizing is currently referred to as Dynamic Systems Theory (DST, De Bot et al. 2007).

In general, a system can be regarded as an ensemble of cooperating parts that form a whole. The core characteristic of a *dynamic* system is change over time; DST is thus an inherently developmental theory. As such, dynamic systems exhibit a number of distinctive properties. First, their development is dependent on initial conditions, with differences in the initial state possibly having far-reaching consequences (De Bot & Larsen-Freeman 2011). A further characteristic is the complete interconnectedness of its components. Changes in one component may impact other parts of the system. Development is assumed to be non-linear with fluctuating learning curves reflecting progress and backsliding.

Changes in the system and its subcomponents are assumed to be driven by two forces: internal reorganization and interaction with the environment. Over time, the system is drawn toward attractor states, which are preferred states resulting from developments in the system. These can be considered natural or unmarked states (Larsen-Freeman 1997). The strength of attraction determines the amount of energy needed to set the system in motion to settle in another temporary attractor state. Fossilization is argued to reflect such an attractor state (Larsen-Freeman 2005). Repeller states, on the other hand, reflect disfavored conditions and are thus unlikely to be observed. While individual developmental paths may vary substantially, there will be considerable similarities between them, which may arise from limitations in change and variability (De Bot et al. 2007).

According to DST, IL sound patterns emerge from the interaction of a complex system comprising a large set of interacting components. The theory introduces a number of concepts that describe characteristics of dynamic systems. As these terms and ideas are borrowed from complexity theory, it is not immediately clear how they can be mapped onto linguistic phenomena. In fact, from the perspective of L2 phonology, DST may be considered a theory-neutral framework, offering terminology and concepts rather than substantive claims that directly translate into research hypotheses. Nevertheless, DST encourages researchers to emancipate from a reductionist approach to isolated structures toward more comprehensive accounts of the dynamic interplay between different IL phenomena.[30]

[30] It should be noted, however, that Major's (2001) OPM and James' (1988) LTD are also genuinely dynamic frameworks. Both models focus on the interplay or interaction between different factors or levels of representation and make specific statements on the nature of systematic changes over time. In contrast to DST, both models operate on the basis of linguistic constraints and thus invoke factors and mechanisms that are well-documented and familiar to SLA researchers. Their tenets can thus be more readily applied in empirical work.

2.5 *Summary*

Our current understanding of non-native speech receives support from a variety of theoretical contributions. These differ in scope and have brought forward various constraints to account for IL phenomena. The purpose of this section is to provide a comparative summary of these accounts and to highlight unifying themes and implications for empirical work. Table 2.8 provides a summary and recapitulates the linguistic factors that are put forward as well as the predicted areas of difficulty and expected types of deviant productions. Further, the overview aims to show how these pieces of theoretical knowledge relate and contrast. To this end, four aspects are highlighted:

- **Theme.** The three major themes that organized this chapter provide a bird's-eye perspective on the type of reasoning adopted. An important distinction is that between cross-linguistic and universal accounts. While the former focus on L1 interference as an explanation of IL phenomena, the latter highlight the relevance of constraints that operate independently of the L1 and TL. The distinctive feature of the third group is the emphasis on variation, especially in terms of IL development.
- **Mode.** Theories also differ in the importance they attribute to perception

and production constraints. While a dominant stream of research is exclusively concerned with speech perception, other accounts restrict their focus to speech production without reference to perception. Finally, a considerable number of contributions take into account both modes, assuming that both channels are constrained in systematic ways.

- **Scope.** An aspect that is particularly relevant to the empirical application of theoretical accounts is their scope – specifically, the types of structures they are concerned with. A general distinction is that between segmental and suprasegmental structures. With most contributions extending to both levels of analysis, theoretical work can provide guidance in a variety of research contexts. It is worth noting, however, that perceptual models are exclusively devoted to the segmental level of analysis.
- **Variation.** A further distinction that is fruitful for empirical research concerns the type of variation that is predicted and explained. As illustrated above, the nature of the phenomenon of interest suggests a division into variation between and within learners and structures. Between-structure perspectives allow the researcher to formulate hypotheses about the relative degree of difficulty associated with a given structure. This information may be based on perceptual similarity assessments or intrinsic properties of the structure such as the relative degree of markedness or articulatory complexity. The between-structure point of view thus translates into predictions about the relative accuracy rate of a given structure. Within-structure variation, on the other hand, is concerned with the variable accuracy of a given structure, generating predictions about production rates in certain contexts. These can be defined in linguistic terms by focusing attention on phonetic and prosodic constraints. An awareness of these different sources of variation in learner speech allows for a systematic investigation of IL phenomena, which can build on a theoretically grounded set of constraints that are either intrinsic to a structure (foregrounding between-structure comparisons) or extrinsic (encouraging within-structure comparisons).

Table 2.8: (Opposite page) Overview of theoretical contributions.

A final point that should be noted is the indispensability of assumptions when contemplating the implications of theoretical accounts on empirical work. As the discussion of contributions has illustrated, predictions often hinge on one or several premises regarding the underlying linguistic constraints. Predictions on the basis of the SLM, for example, rest on the identification of the L1 sound closest to the focal TL segment as well as the degree of perceived phonetic dissimilarity between the two. A Natural Phonology approach to IL, on the other hand, requires the researcher to make assumptions about which natural processes may affect the acquisition of a particular structure. Similarly, taking an MDH perspective necessitates the evaluation of the degree of markedness associated with a particular structure or context. In consequence, it is critical for researchers to clearly formulate assumptions that perform a bridging function between theory and empirical predictions.

		Mode			Scope		Variation			
Contribution	Reference	Perception	Production	Linguistic factors	Segmental	Prosodic	Between	Within	Development	Stylistic
Cross-linguistic influence										
Contrastive Analysis Hypothesis	Lado 1957		●	Contrasts; functional status of structures in L1/TL	●	●	●			
Articulatory Settings	Honikman 1964		●	Contrasts in articulatory settings	●		●			
Speech Learning Model	Flege 1995	●	○	Perceived phonetic dissimilarity	●		●		●	
Similarity Differential Rate Hypothesis	Major & Kim 1996	●	○	Perceived phonetic dissimilarity	●		●		●	
Perceptual Assimilation Model	Best 1995	●	○	Perceived gestural similarity; frequency; functional load	●		●			
Desensitization Hypothesis	Bohn 1995	●	○	Perceptual bias toward vowel length	●		●			
Perceptual Interference Model	Bohn 1995	●	○	Presence/absence of phonological features in L1	●		●			
Feature Competition Model	Brown 1998	●	○	Prominence of phonological features in L1	●		●			
Universals										
Markedness Differential Hypothesis	Hancin-Bhatt 1994		●	Contrasts; relative degree of markedness	●	●	●	●		
Structural Conformity Hypothesis	Eckman 1977	●	●	Universal generalizations	●	●	●	●	●	●
UG Model of Stress Acquisition	Eckman 1991		●	Contrasts and UG constraints in parameter settings		●	?	?	●	
Natural Model of L2 Phon. Acquisition	Dziubalska-Kołaczyk 1990	●	●	Natural processes; status in L1 vs. TL; frequency	●	●	●	●	●	
Naturalness Differential Hypothesis	Schmid 1997		●	Contrasts in the status of natural processes	●	●	●	●	●	
Functional Model of Phon. Acquisition	Boersma 1998	●	●	Articulatory constraints; communicative demands; freq.	●	●	●	●	●	●
Model of Segmental Acquisition	Colantoni & Steele 2008	●	●	Articulatory constraints; perceived phonetic dissimilarity	●		●	●	?	
Developoment and variation										
Gradual Diffusion Model	Gatbonton 1978		●	Sonority of adjacent segment(s)	●			●	●	
Sociolinguistic Model	Fasold & Preston 2007		●	Phonetic and prosodic context in general	●	●	○	●	●	●
Linguistic Theory of L2 Phon. Development	James 1988		●	Structural strength; level of representation	●	●	●	●	●	
Ontogeny Phylogeny Model	Major 2001	●	●	Contrasts; markedness; similarity	●	●	●	●	●	●
Dynamic Systems Theory	De Bot et al. 2007	?	●		●	●	●	○	●	

Note. Key to symbols: ● explicitly stated, ○ implied, ? unclear

3
Method and data

This chapter provides details about methods and data.[1] Sections 3.1 and 3.2 give an overview of the materials, procedures, and participants. §3.3 describes supplementary data sets used in this study. These include additional data on German Learner English from previous research and empirical evidence on English L1 acquisition and English phoneme frequencies. §4.6 gives an outline of key aspects of the statistical analysis.

[1] The data sets analyzed in the following chapters are openly available via the *Tromsø Repository of Language and Linguistics* (*TROLLing*). Links to the repositories are provided in the method and data section of each chapter.

3.1 Materials and procedures

Three reading tasks were used for data elicitation:

- a word list (n = 42 items),
- short phrases (n = 33 phrases, each including 2 target words), and
- mini-dialogues (n = 10).

The materials used are listed in Appendix A.1.1. Recordings took place in a quiet location using a portable H2 Zoom device. Prior to the recordings, participants were given time to familiarize themselves with the materials, which were printed on cards. All speakers were asked to read at a comfortable pace and were given the opportunity to correct themselves. In terms of vocabulary and syntax, the stimulus design was based on the level of English after one year of instructed learning.[2]

[2] To this end, the school book *Green Line New 1* (Hellyer-Jones et al. 2003) was consulted to ensure familiarity with lexical and grammatical structures.

The tasks were designed to allow for an analysis of segmental and suprasegmental aspects of speech production. The objective of the word list was to elicit a number of segmental features. Items were divided into two sets in order to separate a number of minimal pairs (e.g. *plays-place*, *bag-back*). This was in many cases necessary in order for learner productions to be assigned to the correct target item if no difference between the renditions was discernible. A partition of the word list also ensured that minimal pair constituents did not occur in proximity, as this might have influenced their pronunciation.

The purpose of the short phrases task was to elicit 11 English monophthongs /iː ɪ e æ ʌ ɜː ɑː ɒ ɔː ʊ uː/, which were used to map each speaker's acoustic vowel space. The task was designed to yield 5 to 7 tokens per monophthong. All items (n = 2 per phrase) occurred in the same contrastive context: *I said* "…", *not* "…". The choice of items struck a balance between using familiar material and using contexts that exhibit minimal coarticulatory influence of adjacent consonants. A classic context for the elicitation of canonical vowels is the neutral hVt or hVd frame (e.g. *heat, hit, hat*; Peterson & Barney 1952). In the current study, unfamiliar hVd- and hVt-words were elicited by embedding them as the second item in short rhymes so that their pronunciation could be inferred from the first (e.g. *I said "pet", not "het".*). Overall, there were three different types of short phrases:

- $n = 12$ contrasts of familiar words (e.g. *I said "sit", not "sat".*)
- $n = 13$ rhymes of familiar words (e.g. *I said "bed", not "head".*)
- $n = 8$ rhymes including nonce words (e.g. *I said "bought", not "hought".*)

The short dialogues consisted of question-answer sequences, which aimed at eliciting consistent accentual patterns. Subjects read both turns of the dialogue and were given the opportunity to re-read a sequence if they were unsatisfied with their rendition.

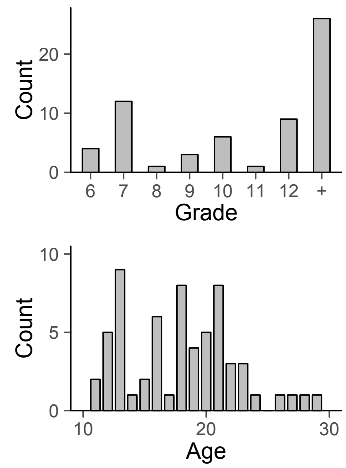


Figure 3.1: Age distribution of the $n = 62$ German learners (bottom) and distribution across grades (top; '+' denotes completed secondary education) and age.

3.2 *Subjects*

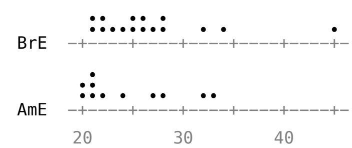
A total of $n = 89$ subjects were recorded in the present study:

- $n = 62$ German learners (39 female),
- $n = 11$ AmE native speakers (4 female), and
- $n = 16$ BrE native speakers (11 female)

Before the recordings, subjects filled out a questionnaire eliciting biographical information.[3] All British English (BrE) and American English (AmE) informants reported that both of their parents' native language was English. Neither group can be considered as representing a well-defined variety of English, as resource limitations made it unfeasible to consider exclusionary variables for participant recruitment. As the reported minimal education level of all subjects was a bachelor's degree, however, the sample of native speakers recorded in the present study may be described as speaking an educated variety of English. BrE informants were between 21 and 45 years of age[4], and most had grown up in the London area and the Midlands. Native speakers of AmE were between 21 and 33 years old and were predominantly from the northeastern part of the US.

[3] The form used for the native speakers was based on the *Bamberg questionnaire for lexical and morphosyntactic variation in English* (Krug & Sell 2013). All questionnaire forms are available from the OSF.

[4] The age distribution for the two native speaker groups is given below:



The learners recorded for the present study are best described as a convenience sample. The aim was to capture a range of proficiency levels that may be considered as representative of instructional-setting learners. Figure 3.1 gives biographical details about the $n = 62$ learners. All informants had learned English at school and reported that both of their parents spoke German as a native language. Learners ranged from grade 6 (having completed at least 1 year of formal English instruction) to university level. Subjects were also asked to indicate where they had gone to primary school. Figure 3.2 shows that most of the informants grew up in northern Bavaria (Bamberg), where the recordings took place. The age of the subjects ranged from 11 to 30 (*Mean* = 18) and the AOL, the average age at which English instruction began, was 10 (*SD* = 1.5).

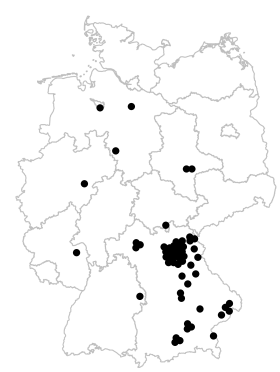


Figure 3.2: Regional background of the learners in this study. The map was drawn using the R package *tmap* (Tennekes 2020) using data from the GADM database.

With comparisons between different proficiency levels being a key concern of the present study, a foreign accent rating (FAR) was used to obtain a global accentedness score for each learner (see Jesney 2004 for an overview). To this end, four mini-dialogues (dialogues 2, 7, 9, 10; see Appendix A.1.1) spoken by each learner were presented in randomized order to two BrE native speakers, who rated the overall degree of foreign accentedness on a 12-point Likert scale (from "strong foreign accent" to "native speaker level"). The software *OpenSesame* (Mathôt et al. 2012) was used for the implementation of the FAR task and ratings were given via the keyboard. The scores for the four stimuli were averaged for each subject. The distribution of the average ratings is shown in Figure 3.3. The learners are distributed across the entire scale and therefore represent a broad range of proficiency levels (*Mean* = 6.0, *SD* = 2.5, *Min* =1.6, *Max* = 10.8).

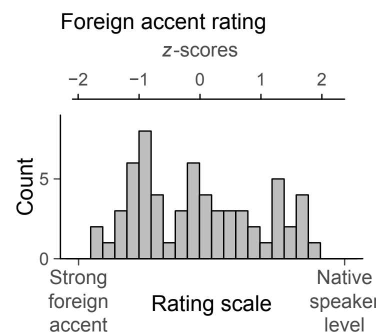


Figure 3.3: Results of the foreign accent rating: Original scale and *z*-transformed ratings.

For the analyses in the present study, the average ratings were converted to *z*-scores (i.e. standardized to have mean 0 and standard deviation 1). These will be used hereafter to express L2 pronunciation proficiency. For purposes of exposition, the discussion of findings will usually refer to three proficiency levels: beginner, intermediate, and advanced. These labels correspond to the following ranges of *z*-scores:

- Beginner: *Range* [−2, −1], *Mean* = −1.5 (*n* = 12 learners)
- Intermediate: *Range* [−1, +1] , *Mean* = 0 (*n* = 37 learners)
- Advanced: *Range* [+1, +2] , *Mean* = +1.5 (*n* = 13 learners)

For some of the structures investigated in the present study, the two standard varieties offer different targets. Thus, allophones may differ in distribution (e.g. for rhotics and laterals) and/or acoustic features (e.g. vowel quality and duration). In such cases, German learners with an orientation toward BrE or AmE should be treated separately since they form two different populations of speakers. Further, they must be compared to different reference values – those for BrE and AmE. We therefore assigned to each learner a variety score that ranges from –1 (indicating perfect alignment with BrE) to + 1 (AmE). To determine these scores, five segmental features that contrast between the standard varieties were selected and analyzed:[5]

[5] Some features could not be measured: Yod-dropping (*news, student*) and the /æ/ vs. /ɑː/ contrast (*path, dance*) did not occur in the recorded material. *T*-flapping (as in *better, writer*), on the other hand, could not be analyzed since, in the population of German learners represented by our sample, [ɾ] may surface due to L1 transfer: Like AmE, the Franconian accent shows *t*-flapping.

- The realization of laterals ([l] vs. [ɫ]) as in *lunch, willing*
- The LOT vowel: /ɒ/ vs. /ɑː/ as in *stop, pot*
- The GOAT vowel: /əʊ/ vs. /oʊ/ as in *show, alone*
- The NURSE vowel: /ɜː/ vs. /ɝː/ as in *bird, urge*
- Post-vocalic /r/ as in *year, tour*

In Appendix A.2, we describe how these these features were combined into a variety score for each learner. Figure 3.4 shows the distribution of scores by proficiency. At upper-intermediate levels (around $z = +0.5$) learners begin to form two groups, which become clearly distinct at advanced proficiency levels. For the analysis of features where BrE and AmE offer different targets, we will choose $z = 0$ as the point from which on we begin to describe two different populations of learners. This means that beginner and lower-intermediate learners are treated as a homogenous

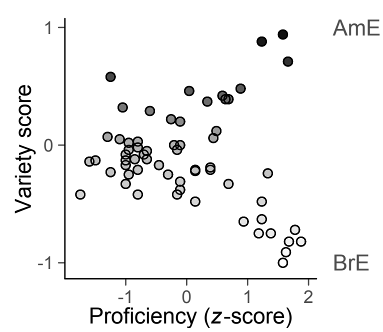


Figure 3.4: Variety score by proficiency level.

group and described with, say, a single trend line. For upper-intermediate and advanced learners, this trend line is allowed to branch out to trace the behavior of the two populations that gradually emerge. In Figure 3.4, grey shading is used as an additional indication of the target accent: Black signals AmE, white BrE. This visual cue will be encountered again in later chapters, where it allows us to enrich graphs with information about the target variety.

3.3 Supplementary data

This section introduces supplementary data used throughout the following chapters. These include data from previous work on German Learner English which were reanalyzed for the purposes of the present study (§3.3.1). §3.3.2 describes the set of studies used for a survey of the L1 acquisition of English consonants, and §3.3.3 offers a review of studies on English phoneme frequencies.

3.3.1 German Learner English

Our analyses of segmental features in GLE make use of supplementary data sets that were taken from the literature or kindly provided by other researchers.

Pascoe (1987)

Pascoe (1987) carried out an extensive auditory analysis of the productions of $n = 26$ instructional-setting learners from southern Bavaria (age: 14–16 years). Informants were sampled from grade 8 to 10 from different school types in the area of Munich and had been learning English for approximately 3 to 6 years. The analysis was based on semi-free speech: the retelling of a picture story and a guided telephone conversation. Detailed phonetic transcriptions are provided for all subjects, and we will refer to this data base as the *SpIL* corpus.[6] For the purpose of the present study, these data were reanalyzed with respect to factors not considered in the original investigation.[7] In order to establish comparability to the present study, the L2 pronunciation proficiency of the subjects was determined based on the overall accuracy rate of the 49 segmental and suprasegmental parameters reported by Pascoe (1987). These scores ranged from .70 to .95. For use in the present investigation, it was assumed that these figures give a comparable reflection of pronunciation proficiency. For the analysis, these were converted to *z*-scores. Figure 3.5 shows the correspondence between these and the original proportions.

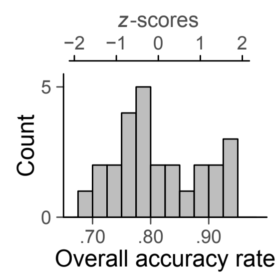


Figure 3.5: Correspondence between the overall accuracy rate reported by Pascoe (1987) and the FAR scores assigned for purposes of comparison.

[6] The research project was titled *Spoken Interlanguage*.

[7] To this end, the phonetic transcriptions, which are provided in the appendix of Pascoe (1987), were searched manually for relevant occurrences.

Wunder (2012)

A study[8] by Wunder (2012) investigated the production of coda obstruents by $n = 30$ German learners (age: 11–20; 15 female) from a grammar school in northern Bavaria (Bamberg). Subjects were sampled from grades 6, 9, and 12 ($n = 10$ each) and a reading passage was used for data elicitation

[8] These data were collected for an MA thesis at the University of Bamberg.

(see Appendix A.1.2). These recordings were kindly provided by the author for further analysis. To assess the proficiency of the informants, I rated their productions on a scale from 1 to 12. While these ratings are not directly comparable to those from native speakers, this serves to establish at least some degree of comparability between the data sets.

Rank (2016)

The third supplementary data are from a study[9] on /v/- and /w/-production by $n = 26$ instructional-setting learners. Subjects were mostly from southern Germany (Bavaria) and had learned English in an instructional context for at least 7 years. The sample may be considered as representing intermediate to advanced proficiency levels. Two tasks were used for data elicitation, a word list and a reading passage (see Appendix A.1.3). The author kindly provided her primary data for reanalysis. The original recordings were not available.

[9] These data were collected for a term paper supervised by the author of the present study.

Rank (2018)

The final supplementary data are from a study[10] on the production of postvocalic /r/ by $n = 21$ advanced learners. Most of the subjects were from southern Germany (Bavaria) and studied English at the University of Bamberg. Three tasks were used for data elicitation: a word list, a reading passage, and a casual conversation with the author of that study (see Appendix A.1.4). The primary data (but not the original recordings) were made available for reanalysis.

[10] These data were collected for a BA thesis co-supervised by the author of the present study.

3.3.2 English L1 acquisition

Our review of L2 phonological theories in Chapter 2 revealed that the L1 phonological development of English-learning children may inform L2 speech research. Relevant insights include the sequence of acquisition of segmental structures as well as typical error patterns. A number of large-scale, cross-sectional analyses have shed light on the L1 acquisition of English consonants. Attributes of these studies[11] are listed in Table 3.1. We observe differences in sample size, age range, and the positions investigated. Each provides detailed information about individual segments, sample sizes, and accuracy rates, which allows us to combine the information in a joint analysis.[12]

The estimates from this survey are shown in Figure 3.6. The plotted values are thresholds and indicate the age at which a certain proportion of

[11] The data are available on the OSF (https://osf.io/yv572/).

[12] We ran a segment-by-segment series of beta-binomial regression models using the *R* package *aod* (Lesnoff & Lancelot 2019). Each model estimated, separately for initial and final position, the average trend line across the four studies, i.e. the increase in accuracy rate with age. The code can be found in the OSF project (https://osf.io/sk5wd/).

Study	*N*	Position	Age range	Place
Wellman et al. 1931	204	I, M, F	2;0 – 6;0	US (Iowa)
Templin 1957	480	I, M, F	3;0 – 8;0	US (Minnesota)
Prather et al. 1975	147	I, F	2;0 – 4;0	US (Oregon)
Smit et al. 1990	997	I, F	3;0 – 9;0	US (Iowa, Nebraska)

Table 3.1: Studies on English L1 acquisition. Positions refer to I(nitial), M(edial), and F(inal). *N* denotes the number of children analyzed.

children produce the sound accurately (cf. Amayreh & Dyson 1998):

- Age of customary production: .50 (lower end of the bar)
- Age of acquisition: .75 (dot)
- Age of mastery: .90 (upper end of the bar)

Points sitting on the horizontal axis indicate that a threshold is reached by the age of 2. Lines without cross-bars at the upper end indicate that mastery (i.e. accuracy rates of .90 or higher) is expected at the age of 9 or higher.

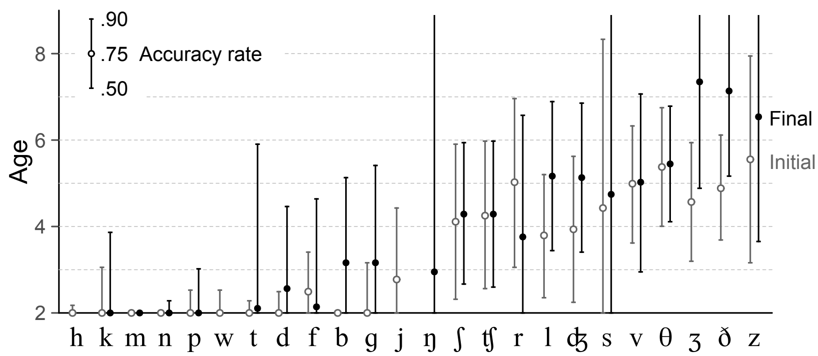


Figure 3.6: Age and order of acquisition of consonants in English L1 acquisition. Results from a survey of 4 studies (see Table 3.1 for details). The points denote the age of acquisition, where .75 of children produce the sound correctly (grey open circles: initial position; filled circles: final position). The lower end of the bar indicates the age of customary production (.50), the upper end the age of mastery (.90).

3.3.3 *English phoneme frequencies*

A factor highlighted by several theoretical contributions is input frequency. To estimate the rate of occurrence of English phonemes in the input, findings from $n = 9$ studies dealing with different genres were combined.[13] Details about these are summarized in Table 3.2.

[13] The data can be accessed via the OSF (https://osf.io/d8zjy/).

Study	Variety	Genre	*N*
Denes 1964	BrE	Teaching material	72,000
Fry 1974	BrE	—	—
Dewey 1923	BrE, AmE	Prose	370,000
Hayden 1950	AmE	University lectures	65,000
Delattre 1965	AmE	Prose, drama	6,500
Roberts 1965	AmE	Basic writing vocabulary	67,000
Mines et al. 1978	AmE	Interview	104,000
Blumeyer 2012	AmE	BNC	—
Carterette & Jones 1974	AmE	Natural speech	252,000

Table 3.2: Studies on English phoneme frequencies. *N* refers to the number of phoneme occurrences.

As the use of symbols for vowel categories was too variable to yield reliable comparisons, the analysis is restricted to consonants. Figure 3.7 summarizes the distribution of the log-transformed frequencies (per 100 segments) across the studies. Overall, results are in fairly good agreement.[14] The boxplot at the right margin, which is highlighted with a grey fill, summarizes the distribution of the median frequencies (i.e. the

[14] There is considerable variation in /r/, which is presumably due to the pooling of rhotic and non-rhotic varieties. Variation is also evident for /j/, which might be attributable to genre-related variation in rate of the high-frequency pronouns *you* and *your*.

filled circles in the display). The distribution is symmetric on the log scale. The overall median frequency is 1.9 per 100 segments and the interquartile range extends from 1.0 to 3.2; the extremely infrequent phoneme /ʒ/ is a clear outlier (*Mdn* = 0.1 per 100 segments). For the purposes of the present study, the frequency of a segment will be assessed by comparison to the log frequency distribution described by the boxplot at the right margin of Figure 3.7.

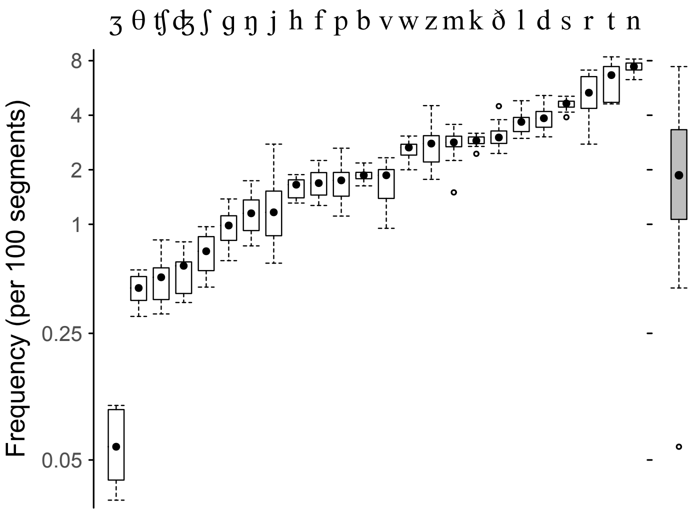


Figure 3.7: Comparison of estimates for the frequency of English consonant phonemes across 9 studies. The boxplot at the right margin summarizes the distribution of the estimated medians.

3.4 *Data analysis*

This section gives an outline of the methods used for data analysis. §3.4.1 starts with a discussion of hierarchical data structures and the implications for analysis. We then move on to a discussion of key features of Bayesian inference (§3.4.2), the paradigm we adopt for statistical analysis. §3.4.3 then outlines the strategies we use for the specification of statistical models.

3.4.1 *Hierarchical data structures and multilevel models*

Empirical research in the field of L2 phonology typically involves multiple observations per subject. To illustrate, in a study on the production of final voiced obstruents each speaker may produce several items and thus contribute multiple data points. This sampling situation is shown schematically in Figure 3.8, where observations (•) are nested within subjects (s). In this case, there are $n = 5$ subjects, who produced $n = 4$ items each.

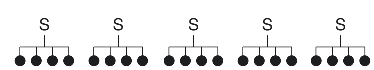


Figure 3.8: Illustration of hierarchically structured data, where observations are nested within subjects.

When drawing conclusions from these data, it is important to take into account this hierarchical structure. In simple settings such as the one shown in Figure 3.8, there are two levels at which inferences can be made. The first level is that of the subject. In this example, each subject has a certain ability to produce FVOs, which we may express as a percentage.

At this level, a sample of size $n = 4$ is used to draw inferences about the subjects' accuracy rate. At the second level, the subjects constitute a sample of size $n = 5$ from the population of German learners. We can similarly draw inferences about this population based on $n = 5$ observations. In essence, therefore, the data arose from a two-stage sampling procedure: $n = 5$ speakers (the primary sampling units) were sampled from the population of German learners, and then $n = 4$ cases (the secondary sampling units) were sampled from the population of FVO renditions of each speaker. If this structure is ignored, the analysis may produce overconfident results. This is the case if we use $n = 20$ observations to (directly) draw inferences about the population of German learners, skipping the second level at which observations are grouped within subjects.

From a statistical perspective, the assumption of independence[15] between the $n = 20$ data points is not met, because observations from a single learner can be expected to be similar and therefore correlated; they do not contribute independent pieces of information to the analysis. As the accuracy of an estimate depends on sample size, inferences about the population of German learners based on $n = 20$ data points without accounting for the structure in the data will inevitably produce overconfident summaries. This is to say that uncertainty estimates such as confidence intervals and p-values will be too small.

One way of dealing with structured data is to compute a summary measure for each subject and then (in this case) continue with $n = 5$ data points (e.g. a percentage for each subject). While this guards against overconfidence in the results, multilevel models[16] can do better, especially in L2 speech research and studies on language variation in general. The reason for this is that linguistic data are typically structured in multiple ways. In our exemplary FVO production study, there is in fact a second way in which observations are organized. Let us assume, for instance, that each subject was given the same word list (with $n = 4$ items). There is reason to believe that the productions of each word across learners are correlated. This may be due, say, to familiarity with the item or properties of the final voiced obstruent (e.g. fricative vs. stop). Again, the $n = 20$ observations are not independent due to the fact that the productions of a given item are likely to be similar. Each observation can therefore be assigned to two different clusters of data points: to a specific subject and to a specific item. This is illustrated in Figure 3.9, where observations are grouped by s(ubject) and w(ord). The connecting lines illustrate why such designs are often referred to as *crossed* or *cross-classified*.[17] In this sense, linguistic data are in many research contexts naturally crossed.[18]

Our default analysis tool throughout this study will be multilevel regression, with varying effects on subjects and words. Our analyses thereby take into account the crossed clustering of observations. Before we go further, we discuss two issues that are involved when specifying and interpreting multilevel regression models.

[15] For a discussion of the statistical independence assumption in language data analysis, see Winter (2019: 232–233) and Winter & Grice (2020).

[16] Multilevel models are also referred to as *mixed-effects* models, *hierarchical* models, or *random coefficient* models. The standard reference for these models in language data analysis is Baayen et al. (2008).

[17] More specifically, it is the factors (or classificatory variables) subject and word that are crossed.

[18] It has been convincingly argued that hierarchically structured data should be the default assumption in language data analysis (see Johnson 2014; Winter & Grice 2020).

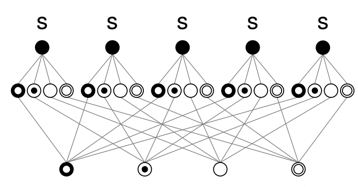


Figure 3.9: Illustration of hierarchically structured data, where observations are grouped by subject and word.

Specification of the variational structure in multilevel models

When formulating a multilevel regression model, we need to give some thought to the variational structure of the model (see Cox & Donnelly 2011: 112–113), which, for our present purposes, concerns the question of which varying (or "random") effects to incorporate into the analysis. Different strategies may be pursued, and decisions should be informed by (i) the purpose of the analysis and (ii) our background knowledge about the data and phenomenon. Our default strategy will be to specify a "maximal" variational structure (Barr et al. 2013). This structure reflects our preference for a model that allows the association between a given predictor and the outcome to vary across subjects and/or words.[19] We will therefore include varying (or "random") slopes for all predictors that, based on our background knowledge, can be expected to vary in strength across subjects and/or words. Since the variational structure is motivated by linguistic considerations, we forego a model-comparison-based elimination of random components and stick to our pre-data specification.

[19] The alternative, i.e. not including varying slopes on predictors of interest, would require us to assume constant effects across clusters. This is to say that, for a given predictor variable, each lexical item and each learner shows the same directionality and magnitude of association (with observable variation being attributable to sampling variation). We usually do not feel comfortable with this assumption and therefore default to including varying slopes to capture and accomodate variation between clusters.

Population-averaged vs. subject-specific estimates

When it comes to interpreting multilevel regression models for categorical outcomes, complications arise. Categorical outcomes in the current study are of two types: binary responses (binary logistic regression) and nominal outcomes with 3 or more unordered categories (multinomial logistic regression). In these settings, we need to distinguish between two types of estimates a multilevel model can offer. This issue arises from the fact that these regression models operate on a transformed, non-linear scale. The two meanings that can be communicated for hierarchically structured categorical outcomes are usually referred to as a *subject-specific* (or conditional) interpretation (the default for multilevel/mixed-effects models) and a *population-averaged* (or marginal) interpretation.[20] These labels are unfortunately not very helpful, so let us consider an example. We will stick to our illustrative FVO setting, now with a hypothetical scenario of $n = 20$ subjects (of low proficiency) and $n = 100$ observations each. We have measured the FVO accuracy rate for each learner and express it on the proportion scale, which expresses the share of target-like realizations. These range between roughly 0 and .30 (i.e. 0 and 30%). The distribution is shown in Figure 3.10. It is symmetric and centered at .15 (or 15%). This is the average on the proportion scale.

[20] This issue is discussed in Stroup (2013: 99–114); see also Molenberghs & Verbeke (2005: 297–306), Rabe-Hesketh & Skrondal (2012: 529–532), and Agresti (2013: 495–498).

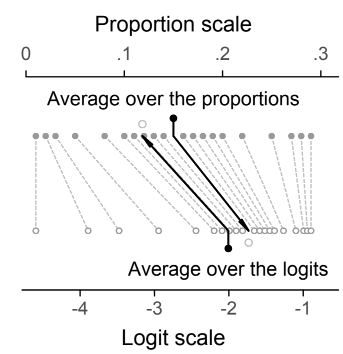


Figure 3.10: Illustration of the mismatch between the average computed over the proportions, (filled circle at the top), which is population-averaged estimate, and the average over the logits (filled circle at the bottom), the subject-specific estimate. Back-transformation of the average on the logit scale leads to a different proportion (grey empty circle at the top), which is the subject-specific estimate on the proportion scale. The arrows reflect the (re-)transformation bias.

Categorical regression models don't work directly with these proportions, but operate on a transformed scale, in our binary illustration the logit (i.e. log-odds) scale. Figure 3.10 shows how the proportions map onto this scale. Note that the distribution changes shape: It is now asymmetric. The mean of this distribution of logits is also indicated. This is the default mean returned by a multilevel logistic regression model. If we translate this average back to the proportion scale we arrive at .12 (or 12%), which is different from the average over the original proportions. In the current study, we are typically interested in the average over proportions, and we therefore always post-process the regression output to approximate these values (i.e. population-averaged estimates). To this end, we exploit the flexibility of the Bayesian paradigm[21] for data analysis, to which we now turn.

[21] Specifically, we follow the procedure described by Molenberghs & Verbeke (2005: 301). For each cluster variable (i.e. subject and word), we simulate, for each line in the posterior distribution, a random sample of 1,000 values from a (usually multivariate) varying effects distribution. We use the parameter values in the respective line of the posterior to propagate the uncertainty into the approximation. We then add the relevant parts of this simluation to the linear predictor, then convert to proportions, and then take the average. This leaves us with a posterior distribution of proportions, which we then summarize and interpret.

3.4.2 *Bayesian inference*

The current study relies on Bayesian methods for statistical inference. In this section we briefly describe what we consider the key advantages of Bayesian estimation, and discuss and illustrate the use of prior distributions (or priors, for short) in data analysis. Finally, we discuss our methods for assessing the reliability of our computational procedures.

Advantages

While Bayesian data analysis is becoming increasingly common in empirical research, frequentist (or ”classical”) methods still form the backbone of most quantitative work. In the present study, priority was given to Bayesian methods, mostly for two reasons:

- In certain settings, classical procedures run into computational dead-ends. We confronted this issue on numerous occasions, where software quit with a warning message (failure to ”converge”), forcing us to make compromises in our analysis.[22] Bayesian inference can handle these situations, which is why we opted for this paradigm in the first place. Further, Bayesian methods allow us to fit models which are currently not routinely available[23] in the frequentist facilities offered in *R* (R Core Team 2020).
- The output of a Bayesian analysis offers great flexibility for the summary and presentation of results. This allows us to communicate results using quantities that are familiar to every linguist. For instance, the output of categorical regression models (e.g. binary or multinomial logistic regression) can be back-transformed to the natural and more intuitive scale of proportions (or percentages), which can further be enriched with measures of statistical uncertainty (and, in the case of categorical outcomes, translated to population-averaged quantities).

[22] Such compromises usually entail a reduction of the variational structure of the model (see above), which would condition our results on a set of premises that do not sit well with our understanding of the phenomenon under study and the way in which we expect it the vary across speakers and lexical items.

[23] Examples are the use of multilevel models for multinomial outcomes and the use of breakpoint regression to enable our model to describe, at advanced stages of L2 acquisition, diverging populations of learners that reflect orientation toward different target accents.

Priors

Probably the most salient feature of Bayesian estimation is its ability to incorporate into the analysis knowledge about the phenomenon of interest. This pre-data state of information enters the model via prior distributions over the parameters. The statistical inferences produced by a Bayesian analysis therefore combine this prior information with that in the data. The output is referred to as a posterior distribution, a compromise between these two sources of information.[24]

We can usually build on previous research to arrive at a preliminary state of information about certain quantities of interest. Rarely is it reasonable to assume every parameter value from negative to positive infinity to be equally plausible before having seen the data. It is crucial for researchers, however, to make transparent and justify the specification of

[24] It is of interest to note that frequentist procedures may in fact be considered a special case of Bayesian analysis: The inferences generated by the two approaches converge as the amount of prior information added to a model approaches zero, by which we mean a flat prior distribution extending from negative to positive infinity.

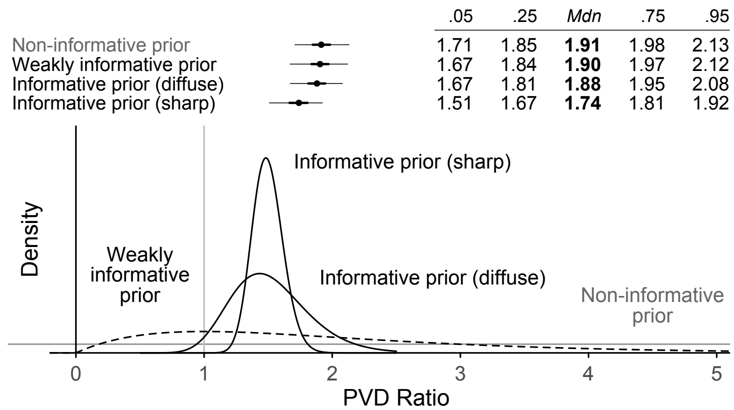


Figure 3.11: Shape and effect of different priors for the estimation of the PVD ratio. Illustration of non-informative, weakly informative, and informative prior distributions and their effect on the estimate (see Chapter 10 for details).

prior distributions, which should be based on accepted knowledge and/or empirical evidence in the literature.

Let us consider the effect of prior distributions in more detail by showing how linguistic knowledge may be integrated into an analysis by means of an example taken from this study. As such, prior distributions differ in the amount and type of information they offer. Accordingly, a distinction can be made between[25]

- non-informative,
- weakly informative,
- informative and diffuse,
- informative and sharp, and
- regularizing priors.

To illustrate the application of different priors, let us look at data from Chapter 10, which deals with German learners' production of final voiced obstruents. In native speech, final obstruent voicing affects the duration of the preceding vowel. An important cue to the distinction of minimal pairs such as *bat-bad* and *back-bag* is the duration of the nucleus, which is longer before voiced obstruents. The difference in preceding vowel duration (PVD) may be expressed in relative terms (as a ratio). Previous research has shown that AmE native speakers typically produce PVD ratios of about 1.50, meaning that vowels are on average 50% longer before voiced (vs. voiceless) obstruents. Let us assume (as is the case in this study) that we have recorded $n = 11$ native speakers of AmE and are interested in the average PVD ratio they produce. In this case, we estimate an average.

We could choose a non-informative prior, which means that we "let the data speak for themselves". We are then making the assumption that PVD ratios of, say, 0.1, 1.5 and 30.0 are equally plausible. In Figure 3.11, this non-informative prior distribution is represented as the dotted horizontal line, which assigns equal probability to all values from negative to positive infinity. The resulting estimate is a duration ratio of 1.91.

In this context, some thought reveals that we are not completely ignorant about the likelihood of different PVD ratios. For instance, negative values are impossible, and we can therefore truncate the prior at zero. Reflection

[25] These labels are largely adopted from the literature (see, e.g. Gelman et al. 2014). It is important to note, however, that prior distributions mainly differ along two dimensions of information: (i) location and (ii) spread. The location of a prior controls which values receive the greatest pre-data weight. The spread of the prior, on the other hand, determines how tightly the prior probability mass is distributed around these values. Put simply, the location expresses our state of information about (or best guess at) the parameter value, and the spread indicates the precision of this information (or how confident we are in our guess). It follows that *informative* priors can be informative in two ways, namely through (i) the choice of the location of the prior and (ii) the concentration of the probability mass close to this value. We will therefore make a distinction between *sharp* and *diffuse* informative priors.

about durational characteristics of speech might further suggest that ratios exceeding 5:1 are implausible and we may reasonably assume that the duration ratio is likely somewhere between 0.2 and 5.0. This is an example of a weakly informative prior constructed on the basis of our general knowledge about language and the measurement scale (i.e. ratios, which must be non-negative). The corresponding distribution in Figure 3.11 assigns 90% probability mass to values between 0.66 and 3.40. Using this prior, the estimate is 1.90, and thus virtually identical to the non-informative case.[26]

[26] Weakly informative priors may not affect inferences much, but they can save computational time, as they mildly guide the computational algorithm that modern Bayesian inference relies on to approximate the posterior distribution of model parameters.

In our present example, the literature offers more guidance on likely values for the PVD ratio in AmE. As discussed in detail in Chapter 10, at least $n = 20$ studies have looked at this phenomenon in AmE, and the distribution of estimates in the literature can be used to construct an informative prior distribution. Based on previous instrumental work, the PVD ratio in this population of speakers is expected to be somewhere around 1.50. If our analysis can be informed by previous empirical work that is directly relevant to the estimation task at hand, we will, in this study, talk about informative priors. These are centered at a value that is considered most likely in the light of the literature. Informative priors may differ in sharpness, however. Sharpness indicates the prior uncertainty in the value of interest. Figure 3.11 shows two informative prior distributions, one diffuse and one sharp. Sharp priors require solid evidence in the literature, that is, a large number of studies producing highly consistent estimates about the same phenomenon in the same context(s) and for the same population of interest. As this is rarely the case, we will exclusively be dealing with diffuse informative priors in this study. These are grounded in previous research findings, but give enough leeway for deviations from the empirical record.

In our case, a diffuse informative prior with 90% of its probability between 1.07 and 2.07 would give an appropriate expression of the current state of information.[27] The resulting estimate of 1.88 is a compromise between the information in the data (1.92) and the literature (1.50). The sharp informative prior, which assigns 90% probability mass to the interval [1.31; 1.70] and is in any case difficult to motivate given the heterogeneous estimates in the literature, puts more weight on the a priori information and produces an estimate of 1.74.

[27] This estimate ignores systematic variation among study estimates that may be linked to differences in elicitation task, vowel type and obstruent class.

The final type of pre-data information, regularizing priors (McElreath 2020: 214–216), may be specified so as to make an analysis more cautious or conservative. This is the case when we encode skepticism by considering values of zero to be most likely, a priori. In essence, regularization makes a model skeptical toward estimates that do not receive much support from the data. The amount of regularization is determined by (i) the spread of the regularizing prior and (ii) the amount of evidence in the data. This is illustrated in Figure 3.12, which shows cases in which the information in the data is vague (left) and solid (right). The dashed curve shows the regularizing prior, which is the same in both examples. The dotted curve represents the information in the data. In the left panel, this information is vague, which may be due, say, to a relatively small sample size. This vagueness is reflected in the wide range of plausible parameter

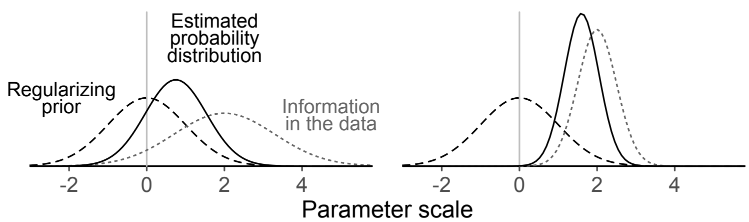


Figure 3.12: Illustration of the effect of regularizing priors. The solid density curve represents a weighted combination, or compromise, between the information in the data (dotted line) and the regularizing prior centered about zero (dashed line). The regularizing effect is more pronounced when the data offer only sparse information about a parameter (left panel).

values suggested by the data. The solid curve represents the compromise between data and regularizing prior. In this case, the estimated probability distribution for the parameter is pulled toward zero, as the model was told to be skeptical of vague estimates. In the right panel, there is solid information in the data for the parameter of interest. The estimated probability distribution is therefore dominated by the data and only mildly drawn toward zero.

When we analyze a set of data, our objectives allow us to distinguish between parameters of primary interest and auxiliary parameters.[28] Parameters of primary interest are those that relate to or directly answer our research question. Auxiliary parameters, on the other hand, are part of the statistical model but not interpreted. Our strategy for specifying priors will differ for these two groups of parameters. For parameters of primary interest, the default choice will be regularizing priors. The amount of skepticism they express will be made transparent. If the literature suggests certain parameter values to be more likely, we will move toward diffuse informative priors. For auxiliary parameters, we will proceed in the same way, but more often default to regularizing specifications. This is because (i) our literature review concentrates on the quantities of direct interest and (ii) auxiliary measures are rarely reported in empirical publications. In any case, the location and sharpness will be explained and justified by reference to the scale of measurement and/or quantitative evidence gathered in our survey of previous work.

[28] In the statistical literature, auxiliary parameters are also referred to as ”ancillary” or ”nuisance” parameters.

The statistical analyses in this study were carried using the open-source software *R* (R Core Team 2020), the *R* package *rstan* (Stan Development Team 2020), and the Bayesian inference engine *Stan* (Carpenter et al. 2017; Stan Development Team 2018). The posterior distributions generated by the models were processed in *R* and visualized using the *R* package *lattice* (Sarkar 2008).

MCMC diagnostics

Bayesian procedures also differ from classical ones in the way inferences are generated. Modern Bayesian data analysis relies on the use of Markov chain Monte Carlo (MCMC) algorithms, which are computationally expensive procedures that approximate the posterior distribution of model parameters (see McElreath 2020: 263–296). This approximation is not guaranteed to be reliable and the researcher must check for signs of malperformance. The key strategy is to run several MCMC chains in parallel[29] and then compare them to see if they produce similar approximations. If

[29] Each analysis in the present study used four chains. The *Stan* defaults of a warm-up period of 1,000 draws plus 1,000 post-warm-up iterations were used, yielding n = 4,000 samples of the posterior distribution. If there were signs of non-convergence, the number of iterations was increased until all diagnostics were satisfactory.

they do, the chains are said to have "converged". Several diagnostics are available to assess convergence:

- $\hat{R}$ values: These scores offer a rough measure of agreement among the chains. Values above 1.01 indicate that that the chains may be offering different approximations. As a minimal requirement, then, all $\hat{R}$ values must be smaller than 1.01. As the number of parameters in a model can be quite large, we will inspect histograms of $\hat{R}$ values.
- Rank plots: These graphical aids allow us to compare chains in more detail.[30] If the collection of profiles in these displays deviates from a horizontal band, there is reason for concern and we must identify the problematic parameter(s).
- Effective sample size (ESS): This is a measure of the amount of information the MCMC samples offer about the posterior distribution. Vehtari et al. (2020) note that this measure should be higher than 400. For a given parameter, an ESS can be determined for the center and the tails of the posterior distribution. When summarizing and interpreting the output of a model, we use the center of the posterior to produce a point estimate and look into the tails to derive an uncertainty estimate. We use quantile plots to inspect, for each parameter, the ESS at representative quantiles.[31] In these graphs, then, all ESS estimates should exceed 400.

For each model we present a graphical summary of MCMC convergence diagnostics, as illustrated in Figure 3.13. The left panel shows the distribution of the $\hat{R}$ values; the 1.01-threshold is marked by a vertical line.[32] The middle panel shows a rank plot. The profiles should form a horizontal band that doesn't fan out at the margins. Finally, the right panel indicates, for the parameters of primary interest, the effective sample size. The 400-threshold is marked by a grey horizontal line. The percentiles we will refer to in our summaries (i.e. the 5th, 25th, 50th, 75th, and 95th) are marked on the horizontal axis. Based on the graphical diagnostics shown in Figure 3.13, then, the MCMC algorithm appears to have converged. The $\hat{R}$ values (left display) are all well under the 1.01-threshold, the rank plot (middle) shows relatively flat profiles, and the quantile plot indicates that the effective sample size for the quantiles of interest is well above the minimal requirement of $n = 400$ for all parameters.

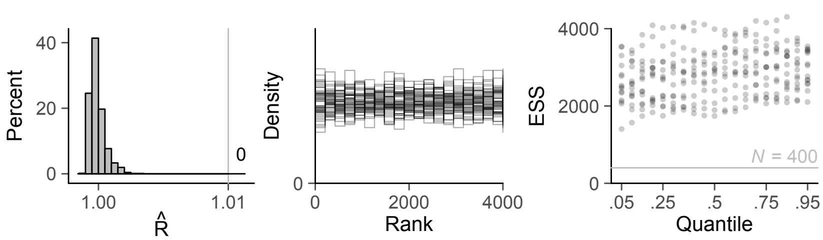


Figure 3.13: Graphical inspection of convergence diagnostics. From left to right: histogram of $\hat{R}$ values, rank plot, and a quantile plot.

[30] Put simply, we combine the samples from all chains and rank-order the values from lowest to highest. We note down the rank of each sample and then split the chains again. Each chain should then have a similar rank profile, i.e. the proportion of low ranks (or high ranks) as the other chains. Rank plots offer a visual comparison of these rank profiles. Vehtari et al. (2020) suggest plotting a histogram for each chain; the distributions should be flat and cover the entire range of the horizontal axis. To save space and directly compare distributions, we will superpose histogram profiles (see also McElreath 2020: 284). Moreover, we will combine into a single display the chains of all key parameters in the model.

[31] These include the .05, .25, .75, and .95 quantiles, which form the basis of the 50% and 90% percentile intervals we report.

[32] Ideally, all values are to the left of the line and "0" occurs to the right of the line (indicating that no value was higher than 1.01). If there were values above 1.01, these cases are discussed.

3.4.3 Modeling strategies

Deductive modeling

Our strategy for building regression models for statistical inference may best be described as a form of deductive modeling (see Tizón-Couto & Lorenz 2020): Our decisions are primarily informed by the context and linguistic background of the analysis. The choice of predictor variables, for instance, is motivated by theoretical considerations, which earnes them their place in the analysis. We therefore see no point in applying model comparison to reduce the set of variables in our analysis. The same applies to the variational structure, as discussed above.

Robust estimation

Outliers are a concern in applied research, since classical procedures are sensitive to deviations from normality. In practical terms, this means that a few aberrant data points (or subjects or lexical items) can have an undue influence on our data summaries. This study makes use of a simple technique to guard against the influence of outliers. We slightly modify our statistical models and replace normal error distributions by Student-t distributions, which allows for extreme scores without affecting inference for the more typical observations in the sample (Gelman et al. 2014: 435–445; Kruschke 2015: 458–461; McElreath 2020: 233). For the regression models with continuous outcomes, we use t-distributed errors both for the individual measurements and for for the varying effects distributions. For our categorical regression models (binomial and multinomial logistic regression), we only apply this feature to the varying effects. We set the degrees-of-freedom parameter ν, which controls the thickness of the tails, to 7.[33]

[33] See Appendix A.3.1 and Figure A.3 (p. 253) for an illustration.

3.5 Constraints on generality statements

Similar to most empirical work, the present study aims to offer insights that apply not only to the limited set of observations studied, but to a larger population of entities. From a statistical perspective, inferential elements in our analysis contribute to this aim by indicating the statistical uncertainty associated with our estimates. Of course, these quantities can only be valid for generalizations to the broader context of entities that our sample actually represents. It is therefore critical to think about and clearly delimit the scope of our inferences (see Yarkoni 2019). In other words, we must distinguish the sampled population from the target population. In the present study, for instance, we may wish to extend our findings about the PVD ratio to the way in which German learners contrast final voiced and voiceless obstruents in general. However, it is unlikely that our data cover the full range of contexts in which this contrast may be signaled. Our inferences are therefore limited to the set of circumstances we have elicited. We must reflect on the degree to which the sampled instances allow us to extend our conclusions to other, unobserved contexts.

Before we consider the results of our analyses in each chapter, we will therefore pause to reflect on constraints on generality (Simons et al. 2016). We will distinguish three directions into which we may wish to generalize the indications in our data. Thus, *language-externally*, we define the population of German learners to which our statements can be applied. This language-external scope may vary depending on the segmental structure, since dialectal idiosyncracies may narrow the generalizability for some categories of speech. A *language-internal* orientation, on the other hand, seeks to delimit the scope of our inferences across all possible instances of a language structure. As our analyses rests on a limited set of lexical items, we state the extent to which we may generalize to other words, or instances of the target structure. If the items in our sample represent a narrowly defined set of contexts, our findings may not extend to instances that differ systematically from these. Finally, we will consider to which *speaking styles* we can extend our insights.

3.6 Presentation of results

When presenting the results of regression analyses, we will focus on directly interpretable, meaningful quantities in the text and relegate technical details to the appendix.[34]. Two key features of our presentation strategy are (i) the use of transparent measures such as proportions (instead of log-odds or odds ratios) when dealing with categorical outcomes, and (ii) the use of uncertainty intervals instead of *p*-values when communicating inferential estimates.[35] We will default to 50% and 90% uncertainty bounds, which we determine based on relevant quantiles from the posterior distribution of the quantity of interest. Findings will primarily be presented in graphs and corresponding estimates will be listed in tables.[36] Technical details, which include information about the model structure, MCMC diagnostics, and the posterior distribution of parameters, are deferred to the Appendix and the online supplements. The regression models we fit deviate in some ways from default implementations and we therefore describe their structure in detail in the Appendix, including comments about the formulation of priors.

[34] See Vanhove (2020), who advocates the liberal use of appendices and supplementary materials to unclutter the linguistic line of argumentation.

[35] See Sönning (2019) for a discussion and Vanhove (2019) for practical advice.

[36] The general strategy is influenced and inspired by Long (1997), Long & Freese (2014), and Long & Mustillo (2018). Two graph types that are frequently applied throughout the following chapters are the dot plot (see Sönning 2016) and the line plot (see Sönning 2020b).

4
The *trap-dress* contrast

This chapter is concerned with the acquisition of a new vowel contrast by German learners: the front mid vowel in *bed* vs. the front open vowel in *bad*. After a contrastive analysis of English and German monophthongs in §4.1, §4.2 considers linguistic factors constraining the L2 acquisition of the *bed-bad* distinction. Guidance offered by theoretical work is discussed in §4.3, which derives predictions from the contributions covered in Chapter 2. The current state of empirical knowledge is reviewed in §4.4, followed by an outline of the aims (§4.5) and methods (§4.6) of this study. Following constraints on generality (§4.7), §4.8 presents the results. §4.9 concludes with a summary and discussion.

4.1 Contrastive analysis

The description of vowels involves two key parameters: length and quality. In IPA notation, quality is denoted by different symbols and length is indicated by diacritics (e.g. long [eː] and half-long [eˑ]). It is common practice to describe vowel quality using vowel charts, where location reflects auditory quality. To this end, the cardinal vowel diagram (Jones 1918) is used as a reference grid. Its margins span the physiological limits of the vowel space, with 16 cardinal vowels providing auditory anchors for the classification of vocalic sounds. The top panel of Figure 4.1 shows this conventionalized chart with the 8 primary cardinal vowels [i e ɛ a ɑ ɔ o u]. Such vowel charts are partially physically grounded, reflecting tongue height and advancement, which is shown in the lower panel. However, these articulatory labels should be considered a convenient heuristic only, since the same auditory effect can in fact be achieved by different tongue gestures. As such, auditory vowel descriptions are therefore argued to be more reliable (Cruttenden 2014: 39).

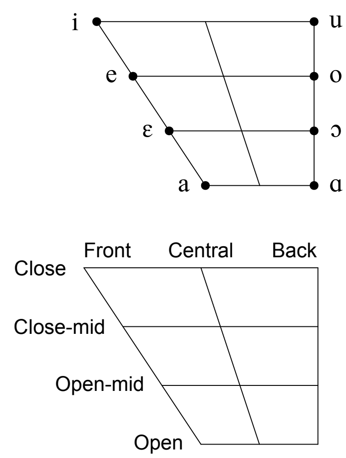


Figure 4.1: The cardinal vowel diagram showing the eight primary cardinal vowels (top) and their articulatory correspondence (bottom).

The number of distinctive vowel categories (i.e. phonemes) differs across languages. Figure 4.2 shows the distribution of inventory sizes in $n = 317$ languages in the UPSID database (Maddieson 1984: 126); counts range from 3 to 24. From a typological perspective, then, English and German have complex vowel systems. For General American (GA), the number of vowels has been put at 15, for General British (GB)[1] at 20 (e.g. Jones et al. 2011; Wells 2008); German falls in between with 18 categories (e.g. Kohler 1995).

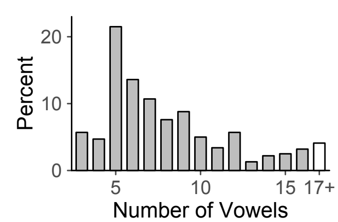


Figure 4.2: Distribution of the number of distinctive vowel qualities in the UPSID database (Maddieson 1984).

[1] We will follow Cruttenden (2014) and use the term *General British* to refer to the supra-regional variety that may be considered the British standard accent.

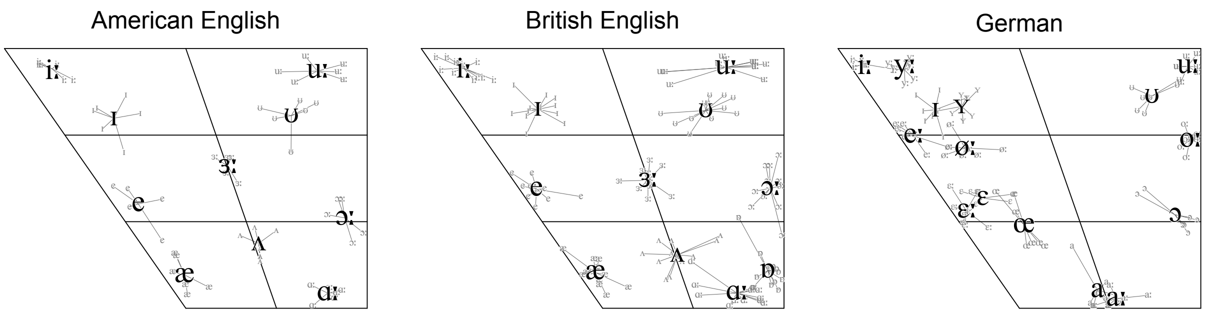


Figure 4.3: Vowel quality in American English, British English, and German: Survey of auditory descriptions of monophthongs. Small grey IPA symbols show the locations given by different sources. Black IPA symbols mark the average location across all studies. The graphs summarize the results of studies on BrE (n = 12), AmE (n = 6) and German (n = 6).

While similar in complexity, the vowel systems of English and German differ noticeably. Our contrastive analysis starts out with a comparison of auditory descriptions in the literature. The fact that vowel quality is indicated relative to the cardinal vowels makes it possible to establish comparability across different accounts of the same language or accent. To this end, a normalization procedure was devised, which allows us to directly compare vowel chart locations reported by different authors.[2] As for terminology, we will resort to the labels *British English* (BrE) and *American English* (AmE) instead of GB and GA when discussing data that may not be purely reflective of either standard variety. Figure 4.3 provides a summary of the auditory descriptions given in previous studies[3] on

- British English (Scherer & Wollmann 1986; Giegerich 1992; Kortmann 1999; Roach 2004; Eckert & Barry 2005; Davenport & Hannahs 2005; Clark et al. 2007; Wells 2008; König & Gast 2009; Jones et al. 2011; Ladefoged & Johnson 2011; Collins & Mees 2013),
- American English (Giegerich 1992; Ladefoged 1999; Davenport & Hannahs 2005; Wells 2008; König & Gast 2009; Ladefoged & Johnson 2011) and
- German (Scherer & Wollmann 1986; Kohler 1995; Becker 1998; Kortmann 1999; Eckert & Barry 2005; König & Gast 2009).

Among other contrasts, German lacks the open front vowel /æ/ (hereafter referred to as TRAP), which may be described as an unrounded front mid-open to open vowel. Neighboring vowels are unrounded short central open to mid-open /ʌ/ (hereafter STRUT) and unrounded short front mid /e/ (hereafter DRESS). The unrounded long open central to back /ɑː/ (hereafter PALM) is further back and slightly more open than TRAP.

The closest counterparts in the German inventory are unrounded mid front /ɛ ɛː/ (*hätte, Mädchen*) and unrounded central open /a aː/ (*hat, mahnen*). German front mid-open /œ/ is rounded and its quality is therefore very different from that of adjacent unrounded vowels. TRAP, DRESS and STRUT are checked vowels – in contrast to PALM, they do not occur in open syllables. In German, the short monophthongs /ɛ a/ are similarly restricted to closed syllables; only the tense vowels /ɛː aː/ are licensed in open syllables. On a phonotactic level of comparison, TRAP is thus more similar to the German short monophthongs /ɛ a/.

[2] The edges of the vowel diagram are used as reference points: By measuring the vertical distance from the top and bottom edge and the horizontal distance from the left and right edge, the position of each vowel can be expressed relative to these margins. Figure 4.3 shows these locations, which are directly comparable conditional on the assumption that the studies in our survey express vowel quality relative to the cardinal vowels.

[3] The data are available at https://osf.io/a98hg/.

TRAP and DRESS are subject to regional and social variation. A sound change taking place in present-day GB is the lowering and retraction of TRAP (Wells 1982; Henton 1983; Bauer 1985; Harrington et al. 2000; Hawkins & Midgley 2005; Mair 2006: 165). In the north of England, TRAP is generally shorter and fully open (Cruttenden 2014: 121).[4] As for DRESS, relatively high (or close) qualities tend to be associated with older speakers; more open qualities are typical of younger speakers (Wells 1982: 291; Harrington et al. 2000; Hawkins & Midgley 2005; Hughes et al. 2005: 48). While a more open quality is also typical for DRESS in northern English accents, a closer variant is found in London (Cruttenden 2014: 117).

In AmE dialects, TRAP and DRESS are affected by several regionally confined vowel shifts. Metropolitan areas from New England to Wisconsin are affected by the Northern Cities Vowel Shift (Labov et al. 2006: 122), which describes a rotation in certain areas of the vowel space leading to fronting and raising of TRAP (and partly diphthongization, e.g. *man* [eɐ ɪə]) and backing of DRESS. A common development is the raising of TRAP only before nasals. In the Southern Vowel Shift, DRESS raises and diphthongizes to [eɪ]. Breaking of TRAP is observed in large parts of the South, where the vowel begins and ends in open front position and shows an intermediate [j]-transition. The California Vowel Shift also involves systematic changes in vowel quality. In northern California, DRESS is lowering and shifting toward TRAP, which in turn splits into two variants: It diphthongizes to [ij] before nasals and shifts toward [ɑː] elsewhere (Wolfram & Schilling 2016: 372f.). We thus see divergent tendencies in TRAP in BrE and AmE dialects. In this study, the neutral designations TRAP and DRESS will be employed whenever possible. When shorthand symbols are needed, ″æ″ and ″e″ will be used. This is in accordance with Jones et al. (2011) and Wells (2008) and allows us to distinguish TRAP from German /a aː/ and to distinguish DRESS from German /ɛ ɛː/.[5]

According to the auditory descriptions summarized in Figure 4.3 above, AmE and BrE both distinguish DRESS and TRAP in terms of vowel quality. We can quantify the quality contrast with the help of a Euclidean distance[6], which reflects in the distance between two categories in the auditory vowel space. To make these quantities comparable across studies, the auditory coordinates shown in Figure 4.3 were first converted to *z*-scores.[7] This procedure allows us to express the auditory distance between TRAP and DRESS for each of the studies listed above. Figure 4.4 compares these scores across studies, which are grouped by variety (AmE at top, BrE at bottom). The box plots in the middle summarize the distribution of distances for each variety. We see a fair amount of variation within each variety, but on average, the auditory distance is greater in BrE ($D = 0.91$) than in AmE ($D = 1.08$).[8] The dispersion of scores within each variety at least partly reflects social, regional and diachronic trends.[9]

Auditory vowel descriptions of BrE and AmE have received support from instrumental studies. The key acoustic correlate of vowel quality is the formant structure of the vowel. Formants are areas of high intensity in the frequency spectrum and can be measured using speech analysis software. Vowel measurements are often visualized using the first two formants. A scatterplot, with the first formant (F_1) on the vertical axis (increasing

[4] Some reference works therefore prefer the IPA symbol /a/ (e.g. Upton et al. 2001; Cruttenden 2014).

[5] However, see Schmitt (2007) for arguments in favor of /ɛ/ to denote DRESS.

[6] It may be unusual to see Euclidean distance measures applied to auditory vowel descriptions. This may be due to the fact that ear-phonetic vowel charts do not lend themselves as easily to quantification as acoustic measurements. For the latter, the calculation of distances is a widely used technique (e.g. Harrington & Cassidy 1999: 243). As we have quantified the location of vowels in the auditory space, it is straightforward to subject these measurements to the same set of statistical procedures as acoustic data.

[7] A method similar to that proposed by Lobanov (1971) was used (see §4.6 for methodological details).

[8] We will use the symbol D to refer to a Euclidean distance.

[9] Thus, the small distance derived from Ladefoged (1999), who models his description on a young Californian speaker, is due to lowered DRESS. On the other hand, BrE accounts with the largest distances between DRESS and TRAP (Eckert & Barry 2005; Collins & Mees 2013) capture current sound changes, most notably lowering and retraction of TRAP.

from top to bottom) and the second formant (F_2) on the horizontal axis (increasing from right to left), yields a remarkably close correspondence to ear-phonetic vowel charts (Joos 1948).

Figure 4.5 gives an overview of the formant measurements[10] reported in previous work[11] on

- British English (Wells 1962; Wiik 1965; Henton 1983; Nolan 1983; Deterding 1997; Deterding 1990; Hawkins & Midgley 2005; Steinlen 2005; Ferragne & Pellegrino 2010; Williams & Escudero 2014),
- American English (Peterson & Barney 1952; Delattre 1965; Childers & Wu 1991; Bradlow 1995; Hillenbrand et al. 1995; Yang 1996; Lee et al. 1999; Assmann & Katz 2000; Wang 2007; Strange et al. 2007), and
- German (Delattre 1965; Rausch 1972; Wängler 1981; Ramers 1988; Claßen et al. 1998; Steinlen 2005; Sendlmeier & Seebode 2007; Strange et al. 2007).

Looking at Figure 4.5, we see that acoustic studies draw a somewhat different picture of the DRESS-TRAP contrast, suggesting that the distance in the F_1-by-F_2 vowel space is smaller for AmE. Figure 4.5 shows that the varieties primarily differ in the spectral properties of TRAP, which is produced with higher F_1 in BrE, that is, it assumes a lower position in the acoustic vowel chart. The location of DRESS, on the other hand, is virtually identical.

Again, we can directly compare the quality contrast across studies by using a Euclidean measure, which gives the acoustic distance between TRAP and DRESS.[12] Values from the studies[13] in our survey are graphed in Figure 4.6, where they are grouped by variety. Similar to the auditory distances shown in Figure 4.4, there is considerable between-study variation. However, the same systematic contrast between the varieties emerges: vowel quality differences are larger in BrE speech ($D_{zBk} = 0.89$) than in AmE speech ($D_{zBk} = 1.19$).[14] The measurements for AmE are highly dispersed,

[10] Measurements were transformed from Hertz to Bark and then normalized to *z*-scores (see §4.6 for methodological details). The scores are therefore on the same scale as those derived from auditory work.

[11] The data can be accessed via the OSF: AmE (https://osf.io/r86dv/), BrE (https://osf.io/b9gyv/), German (https://osf.io/xbr94/).

[12] The Euclidean distances are calculated from the Lobanov-normalized F_1 and F_2 Bark values (see https://osf.io/tvmnd/ for details).

[13] Details can be found in the online supplements: https://osf.io/tvmnd/.

[14] The subscript *zBk* is meant to clarify that the Euclidean distance is based on *z*-transformed (i.e. Lobanov-normalized) Bark measurements.

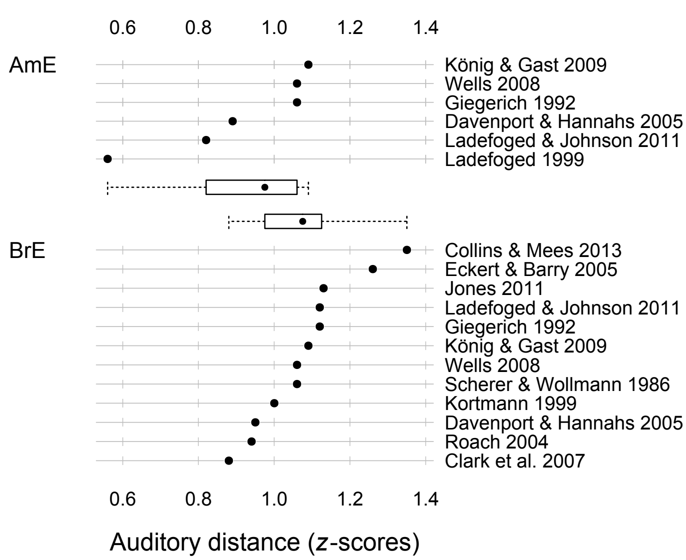


Figure 4.4: Auditory vowel quality contrast between TRAP and DRESS: Auditory distances for AmE and BrE. The plotted values are Euclidean distances on the scale of *z*-scores. These were calculated based on the location of TRAP and DRESS in Figure 4.3. The boxplots in the middle summarize and compare the distribution of scores.

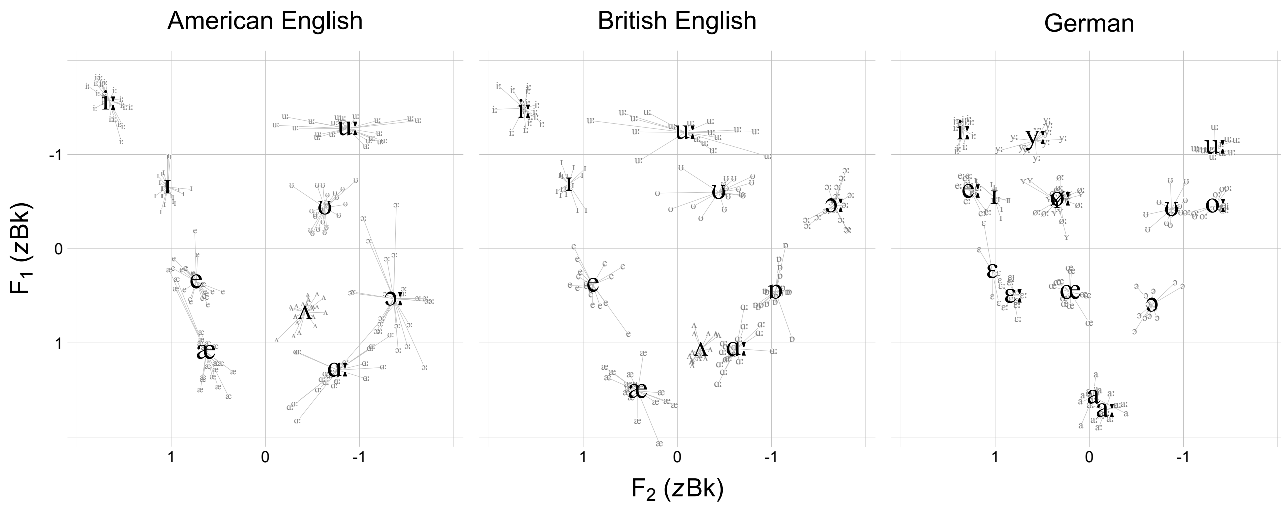


Figure 4.5: Acoustic vowel quality in American English, British English, and German: Survey of F_1-by-F_2 measurements of monophthongs. Coordinates are arranged to mirror position in the vowel chart. Plotted values are *z*-scores based on Bark-transformed measurements. Small grey IPA symbols show measurements given in references. Black IPA symbols mark the average location across all studies. The graphs summarize the results of studies on BrE ($n = 10$), AmE ($n = 10$) and German ($n = 8$). If you are holding the print version of this book, you can lift the left-hand page for a side-by-side comparison with the auditory vowel spaces in Figure 4.3 (p. 54).

which is in part due to regional variation.[15] The fact that the auditory and acoustic procedures yield very similar average distance scores (BrE: 1.08 and 1.19; AmE: 0.91 and 0.89) throws into relief the external validity of formant measurements, on which the subsequent instrumental analyses rely.

Let us finally consider the parameter length. While TRAP is traditionally classed as a short vowel, acoustic studies indicate that it is generally longer than the other short monophthongs, both in British and American speech. The measurements reported in previous work[16] are summarized in Figure 4.7. Vowel duration is plotted on the log scale and the original values (in ms) are shown at the top margin. We see that TRAP clearly patterns with

[15] Indeed, the studies with the smallest values (Hillenbrand et al. 1995; Bradlow 1995; Lee et al. 1999) recorded speakers from the Inland north dialects.

[16] The data are at https://osf.io/r2fy5/ and further details are given in the online supplementary files: https://osf.io/tvmnd/

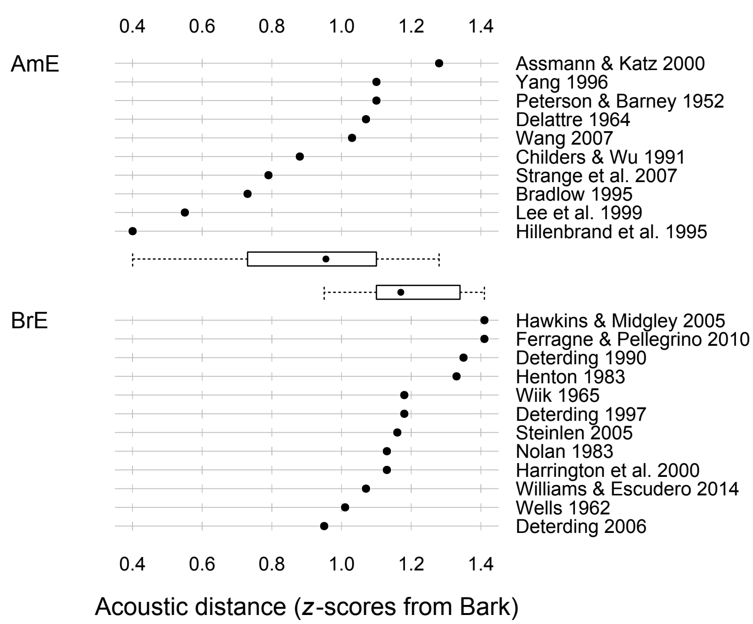


Figure 4.6: Acoustic vowel quality contrast between TRAP and DRESS: Acoustic distances for AmE and BrE. The plotted values are Euclidean distances on the scale of *z*-scores (which are calculated based on Bark-transformed formant measurements). These correspond to the distance between the vowel categories in Figure 4.5. For all studies, average values are reported. The boxplots in the middle summarize and compare the distribution of scores.

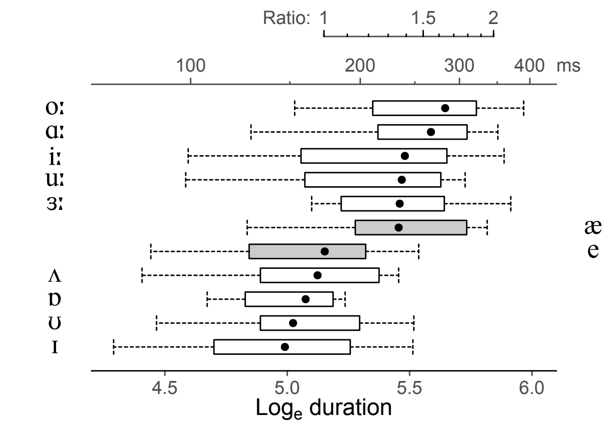


Figure 4.7: Duration of TRAP and DRESS in comparison to other English monophthongs: Survey of 8 studies (Rositzke 1939; Black 1949; Peterson & Lehiste 1960; Wiik 1965; Hillenbrand et al. 1995, 2001; Strange et al. 2007; Liu et al. 2014). The graph is based on log-transformed average durations; the axis at the top translates these back into the original ms values. The floating axis at the very top shows the relative differences (duration ratios) signaled by the log-transformed values.

the long monophthongs. An advantage of using logarithmic scaling is that distances in the graph signal relative (rather than absolute) differences.[17] It appears that, on average, TRAP is about 30% to 40% longer than DRESS. We will now turn to a more detailed survey of temporal differences between these vowels.

Throughout the present study, we will express durational differences between vowels in relative terms, using ratios, as this allows us to directly compare findings in the literature.[18] These ratios will express the duration of TRAP divided by that of DRESS. A ratio of 1.50 would tell us that TRAP is 50% longer than DRESS (e.g. 300 ms vs. 200 ms.). Figure 4.8 shows the ratios derived from instrumental work on AmE (top; $n = 11$ studies) and BrE (bottom; $n = 6$ studies).[19] On average, TRAP is about 35% longer than DRESS in AmE and 25% longer in BrE ($M_{AmE} = 1.36$; $M_{BrE} = 1.25$).

Overall, the acoustic results indicate a trading relationship between duration and spectral distance: In AmE, a greater contrast in duration

[17] The floating scale at the top can serve as a measuring stick to read relative differences, expressed as ratios, from the graph. For convenience, it is anchored at the median duration of DRESS.

[18] Absolute differences appear less suitable because they are more sensitive to speech rate – with increased tempo, absolute differences between segment durations become smaller. There are also perceptual reasons in favor of a relative measure. Thus, across different domains of human sensation, it has been observed that perception operates in relative (rather than absolute) terms (Dehaene 2003; Varshney & Sun 2013).

[19] Details can be found in the online supplementary files: https://osf.io/tvmnd/.

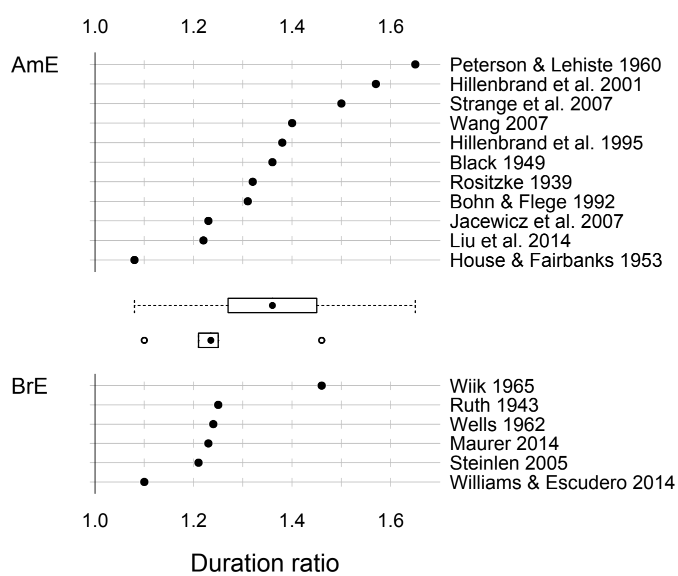


Figure 4.8: Vowel length contrast: Survey of TRAP-to-DRESS duration ratios reported across different studies for BrE and AmE (see web supplement for details: https://osf.io/tvmnd/). The boxplots in the middle summarize and compare the distribution of scores.

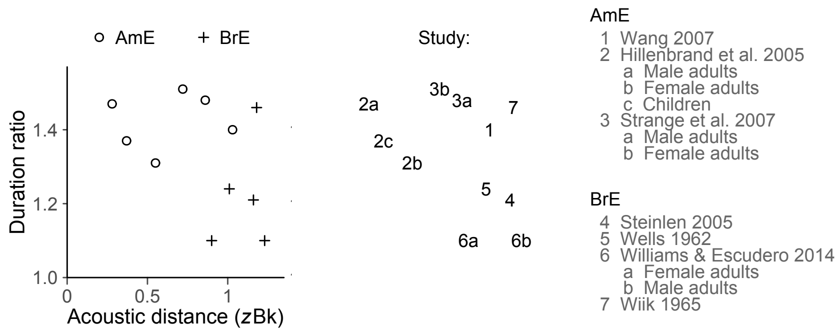


Figure 4.9: Relationship between quality and length as contrastive features: Duration ratio by acoustic distance for TRAP/DRESS in BrE and AmE. While AmE speakers exploit duration to contrast TRAP and DRESS, BrE speakers more strongly rely on quality.

compensates for a smaller degree of spectral separation. The trade-off between these signals emerges clearly when we draw into a single display the duration ratio and the acoustic distance between TRAP and DRESS. Figure 4.9 shows the relationship between quality and length for the subset of n = 7 studies that report both acoustic properties. The shape of the point cloud indicates that speakers who do not differentiate the vowels in terms of duration (lower half of the display), distinguish them in terms of quality. Vice versa, speakers who do no not exploit quality to mark the category contrast (left half of the display) make use of temporal cues. The data points in Figure 4.9 furthermore clearly pattern by variety: AmE speakers (○) are biased toward temporal cues, whereas BrE speakers (+) maintain a clear quality contrast. No points occupy the lower left corner of the plot, which is where speakers with merged categories would fall.

In summary, auditory and instrumental evidence suggests that the phonological contrast between TRAP and DRESS is implemented in different ways in BrE and AmE. In terms of vowel quality, the distance is greater in BrE, which is mainly due to a more open articulation of TRAP. In AmE, the durational contrast between the vowels is greater, which is largely attributable to the longer duration of TRAP. The auditory and acoustic structure of the German vowel space shows no equivalent to English TRAP. In contrast to English, however, German has two unrounded mid front vowels /ɛ/ as in *hätte* and /ɛː/ as in *Mädchen*, which differ in length. Acoustic studies suggest that vowel quality does not play a role in contrasting these categories (cf. Figure 4.3); vowel duration therefore serves as the primary cue distinguishing /ɛ/ from /ɛː/. Table 4.1 summarizes duration ratios reported in instrumental studies; these range from roughly 1.50 to 2.50.

Study	[ɛː]	[ɛ]	Ratio	[aː]	[a]	Ratio
Antoniadis & Strube 1984	175	81	2.16	184	78	2.36
Ramers 1988	188	89	2.11	186	91	2.04
Steinlen 2005	187	87	2.15	198	84	2.36
Hoffmann 2011 (Kiel Corpus)	125	82	1.52	154	81	1.90
Hoffmann 2011 (LeaP Corpus)	150	61	2.46	97	73	1.33

Table 4.1: Acoustic properties of German /ɛː ɛ/ and /aː a/: Mean duration (in ms) and ratio.

4.2 L2 acquisition of English TRAP: Linguistic factors

4.2.1 Markedness

Cross-linguistically, certain vowel qualities are favored over others. Figure 4.10 gives a summary of the distribution of the roughly 2,500 vowels in the UPSID database (Maddieson 1984: 124). To provide an overview of the prevalence of different auditory features, the vowel space is divided into nine regions, with black circles denoting unrounded vowels and grey circles rounded vowels. The area of the circles in Figure 4.10 is proportional to the number of vowels attested in the respective region. We see a clear bias toward 5 positions and the front low area of English TRAP appears to be generally disfavored. While lower mid front unrounded [ɛ] occurs in .41 of the languages, the figures are smaller for low front unrounded [a] (.06) and raised low front unrounded [æ] (.11).

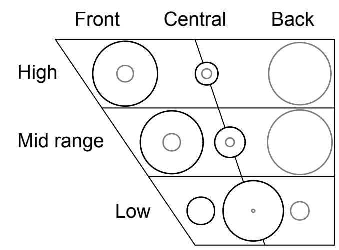


Figure 4.10: Typological distribution of vowel quality; black circles denote unrounded vowels, grey circles show rounded vowels. The area of the circles is proportional to the number of vowels attested in the UPSID database (Maddieson 1984). 🅭🅯

The typological distribution of vowel quality is in good agreement with a set of natural processes proposed by Donegan (1978). She classifies vowels in terms of three features: palatality [±pal], labiality [±lab], and sonority [±high ±low]. These labels correspond to the articulatory features given in Figure 4.1. Front vowels are [+pal], back vowels are [+lab], central vowels are [–pal, –lab]. Sonority is linked to vowel height: Lower (or more open) vowels are more sonorous. Two universal pairs of processes proposed by Donegan are (i) raising vs. lowering and (ii) bleaching (centralization) vs. coloring (fronting/backing). Their effects are illustrated schematically in Figure 4.11. Raising predominantly affects front and back vowels. Coloring is most likely to be observed for high vowels and least likely for low vowels. Bleaching, on the other hand, is assumed to primarily affect low vowels. Open front vowels such as TRAP therefore tend to be affected by raising and bleaching, which is consistent with their cross-linguistic rarity and diachronic instability.[20]

[20] Indeed, vowel shifts that are currently observable in BrE and AmE dialects are in line with these processes.

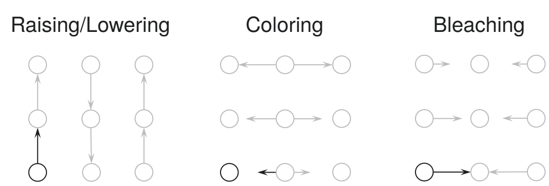


Figure 4.11: Natural processes affecting vowel quality (Donegan 1978): Arrows show process direction and strength at each location in the vowel space; adapted from Smith (1980). TRAP, which is shown in grey, is affected by raising and bleaching.

In L1 acquisition, vowels are acquired early. Even the relatively complex system of English monophthongs is mastered by most children around the age of 2;6 (Cruttenden 2014: 109).[21] The last vowel distinctions to emerge in English L1 acquisition are those between articulatorily similar categories such as TRAP, DRESS, and STRUT (ibid.: 109).

[21] The full range of vowel phonemes is thus established at an age where most consonants have not been acquired (cf. Figure 3.6).

In sum, there is a discrepancy between evidence from typology, language change, and natural phonology on the one hand, and L1 acquisition on the other. While TRAP appears to be a marked and unnatural structure, it is acquired early by English-learning children. Relative to other vowels, however, it emerges late due to its articulatory similarity to DRESS and STRUT. The summary in Table 4.2 indicates that the evidence supports a classification of [æ] as a marked structure.

Criterion	[æ]
Implicational relationship	○
Cross-linguistic frequency	++
L1 acquisition	○
Synchronic/diachronic instability	+
Articulatory complexity	○

Table 4.2: Summary of markedness criteria for [æ]: Open circles ○ denote no (or inconclusive) evidence; (+)+ refers to (strong) evidence for marked status; (–)– refers to (strong) evidence for unmarked status.

4.2.2 *Frequency and orthography*

Looking at the frequency distribution of English vowel phonemes, we find that the rate for TRAP, about 1.6 per 100 segments, is close to the average (figures based on Cruttenden 2014: 159). The grapheme-to-phoneme correspondence for TRAP in English is unambiguous: Its orthographic representation is almost exclusively <a> (Cruttenden 2014: 119). The grapheme <a>, however, may also denote DRESS in a few high-frequency words (*many, any*). DRESS is most frequently spelled <e ea> (*get, head*).

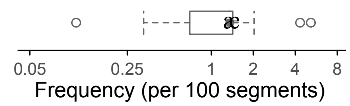


Figure 4.12: Frequency of TRAP relative to other English vowels (n = 20), expressed as the number of instances per 100 segments. Rates are shown on the log scale, but tick mark labels give the absolute values of the rates. Data are from Cruttenden (2014).

4.2.3 *Similarity*

A number of studies have shed light on German learners' perceptual sensitivity to cues relevant for the TRAP-DRESS distinction. Weiher (1975) investigated the perception of this contrast by $n = 12$ early-stage instructional-setting learners from northern Germany.[22] A discrimination and an identification task with nonce words (n = 132) were used. In the discrimination task, subjects heard sets of 4 stimuli. Three of these were identical and the task was to sort the odd one out (e.g. from the set [emps]–[æmps]–[emps]). In the identification task, an X-ABC design was used. Subjects heard a stimulus and had to match it with one of the three forms that followed (e.g. [sleŋz]—-[slæŋz]–[slɪŋz]–[sleŋz]). The learners' performance on both tasks was excellent, with an overall error rate of only .04.

[22] Subjects study were between 10 and 11 years of age, with an average of 0.5 years of instruction. Weiher (1975) reported findings by school type. For our present purpose, however, these learners represent the same population (early-stage instructional-setting learners) and figures were therefore pooled.

Further evidence for German children's sensitivity to front open vowel contrasts was reported by Butcher (1976). The study assessed German, French and BrE listeners' perception of a number of language-independent vocalic contrasts. Among the contrasts tested was that between the cardinal vowels [ɛ] and [a]. We may interpret this contrast as reflecting perceptual sensitivity in the open front region of the vowel space. Perceptual dissimilarity ratings were obtained for two age groups (10–12 vs. 17–22 years) in each L1 (n = 26–34 in each of the 6 groups). The results are shown in Figure 4.13, where the vertical axis denotes the amount of perceptual sensitivity in the open front region of the vowel space. There was a clear effect of age: Children judged [ɛ] and [a] to be less similar than adults. Perceptual desensitization also varied by L1. Overall, L1 French and English listeners, whose native languages have contrasts in this region of the vowel space, showed greater sensitivity. Even though age effects surface in each L1 group, the decline was most apparent among German listeners: While German children showed good perceptual sensitivity, adults perceived almost no difference between the vowels.

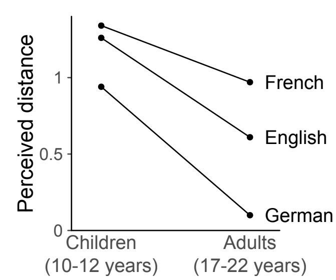


Figure 4.13: Decline of perceptual sensitivity in the front low area of the vowel space: Dissimilarity ratings for cardinal [ɛ a] by age and L1; data from Butcher (1976); values from Tables VII, VIII, IX, X, XI, XII.

Barry (1977) used synthesized stimuli to test $n = 26$ German learners'[23] perceptual sensitivity to spectral contrasts along the acoustic continuum crossing the TRAP-DRESS-STRUT category boundaries. Learners were compared to a control group of $n = 6$ native speakers. A set of $n = 14$ stimuli represented equal steps along the *bet-bat-but* continuum and the participants were asked to categorize each token as either *bet*, *bat*, or *but*. The results are shown in Figure 4.14, where the horizontal axis represents the acoustic continuum and the vertical axis shows the share of tokens for each stimulus that was identified as DRESS (dotted line), TRAP (solid line), or STRUT (dashed line). Responses by the native listeners produced clear boundaries between the vowel categories, which are marked in the diagram using light grey lines. Findings for the German learners revealed the following patterns: (i) the TRAP-DRESS boundary assumed a similar location on the continuum but was weakly defined; (ii) the STRUT-TRAP boundary was well-defined but shifted to the left – that is, compared to native listeners, STRUT occupied a larger portion of the perceptual space. These findings show that the differentiation between DRESS and TRAP is more difficult for German learners than that between TRAP and STRUT.

A cross-sectional study by Wieden & Nemser (1991: 160) reported on the development of L2 speech perception in $n = 384$ Austrian learners at 6 proficiency levels ($n = 64$ each). The authors used years of instruction as an indicator of developmental stage, which ranged from 0 years (grade 3; 8–9 years of age) to 8 years (grade 11; 16–17 years old). In an X-ABC matching task, [fæps] was contrasted with two sets of distractors: English [fʌps], [feps] and German [faps], [faːps].[24] Perceptual matchings are summarized in Figure 4.15, which shows how the categorization of [fæps]-stimuli varied across developmental stages. The top panel gives the results for the [æ ʌ e]-set. The accuracy rate was .60 at early stages. TRAP was slightly more likely to be misheard as [ʌ] than as [e] at most stages. Similar patterns emerge in the lower panel, which plots the [æ a aː]-set. The initial accuracy rate was .50. While [a] did not figure prominently as a perceptual substitute, [aː] was a frequent mismatch for TRAP. The fact that TRAP is more likely to be confused with [aː] than [a] may point to reliance on duration in discrimination tasks involving this unfamiliar vowel.

Flege et al. (1997) provide further insights into cue reliance by German listeners. The study included three groups of $n = 10$ adult subjects each: native speakers of AmE and two L1 German groups differing in English-language experience.[25] The authors' findings are summarized in Figure

[23] The learners in this study were between 17 and 24 years of age and had received between 6 and 9 years of instruction. Barry (1977) reports data for a test group ($n = 10$) and a control group ($n = 16$). For the present review, these were combined by computing a weighted average.

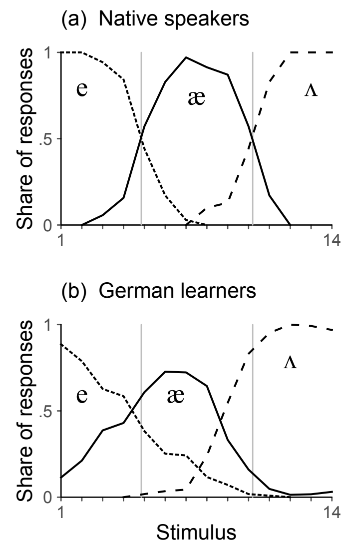


Figure 4.14: Perceptual sensitivity to spectral information: (a) native speakers and (b) German learners. The graph shows identification curves for 14 synthetic stimuli covering the vowel qualities for DRESS, TRAP, and STRUT; adapted from Barry (1977: 130-131) with permission. For the German learners, the data for the test group ($n = 10$) and the control group ($n = 16$) were pooled and panel (b) shows the weighted average over the two groups.

[24] The vowel symbols in these phonetic transcriptions are meant to represent those vowel qualities that are typical for the respective phonemes in English and German.

[25] The experienced group had been living in an English-speaking environment for an average of 7.5 years, compared to 0.6 years for the inexperienced group.

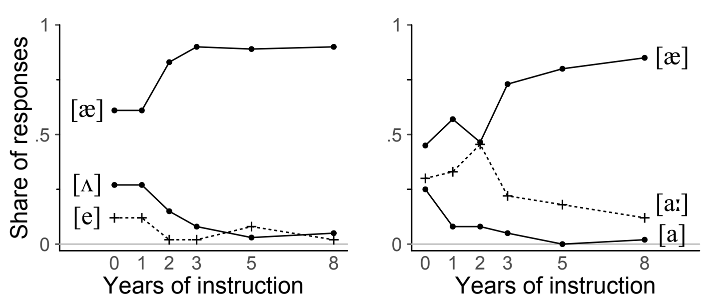


Figure 4.15: Perception accuracy of TRAP against two English distractors DRESS and STRUT (left) and two German distractors (right); adapted from Wieden & Nemser (1991: 160) with permission.

4.16, which shows the degree to which each group relied on spectral and temporal cues to distinguish TRAP and DRESS. Whereas native speakers mainly resorted to spectral information, the inexperienced German listeners relied strongly on length. The experienced group was more similar to the native speakers in terms of cue weighting, but differences remained. Compared to inexperienced listeners, they made greater use of spectral cues and less use of temporal signals. These findings suggests that (i) adult learners can learn to adjust their L2 perceptual weightings, and (ii) more experienced learners differ from less experienced ones in cue reliance for the TRAP-DRESS contrast.

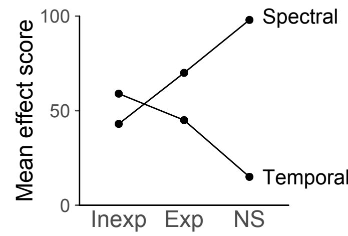


Figure 4.16: Cue reliance for the TRAP-DRESS distinction by Inexp(erienced) and Exp(erienced) German learners in comparison to native speakers; data from Flege et al. (1997).

To summarize, there is evidence that young learners show good perceptual discrimination between DRESS and TRAP, and that spectral sensitivity in the auditory region of TRAP appears to decrease after the age of 12. Overall, German listeners show greater sensitivity along the TRAP-STRUT compared to the TRAP-DRESS continuum. However, considerable variation in perceptual sensitivity to quality contrasts has been observed in adults. It appears that German learners are biased toward temporal signals for discriminating DRESS and TRAP, both as students in an instructional setting and adults with a limited amount of language experience. On average, perception skills improve for students in instructional settings. Adult learners may adjust cue reliance as a result of language experience, with the balance tipping in favor of spectral information (vs. duration). Nevertheless, they may still show perceptual foreign accents in comparison to native speakers.

4.3 L2 acquisition of TRAP: Theoretical predictions

We now consider German learners' acquisition of the TRAP-DRESS contrast from the viewpoint of L2 speech theories. Each framework is consulted to arrive at predictions about expected patterns in the acquisition process. Following this, theory-based statements are summarized in Table 4.5.

CAH For L1 German learners, TRAP is a new allophone of a new phoneme, which suggests relative ease of acquisition. Based on a contrastive analysis, German /ɛ ɛː a aː/ are the most likely substitutes.

SLM With German [a aː ɛ ɛː] as perceptually close L1 sounds, the degree of perceived dissimilarity determines whether a learner identifies TRAP as a new sound or equivalence-classifies it with an L1 vowel. While there is evidence to suggest that German learners are sensitive to the TRAP-DRESS contrast (Weiher 1975), they have also been observed to experience greater difficulty at the TRAP-DRESS boundary than at the TRAP-STRUT boundary (Barry 1977), suggesting that [ɛ ɛː] are perceived as closer in terms of vowel quality. As for vowel length, Wieden & Nemser's (1991) findings indicate that TRAP is perceived as more similar to German [aː] (vs. [a]) and, extrapolating from this finding, also to long [ɛː] (vs. short [ɛ]). Flege et al. (1997) also observe an overreliance on duration for the TRAP-DRESS contrast. Nevertheless, the duration contrast for the German tense-lax vowel pairs /ɛ ɛː/ and /a aː/ is typically much larger than that between English TRAP and

DRESS (cf. Table 4.1 and Figure 4.8). Based on the empirical evidence, we may therefore assume that (i) [ɛ ɛː] are the L1 sounds perceptually closest to TRAP; and (ii) the degree of perceived similarity between [ɛ ɛː] and TRAP is moderate to high.

According to the SLM, the distinction between DRESS and TRAP in speech production hinges on the underlying perceptual categorization of both TL sounds. Based on existing evidence, TRAP may be perceived by learners as a new sound or as an exemplar of German /ɛ/ or /ɛː/. The same is true for DRESS, for which equivalence classification with /ɛ/ or /ɛː/ appears to be more likely, however. These categorization options produce six scenarios[26], which are listed in Table 4.3. These patterns translate into different cue weights in speech production. Thus, the employment of length and quality cues depends on which categories the vowels are assigned to. Due to a consistent length contrast between TRAP and DRESS, pattern 6 is unlikely to occur and thus excluded from the following discussion.

[26] Since the aim of the present study is to shed light on the contrast between two vowel categories, we do not go into detail about within-category variability in categorization patterns. Thus, it may in fact be more appropriate to consider, at least for temporal cues, the position-sensitive nature of vowel duration, which varies systematically as a function of the voicing of the following consonant (see Chapter 10).

	Categorization		Contrast		
	TRAP	DRESS	Length	Quality	Comment
1	ɛː	ɛː	–	–	Merged categories
2	ɛ	ɛ	–	–	Merged categories
3	ɛː	ɛ	+	–	
4	New	ɛː	(+)	(+)	Depends on cue weights of new category
5	New	ɛ	+(+)	(+)	Depends on cue weights of new category
6	ɛ	ɛː	+	–	Unlikely

Table 4.3: Different patterns of perceptual categorization of DRESS and TRAP and corresponding use of length and quality as cues to the contrast in speech production.

Figure 4.17 shows the expected type of contrast in speech production graphically. The arrangement corresponds to that shown in Figure 4.9: The horizontal axis marks the degree of quality contrast and the vertical axis reflects the magnitude of the temporal contrast. The SLM states that a newly established IL category may differ from the TL category in cue weighting. The diagram attempts to represent this uncertainty with dashed regions for patterns 4 and 5. The hypothesized categorization patterns rule out cases in which a category contrast is signaled by quality only.

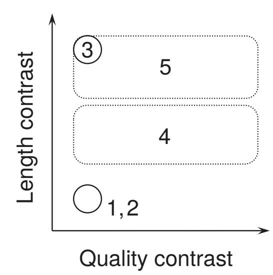


Figure 4.17: Implication of different categorization patterns for the implementation of length and quality for the TRAP-DRESS contrast in speech production. Numbers correspond to those given in Table 4.3. Dotted lines indicate uncertainty in the cue weights associated with the newly established category for TRAP.

FCM Stipulating German [ɛ ɛː a aː] as likely L1-derived perceptual substitutes, the relevant phonological features for FCM-based predictions are: long (0.50) > back (0.44) > low (0.13). Table 4.4 shows that German [ɛː] is predicted as the most likely substitute.

Vowel	Rank	[long]	>	[back]	>	[low]
[ɛː]	1	+		–		–
[aː]	2	+		+		+
[ɛ]	3	–		–		–
[a]	4	–		+		+
[æ]		+		–		+

Table 4.4: Substitutes for [æ]: Predictions of the FCM.

SCH/MDH The following assumptions find support from the literature review: TRAP is (i) a marked segment and (ii) relatively more marked than [ɛ ɛː a aː]. It is thus predicted to be a difficult vowel in L2 acquisition. As for the TRAP-DRESS contrast, acoustic evidence from BrE and AmE indicates that length and quality engage in a trading relationship, with a tendency for vowel length to compensate for relative proximity in quality. Learners may thus be expected to show similar strategies for maintaining a category contrast, with a possible (universal) bias toward duration, as suggested by Bohn (1995).

NM/NDH Following Donegan (1978), we will assume that raising and bleaching are natural processes that have to be suppressed in the acquisition of TRAP. These processes are active in German, but suppressed in English.[27] The L2 acquisition of TRAP thus requires German learners to suppress two natural processes. The NDH predicts difficulty for this type of acquisition scenario, with substitutes reflecting either raising toward [ɛ] or bleaching toward [a]. While natural processes interfere with the production of target-like vowel quality, no such limitations exist for the acquisition of length features. Length contrasts are thus expected to emerge before quality contrasts. The NM further takes into account frequency effects. Based on the figures provided by Cruttenden (2014: 119), TRAP is a moderately frequent structure and occurs at an average rate in learner input and output. Frequency is therefore expected to be neither facilitative nor inhibitory in the acquisition of TRAP.

[27] Bleaching is arguably resurfacing in present-day sound changes in both BrE and AmE dialects (see §4.1).

FM The FM holds that once [æ] is perceived as /æ/, a new faithfulness constraint is established. Sensorimotor learning produces a new undominated articulatory constraint. At stage 4 of the FM, the learner's speech production first evidences learning of the new vowel, where the model posits variable production as a function of focus on form. Once the correct ordering of the constraints has been established, the output will be overly faithful to the underlying representation. The final stages are characterized by an increased interplay of functional principles that operate at the post-lexical level. In casual utterances and unstressed contexts, less canonical realizations are expected to surface, with a more centralized vowel quality (vowel undershoot) and possibly shorter duration. Vowel duration at this stage may reflect context-dependent lengthening (before voiced obstruents) and clipping (before voiceless obstruents). Given the FM's statistical view on learning, input and output frequency are expected to affect perceptual and sensorimotor learning in a way that is neither debilitative nor facilitative.

MSA To the best of my knowledge, there is no evidence indicating natural physiological constraints on the production of front open vowels. Given this state of knowledge, the MSA reduces to the SLM.

LTD The LTD predicts target-like [æ] to first emerge in structurally strong syllables. While this affects stressed syllables at the lexical level, the acquisition of prosodic representation yields finer-grained strength asymmetries in connected speech. Based on these strength values, segments are predicted to differ in distinctness of the features [front], [low] and [quantity]. Further, the LTD predicts variability in TRAP-production to become increasingly systematic over the course of acquisition, reflecting the influence of phonetic and prosodic context.

OPM The discussion above has shown that two of the linguistic factors highlighted in the OPM, similarity and markedness, are likely to play a role in the acquisition of TRAP. The OPM does not explicitly address the acquisition of structures that are both similar and marked. However, based on the assumption that perception precedes production, we can extend the OPM to describe the acquisition of structures that are similar and marked. The hypothesized interplay of L1, L2, and U is shown in Figure 4.18. The acquisition of a structure that is both similar and marked may thus be assumed to proceed at a very slow rate, expanding over 9 developmental stages.[28] Assuming that the perception of differences between TRAP and similar L1 categories are a prerequisite for learning, the initial stages resemble those of the Similarity Corollary. Transfer thus dominates up to the intermediate stages. Once a new category is established, IL development is shaped in analogy with the Markedness Corollary, with U emerging as an influential force. TL structures emerge at a very late stage when other normal, similar or marked structures have already been acquired.

[28] This is in contrast to 7 stages for marked or similar structures and 5 for normal structures.

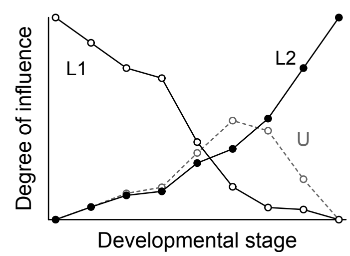


Figure 4.18: Extrapolation of the OPM to structures that are both similar and marked. This graph should be compared to Figure 2.6.

The effect of similarity on the acquisition of the TRAP-DRESS contrast was discussed above. Based on the OPM, at later stages U is expected to exert influence on both vowel quality and vowel duration. In particular, vowel quality is expected to show a bias toward unmarked variants – that is, toward a front mid or central open quality. Acquisition of the front low vowel quality may thus be impeded further by a universal magnet-effect of the front mid area, and later deflected toward a central low quality. Assuming that vowel duration is indeed a universally accessible feature, we expect U to lead to overreliance on vowel duration at later stages.

Table 4.5 offers a summary of the set of predictions we have arrived at. Reviewed next are empirical insights on TRAP-production in GLE.

4.4 *English* TRAP *in German Learner English: Previous work*

This section gives a survey of previous work on the acquisition of TRAP by German learners. After a discussion of auditory investigations, we will turn to acoustic studies.

Auditory studies

Insights into TRAP-production by German learners have been reported by a number of auditory studies. Sieg (2004, quoted in Wode 2009) investigated speech production development in immersion-setting learners between 6 and 10 years of age. Correct production rates for TRAP remained low throughout grades 1 to 4 (.10/.25/.20/.15). Learners almost categorically

Theory	Predicted patterns in L2 acquisition by German learners
CAH	Early acquisition of TRAP Substitutes: [a aː ɛ ɛː]
SLM	Late acquisition by the majority of learners Substitutes: [ɛ ɛː]
FCM	Perception-based substitute in speech production: [ɛː]
MDH	Late acquisition of TRAP
SCH	Substitutes: Bias toward length features
NM/NDH	No frequency effect for TRAP Late acquisition of TRAP Substitutes: [a aː ɛ ɛː]
FM	No frequency effect for TRAP Development: Deviant production > Stylistic variability > Overly faithful production > Variability reflecting connected speech processes Effect of speaking style (focus on form) at early stages of acquisition Style-dependent vowel undershoot at late stages
LTD	Acquisition order: Stressed > unstressed syllables
OPM	Late acquisition of TRAP Substitutes: [ɛ ɛː] at earlier stages, [a aː ɛ ɛː] at later stages More formal speech: Higher degree of TL-like production Less formal speech: Higher degree of L1 transfer

Table 4.5: The acquisition of TRAP by German learners: Summary of predictions based on theoretical work.

used [ɛ] as a substitute (> .98); [a] was also observed, but with a share of only .01 or less.

Another study including learners at an early stage of L2 acquisition was conducted by Weiher (1975), who analyzed the production of [ɪ e æ] by $n = 12$ northern German 5th-graders.[29] Three tasks were used: the repetition of nonce words, the repetition of English words and the reading and repetition of English sentences. The findings show that the correct production rate and the predominant direction of deviation varies with elicitation method and speaking style. For nonce words, the accuracy rate was .64; the majority of deviant productions was toward [a aː]. For the English words, accuracy was at .60, with no preferred direction of deviation. In the English sentences, the rate was .47 and most deviations were too close – that is, deflected toward DRESS.

These findings suggest an influence of speaking style and word familiarity on TRAP realization. More specifically, unfamiliar words (i.e. nonce words) triggered a preference for [a aː]; in contrast, no preference was found for familiar items. When pronouncing unfamiliar items, learners apparently rely more strongly on perception. Familiar English words, on the other hand, seem to have undergone merging with DRESS. In sentences, deviations were almost exclusively too close (.86) and thus similar to [ɛ]. This may be due to the fact in connected speech, less monitoring results in a stronger tendency toward L1-derived [ɛ].

An auditory study by Pascoe (1987)[30] explored the acquisition of a wide range of segmental features by learners from southern Bavaria. Among

[29] Learners were 10 to 11 years of age, with 0.5 years of instruction on average.

[30] See §3.3.1 for details.

the $n = 49$ features investigated, TRAP ranked lowest in terms of correct production rate (only .06).

In a further study on instructional-setting learners from southern Bavaria, Kucharek (1988: 132, 145) investigated TRAP-production in four groups: learners in grade 6 ($n = 169$), 7 ($n = 138$), 8 ($n = 117$), and 12 ($n = 94$) and teachers ($n = 62$), adding to a total of $n = 580$ participants. Production accuracy was assessed with 4 items (*rat, Dad, accident, bad*), which were elicited with a reading passage. The results, which are summarized in Figure 4.19, showed accuracy rates below .25 in all grades, with only moderate improvement from around .10 to .25. Among teachers, production accuracy was at .60.[31]

[31] This indicates that inaccuracies among instructional-setting learners may be partly attributable to faulty teaching.

Langguth (2009) reported on the production of TRAP by German learners at a grammar school in northern Bavaria (Bamberg). The study included $n = 7$ students each from grades 5 (age: 10 to 11), 9 (14 to 15) and 12 (17 to 18). The tasks included a word list and a (near-)minimal pair list. TRAP-accuracy was .30, with a moderate improvement across age groups (.25/.30/.40). Performance was better in the word pair task, where most pairs contrasted TRAP with DRESS (.40 vs. .20 for the word list). The most frequently observed substitutes were [ɛ] (.40) and [ɛː] (.20).

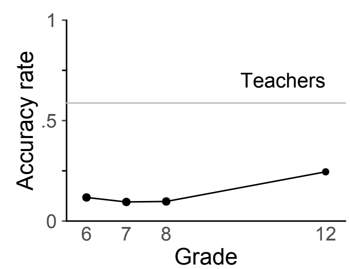


Figure 4.19: TRAP-accuracy in the productions of German instructional-setting learners in grade 6,7,8, and 12 in comparison to teachers; data from Kucharek (1988).

In sum, the evidence reported in auditory studies indicates that TRAP is acquired relatively late by German learners. Table 4.8 gives a summary of empirical work. Accuracy rates have been found to vary as a function of proficiency, task type, and speaking style. The predominant substitute for TRAP is [ɛ(ː)], but [a aː e eː ɛ̝] are also attested. Factors that have been observed to influence production accuracy and the distribution of substitutes are also listed. Values reflect absolute differences (Δ) on the proportion scale.

Table 4.6: Empirical evidence from auditory studies on TRAP in German Learner English: Reported accuracy rates, constraints on accuracy, and constraints on substitutes.

Results		Notes	Study
Accuracy rate	.47–.64	Grade 5; age: 10–11 years	Weiher 1975
	.25–.40	Grades 5, 9, 12; age: 10–18 years	Langguth 2009
	.10–.25	Grades 1–4; age: 6–10 years	Sieg 2004
	.10–.25	Grades 6–12; age: 11–18 years	Kucharek 1988
	.06	Grades 8–10; age: 14–16 years	Pascoe 1987
Constraints on accuracy rate			
Task type	Δ = .20	Minimal pairs (.40) > single words (.20)	Langguth 2009
Speaking style	Δ = .17	Isolated words (.60) > sentences (.47)	Weiher 1975
Constraints on substitutes			
Familiarity		Unfamiliar: [a aː], familiar: [ɛ a aː]	Weiher 1975
Speaking style		Sentences: [ɛ], single words: [ɛ a aː]	Weiher 1975

Acoustic studies

The acquisition of the TRAP-DRESS distinction by German learners has also been addressed using acoustic methods. Barry (1977) investigated the production of DRESS, TRAP and STRUT by $n = 10$ German learners from northern Germany (age: 17–19 years), who had been learning English at school for 6 years. Based on a perceptual assessment (see §4.2.3) two groups were formed: learners with ($n = 4$) and without ($n = 6$) established perceptual categories. Subjects with established vowel categories produced a greater spectral distance ($D_{Bk} = 1.68$)[32] than the perceptually less sensitive learners ($D_{Bk} = 0.33$).

[32] D_{Bk} will be used as a shorthand symbol to denote the Euclidean distance between vowels measured on the Bark scale.

A study by Kautzsch (2010) focused on university students of English who had been taught in an instructional setting for a minimum of 7 years. Two groups ($n = 20$ each) were compared: students in their first or second year at university, and more advanced learners in their third year or later. The vowels were acoustically more similar in the productions of beginners ($D_{Bk} = 0.27$) than in the advanced group ($D_{Bk} = 0.65$).

Maurer (2014) examined advanced university students who had completed a stay abroad in an English-speaking country for at least 4 months. Subjects were divided into two groups depending on the pronunciation model chosen (BrE vs. AmE; $n = 30$ each). The study also included a control group of less advanced students who had not been abroad ($n = 6$), and two native speaker groups (BrE and AmE; $n = 10$ each). For each vowel, $n = 11$ tokens from a word list task were analyzed. Degree of overlap in the F_1-by-F_2 vowel space (units: Hertz) was measured using Pillai scores (Olson 1976; Hay et al. 2006), which express the degree of spectral separation (0 = complete overlap; 1 = perfect separation). The author observed that learners differed markedly from native speakers in the degree of spectral separation between DRESS and TRAP. Nevertheless, a clear difference emerged between the advanced learners and the control group. The fact that Maurer reported Pillai scores makes it difficult to compare her findings to those obtained in other studies – no formant measurements were given. However, the appendix includes F_1-by-F_2 scatterplots for each speaker (Figures D.1–1 to D.4–10). These allow us to eyeball the average F_1 and F_2 produced by each subject. These rough estimates were transformed to Bark and the distance between DRESS and TRAP was determined for each speaker. Figure 4.20 shows the distribution of these distances by group. Overall, the control group shows the smallest spectral distance ($D_{Bk} = 0.31$) and among more advanced students, AmE-oriented learners produced smaller acoustic distances ($D_{Bk} = 0.74$) compared to the BrE group ($D_{Bk} = 1.16$). Figure 4.20 suggests considerable variability among advanced learners.

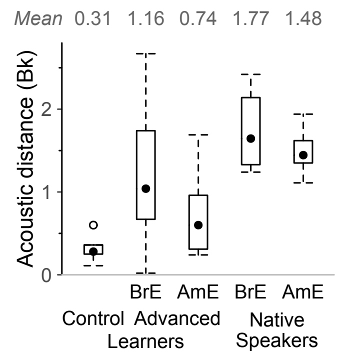


Figure 4.20: Spectral distance between TRAP and DRESS in German learners compared to native speakers; the plotted values were approximated by eye based on the scatterplots provided in the appendix of Maurer (2014).

Further insight is provided by Steinlen (2005), who studied $n = 10$ native speakers of BrE and $n = 10$ learners from northern Germany who had learned English in an instructional setting for about 9 years. Their average age was 26 years and none had been to an English-speaking country. The learners almost completely merged the two vowels in the F_1-by-F_2 space ($D_{Bk} = 0.11$; $D_{zBk} = 0.09$). The native speakers' TRAP-to-DRESS duration ratio was larger than that of the learners (1.22 vs. 1.09). Figure 4.21 compares the findings by Steinlen (2005) to the measurements for native speakers.[33]

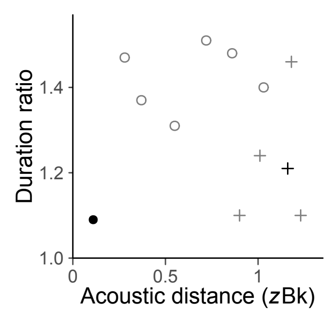


Figure 4.21: Acoustic cues for the TRAP-DRESS contrast: comparison of Steinlen's (2005) measurements (black) with those reported for BrE (+) and AmE (○) native speakers (cf. Figure 4.9).

[33] Note that the Euclidean distances in this figure are based on z-transformed Bark measurements (D_{zBk}). For the source of the reference values, see Figure 4.9.

Flege et al. (1997) studied the production, perception and intelligibility of front vowels in three groups of $n = 10$ adult subjects each: native speakers of AmE and two L1 German groups differing in English-language experience.[34] The results showed consistent differences between the learner

[34] The experienced group had lived in an English-speaking environment for a longer time (an average of 7.5 vs. 0.6 years).

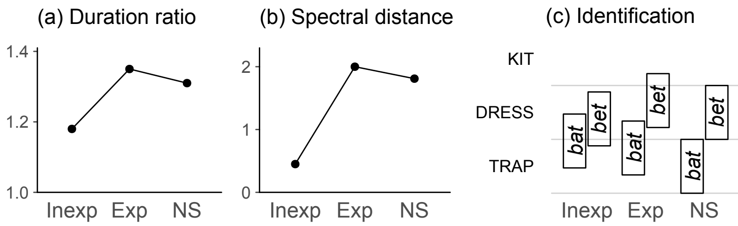


Figure 4.22: Cue reliance by English-language experience. Right panel: Location of the rectangle indicates identification rate as KIT, DRESS and TRAP. Data are from Flege et al. (1997).

groups, which are summarized in Figure 4.22a and b. The experienced group produced a spectral and durational contrast similar to that of native speakers. The inexperienced group did not contrast the vowels to the same degree. The difference between the groups is reflected in the duration ratios (1.18, 1.35, 1.31) and spectral distance in the Bark-difference space, expressed as a Euclidean distance (0.45, 2.00, 1.84).[35] The intelligibility scores show that the learners' renditions of *bet* and *bat* were often misidentified by native speakers. The identification patterns are summarized in Figure 4.22c. In the inexperienced group, .47 of the *bat*-tokens were misheard as *bet*. In the experienced group, the confusion rate was .34. It was further observed that .22 of the *bet*-tokens produced by the experienced group were identified as *bit*. This may reflect a change of the articulation of DRESS as a strategy for category separation.[36] None of the *bed*-tokens in the inexperienced group were misidentified as *bid*.

To sum up, acoustic studies corroborate auditory work in showing that TRAP is acquired late. German learners typically produce an insufficient spectral contrast between DRESS and TRAP. Nevertheless, certain learners are able to achieve a native-speaker level. Figure 4.23 summarizes the acoustic distance scores derived from the literature and Table 4.7) lists formant measurements. In terms of the realizations of target TRAP that are observable in German learner productions, acoustic data suggests a gradual development toward a native-speaker like contrast. Further, learner productions show sensitivity to the target accent, with learners aiming for BrE on average producing larger spectral contrasts than those opting for the AmE model. Studies investigating both spectral and temporal cues suggest that the duration contrast is acquired before the spectral contrast.

[35] This is yet another scale on which the acoustics of vowels can be expressed. Unfortunately, these values are not directly comparable to the D_{Bk} and D_{zBk} scores we are using.

[36] These findings are consistent with observations by Arnold & Hansen (1974: 58), who note that DRESS may be raised to distinguish it from TRAP.

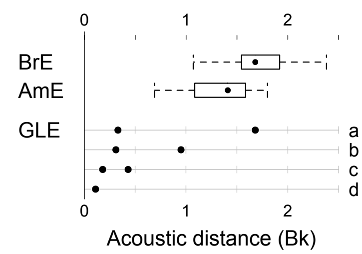


Figure 4.23: Summary of the instrumental evidence on the spectral TRAP-DRESS contrast in German Learner English in comparison to properties of the target language varieties (top). The plotted values are Euclidean distances on the Bark scale. Boxplots show the distribution of scores in native speech. Key to the studies: (a) Barry 1977, (b) Maurer 2014, (c) Kautzsch 2010, (d) Steinlen 2005.

Table 4.7: Acoustic evidence for the spectral contrast (Bk) between TRAP and DRESS in German Learner English.

			DRESS		TRAP		Distance	
Study	*N*	Proficiency	F_1	F_2	F_1	F_2	Bk	*z*Bk
Barry 1977	6	Lower	5.3	12.0	5.6	12.2	0.33	—
	4	Higher	5.2	12.4	6.8	11.9	1.68	—
Steinlen 2005	10	—	5.3	11.9	5.4	11.9	0.11	0.09
Kautzsch 2010	20	Lower	5.1	11.9	5.4	11.8	0.27	—
	20	Higher	5.2	12.1	5.8	11.8	0.64	—
Maurer 2014	6	Lower	5.9	12.2	6.2	12.2	0.31	—
	60	Higher	6.2	12.8	7.0	12.3	0.95	—

4.5 *Aims of this study*

The objective of this study is to jointly consider vowel length and quality in the acquisition of the TRAP-DRESS contrast by German learners. The analysis is motivated by the following questions:

- Are the initial stages of IL development characterized by merging of the vowels in terms of both length and quality?
- In the course of IL development, are length and quality cues learned simultaneously or are learners biased toward one of the two?
- How do learners at high proficiency levels implement this contrast? Do BrE- and AmE-oriented learners align with the acoustic values for their target varieties?

Category separation among lower-proficiency learners

As illustrated in Figure 4.17, the implementation of the TRAP-DRESS contrast in speech production may vary depending on how learners perceptually categorize these vowels. Focusing on vowel length and quality as (potentially) distinctive cues, the contrast may be signaled by both (categorization patterns 4 and 5), only length (3 and 6) or neither (1 and 2). While acoustic studies have shed light on the TRAP-DRESS contrast in learners aged 16 or older, evidence on early-stage instructional-setting learners is lacking. This study aims to address this gap by investigating whether the lack of length and quality contrast that has been reported for less proficient subjects (Barry 1977; Bohn & Flege 1992; Flege et al. 1997; Steinlen 2005; Kautzsch 2010; Maurer 2014) holds when looking at a larger cross-section of developmental stages. With perception studies indicating that German learners aged 12 or younger are more sensitive to quality contrasts in the open front portion of the vowel space, evidence for merging of DRESS and TRAP by adult speakers may not generalize to this population of learners.

Cross-sectional patterns in cue development: Length vs. quality

Previous experimental work indicates strong tendencies toward TRAP-DRESS merging in lower-proficiency learners. The acquisition of the contrast thus appears to follow after a period insufficient category separation on both dimensions. In Figure 4.24, this phase is reflected by position in the lower left corner, which denotes the absence of a length and quality contrast. Following this stage, two developmental paths along the length and quality grid appear particularly plausible. If length and quality are acquired simultaneously, this would be reflected in a straight line, which is shown as (a). In case of a bias toward length, development on the grid should be curvilinear (as in (b)), reflecting the acquisition order "length before quality". A bias toward length is predicted by the SCH and Bohn's (1995) DH. The few indications in the literature suggest an earlier emergence of a length contrast (Flege et al. 1997; Steinlen 2005). This study shall throw further light on the relationship between these cues in the productions of learners at different levels of L2 pronunciation ability.

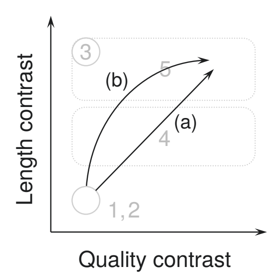


Figure 4.24: Emergence of length and quality cues in the acquisition of the TRAP-DRESS contrast after a phase of merging. TRAP-DRESS merging is denoted by position in the lower left area signaling insufficient length and quality distinction between the vowels. Two hypothetical paths along the length and quality grid are shown, marking (a) the simultaneous acquisition of both cues, and (b) an initial bias toward length to signal the contrast between the vowels.

Category separation by proficient learners

With the two standard varieties providing different targets for the TRAP-DRESS contrast, the relative weight given to length and quality cues is expected to vary depending on the target variety toward which they are oriented. Further, motivated by economy principles, learner speech may generally show a trading relationship similar to that observable in native speech, with lack of contrast in one being compensated for by the other.

4.6 Method and data

Informants, materials and procedures

The following analyses are based on vowels elicited in the short-phrases task.[37] Each subject produced $n = 12$ tokens (6 for each vowel), which sums to $n = 654$ vowels for the learners ($n = 26$ missing cases) and $n = 278$ for the native speakers ($n = 8$ missing). Each speaker produced the following items:

- DRESS: *get, set, pet, het, bed, head*
- TRAP: *sat* (2x), *cat, hat, bad, had*

[37] The data sets analyzed in the present chapter are openly available via the *Tromsø Repository of Language and Linguistics* (*TROLLing*) at https://doi.org/10.18710/ATIRRV (Sönning 2020c).

The acoustic analysis was carried out in Praat (Boersma & Weenink 2019) and the data were segmented manually following the principles outlined in Machač & Skarnitzl (2009). Vowel duration was determined using the onset and offset of F_2 and the vowel target was determined visually at the point of maximal displacement (Di Paolo et al. 2011). The F_1 and F_2 measurements were transformed from Hertz to Bark using the formula by Traunmüller (1990) and then normalized to *z*-scores using the method proposed by Lobanov (1971). The label *z*Bk will be used as a shorthand for these transformed and normalized formant frequencies.

Using the Bark instead of the Hertz scale to measure vowel quality gives a closer reflection of auditory differences between vowels (Thomas 2011: 158). The aim of normalization, on the other hand, is to make vowel measurements directly comparable between speakers. Thus, on the Bark scale, the location of vowels is influenced by physiological differences between speakers – that is, it varies with vocal tract length (ibid.: 160–171). Normalization techniques aim to control for this variation and allow for comparison of vowels produced by different speakers on the same scale and in the same display. The method proposed by Lobanov (1971) has been observed to yield good results (e.g. Adank et al. 2004; Flynn 2011). In order to apply this normalization technique, 5 to 7 tokens of each monophthong were elicited per speaker (i.e. a total of about 6,000 tokens) and the average F_1 and F_2 were calculated for each vowel. These averages were then used for normalization.

The ensuing analyses are carried out on the same scale as the literature survey presented above, which allows for direct comparison of results. Vowel durations (in ms) were transformed to the (natural) logarithmic scale for analysis. For the presentation of findings, scores are presented as ratios for ease of interpretation and comparability to the literature. The next section outlines the statistical analysis and may be skipped without loss of continuity.

Statistical analysis

Analyses were carried out separately for learners and native speakers, which means that two models were fit.[38] As there are three acoustic values describing a single vowel token uttered by a speaker (duration, F_1, and F_2), a multivariate regression model was used, which simultaneously addresses three outcomes. To guard against potential outliers on these outcome dimensions, a robust Student-*t* distribution was chosen to represent these acoustic measurements. The models also included, for each outcome, varying intercepts on subject and word, which were also given a joint multivariate distribution. This is to say that each subject was assigned a unique duration and location for each vowel in the F_1-by-F_2 space. The acoustic values of DRESS and TRAP were thus allowed to vary between subjects, and the objective of the model was to detect systematic differences in vowel location and duration between speaker groups. In the level-2 part of the model, the varying intercepts for subjects, that is, the speaker-specific duration and formant frequencies for the vowel categories, were regressed on predictors of interest:

[38] The models are described in more detail in Appendix A.3.

- For learners, these were *target variety* and *proficiency level*: The model describes, for each vowel category (DRESS and TRAP), how duration, F_1, and F_2, considered jointly, vary depending on the proficiency level and the target variety of the speaker. From intermediate stages onwards, the model is open to the possibility of divergent patterns for learners oriented toward different standard varietes. Both predictors were enriched with varying slopes for the lexical items.
- For native speakers, *variety* was the focal predictor. This means that the three acoustic parameters were allowed to vary between BrE and AmE speakers. The model also included by-word varying slopes for variety.

The phonetic context, more specifically the voicing of the following segment, is known to influence the duration of a vowel (see Chapter 10). Even though voicing contexts were balanced across the two vowel categories, the voicing of the following consonant was included in the model to partial out this source of variation in vowel duration. The voicing of the following context was therefore included as a predictor in the level-2 model for varying duration intercepts on word.

As for the prior distributions, previous instrumental work on DRESS and TRAP in BrE and AmE, which was summarized in §4.4, was used to construct prior distributions for parameters. Figure 4.25 shows, for the F_1-by-F_2 vowel plane and the log duration scale, how the acoustic values reported in previous work map onto the priors. For GLE, there is little quantitative work to build on. Accordingly, priors were vague but reasonable, informing the model that TRAP and DRESS should be somewhere in the central to front and mid to low area of the vowel space. As for duration, the measurements for [a aː ɛ ɛː] listed in Table 4.1 were used as reference values to construct a prior that expects vowel duration in the

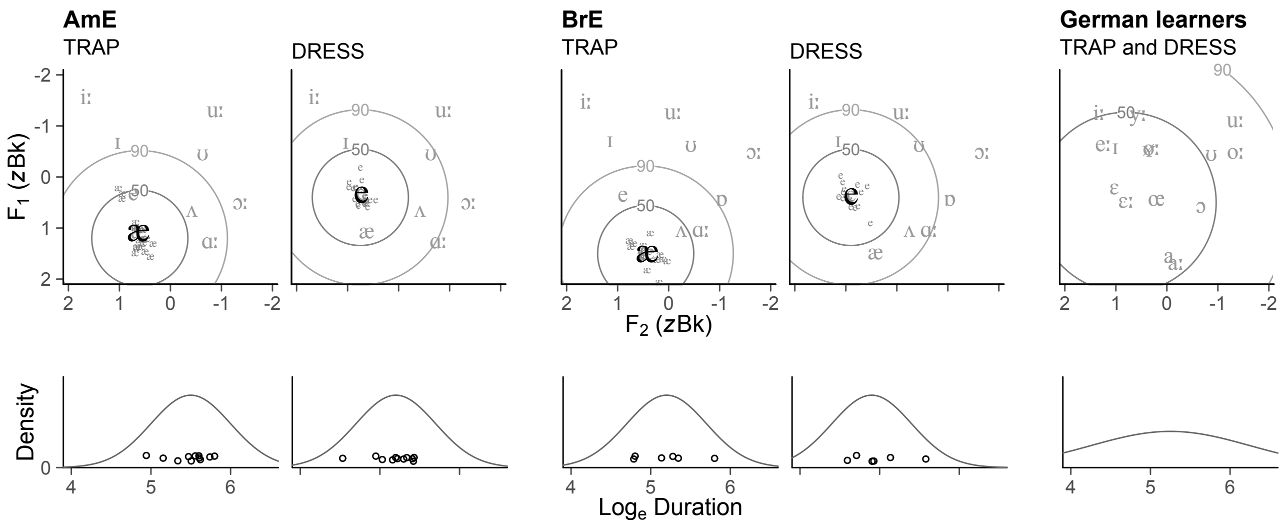


Figure 4.25: Prior information on the acoustic properties of TRAP and DRESS in AmE (left), BrE (center), and German Learner English (right). In the top row, 50% and 90% density ellipses show the prior distribution in the F_1-by-F_2 plane. The IPA symbols show the average across studies in the literature (large grey symbols) and, for DRESS and TRAP, the scatter of values across studies (small grey symbols). In the bottom row, density curves represent the prior distributions for log duration, overlaid on the values reported in previous work, where each circle denotes a study. For German learners, there is little quantitative work to build on and the priors are therefore more diffuse.

range from 60 to 250 ms with probability .68, thus easily allowing for more extreme parameter values.

The prior specification for dispersion parameters relied on the fact that formant measurements are expressed as *z*-scores. Since a standard deviation of 1 describes the scale of the scatter of vowel categories in the F_1-by-F_2 plane, the dispersion for a single category, whether assessed across words, subjects, or measurements, is certainly expected to be (well) below 1. Our priors therefore assign probability .86 to the interval $[0, 1]$.

Appendices A.3.1 (native speakers) and A.3.2 (learners) describe the model specification (including all priors) in detail, and report convergence diagnostics and the posterior distribution of all model parameters.

4.7 *Constraints on generality*

Before we turn to the results of the present investigation, let us delimit the scope of our inferences.

- *Language-external scope*: The sampled population may be considered as largely representative of German Learner English as spoken predominantly in southern Germany, more specifically northern Bavaria. It appears that there are no dialectal features that would limit the generalizability of the present findings to other, especially Standard High German, populations of instructional-setting learners. Especially the lower proficiency levels represent the type of English that may be considered typical of instructional-setting learners. Our findings therefore do not extend to naturalistic learners.
- *Language-internal scope*: The degree of generalizability to other instances of these vowel categories is limited. Where possible, the test words were chosen with the aim of (i) minimizing coarticulatory effects (hence a preference for /h/-onsets and alveolar onsets and codas) and (ii) maximizing segmentability (hence a dispreference for sonorants). The

findings, then, only apply to such phonetic contexts. Furthermore, all instances were read with a nuclear accent, so our insights are limited to prosodically prominent environments.

- *Speaking style*: As we relied on a reading task for data elicitation, the results extend to this type of speech production only. We must also take into consideration the possibility of orthography-induced speech behavior, which suggests that these data should be considered as representative of careful reading style only.

4.8 Results

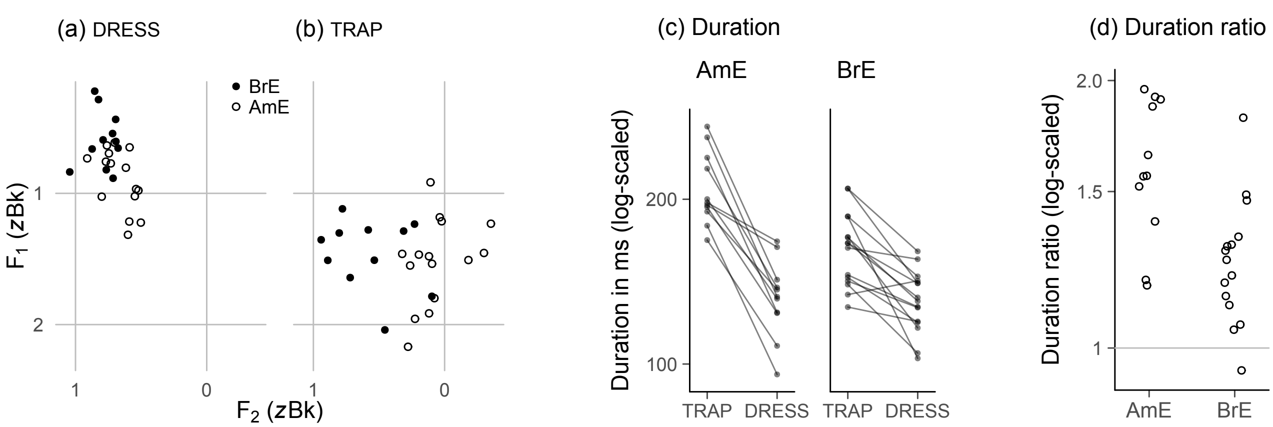
We report our findings for the natives speakers (§4.8.1) and the learners (§4.8.2) in turn.

4.8.1 Native speakers

Before we consider the output of the regression analysis, we will take a look at the distribution of our measurements. To this end, we first average the formant and duration measurements for each speaker and then inspect these averages visually. Figure 4.26 offers a descriptive summary of the native speaker data. The left-hand panels show the distribution of speakers in the vowel space, where each point denotes, for a single speaker, the average F_1-by-F_2 calculated over all tokens for the vowel category. The average readings for DRESS are less dispersed than those for TRAP. Both vowels seem to pattern differently in the two groups: For DRESS, BrE subjects show lower F_1 values, indicating a slightly higher position in the acoustic vowel space. For TRAP, BrE speakers show lower F_1, on average, which reflects a retracted vowel quality.

Figure 4.26c graphs the duration averages for each vowel category. The averages for a speaker are connected by a line, and the slope of the line indicates the direction and magnitude of the durational contrast. Lines are steeper for the AmE subjects, who produce longer TRAP-tokens. We can express the durational contrast for each speaker with a ratio, which gives us a single value per speaker. Figure 4.26d compares the ratios for BrE subjects

Figure 4.26: Overview of the acoustic measurements for the $n = 26$ native speakers: (a) Average formant measurements for each vowel category: Open circles denote AmE subjects ($n = 11$), filled circles BrE subjects ($n = 15$); (b) Average durations for each vowel category by speaker and variety: Each line denotes a speaker; and (c) Average duration ratios by variety: Each point represents a speaker.



to those of AmE subjects. We see considerable variation within each group, and AmE subjects appear to show higher ratios, on average.

Let us now summarize these measurements more compactly so we can compare them to the values reported in the literature. 4.27a shows the estimates from our model in comparison to the ratios observed in previous work, which are summarized with boxplots (cf. Figure 4.8). Our results agree with those in the literature, with BrE subjects on average producing smaller TRAP-to-DRESS duration ratios (1.21 vs. 1.54 for AmE speakers). The wide uncertainty intervals reveal the imprecision of the estimates, which is due to (i) the small number of speakers recorded (n = 11 for AmE; n = 15 for BrE), and (ii) the large variation among speakers within each group, which was evident from Figure 4.26d.

Figure 4.27b locates the F_1-by-F_2 measurements for TRAP and DRESS relative to those reported in previous studies (grey; see Figure 4.5), with 50% and 90% uncertainty contours. AmE TRAP and DRESS agree with previous work: While both are slightly lower in this study, the distance between, and relative positioning of, the vowels is similar. BrE TRAP appears to be more retracted, and BrE DRESS more open compared to previous findings. As most informants were in their twenties, this could be attributable to recent sound changes – that is, opening of DRESS and lowering and retraction of TRAP (see §4.1). Estimates for the acoustic properties of TRAP and DRESS produced by AmE and BrE subjects are reported in Table 4.8.

Let us finally consider the relationship between temporal and spectral cues. Figure 4.27c compares the speaker groups in this study to the values documented in the literature (cf. Figure 4.9), with 50% and 90% uncertainty ellipses. The positioning of the native speaker groups conforms to the familiar clustering by variety: AmE speakers make greater use of length as a contrastive feature. The speaker groups recorded in this study make rather similar use of vowel quality, at least as far the distance in the vowel space is concerned. Figure 4.27b, however, demonstrates that these absolute distances arise from different spatial patterns. To the extent that F_1 and F_2

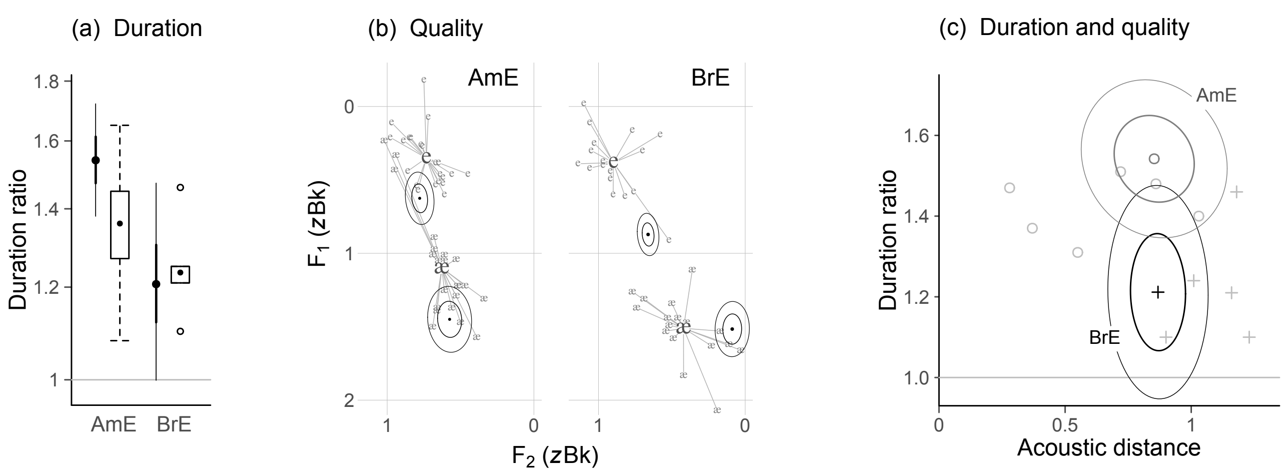


Figure 4.27: Estimates for the AmE and BrE native speakers in the present study in comparison to values in the literature: (a) Duration ratio - boxplots show values reported in the literature (n = 11 for AmE, n = 6 for BrE; see Figure 4.8); (b) Location in the F_1-by-F_2 plane; (c) Relationship between quality and length as contrastive features (cf. Figure 4.9). Both speaker groups pattern with the attested values for their respective varieties. Error bars and contours denote 50% and 90% posterior intervals.

	AmE					BrE				
		Posterior quantiles					Posterior quantiles			
Parameter	*Mdn*	.05	.25	.75	.95	*Mdn*	.05	.25	.75	.95
Duration ratio	1.54	1.38	1.47	1.61	1.72	1.21	1.00	1.12	1.30	1.47
DRESS										
F_1	0.63	0.47	0.56	0.69	0.78	0.87	0.74	0.82	0.92	1.00
F_2	0.78	0.69	0.75	0.81	0.87	0.66	0.60	0.64	0.69	0.73
TRAP										
F_1	1.45	1.25	1.37	1.53	1.65	1.52	1.35	1.45	1.58	1.69
F_2	0.57	0.44	0.52	0.63	0.71	0.09	–0.01	0.05	0.13	0.19

Table 4.8: Estimates of the duration ratio, F_1 and F_2 (zBk) for AmE and BrE native speakers.

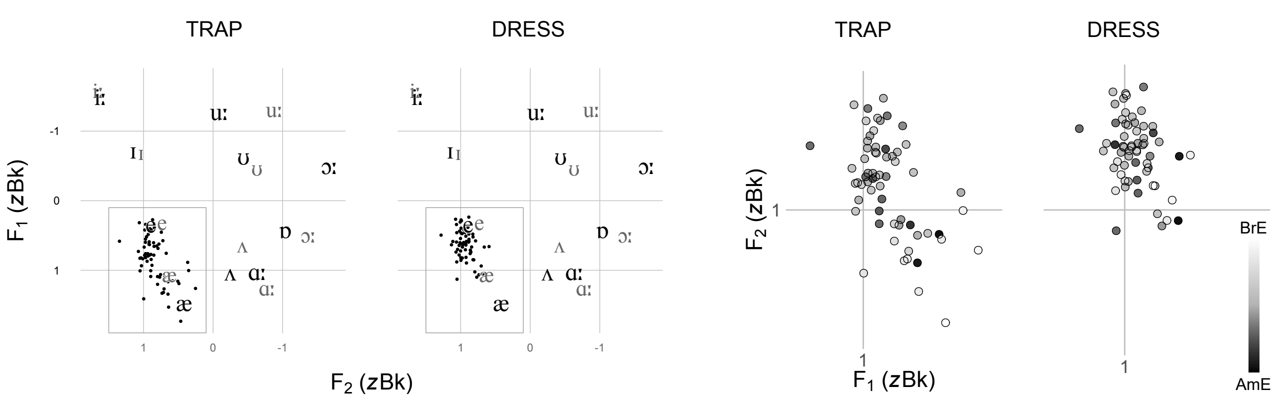
correspond to auditory vowel locations, AmE speakers primarily signal the contrast in terms of vowel height (F_1); BrE speakers, on the other hand, make use of both height (F_1) and advancement (F_2). This information is lost when considering only the absolute distance in the F_1-by-F_2 vowel space.

4.8.2 Learners

For the learner data, we also start with a closer look at the distribution of our measurements. To this end, we take averages of our F_1, F_2, and duration readings and graph them. Figure 4.28 shows the dispersion of our formant measurements in the acoustic vowel space, with native speaker values added for comparison (black: BrE, grey: AmE; these are the averages over studies shown in Figure 4.5). The right-hand panels zoom in to the two vowel categories and add information about the target variety orientation of learners (black: AmE, white: BrE). We note that, for TRAP, values that approximate those found in native speech tend to stem from learners with a clear orientation toward one variety. For DRESS, we see that these speakers also surface at the lower right end of the point cloud.

As for the duration of the vowel categories, Figure 4.29a graphs, for

Figure 4.28: Overview of the F_1 and F_2 measurements for the $n = 62$ German learners: The panels on the left show measurements in the context of the acoustic vowel systems for native speakers (black: BrE, grey: AmE; IPA symbols denote averages across studies surveyed above, see Figure 4.5). The rectangles show the area in the vowel space that is enlarged in the right-hand panels. The panels on the right show the dispersion of the measurements in the F_1-by-F_2 plane, where grey shading is used to indicate target variety orientation.



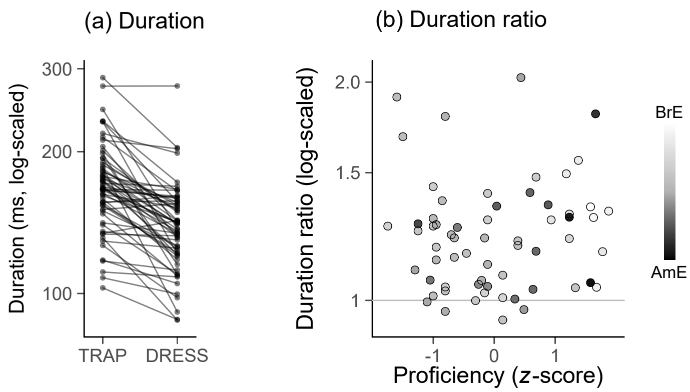


Figure 4.29: Overview of the acoustic measurements for the $n = 62$ German learners: (a) Average duration for each vowel category by speaker: Each line denotes a speaker; (b) Average duration ratios by proficiency: Each point represents a speaker and grey shading indicates target variety orientation.

each speaker, the average duration for each vowel, with the lines signaling the difference. On average, TRAP is longer than DRESS, but the magnitude of the difference varies markedly across learners. Figure 4.29b shows the TRAP-to-DRESS ratio by proficiency. There is considerable variation across the entire scale, and target variety orientation does not seem to be linked to differences in durational contrast.

A model-based summary of these patterns is shown in Figure 4.30, which distinguishes, at higher proficiency levels, between AmE-oriented (left panel) and BrE-oriented learners (right panel). The graphs show how the TRAP-to-DRESS duration ratio varies across proficiency levels, with 50% and 90% uncertainty intervals. At the right margin of each graph, our estimate for the respective native speaker group is displayed, along with the distribution of values from our literature survey (see Figure 4.27a). Given the high level of uncertainty, the information value of the duration data is low. We can summarize the patterns as follows:

- Average ratios vary between about 1.15 and about 1.30.
- There is a hint at a U-shaped trend, with learners of low and high proficiency averaging at higher ratios than intermediate learners.
- There is massive uncertainty in the estimates at advanced levels, but BrE-oriented and AmE-oriented learners do not diverge systematically from one another.

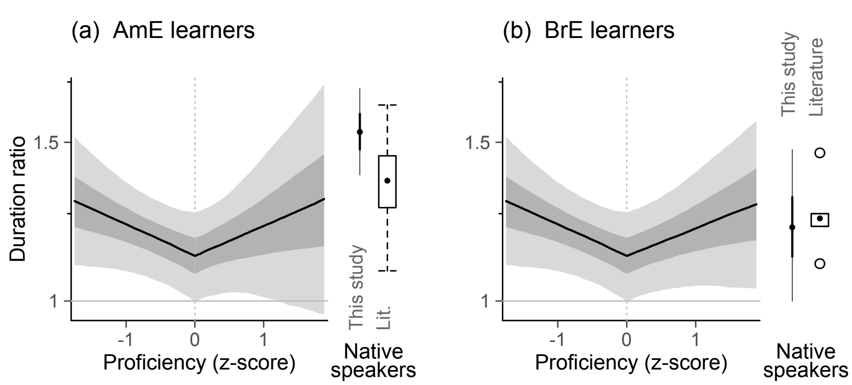


Figure 4.30: Use of length as a contrastive cue by German learners: Duration ratio by proficiency level. At higher proficiency levels, two populations of learners are distinguished: Those oriented toward AmE (left) and those oriented toward BrE (right). Estimates for the native speakers in this study are added at the margins. Error bars/bands reflect 50% and 90% posterior intervals.

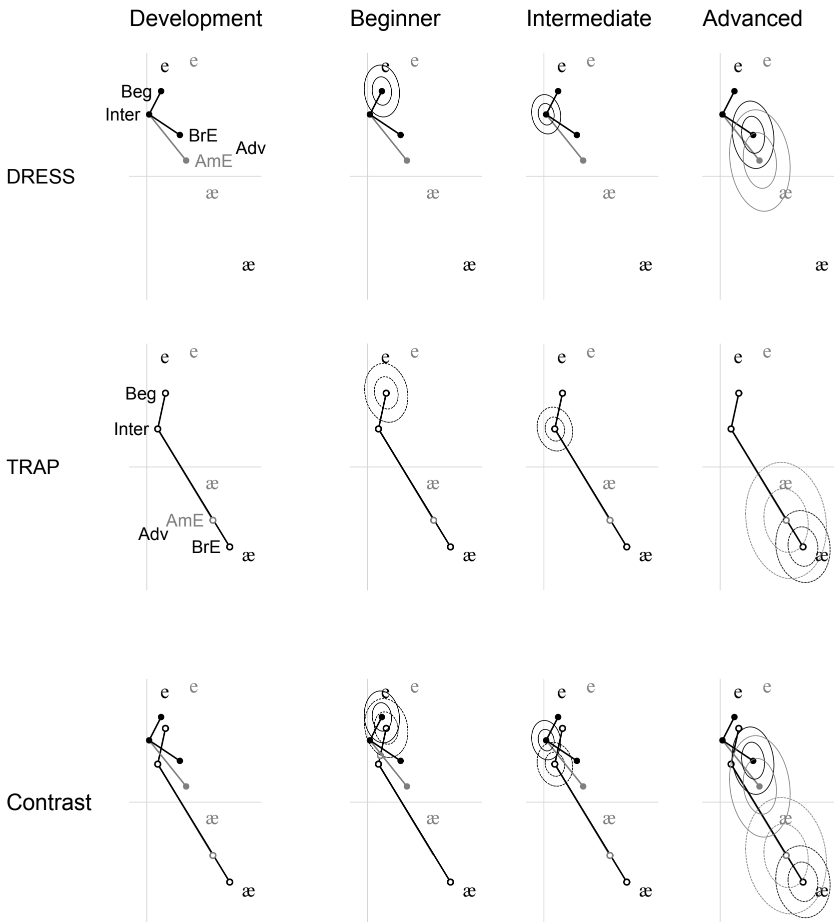


Figure 4.31: F_1-by-F_2 location of TRAP and DRESS in German learner productions: DRESS (top row, filled circles), TRAP (middle row, empty circles), and a comparison of the two (bottom row). Advanced learners are divided into AmE-oriented speakers (grey) and BrE-oriented learners (black). IPA sybols denote native speaker values from the literature (grey: AmE, black: BrE; cf. Figure 4.5). The plotted values are Lobanov-normalized (i.e. z-transformed) Bark measurements. Error bars reflect 50% and 90% posterior uncertainty contours.

The location of TRAP and DRESS in the F_1-by-F_2 plane also changes across proficiency levels. Figure 4.31 shows findings separately for DRESS (top row) and TRAP (middle row). The left-most column of panels summarizes the cross-sectional patterns by showing three proficiency levels: beginners, intermediate and advanced learners. These estimates are connected with lines, which is to suggest, for each vowel, a developmental path in the vowel plane. For advanced levels of pronunciation ability, we distinguish between BrE-oriented (black) and AmE-oriented learners (grey). The "paths" are therefore allowed to branch out. The IPA symbols denote the native speakers values from our literature survey (grey: AmE, black: BrE; cf. Figure 4.5.). The columns to the right show, for each proficiency level, uncertainty contours for our estimates. The following patterns emerge from Figure 4.31:

- Both vowels in general undergo lowering in the vowel plane. TRAP also undergoes retraction.
- DRESS shows a constant rate of change, consistently shifting by a small amount across proficiency levels.
- TRAP parallels this trend up until intermediate levels of proficiency and it seems that the two categories shift in tandem.

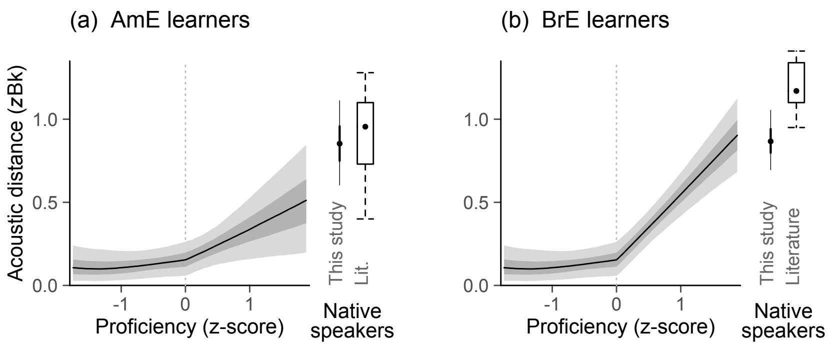


Figure 4.32: German learners' use of vowel quality to contrast trap and dress: Distance in the F_1-by-F_2 space across proficiency levels, for AmE-oriented learners (at left) and BrE-oriented learners (at right). Native speaker values from this study are shown at the margins.

- On average, learners at the beginning and intermediate stages appear to produce no discernible spectral contrast between the vowels.
- From intermediate to advanced levels, trap undergoes lowering and retraction. For both AmE- and BrE-oriented advanced learners, it is further back, consistent with native speaker values.
- It is only at advanced stages, then, that a contrast is signaled in the acoustic vowel plane.
- The two groups of advanced learners differ systematically in that BrE-oriented learners produce a larger quality contrast.
- For trap, both groups show values that may be considered as characteristic of their target variety.
- For dress, however, both advanced learner groups produce acoustic values that differ noticeably from those of native speakers.

The spectral contrast as a measure of category separation is shown in Figure 4.32, which graphs the Euclidean distance between trap and dress by proficiency level for learners in comparison to the figures produced by native speakers in this study. The following points are noteworthy:

- Native speakers produce distances of about 0.85, on average.
- Learners consistently fall short of this value. Lower proficiency levels show distances of about 0.10 to 0.15, which is less than a fifth of the distance observed in native speech.
- At upper intermediate and advanced levels learners begin to approximate the spectral distance in native speech. On average, BrE learners reach values of 0.75 (at $z = 1.5$), and AmE learners of 0.44. While both fall short of the respective targets, the divergence between the learner groups is in line with reported differences between standard British and American English.

4.9 *Summary and discussion*

This chapter was concerned with the acquisition of the English trap-dress contrast by German learners. A key difference between the vowel systems of the two languages is the absence of a front open vowel in German. In English, this region of the vowel space is occupied by trap. Acoustic

Parameter	Proficiency	*Mdn*	Posterior quantiles .05	.25	.75	.95
DRESS						
F_1	Beginner	0.52	0.40	0.47	0.56	0.63
	Intermediate	0.65	0.56	0.61	0.68	0.73
	Advanced AmE	0.91	0.69	0.82	1.00	1.13
	Advanced BrE	0.77	0.62	0.70	0.82	0.91
F_2	Beginner	0.92	0.84	0.89	0.95	1.00
	Intermediate	0.99	0.93	0.96	1.01	1.05
	Advanced AmE	0.78	0.64	0.72	0.83	0.90
	Advanced BrE	0.81	0.72	0.78	0.85	0.90
TRAP						
F_1	Beginner	0.58	0.45	0.53	0.63	0.71
	Intermediate	0.78	0.69	0.75	0.82	0.88
	Advanced AmE	1.30	1.05	1.20	1.41	1.56
	Advanced BrE	1.45	1.29	1.39	1.52	1.60
F_2	Beginner	0.89	0.80	0.86	0.93	0.99
	Intermediate	0.94	0.86	0.91	0.97	1.02
	Advanced AmE	0.62	0.45	0.55	0.70	0.80
	Advanced BrE	0.53	0.41	0.48	0.58	0.65
Duration ratio	Beginner	1.27	1.09	1.19	1.34	1.46
	Intermediate	1.12	1.00	1.07	1.17	1.25
	Advanced AmE	1.26	0.98	1.14	1.39	1.60
	Advanced BrE	1.25	1.04	1.16	1.34	1.49

Table 4.9: Estimates of the duration ratio, F_1 and F_2 (zBk) for learners at representative proficiency levels.

studies indicate that native speakers of English contrast TRAP from DRESS in both length and quality. These cues are in a trading relationship, however, and speakers of BrE typically foreground quality to differentiate these vowels; native speakers of AmE, on the other hand, make greater use of length to cue this category contrast.

Linguistic constraints appear to inhibit the L2 acquisition of the TRAP-DRESS contrast by German learners. Thus, an assessment of TRAP from the viewpoint of markedness indicated that front open vowels may be considered marked based on their cross-linguistic (in)frequency and susceptibility to language change. Further, previous research has shown that the sensitivity of German listeners to vowel quality contrasts in the relevant portion of the acoustic vowel space declines after the age of 12, with a majority of older learners experiencing perceptual difficulties at the TRAP-DRESS boundary. The literature on speech perception further suggests that German learners are biased toward temporal signals for this vowel contrast.

Theoretical frameworks resort to different types of linguistic constraints to hypothesize about the course of TRAP-acquisition by German learners. Except for the CAH, which assumes a facilitating role of the phonemic status of TRAP, theoretical contributions unanimously consider the front open vowel a difficult structure. From the viewpoint of markedness and naturalness, equal likelihood is assigned to [ɛ ɛː] and [a aː] as L1-derived substitutes. Perceptual constraints, however, predict German [ɛ ɛː] as the most likely variant in L2 speech production.

Claims about the difficulty of TRAP and the tendency toward substitution by L1 [ɛ ɛː] find support from empirical work. The production of TRAP by L1 German learners has been addressed by a number of studies, with auditory work typically reporting accuracy rates below .40. Weiher (1975), who studied early-stage instructional-setting learners aged 10 to 11, deviates in this regard, noting accuracy rates above .50. Instrumental studies, which have exclusively dealt with learners aged 16 or older, provide acoustic details about the implementation of the contrast by German learners. While most work is restricted to spectral measurements (Barry 1977; Kautzsch 2010; Maurer 2014), two studies also report on temporal properties (Flege et al. 1997; Steinlen 2005). As for vowel quality, previous work suggests that learners are able to acquire a target-like contrast between DRESS and TRAP. The majority, however, fails to reach native speaker values. The two studies that have reported on temporal measurements noted that inexperienced learners produce an insufficient spectral distinction between the vowels.

This study aimed to provide new insights into German learners' acquisition of the TRAP-DRESS contrast by (i) jointly considering temporal and spectral properties of both vowels and (ii) studying a wide range of proficiency levels, to extend the scope of acoustic evidence to younger instructional-setting learners. Building on theoretical work, experimental insights into German learners' perception of relevant vowel categories, and existing acoustic work, it was hypothesized that German learners may eventually progress from a phase of TRAP-DRESS merging to a more target-like implementation of the contrast. This progression was projected to be characterized either by simultaneous acquisition of both types of cues, or a bias toward vowel length as a distinctive feature. A tendency toward the exclusive use of vowel quality to contrast DRESS and TRAP was precluded on the grounds that the corresponding perceptual categorization pattern appears implausible in the light of theoretical considerations and external information (see Table 4.3).

To provide a basis for the interpretation of the results, a survey of the acoustic correlates of the TRAP-DRESS contrast in BrE and AmE was carried out. This review highlighted systematic differences between the varieties, which are relevant for judging the native-likeness of learner productions. Thus, speakers of AmE tend to rely more strongly on temporal cues (i.e. vowel quantity), whereas BrE native speakers primarily exploit spectral signals (i.e. quality) to distinguish TRAP and DRESS. Findings for the native speakers recorded in this study were consistent with the literature, which points to the adequacy of the materials and procedures used to study variation in the production of both vowels. As for the German learners, the results for the duration ratio as a measure of the temporal contrast between TRAP and DRESS hinted, tentatively, at a U-shaped pattern, with lower and higher proficiency levels showing somewhat greater temporal differentiation. The figures for intermediate to advanced learners are consistent with results presented by Flege et al. (1997), who noted duration ratios of 1.18 and 1.35 in inexperienced vs. experienced German learners. Steinlen's (2005) duration ratio of 1.09 is also consistent with values reported in the present study. The spectral analysis of DRESS and TRAP revealed that both vowels shift with increasing levels of pronunciation ability, relocating to a more open and back position in the F_1-by-F_2 vowel space. Larger movement was observed for TRAP, which primarily occurred at high levels of pronunciation proficiency. As the figures reported in this study are on the *z*Bk (rather than the Bark) scale, a direct comparison to the vowel locations reported in previous studies on GLE is not possible. Using the Euclidean distance between the two vowels, this study found that German learners generally fall short of the type of spectral contrast found in native speech.

The findings of the present study are in line with the theory-based specification of TRAP as a difficult structure and concur with earlier experimental work by showing that a target-like production of TRAP and the TRAP-DRESS contrast only occurs in highly proficient learners. This difficulty, however, was predicted by recourse to different explanations, such as perceptual distortion, markedness, and natural processes disfavoring front open vowel qualities. As these linguistic constraints produce identical predictions, the underlying processes cannot be disentangled on the basis of evidence from speech production. The mechanism(s) underlying TRAP-production by German learners at different proficiency levels thus remain to be more fully explored in future investigations. To this end, perception data must be gathered on the learners studied. As such, the stipulation of constraints that are not perceptual in origin can only be put on a firmer basis once perceptual origins can, at least the level of the individual learner, be ruled out as viable explanations. Based on the indications in the literature, we would expect perceptual sensitivity to vary with age and therefore, perhaps, with the overall proficiency level of instructional-setting learners. This opens up interesting avenues for further study. As to the search for explanatory mechanisms, it appears that progress can only be made via a combination of perception and production data.

The results produced by the present investigation provide tentative evidence that DRESS and TRAP are not merged completely by lower-proficiency learners. The duration ratios reported for this population of learners are consistent with categorization patterns 3 and 5 described in Table 4.3. The tendency toward temporal cues may thus be reflective of the categorization of TRAP as L1 [ɛː] and DRESS as L1 [ɛ]. This interpretation is compatible with the observation that the two vowels are contrasted in length but not (or much less so) in quality. However, the average duration ratio observed for this subgroup (1.30) is smaller than the contrast between German /ɛː/ and /ɛ/ (1.50 to 2.50; see Table 4.1). A second perceptual categorization process that is consistent with the findings presented here is the formation of a new category for TRAP. This new category differs from the IL category for DRESS in having a greater value of the feature length. It also differs from the TL category for DRESS in that spectral cues are backgrounded. Evidence from speech production cannot determine whether the duration ratios observed in lower proficiency learners reflects two-category assimilation (TRAP → /ɛː/, DRESS → /ɛ/) or new category formation, with duration being the more easily accessible cue (as suggested by Bohn 1995). Nevertheless, it seems that categorization patterns 1, 2, and 6, can be excluded as plausible underlying perceptual mechanisms.

An unexpected pattern, which we must interpret cautiously, was the U-shaped trajectory, which indicated that the magnitude of the temporal contrast may deteriorate in intermediate learners. From the viewpoint of linguistic constraints, there does not appear to be a plausible explanation for this finding, at least at the level of depth provided by this study. We are therefore not prepared to believe that this patterns reflects a genuine cross-sectional trend. Should this indication nevertheless be picked up in future work, it would seem necessary to first reflect on how such a pattern may be accounted for on theoretical grounds.[39] A factor that may have given rise to a regress in duration contrast is a change in DRESS-production as a result of overgeneralization. This is to say that features of the newly acquired, or emerging TRAP-category may have been extended to DRESS. In the context of L2 speech, this processes has also been referred to as overcorrection or hypercorrection. A similar explanation was proposed by Major (1987) to account for the productions of more vs. less proficient Portuguese learners of English, whose L1 also lacks TRAP but has /ɛ/, a vowel similar to DRESS. The intelligibility of their DRESS-productions declined while that of TRAP improved. The way in which overgeneralization surfaces in length and quality features does not appear to be well understood, however. Further research is clearly needed to address this issue. What can be stated on the basis of the present data is that, on average, spectral properties of DRESS appear to change in parallel with TRAP.

[39] It is worth noting that a reanalysis of the duration ratio conditional on age showed the same bend as that observed for proficiency. It is thus tempting to speculate on social factors that may be involved, such as speech accommodation as a result of peer influence.

Besides the scope statements formulated earlier, the conclusions that can be drawn from this study are limited in several respects. Thus, a more detailed account of the relationship between the IL categories for DRESS and TRAP and the L1 categories /ɛ/ and /ɛː/ could have been given had the subjects' productions in their native language also been documented. This would have warranted more defensible statements on the degree of equivalence between the L1 and IL categories in speech production. A further limitation concerns the temporal measurements reported in this study. While spectral quantification was considered in the context of the full system of monophthongs – that is, converted to z-scores reflecting position in the learner's IL vowel system – the assessment of length contrasts would have profited from a similar approach. Thus, vowel duration could have been measured relative to other monophthongs in the IL vowel system, which would have yielded a richer account of the TRAP-DRESS duration contrast. This study only considered the relative contrast between the two vowels. This approach lacks sensitivity to changes in the duration of TRAP and DRESS individually. Thus, a complementary measurement of a normalized vowel duration would have yielded insights into the nature of the differential length contrasts observed in learner productions. As was noted above, this generates an interpretative dilemma, as changes in the duration ratio may reflect changes in TRAP duration, DRESS duration, or both.

Apart from these caveats, the present study has made methodological contributions to the field of non-native vowel production research. A comprehensive survey of the auditory and acoustic properties of BrE and AmE monophthongs was carried out. These data, which are available in the web appendix[40], may serve as a point of departure for future studies on L1 and L2 varieties of English. One point that emerged from this review was that vowel duration and quality must be considered jointly. Sole reliance on spectral measurements may miss a category contrast that is signaled (predominantly) by durational cues.

[40] For AmE, see https://osf.io/r86dv/, for BrE, https://osf.io/b9gyv/, and for German, https://osf.io/xbr94/.

A further methodological trait that may be of value to future work is the choice of *z*Bk scores as the unit of analysis. By controlling for physiological differences between speakers, these scores aim at maximizing comparability to other work on native and non-native vowel production. Measurements from this study and the literature may serve as reference values for future research. The data that is made available online also lists Hertz and Bark measurements, as these may be more suitable for certain purposes. Finally, a method for normalizing auditory vowel descriptions was proposed. This procedure bears potential in that it allows for direct comparison of different accounts in the literature (see Figure 4.3). More importantly, from the viewpoint of acoustic work, the derived scores permit external validation of acoustic measures, whose main purpose in most applied work is the quantification of perceptible differences. Relevant to this study, the acoustic measure used to express differences in vowel quality, *z*Bk, was thereby shown to offer a valid reflection of auditory impressions. Auditory *z*-scores can also serve as a frame of reference for the assessment and comparative evaluation of different types of psychoacoustic transformation (e.g. Hertz, Bark, mel) and the multitude of methods that have been proposed for vowel normalization (see Adank et al. 2004 for an overview).

Research on the L2 acquisition of TRAP can draw on a substantial body of literature to map out linguistic constraints that translate into theoretical predictions. A critical issue to be resolved is whether the production of front open vowel qualities is constrained by the general motoric capacities of language users. Knowledge about articulatory constraints could provide more genuine explanations for the markedness of TRAP and complement observations from typology and language change by stipulating underlying processes rooted in the physiology of articulation. Further work is also needed on German learners perceptual (re)organization of phonological categories during the course of L2 development. In particular, evidence on cue reliance by beginning instructional-setting learners aged 10 to 12 is currently lacking. Quantifiable evidence on the relative weight of temporal and spectral cues to the TRAP-DRESS contrast could shed new light on perceptual constraints in this population of learners and would provide a sounder basis for the interpretation of observations in speech production. As the preceding discussion has revealed, it is crucial for future perception studies to investigate reliance on both temporal and spectral signals.

In summary, the present study arrives at the following recommendations for future work on the production of TRAP and DRESS by German learners:

- The acoustic parameters studied should include duration and spectral measurements. With the literature review indicating a trading relationship between these cues in native speech, investigations restricting their focus to only one of these features provide limited insight into the implementation of the contrast in learner speech.

- Given that DRESS-production shows systematic variation in the course of L2 development, assessments of TRAP should be made relative to DRESS, as a change in the former need not result in an increased category contrast. This suggests that studies on TRAP in GLE must (at least) also include DRESS.
- Acoustic information on both L1 and IL productions by each learner is desirable as it permits evaluation of changes relative to both L1 and TL reference points. In combination with temporal measurements, this yields more reliable conclusions about the degree to which equivalence classification and category formation may be evidenced in speech production.
- For direct comparison with native speaker productions, *z*Bk scores have proven a useful measure, as they aim to control for between-speaker variation arising from differences in the physiology of the vocal tract. While the use of normalized formant measurements may not establish perfect comparability between different accents or languages, it allows for comparison of relative locations (e.g. using *z*Bk distances in the vowel space) and provides an acceptable level of comparability for future studies on the same (or similar) learner varieties.
- Similarly, a quantification of vowel duration relative to other monophthongs should offer more detailed insights through comparisons to other vowels. This provides a new perspective for the assessment of changes in the implementation of length features across developmental stages.
- Finally, it is essential for studies to report as much information as possible on the acoustic properties of the vowels investigated. This facilitates surveys and reanalyses of published work and thus the accumulation of quantitative and substantive insight. Ideally, studies should make available the primary data on which their conclusions are based.

5 Laterals: Clear and dark /l/

Following our discussion of vowel contrasts in the preceding chapter, we now turn to German learners' acquisition of English laterals. Together with rhotics, which are the focus of the next chapter, they form the class of liquids. In contrast to prototypical consonants, the acoustic and articulatory properties of liquids are continuous rather than discrete, which means that in many cases, it is difficult to determine clear-cut category boundaries. Further, the same auditory effect may be achieved by different gestures. It is therefore not surprising that, from a synchronic and diachronic perspective, laterals and rhotics are a versatile class of speech sounds. Indeed, they are affected by a number of cross-linguistically attested phonological processes such as lambdacisms, where a rhotic alternates with a lateral, and rhotacisms, the reverse pattern. Both types of liquids may also be subject to vocalization – that is, alternate with or diachronically change into vowels. In terms of their phonotactic properties, they differ from other consonants in the number and range of combinations they form in onset and coda clusters.

This chapter is concerned with the acquisition of English lateral allophones by German learners. §5.1 presents a contrastive analysis of lateral realization in English and German. Linguistic constraints relevant for the L2 acquisition of English lateral allophones are addressed in §5.2. After a consideration of theory-based predictions about English laterals in German Learner English (§5.3), §5.4 reviews relevant empirical work. An outline of the aims (§5.5) and method (§5.6) of this study is followed by limitations as regards the generality of our findings (§5.7). Results are then presented in §5.8 and §5.9 closes with a summary and discussion.

5.1 Contrastive analysis

The lateral /l/ is produced with contact between the tip of the tongue and the upper teeth ridge. The airstream is frictionless and flows laterally, at both sides of the tongue. With the alveolar closure constituting the primary place of articulation, the tongue body may form a secondary constriction. A velarized articulation involves raising of the back of the tongue toward the velum or retraction toward the uvula, creating a secondary constriction in the back of the mouth. This gives a back vowel resonance to the consonant, producing a *dark* lateral, which is phonetically transcribed as [ɫ]. The palatalized variant is referred to as *clear* or *light* and denoted as [l]. It involves raising of the front of the tongue toward the hard palate, which gives a front vowel resonance to the consonant (Cruttenden 2014: 219).

While English and German both have the lateral phoneme /l/, there are differences in phonetic realization: German does not have a dark variant. General British[1] (GB) has three major allophones – clear [l], devoiced clear [l̥], and dark [ɫ]. Clear [l] is typically found before vowels and /j/. It occurs in word- and syllable-initial singletons (*like*, *London*) and clusters (*black*, *glue*), word-medially in intervocalic position (*silly*, *alive*), and (in connected speech) in word-final position if the following word begins with a vowel or /j/ (*will you*, *feel it*). In onset clusters, clear [l] may be fully devoiced after voiceless stops in accented syllables (*play*, *click*) or partially devoiced after voiceless stops in unaccented syllables (*butler*, *helpless*) or voiceless fricatives (*slow, fly*). Dark [ɫ] is found before consonants (*help*, *will she*), in word-final position before a pause (*well*, *all*), or as a syllabic consonant [ɫ̩] (*angle*, *table*), which may be partially devoiced following a voiceless consonant (*bottle*, *apple*) (Cruttenden 2014: 217–218). The complementary distribution of the allophones is thus governed by prosodic constraints at the level of syllable structure. In the following, the terms *prevocalic*, *intervocalic*, and *non-prevocalic* will be used to refer to prosodic position.

There is considerable regional variation in the distribution of clear and dark /l/ in the British Isles. In large parts of the north of England and northern Wales, for instance, dark [ɫ] may occur in all positions. In other dialects, such as those of South Wales and Tyneside, clear [l] tends to be observed in all environments (Wells 1982: 390; Carter 1999; Carter & Local 2007; Cruttenden 2014: 219).[2] One of the sound changes observable in present-day BrE is the spread of /l/-vocalization in non-prevocalic position, especially in younger speakers from London and the southeast (Collins & Mees 2013: 94). With the tongue not touching the teeth ridge, there is no occlusion and the consonant takes on a vowel-like quality. Cruttenden (2014: 219) notes that vocalization is more common after labial consonants. In contrast to GB, GA is often described as having a generally dark allophone in all positions (Ladefoged & Maddieson 1996: 361). Before stressed vowels, however, the degree of velarization is smallest (Wells 1982: 490).

In Standard High German, /l/ is typically realized as a palatalized clear [l] (Moulton 1962: 32; Kufner 1971: 47; Scherer & Wollmann 1986: 98; König & Gast 2009).[3] However, there is dialectal variation, with southern German dialects showing a higher degree of palatalization and speakers from Rhineland having a velarized variant, most notably in word-final position (Kufner 1971: 47; Kohler 1995: 164). Word-finally, southern German and Austrian dialects have been described as exhibiting clear [l] after front vowels and dark [ɫ] after back vowels, with the tip of the tongue curled back during the articulation of the latter (Koziol 1959: 76; Scherer & Wollmann 1986: 98).

Acoustic studies on /l/-variation in English and German have shed more light on the gradience in clearness and darkness as a function of external (e.g. regional) as well as internal factors such as prosodic and phonetic context. Acoustically, the degree of velarization is reflected in the

[1] We follow Cruttenden (2014) and use the term General British for the standard accent also known as "Standard Southern British English", "Received Pronunciation", or "BBC English".

[2] Nevertheless, instrumental studies have shown that such apparently mono-allophonic dialects also show varying degrees of velarization reflecting the influence of syllable structure; thus, coda /l/s are generally darker than onset /l/s (Carter 1999; Carter & Local 2007).

[3] X-ray data reported by Delattre (1965) confirm this, showing no evidence for velarization in prevocalic or postvocalic position.

formant structure of the sounds. Clear [l] is generally agreed to exhibit an F_1 of about 250 to 400 Hz; F_2 may range from 900 to 1600 Hz (Lehiste 1964; Dalston 1975; Nolan 1983). For dark [ɫ], F_1 has similar or slightly higher values, and F_2 is much lower, ranging between 600 and 900 Hz (Lehiste 1964; Dalston 1975). The primary acoustic correlate distinguishing [l] and [ɫ], then, is F_2, which is lower in velarized variants. A metric that serves as an acoustic measure of *darkness* is the difference between F_1 and F_2. Expressed in perceptually more meaningful Bark units, this measure will be referred to as the "$F_2 - F_1$ Bark difference", (hereafter denoted as Δ_{Bk}; read: "Bark difference"). The lower the Δ_{Bk} score, the higher the degree of velarization.[4]

Figure 5.1 shows Δ_{Bk} scores reported in (or derived from) previous work[5]. Scores are grouped by language/accent and prosodic position. Bars indicate further sources of systematic variation, which will be discussed below. A summary of the distribution of measurements is displayed in Figure 5.2. Both figures show no evidence for position-sensitive degrees

[4] For methodological considerations see §5.6.

[5] For details see https://osf.io/ya8pz/.

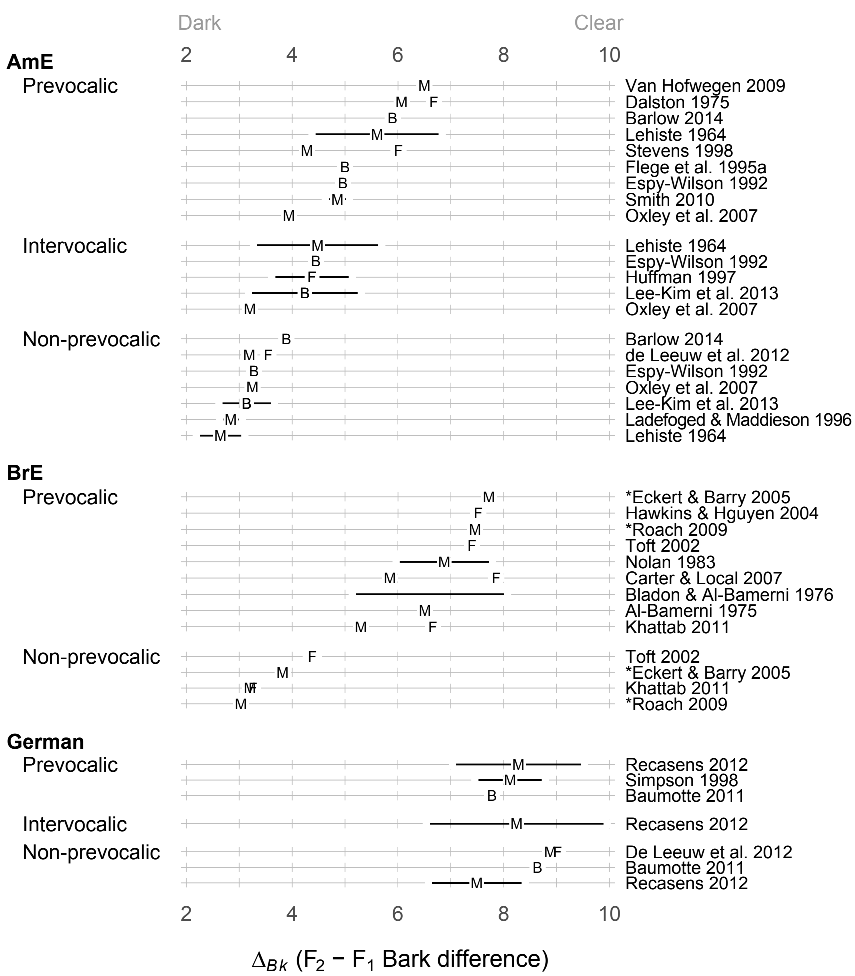


Figure 5.1: *L*-realization by position in German, BrE and AmE: Survey of the $F_2 - F_1$ Bark difference score (Δ_{Bk}) as an acoustic correlate of velarization (M = male, F = female, B = both). Horizontal bars indicate variation induced by the adjacent vowel(s). Asterisks mark sources of recordings used for additional measurements (details are given in the online supplement to Chapter 5: https://osf.io/ya8pz/).

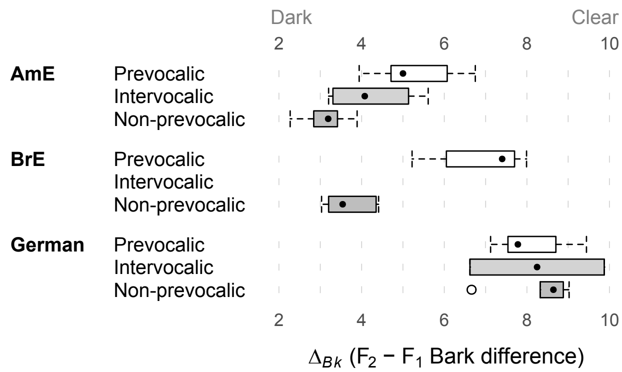


Figure 5.2: Acoustic degree of *l*-darkness by language/accent and position: Boxplots summarizing the distributions of Bark difference scores (Δ_{Bk}) shown in Figure 5.1.

of velarization in German; the measurements for all contexts are around 8 Δ_{Bk}. In BrE and AmE[6], on the other hand, the acoustic structure of the allophones differs considerably in prevocalic vs. non-prevocalic position. Overall, /l/ is confirmed to be somewhat darker in AmE. While the difference in non-prevocalic contexts is small, it is more pronounced prevocalically, which is reflected in a median difference of about 2 Δ_{Bk} (5 Δ_{Bk} for AmE vs. 7 Δ_{Bk} for BrE). However, prevocalic and non-prevocalic /l/ are still clearly separated in AmE, which is evident from the fact that the distributions do not overlap. The acoustic evidence also suggests that prevocalically /l/ is clearer in German than in BrE, which is reflected in a median difference of about 1 Δ_{Bk}.

[6] Here, we use the labels *British English* (BrE) and *American English* (AmE) instead of GB and GA since the estimates reported in the literature may not be purely reflective of the respective standard variety.

Further sources of variation in lateral darkness are reported in the literature. Thus, the vocalic context has been shown to systematically influence the formant structure of laterals. In general, F_1 and F_2 are correlated with the respective formants in the vowel. Coarticulatory effects vary between prevocalic and non-prevocalic laterals, however. Lehiste (1964) investigated the acoustics of /l/-allophones produced by $n = 5$ male speakers of AmE. While word-final /l/ was found to be largely independent of the quality of the preceding vowel, substantial differences were observed in pre- and intervocalic position, which can be seen in the range of values shown in Figure 5.1.

Vowel-induced coarticulation in prevocalic /l/ has also been reported for BrE (Bladon & Al-Bamerni 1976; Nolan 1983) and the effect of the following vowel appears to be similar in both varieties. Figure 5.3 summarizes the findings of Lehiste (1964) and Nolan (1983) and shows how, in prevocalic contexts, the degree of velarization is affected by the following vowel: the darker the shading, the stronger the "darkening effect". The major link appears to be with vowel height rather than advancement. In particular, the open vowels [æ ʌ ɑː ɒ] have a consistent darkening effect.

The realization of German /l/ has been found to vary as a function of vocalic context in all positions (Recasens 2012). For prevocalic /l/, Simpson (1998) reported coarticulatory effects on F_2 in 12 vocalic contexts in the speech of one male adult. The effect seems to be similar to that observed in English.

English dark [ɫ] also reflects coarticulation patterns induced by adjacent vowels, albeit to a much smaller degree. Recasens (2012) compared

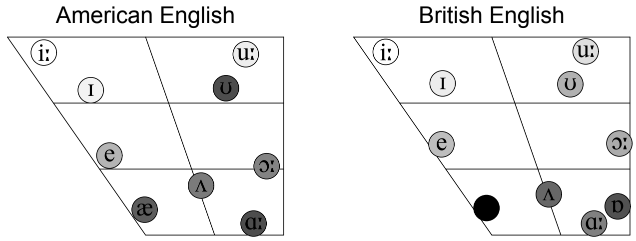


Figure 5.3: Phonetic context effects on prevocalic /l/: Influence of the following vowel on *l*-velarization. Darker shading indicates darker realizations; data for BrE are from Nolan (1983) and for AmE from Lehiste (1964).

coarticulatory influence of [i] and [a] on prevocalic, intervocalic, and non-prevocalic /l/ in a number of languages including AmE and German. Judging from Figure 5.3, [i] and [a] appear to be fairly representative of the range of vowel-induced coarticulation effects observed in laterals. Figure 5.4 summarizes the findings reported by Recasens (2012) and highlights two contrasts between AmE and German:

- As discussed above, AmE shows a pronounced effect of context, which surfaces in the tilted black profiles; no such effect occurs in German, where the profiles are flat.
- In German, there is a consistent effect of the vocalic context in all positions: The i- and a-profiles are displaced by 2 to 3 Δ_{Bk} scores. AmE intervocalic and non-prevocalic /l/, on the other hand, appear to be resistant to coarticulatory influence of the preceding vowel.

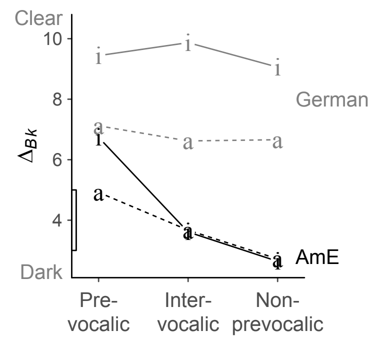


Figure 5.4: Phonetic context effects on prevocalic, intervocalic, and non-prevocalic /l/ in English and German: Bark difference scores (Δ_{Bk}) as an acoustic correlate of darkness in two vocalic contexts; data from Recasens (2012).

Further systematic differences in lateral darkness have been identified in English. Instrumental evidence reported by Lee-Kim et al. (2013) shows morphological structure to play a role in the degree of velarization in intervocalic position. Stimuli were embedded in sentences read by n = 6 AmE native speakers (2 female). Three contexts were compared: (i) word-final /l/ followed by [h] (stem), (ii) intervocalic /l/ preceding a morpheme boundary as in *tall-est* (pre-boundary) and (iii) intervocalic /l/ following a morpheme boundary as in *flaw-less* (post-boundary). Five vowels [ɑː aʊ oʊ aɪ uː] preceded /l/ in each context. The left panel in Figure 5.5 summarizes the results. Note that the vertical axis covers a much smaller range of Δ_{Bk} scores than Figure 5.4; the rectangles attached to the *y*-axes assist in drawing scale comparisons. Expectedly, intervocalic /l/s were lighter than those in non-prevocalic position. A consistent effect of morphological context was also observed: Intervocalic /l/ was darker before than after a morpheme boundary. However, compared to the influence of vocalic context, this difference is rather small (0.5 Δ_{Bk}). Vowel-induced variation was smallest at the dark end of the continuum, where differences span about 1 Δ_{Bk} (vs. 1.5 Δ_{Bk} in intervocalic contexts). This corroborates findings of Lehiste (1964) and Bladon & Al-Bamerni (1976) on the coarticulatory resistance of [ɫ].

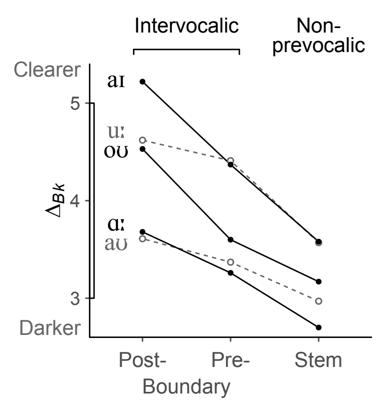


Figure 5.5: Effect of morphological context on *l*-darkness in AmE: Bark difference scores (Δ_{Bk}) by position (intervocalic vs. non-prevocalic) and preceding vowel; data from Lee-Kim et al. (2013).

Past work on the acoustic variation in /l/-allophones also observed a systematic influence of adjacent consonants. The measurements for word-final syllabic /l/ reported by Lehiste (1964) suggest that [ɫ] is clearer after velar consonants (see also Faure 1972). In general, degree of velarization increased for more anterior places of articulation and was highest after

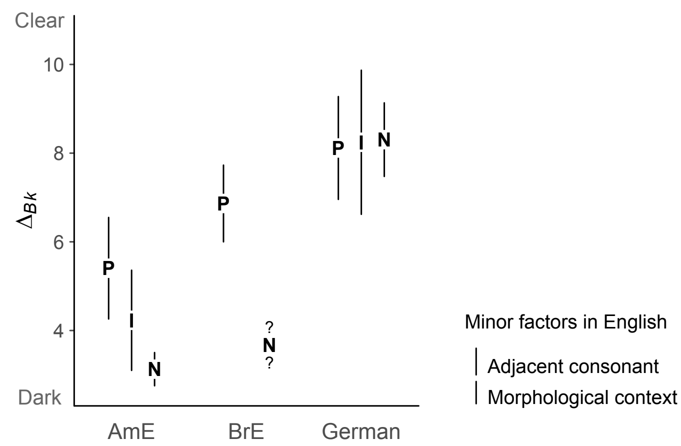


Figure 5.6: Summary of the factors influencing the acoustic structure of laterals: (i) region/language along the *x*-axis, (ii) prosodic position (P = prevocalic, I = intervocalic, N = non-prevocalic) and (iii) vocalic context (indicated with vertical bars). The size of minor effects is shown at the right margin.

dental consonants. The overall effect, however, was small (range: 0.6 Δ_{Bk}).

In summary, auditory and acoustic evidence points to different patterns of variability in English and German laterals. Three major sources of variation have been identified: language/region, prosodic position, and phonetic context. Figure 5.6 summarizes and compares these factors.

The effect of prosodic position differs between BrE, AmE, and German. BrE clearly differentiates laterals in prevocalic (P) and non-prevocalic (N) position. AmE also maintains this contrast, albeit to a smaller degree. This is primarily due to darker prevocalic /l/. The third factor, phonetic context, further contributes to the variability in the acoustic structure of laterals. In Figure 5.6, the effect of adjacent vowels is shown using vertical bars indicating the range of observed acoustic variation. In German, this effect is much larger than that of prosodic position. It appears to operate in all positions but is smallest for non-prevocalic /l/. In AmE, coarticulatory effects appear to be restricted to pre- and intervocalic contexts. As yet, little is known about the context sensitivity of BrE intervocalic and non-prevocalic /l/. The contribution of minor factors discussed above is shown at the right margin, where the length of the segments indicates their influence on Δ_{Bk} scale.

Velarized [ɫ] is a new sound for L1 German learners. To produce [ɫ], the back of the tongue needs to form a secondary constriction, a vocalic gesture similar to [uː]. Acoustically, a distance of roughly 5 to 6 Δ_{Bk} distinguishes German and English non-prevocalic /l/, which gives a quantitative reflection of the learning task. German clear [l] appears to be similar to English clear [l]; however, acoustic evidence suggests that the BrE allophone is slightly darker. Prevocalically, the acoustic distance to BrE /l/ is about 1 Δ_{Bk}, that to AmE /l/ roughly 3 Δ_{Bk}. Acquisition of English laterals not only requires mastering the speech motor differentiation between clear and dark /l/, but also learning their position-sensitive distribution. As for the second major factor contributing to /l/-variation, transfer of vocalic context effects in prevocalic (and intervocalic) position is expected to yield target-like coarticulation patterns. However, vowel-induced assimilation must be unlearned for non-prevocalic /l/.

5.2 L2 acquisition of clear and dark /l/: Linguistic factors

5.2.1 Markedness

Typological comparisons show that clear and dark /l/ differ in their rate of occurrence. In the UPSID database, a proportion of .80 of the documented laterals are approximants, by far the most common type being plain voiced approximants (.75). Considering place of articulation, .95 of the laterals do not involve a secondary constriction and only .03 of the 418 laterals in the database are velarized (Maddieson 1984: 77). In terms of cross-linguistic frequency, it appears that a secondary place of articulation is disfavored. An asymmetry between clear and dark /l/ is also suggested by evidence from L1 acquisition of English. Figure 5.7b shows accuracy rates reported in different studies.[7] Word-initial and -final contexts are distinguished and the curves summarize the trends. A comparison of /l/-allophones to other English consonants , which is shown in the left-hand panel, reveals that /l/ emerges relatively late, especially in word-final position. The age of acquisition, which is indicated by the circles is 3;9 in initial position (grey) and 5;2 in final position (black). The positional allophones thus differ appreciably in their development.

[7] See §3.3.2 for details.

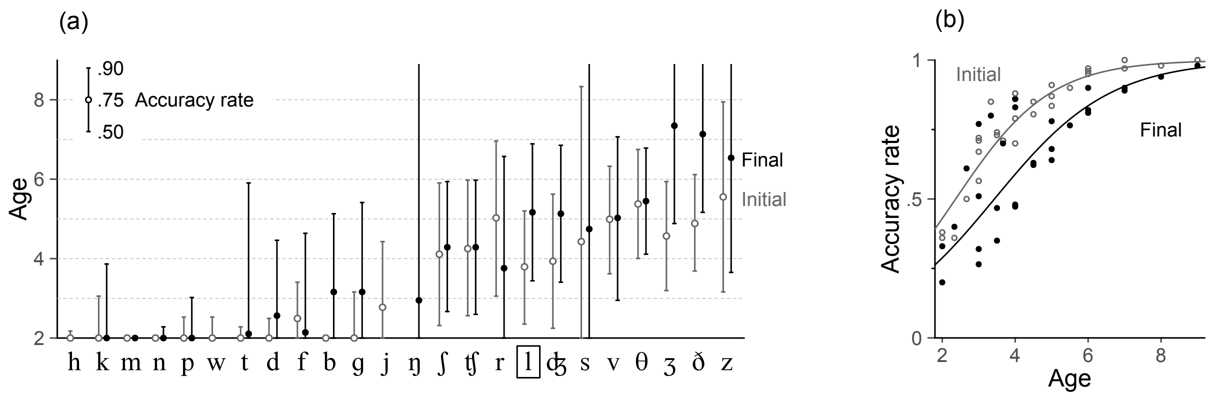


Figure 5.7: Laterals in English L1 acquisition in comparison to other consonants. Panel (a): Points mark the age of acquisition (.75), bars indicate the age of customary production (.50 accuracy, lower end), and age of mastery (.90, upper end). Estimates are based on a joint analysis of 4 studies (see §3.3.2). Panel (b): Accuracy rate in initial (grey) and final position (black) by age. The points denote estimates reported in the individual studies, the trend lines are averages based on the joint analysis.

In L1 acquisition, liquids are affected by the phonological process of gliding, a regular error pattern whereby the glides [w j] replace liquids. Onset /l/ is typically glided to [j], coda /l/ is reported to be substituted by [w] or affected by vocalization – that is, replacement by a (back) vowel. Olmsted (1971), for instance, reported [w] as the most frequent substitute for /l/ (with a share of .56).

One factor accounting for the delayed acquisition of [ɫ] may be articulatory complexity. Its production requires two lingual constrictions, an anterior closure of the tongue tip against the alveolar ridge and a secondary constriction in the back of the mouth. Typical substitutes may result from a lack of sufficient motor differentiation of anatomically coupled articulators (Gick et al. 2008). Lingual complexity may thus be reduced by gestural omission, such as the loss of the anterior gesture in [w]-gliding, which is characteristic of English L1 acquisition. Besides articulatory complexity, prosodic position appears to play a role in explaining the asymmetry between clear and dark /l/. Other things being equal, syllable codas are structurally more marked than onsets. This is corroborated by the following observations:

- In L1 acquisition, open syllables are acquired before closed ones (Jakobson 1968; Levelt et al. 1999/2000; Lleó et al. 2003).
- CV syllables are maximally unmarked; they occur in every language (Maddieson 2013).
- An implicational relationship holds between CV and CVC syllables, with the presence of the latter implying that of the former (Rice 2007).
- Cross-linguistically, onsets bear a higher functional load; the number of contrasts in coda position is more restricted (Colantoni et al. 2015: 186).
- Perception accuracy is lower in coda position (Ohala & Kawasaki 1984; Strange 1992).

Regarding the structural properties of syllable codas, Vennemann (1988: 21) proposed a number of universal preference laws. Thus, natural languages seem to prefer codas with fewer consonants and steeper sonority clines[8] between the nucleus and the coda offset. Singleton postvocalic /l/ thus resembles a disfavored coda due to its proximity to vowels on the sonority scale (cf. Figure 2.2).

[8] Vennemann (1988: 21) talks about consonantal strength, which, for our purposes, can be considered the inverse of sonority.

Further evidence indicating the status of dark /l/ as a marked structure is its historical instability, which can be observed in ongoing sound changes in present-day English. In a wide range of dialects, dark [ɫ] is affected by *l*-vocalization (cf. Johnson & Britain 2007). Through omission of the alveolar constriction, [ɫ] is replaced by a structurally and articulatorily simpler vocalic gesture, a phenomenon familiar from L1 acquisition.

Markedness statements may further be derived for the structural characteristics of onsets and codas. One factor that applies to both *l*-allophones is complexity: Onsets and codas of length $n + 1$ imply those of length n (Anderson 1987; Yavaş 2005). As degree of markedness increases with complexity, singleton laterals are less marked than those occurring in clusters.

To sum up, the available evidence suggests that [ɫ] is relatively more marked than [l] – it is typologically rare, acquired late in L1 acquisition and subject to synchronic and diachronic variation. This is presumably attributable to its higher degree of articulatory complexity, as it involves two simultaneous gestures of the tongue. The markedness of English dark [ɫ] is further due to its distribution: Codas differ from onsets in perceptual robustness and articulatory distinctness and are thus the disfavored context.

Criterion	[l]	[ɫ]
Implicational relationship	○	○
Cross-linguistic frequency	−	++
L1 acquisition	+	++
Synchronic/diachronic instability	○	+
Articulatory complexity	−	++

Table 5.1: Summary of markedness criteria for lateral allophones. Open circles ○ denote no (or inconclusive) evidence; (+)+ refers to (strong) evidence for marked status; (−)− refers to (strong) evidence for unmarked status.

5.2.2 *Frequency and orthography*

Frequency estimates for /l/ range from 3 to 5 occurrences per 100 segments, with a median of just under 4 (see §3.3.3). /l/ is therefore among the most frequent English phonemes. Allophone frequencies, on the other hand, are rarely reported. An exception is Delattre (1965), who observed rates of 2.6 for [l] and 2.3 for [ɫ] in AmE. Both figures are above the overall median consonant frequency (1.9). Syllabic /l/ appears to be an infrequent sound, with Mines et al. (1978) reporting a rate of 0.4. Apart from silent *l* as in *salmon*, *walk*, and *would*, English orthography provides an unambiguous representation for the lateral /l/. However, there is no graphemic cue that points to the appropriate allophonic variant.

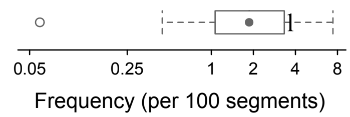


Figure 5.8: Frequency of /l/ relative to other English consonants, expressed as the number of instances per 100 segments. Rates are shown on the log scale, but tick mark labels give the absolute values of the rates. Data are pooled across 9 studies (see §3.3.3 for details).

5.2.3 *Similarity*

L2 studies on the perception of laterals have typically focused on phonemic differences such as the contrast between English /r/ and /l/, because learners whose L1 shows no liquid contrast usually experience difficulties with this opposition. It appears that, to date, no study has investigated L2 perceptual discrimination between /l/-allophones. Bohn & Best (2012), however, shed some light on L1 German listeners' perceptual sensitivity to English approximants. The authors investigated the perception of the AmE /r/-/l/ contrast using a two-choice identification and an AXB discrimination task. Participants were $n = 18$ students at Kiel University in northern Germany (average age: 21). The authors observed equally categorical perception by learners and native speakers – no statistical differences surfaced for boundary location and slope. The same is true for the overall performance on the discrimination task (German: .73; English: .78). However, identification and discrimination were statistically poorer for German listeners toward the /l/-end of the continuum. While these findings do not provide information on the distinction between clear and dark /l/, they may be interpreted as indirect evidence for a lack of perceptual sensitivity in contrasts involving dark [ɫ]. Bohn & Best (2012) reason that due to "modest" phonetic distances, German learners may perceptually link velarized [ɫ] and their native [l], processing them using a single phonetic category.

Knowledge about the perception of lateral allophones by L1 speakers is also scarce. One exception is Müller (2015), who investigated cue weighting by speakers of Greek and Albanian to differentiate between [l] and [ɫ]. While both languages have clear and dark laterals, the contrast is phonemic in Albanian, but not in Greek. Synthesized VCə stimuli were varied along two dimensions, the $F_2 - F_1$ difference and the duration of the vowel-to-consonant transition. Five vowels [a e i o u] preceded the lateral. The stimuli varied from 3.9 to 8.4 Δ_{Bk}, and therefore cover the range of values attested in English and German (cf. Figure 5.2). Transition duration had no effect on the perception of the [l]-[ɫ] contrast, but speakers of both languages made use of formant structure. For the Albanian listeners, whose L1 has a phonemic /l/-/ɫ/ contrast, the discriminating effect of this cue was much higher: The difference in the share of [ɫ]-categorizations along the $F_2 - F_1$ dimension was around 40 percentage points compared to 15 points for the L1 Greek listeners. A tentative conclusion from these findings may be that speakers with [l] and [ɫ] as allophones of the same L1 phoneme have no robust perceptual categories for the two sounds, leading to relatively poor discrimination of non-familiar stimuli. Yet even those speakers whose L1 has two phonemes /l/ and /ɫ/ did not perform exceptionally well in this experiment. Müller's (2015) results thus suggest that, perceptually, laterals may be inherently difficult segments.

5.3 L2 acquisition of clear and dark /l/: Theoretical predictions

The linguistic factors outlined in the preceding section provide a basis for theoretical work on L2 phonology to generate predictions about systematic variation in the acquisition of /l/-allophones by German learners. Table 5.3 at the end of this section offers a summary.

CAH For German learners, velarized [ɫ] is a new allophone of a shared phoneme. Acquisition is therefore expected to be relatively difficult. English clear [l], on the other hand, is considered as largely equivalent to German [l], which suggests ease of acquisition. Learners are expected to show transfer of the L1 clear variant and its coarticulation patterns – that is, a notable influence of vocalic context on the acoustic structure of /l/ – in all positions. In prevocalic position, this yields variants that are slightly clearer compared to BrE and markedly clearer than in AmE. Coarticulation effects are expected to be similar. In non-prevocalic position, this yields a general lack of velarization and a lack of coarticulatory resistance to adjacent vowels.

SLM While there is no empirical work on the perception of English [l ɫ] by German learners, we can assume that German [l] is perceived as the L1 sound closest to both TL allophones. The acoustic evidence reviewed above suggests that English [l] is more similar but not identical to German [l]. While BrE and AmE dark [ɫ] is much darker (difference: 5 Δ_{Bk}), prevocalic /l/ is only slightly darker in BrE (1 Δ_{Bk}), but considerably darker in AmE (3 Δ_{Bk}). We may thus formulate the following assumptions: (i) German [l] is the perceptually closest L1 sound to English [l ɫ]; (ii) the degree of perceived similarity between German [l] and English [l] is high; and (iii) the degree of perceived similarity between German [l] and English [ɫ] is moderate to high. The SLM then predicts persistent L1 transfer of [l] by the majority of German learners. As it posits the position-sensitive allophone as the level of analysis, it predicts category formation to emerge earlier for allophones that are perceived as most dissimilar from German /l/, thus predicting the acquisition sequence non-prevocalic → intervocalic → prevocalic. Target-like /l/-production is thus expected to emerge last in prevocalic position.

FCM While the FCM takes a phonological approach to IL speech perception, feature prominence hierarchies may also serve as a basis for hypothesis-formulation on allophonic variants, if these differ in features that are distinctive in the L1. Following observations in L1 acquisition, we can postulate [l w] as potential substitutes in IL speech production. While the former is available as an L1-derived substitute to learners at all developmental stages, [w] is a new sound that enters the IL grammar during the course of L2 acquisition (see Chapter 7). The two consonants differ in five features with the following prominence ranking: consonantal (0.95) > anterior (0.50) > coronal (0.45) > dorsal, labial (0.23). Classifying [ɫ] as [+cons, ±ant, +COR, +DOR, –LAB], the most likely substitute is [l].

Consonant	Rank	[cons]	>	[ant]	>	[COR]	>	[DOR, LAB]
[l]	1	+		+		+		– –
[w]	2	–		–		–		++
[ɫ]		+		±		+		+ –

Table 5.2: Substitutes for [ɫ]: Predictions of the FCM.

SCH/MDH The literature review of [l ɫ] suggests the following assumptions: (i) [ɫ] is a marked structure; and (ii) [ɫ] is relatively more marked than [l]. The MDH thus predicts English velarized [ɫ] to be a difficult structure in L2 acquisition. While a bias toward less marked [l] is expected, systematic patterns of variability observed in natural languages also suggest [w] as a potential substitute for [ɫ]. Further, given that onsets are both perceptually and articulatorily more robust, velarized [ɫ] is expected to be easier to acquire in onsets. Finally, based on the association between markedness and coda complexity, dark [ɫ] is expected to emerge in coda singletons before clusters. As for syllabic [l̩], its degree of difficulty relative to singletons and clusters is less clear, as it may be considered a cluster element or an autonomous nucleus.

NM/NDH It seems that lateral approximants have not received much attention from natural phonologists. Given the natural tendency toward simple articulatory gestures (principle of economy), it appears reasonable to assume that the production of velarized [ɫ], which involves a complex tongue maneuver, may require suppression of two types of lenitions. Simplification may thus be achieved by omission of the secondary dorsal gesture [ɫ] → [l] or the primary alveolar gesture [ɫ] → [w]. The following tentative assumptions may thus be formulated: (i) Gliding to [w] is a natural process that has to be suppressed in the acquisition of [ɫ]; (ii) simplification to [l] is a natural process that has to be suppressed in the acquisition of [ɫ]. It should be noted, however, that only gliding to [w] is supported by data from L1 acquisition. While simplification to [l] may be considered active in L1 German, coda /l/-gliding is latent. The NDH thus hypothesizes the former process, [ɫ] → [l], to be difficult to suppress. The second process, [ɫ] → [w], which is latent in German, is expected to surface in learner speech. Finally, with the NM upholding input frequency as a relevant factor, we expect no facilitating or inhibitory effect of frequency on the L2 acquisition of [ɫ]. The acquisition of syllabic [l̩] on the other hand, may be delayed.

FM The FM postulates that correct perception of [ɫ] generates a new faithfulness constraint in the grammar, followed by sensorimotor learning and generation of an undominated articulatory constraint. Once the appropriate perception and production constraints are established, output is posited to vary conditional on focus on form. Stylistic variation is thus predicted at early stages of [ɫ]-acquisition. After a phase of overly faithful production, the final stages are characterized by the interaction of functional principles at the post-lexical level. This includes style-dependent modification of the implementation of distinctive features and coarticulation with adjacent segments. This stage may also show style-related gestural weakening resulting in /l/-vocalization (i.e. [ɫ] → [w]). Finally, the FM's focus on stochastic learning suggests no frequency effect for [ɫ] but a delay for [l̩] (cf. the assumptions stated above for the NM/NDH).

MSA As the MSA jointly considers perception and production constraints, its application combines the reasoning outlined above – that is, SLM-based considerations on perceptual difficulty – and the articulatory constraints discussed under the NM. The acquisition of [ɫ] is therefore expected to be doubly constrained and thus particularly difficult. While perceptual difficulties yield substitution by L1 [l], subsequent production constraints similarly favor the gesturally simpler [l], but may also result in gliding to [w].

LTD The LTD hypothesizes target-like variants to be sensitive to *s*-marking at different levels of phonological representation. With codas being *w*-marked relative to onsets, delayed acquisition of [ɫ] is expected in this position. Other things being equal, the following tendencies are predicted: (i) delayed acquisition of [ɫ] in coda position and unstressed syllables, and (ii) no difference between coda singletons and clusters, as [ɫ] is the first segment in coda clusters and thus *s*-marked. The LTD also posits an effect of strength marking on observed variants, with *w*-marked environments favoring weakened segments. Given that, in articulatory terms, [ɫ] is stronger than [l], this may affect the acquisition of [ɫ] in that *w*-marked environments show a bias toward [l] or [w]. Finally, the LTD predicts variability in production to become increasingly systematic over the course of acquisition, reflecting the influence of phonetic and prosodic context. Target [ɫ], however, is largely robust to coarticulatory influence. This type of systematic variation may emerge at advanced stages of L2 acquisition.

OPM Based on the discussion above, [ɫ] may be considered a marked and similar structure. The OPM therefore predicts IL development in terms of its Similarity and Markedness Corollary. Figure 5.9 sketches a synthesis of these corollaries and generates the following predictions.[9]

[9] See §4.3 (Figure 4.18, p. 66) for a discussion of the rationale underlying the fusion of these corollaries.

Initially, acquisition is delayed by a prolonged phase of L1 transfer due to equivalence classification. L1-derived [l] thus permeates the initial stages of L2 acquisition, yielding lack of differentiation between English clear and dark contexts – all variants are expected to be clear. Upon perceptual learning, articulatory and markedness constraints are reflected in a phase of persistent U-dominance. The effect of U may be hypothesized to surface in a preference for gesturally simple segments. The two types of natural tendencies or processes considered above may, under the umbrella of the OPM, be considered as U influence. Therefore, the influence of U may lead to gestural simplification either by omission of the dorsal constriction [ɫ] → [l] or the primary alveolar gesture [ɫ] → [w]. While the result of the former is indistinguishable from L1 influence, the latter may surface independently of the L1 and TL. As shown in Figure 5.9, target-like production is expected to emerge late. The OPM further predicts an effect of speaking style. Formal speech is expected to show a higher degree of target-like production and less formal speech a relatively stronger influence of L1 transfer.

Table 5.3 summarizes the set of theoretical predictions outlined in this section. Next, we take a look at the empirical literature on laterals in GLE.

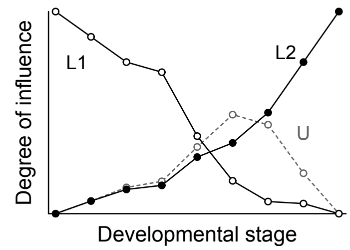


Figure 5.9: Extrapolation of the OPM to structures that are both similar and marked. This graph should be compared to Figure 2.6. 🅭🅯

5.4 *Laterals in German Learner English: Previous work*

The acquisition of English /l/ by German learners has received attention in a number of auditory and instrumental studies, which we will review in this section.

Auditory studies

Wieden and Nemser's (1991: 89, 181) study on phonological features in *n* = 384 Austrian learners[10] investigated /l/-production using two elicitation tasks. In an imitation test, learners were asked to imitate nonce items that were presented to them via earphones. In the second task, the production test, English words were elicited using various methods such German-to-English translation, picture stories, and short conversations. In the production part, /l/ was elicited in prevocalic (*n* = 12) and non-prevocalic (*n* = 8) position. Learners showed higher accuracy rates in prevocalic (.72) vs. non-prevocalic (.16) contexts. Target [ɫ] was typically fronted and realized as [l]. The authors also observed phonetic context effects in non-prevocalic position. Thus, accuracy rates were higher in clusters (*salt*, *cold*, *girls*; .22) than in singletons (*school*, *girl*; .12). Cross-sectional patterns for target [ɫ] realization were also documented. These are shown in Figure 5.10. Correct production rates develop slowly from about .07 to .20. The proportion of substitutes remains roughly constant across all stages, with maximally fronted alveolar [l] at around .30 and less fronted palatalized [ʎ] at a rate of about .40.

[10] The study included learners at 6 proficiency levels, with *n* = 64 subjects in each group. Years of instruction was used as an indicator of developmental stage, which ranged from 0 years (grade 3; 8–9 years of age) to 8 years (grade 11; 16–17 years old).

Theory	Predicted patterns in L2 acquisition by German learners
CAH	Late acquisition of [ɫ] Substitutes in onset position: slightly clearer [l], similar coarticulatory sensitivity to adjacent vowel quality Substitutes in coda position: markedly clearer [l], oversensitivity to adjacent vowel quality
SLM	Late acquisition by the majority of learners Acquisition sequence: [ɫ] > [l]
FCM	Perception-based substitute in speech production: [l]
MDH/SCH	Late acquisition of [ɫ] Acquisition order: onset > coda, singleton > cluster Substitutes: [l w]
NM/NDH	No frequency effect for [ɫ] Low frequency impedes acquisition of syllabic [l̩] Late acquisition of [ɫ] Substitutes: [l w]
FM	No frequency effect for [ɫ] Low frequency impedes acquisition of syllabic [l̩] Development: deviant production > stylistic variability > overly faithful production > variability reflecting connected speech processes Effect of speaking style (focus on form) at early stages of acquisition
MSA	Late acquisition of [ɫ] Substitutes: [l] > [l w]
LTD	Late acquisition of [ɫ] Substitutes: weakened variants in *w*-marked positions: [l w] Acquisition order: stressed > unstressed syllables, coda: singleton = cluster
OPM	Late acquisition of [ɫ] Substitutes: [l] > [l w] Effect of speaking style: higher accuracy in more formal speech, higher rate of transfer in less formal speech

Table 5.3: The acquisition of [ɫ] by German learners: Summary of predictions based on theoretical work.

Sieg (2004, quoted in Wode 2009) reports results on the development of pronunciation in immersion-setting learners at a primary school (grades 1 to 4). The majority of the participants had attended an immersion program in pre-school; only few children did not have any knowledge of English at the beginning of grade 1. Correct production rates of about .50 were observed across the four grades (.55/.40/.50/.55).

Langguth (2009) investigated the production of /l/-allophones by German learners at three proficiency levels ($n = 7$ in each group): 5th-graders (age: 10–11), 9th-graders (14–15) and 12th-graders (17–18). Three speaking styles were elicited: a word list and (near-)minimal pairs, both read from paper, and semi-free speech elicited with questions on a text. Overall, the correct production rate in dark contexts increased steadily from grade 5 to 12 (.70/.80/.90). On average, dark [ɫ] was produced more accurately in single words (word list and minimal pairs) compared to free speech (.82 and .71, respectively). As Figure 5.11 shows, the effect of speaking style was most pronounced at the intermediate stage. Langguth (2009) also observed that velarization increased in general, with the productions of learners in higher grades sounding more "American" due to a higher rate of dark allophones, including in positions where BrE has a clear variant. This tendency was also pointed out by Arnold & Hansen (1974: 28), who note that German learners may believe [ɫ] to by a typical English sound and overuse it to sound more native-like.

Langguth (2010) analyzed the production of the English lateral in 12th-graders at a grammar school in northern Bavaria ($n = 34$; age: 17–20). Subjects read a word list and a text, and then answered questions on the text. Production accuracy in dark contexts varied with speaking style: [ɫ] was produced correctly at a rate of .53 in the word list, .67 in the text, and .75 in free speech. The majority of errors in dark contexts were due to lack of velarization. There were no indications of a difference in accuracy between syllabic and non-syllabic tokens.

An auditory study by Pascoe (1987)[11] explored /l/-realization by $n = 26$ learners from southern Bavaria. Overall, the accuracy rate for dark [ɫ] was .41, with a large difference between subjects with poor pronunciation (.17, $n = 10$) and good pronunciation (.95, $n = 7$).

Further evidence on /l/-production by instructional-setting learners from southern Bavaria was reported by Kucharek (1988: 131, 144), who studied $n = 580$ learners divided into 4 groups: students in grade 6 ($n = 169$), 7 ($n = 138$), 8 ($n = 117$), and 12 ($n = 94$) and teachers ($n = 62$). Production accuracy was assessed with 3 items (*well*, *help*, *school*), which were elicited with a reading passage. As summarized in Figure 5.12, accuracy rates ranged between .50 and .70, with a difference of about .20 in absolute terms between 6th- and 12th-graders. Teachers showed a production accuracy of .93.

In sum, auditory studies on the acquisition of the velarized lateral have produced mixed results. Table 5.4 gives a summary of the empirical literature, showing observed accuracy rates and their variation as a function of different variables. The notable divergence in accuracy rate between Wieden & Nemser (1991) and the other studies may be attributable to

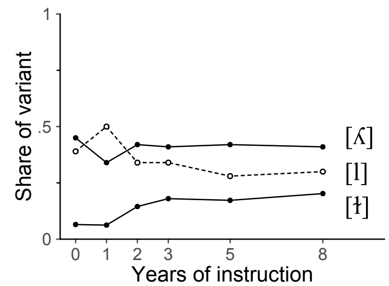


Figure 5.10: Development of dark [ɫ] production accuracy: target-like variants [ɫ] vs. moderately fronted [ʎ] vs. fronted [l] substitutes across different developmental stages; adapted from Wieden & Nemser (1991: 181) with permission. In the original chart, [ɫ]-rates are shown for two regional subgroups. The graph above shows the weighted average over these groups (1/4 Vienna, 3/4 other regions; see Wieden & Nemser 1991: 13).

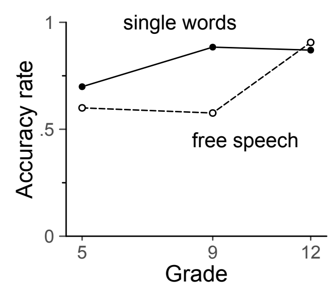


Figure 5.11: Development of dark [ɫ] production accuracy by speaking style: The proportion of target-like variants in a word list task and free speech; data from Langguth (2009).

[11] See §3.3.1 for details.

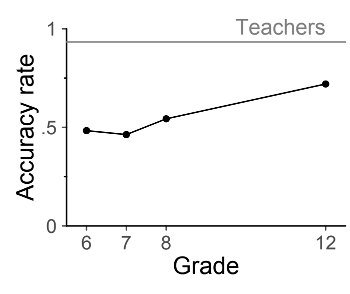


Figure 5.12: Dark [ɫ] accuracy in the productions of German instructional-setting learners in grade 6, 7, 8, and 12 in comparison to teachers; data from Kucharek (1988).

the fact that these authors distinguished 3 (rather than 2) categories: TL [ɫ], moderately fronted [ʎ], and fronted [l]. In fact, addition of the rates of the former two variants yields figures that closely resemble the accuracy rates observed in other studies. As for linguistic constraints on [ɫ]-accuracy, variable patterns have been reported for complexity and speaking style.

In terms of the types of substitutes observed in GLE, studies differ in the level of detail provided. While most authors make a binary distinction between clear [l] and dark [ɫ] (e.g. Pascoe 1987; Kucharek 1988; Langguth 2009), Wieden & Nemser (1991) reported a substantial proportion of palatalized variants [ʎ], which points to the gradient nature of /l/-velarization in learner speech, which may not be captured by a binary categorization into "target" and "non-target" variants.

Instrumental studies, to which we now turn, offer more fine-grained measurements of *l*-velarization.

Results		Notes	Study
Accuracy rate	.70–.90	Grades 5, 9, 12; age: 10–18	Langguth 2009
	.50–.70	Grades 6, 7, 8, 12; age: 11–18	Kucharek 1988
	.40–.55	Grades 1–4; age: 6–10	Sieg 2004
	.07–.20	Grades 3–11; age: 8–17	Wieden & Nemser 1991
	.17–.95	Grades 8–10; age: 14–16	Pascoe 1987
Constraints on accuracy rate			
Complexity	$\Delta = 0$	Syllabic = singleton/clusters	Langguth 2010
	$\Delta = -.10$	Singleton (.12) < cluster (.22)	Wieden & Nemser 1991
Speaking style	$\Delta = .10$	Word list (.82) > free sp. (.71)	Langguth 2009
	$\Delta = -.22$	Word list (.53) < free sp. (.75)	Langguth 2010

Table 5.4: Empirical evidence from auditory studies on dark [ɫ] in German Learner English: Reported accuracy rates and constraints on accuracy.

Acoustic studies

Two studies have produced instrumental evidence on /l/-production by German learners. Baumotte (2011) investigated the degree of velarization in /l/-allophones produced by two groups of German learners with different levels of pronunciation proficiency (n = 5 each; 3 female; M_{Age} = 26). A reading task with 8 words embedded into sentences served to elicit /l/ in three positions: prevocalic (*light*, *lend*), intervocalic (*alike*, *believe*), and non-prevocalic (*will*, *hall*, *halt*, *wild*). A parallel set of German words served as a control context. It is unfortunate that the author collapsed tokens with intervocalic and non-prevocalic /l/ for the analysis, as /l/-velarization differs systematically in these contexts. The results must therefore be interpreted with caution. In intervocalic and non-prevocalic position, proficient learners were observed to produce statistically lower F_2 than less proficient learners (1510 vs. 1880 Hz), which the author interprets as indicating a higher degree of velarization. F_2 for the German control items was intermediate (1780 Hz). However, when measuring the degree of velarization in terms of Δ_{Bk} scores, there is, in fact, no effect of proficiency (7.0 vs. 6.9 Δ_{Bk}).[12] For the analysis of prevocalic /l/, participants were grouped by their adopted standard. The BrE learners

[12] The observed differences in F_1 and F_2 Hertz measurements likely reflect physiological differences between the speaker groups.

produced F_2 values equivalent to their L1 German (1940 vs. 1920 Hz for German) while the AmE learners produced slightly lower F_2 (1850 vs. 1920 Hz); none of the differences were statistical. The L1 German values were apparently pooled across all participants, which makes more detailed within-group comparisons difficult and further covers up possible variation due to anatomical differences. In terms of Δ_{Bk} scores, the AmE learners show a slightly higher degree of velarization (6.8 vs. 7.7 Δ_{Bk} for the BrE group). Although the prosodic effect is weakened by the conflation of intervocalic and non-prevocalic contexts, the results show that the degree of velarization in non-initial position varies with level of language proficiency (see also Baumotte 2009). The lack of a native speaker control group precludes statements about how similar proficient speakers are to native speakers of English. A comparison of the Δ_{Bk} scores with those obtained for intervocalic and non-prevocalic position in previous studies (cf. Figure 5.2), however, indicates large discrepancies between Baumotte's (2011) learners and native speakers.

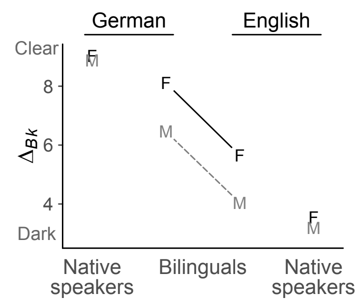


Figure 5.13: Late German-English bilinguals' production of final [l] (phonetic context: /-iːl/) in comparison with English and German native speakers: Bark difference scores (Δ_{Bk}) by language spoken (German vs. English), speaker group (late bilinguals vs. native speakers) and sex (M = male vs. F = female); data from de Leeuw et al. (2012).

de Leeuw et al. (2012) investigated the acoustic characteristics of non-prevocalic /l/ in the productions of $n = 10$ late German-English bilinguals (3 male), who varied in age of arrival in Canada (16–32 years) and length of residence (18–55 years). The study also recorded German and Canadian monolingual control groups ($n = 10$ each). A word list was used to elicit a number of monosyllabic items ending in /-iːl/. A comparison of the control groups showed that German /l/ had statistically lower F_1 (females: 350 vs. 550 Hz; males: 240 vs. 470 Hz) and statistically higher F_2 values (females: 1860 vs. 1060 Hz; males: 1550 vs. 890 Hz). The bilingual subjects produced German and English utterances. The acoustic measurements showed high variation among the late German-English bilinguals. Averaging across all speakers, the Δ_{Bk} scores of their German and English productions were intermediate between those of the native speaker controls. Figure 5.13 summarizes the results and shows that late bilinguals approximate the respective native speakers but fail to reach the values produced by monolinguals (around 3 Δ_{Bk} for Canadian speakers and 9 Δ_{Bk} for German speakers). This divergence from both native speaker groups may reflect L1 attrition as well as lack of complete attainment of Canadian velarization in this context. Table 5.5 summarizes the instrumental evidence in the literature on /l/ in GLE.

Table 5.5: Acoustic evidence for /l/-production in German Learner English.

					Hertz			Bark		
Study	Sex	*N*	Group	Position	F_1	F_2	Δ_{Hz}	F_1	F_2	Δ_{Bk}
Baumotte 2011	—	3	BrE-oriented	Prevocalic	528	1942	1414	5.2	12.8	7.6
	—	3	AmE-oriented	Prevocalic	599	1855	1256	5.8	12.5	6.7
	Female/male	5	More proficient	Intervocalic, Non-prevocalic	423	1513	1090	4.2	11.2	7.0
	Female/male	5	Less proficient	Intervocalic, Non-prevocalic	591	1883	1292	5.7	12.6	6.9
de Leeuw et al. 2012	Male	3	Late bilinguals	Non-prevocalic	443	988	545	4.4	8.5	4.1
	Female	7	Late bilinguals	Non-prevocalic	506	1396	890	5.9	10.6	5.6

5.5 *Aims of this study*

This study aims to shed light on cross-sectional patterns in the realization of laterals by German learners. The focus is on three aspects:

- the realization of /l/ in non-prevocalic vs. prevocalic contexts and the degree to which these are contrasted,
- the variation across proficiency levels, and
- systematic differences between learners linked to their target variety.

Cross-sectional patterns in the realization of non-prevocalic /l/

Theoretical contributions unanimously consider dark /l/ in non-prevocalic contexts a difficult structure (CAH; SLM; MDH/SCH; NM/NDH; MSA; LTD; OPM). It should therefore be acquired late by learners whose L1 does not have velarized allophones. L1-derived [l] is predicted as a substitute, and [w] may be a transitory sound due to articulatory constraints (SCH/MDH; NM/NDH; MSA; LTD; OPM). Research on GLE is generally supportive of the status of dark [ɫ] as a difficult structure: Auditory studies report accuracy rates ranging between .05 and .90 (see §5.4). A limitation of auditory work is its reliance on discrete categories. As velarization comes in degrees, acoustic data should provide a more nuanced account of /l/-production. Existing instrumental studies provide some guidance on /l/-variation in GLE. We build on this work and extend the scope to learners at different levels of pronunciation proficiency.

Cross-sectional patterns in the realization of prevocalic /l/

The two standard varieties of English offer different targets for the production of prevocalic /l/. Acoustically, the difference between AmE and BrE in prevocalic contexts amounts to about 2 Δ_{Bk}. For AmE as a target accent, various frameworks predict the acquisition of prevocalic [ɫ] to proceed with relative ease, due to onsets being either structurally strong environments (LTD) or universally unmarked (SCH/MDH). This suggests the possibility of higher degrees of velarization in pre- vs. non-prevocalic position. Acoustic work on GLE reports Δ_{Bk} scores ranging between 6.5 and 7.5 (Baumotte 2011), which is somewhat intermediate between the Δ_{Bk} span of BrE and German. Auditory work has reported on a general increase in degrees of velarization in prevocalic position in the productions of advanced learners (Langguth 2009). Our objective is to investigate acoustic properties of prevocalic /l/ across different proficiency levels.

Cross-sectional patterns in the contrast between pre- and non-prevocalic /l/

Based on theoretical considerations and existing quantitative work, learners at initial stages are expected to produce acoustically similar laterals in both positions. With learners at more advanced stages usually targeting a specific standard accent, the degree to which non-prevocalic and prevocalic /l/ are contrasted is expected to show an increasingly bimodal distribution. Learners adopting AmE as a model should continue to produce small differences, since all variants are expected to shift toward a darker realization. Subjects aiming at BrE, on the other hand, are expected to show productions with an increased positional contrast. In general, variation among learners in the magnitude of the contrast is therefore expected to increase toward higher proficiency levels.

5.6 Method and data

Informants, materials and procedures

The acoustic analyses of the learners recorded for this study are based on $n = 15$ word list items, including laterals in non-prevocalic ($n = 10$) and prevocalic position ($n = 5$):[13]

- Non-prevocalic /l/: *help, control, called, always, milk, people, animals, apples, bottles, vegetables*
- Prevocalic /l/: *languages, village, alive, usually, television*

The acoustic analyses therefore rely on a total of $n = 898$ measurements produced by the learners ($n = 605$ in non-prevocalic and $n = 293$ in initial position; $n = 17$ missing cases; one learner was excluded from the analysis due to incomplete data). The native speakers produced $n = 383$ /l/-tokens in total ($n = 256$ in non-prevocalic and $n = 127$ in prevocalic position; $n = 7$ missing cases).

There is general consensus that F_2 is the primary acoustic correlate of the clear-dark distinction. Differences between allophones are reflected both in the absolute value of F_2 as well as the difference between the first and second formant.[14] Based on the considerations outlined in §4.6, the Bark scale more closely approximates perceptual values and is therefore preferable to Hertz. While Bark values still reflect physiological differences between speakers, a metric that aims to control for such variation is the Bark difference score (here denoted as Δ_{Bk}). This is the difference between Bark-transformed F_1 and F_2.[15] The analyses in the present study rely on $F_2 - F_1$ Δ_{Bk} scores, a metric that has also been employed in other quantitative studies on /l/-variation (e.g. Espy-Wilson 1992; Van Hofwegen 2009). Formant measurements are usually taken at the point of maximal displacement (Espy-Wilson 1992; Oxley et al. 2007; Lee-Kim et al. 2013) or the midpoint (Sproat & Fujimura 1993; Huffman 1997; Carter & Local 2007; Recasens 2012). Since the segment boundaries between /l/ and an adjacent vowel are difficult to define, measurements for the present study were taken at the point of maximal displacement or steady state, which was determined visually. The final part of this section details the statistical procedures applied and may be omitted by the reader without serious consequences for the general understanding.

Statistical analysis

The acoustic properties of laterals were analyzed separately for learners and native speakers using robust multilevel regression models. The outcome was the $F_2 - F_1$ Δ_{Bk} score for a single /l/-token ($n = 15$ per subject),

[13] The data sets analyzed in the present chapter are openly available via the *Tromsø Repository of Language and Linguistics* (*TROLLing*) at https://doi.org/10.18710/G6PJ5F (Sönning 2020a).

[14] The (number of) formants measured and reported in previous work differs considerably, however:

- F_2 only (e.g. Recasens et al. 1995),
- F_1 and F_2 (e.g. Huffman 1997; Stevens 1998; Hawkins & Nguyen 2004; Oxley et al. 2007; Khattab 2011; de Leeuw et al. 2012; Barlow 2014; Lee-Kim et al. 2013),
- F_1, F_2, and F_3 (e.g. Lehiste 1964; Dalston 1975; Nolan 1983; Flege et al. 1995; Smith 2010; Recasens 2012),
- F_2 and F_3 (Kinnaird & Zapf 2004),
- $F_2 - F_1$ difference in Hertz (Sproat & Fujimura 1993; Jokisch et al. 2011) or Bark (Van Hofwegen 2009), and
- $F_3 - F_2$ difference (Slawinski 1999).

[15] Based on the observation that the distance between adjacent formants tends to be similar across speakers, it was proposed by Syrdal & Gopal (1986) as a vowel-extrinsic normalization method. It should be noted that Thomas (2011) points out that Bark-transformed $F_3 - F_2$ may in fact be the most reliable indicator for the clear vs. dark /l/ distinction. However, since F_3 is rarely reported, the use of this metric would have reduced the body of quantitative work available for comparison.

which expresses the degree of acoustic velarization. The models include varying intercepts on subject and word. For the predictor position, the models included varying slopes on subject. This means that each subject received their own estimate of pre- and non-prevocalic /l/-velarization. The purpose of the analysis was to use these values to describe systematic differences between speakers. For native speakers, differences were expected to surface between the varieties, and for learners, the factors of interest were proficiency and the adopted target variety. Essentially, then, the native speaker analysis reports four quantities: the average Δ_{Bk} scores for each AmE and BrE in pre- and non-prevocalic contexts. The learner model, on the other hand, describes how the degree of velarization varies across proficiency levels. From an FAR score of zero onward, the model describes two groups of learners depending on the adopted target variety. Beginners and lower-intermediate learners are treated as a single population, and for upper intermediate and advanced learners, we allow cross-sectional patterns to fan out for different target accents.

Insights from the literature review on the acoustic properties of English /l/-allophones were translated into prior distributions. For BrE and AmE native speakers, priors for pre- and non-prevocalic contexts were centered around the typical values reported in the literature and then spread out to cover a wide range of Δ_{Bk} values. Figure 5.14 illustrates the correspondence between the Δ_{Bk} scores from the literature survey, which are summarized using box plots (lower panel), and the priors, which are represented as probability curves (upper panel). For German learners, the prior covers a wide range of scores and reflects the lack of information on this learner variety. Based on the results reported by Baumotte (2011), the distributions for both contexts were centered at 6 Δ_{Bk}, but made very diffuse. Since Δ_{Bk} scores cannot be negative, all priors were truncated at zero (shown as a vertical grey line in Figure 5.14. For the GLE model, proficiency was given a regularizing prior centered at zero.

Details about the regression models are deferred to Appendix A.4.1 (native speakers) and A.4.2 (learners), where we define and describe the structure of the models (including all priors), report convergence diagnostics, and list the posterior distribution of all model parameters. In

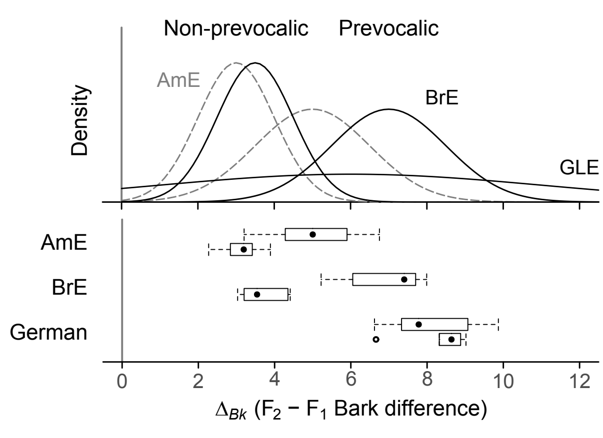


Figure 5.14: Prior information on the acoustic properties of *l*-allophones. Boxplots show Bark difference scores (Δ_{Bk}) from the literature review (cf. Figure 5.1). Curves represent the information that was added to the analysis in the form of prior distributions for AmE (grey dashed), BrE (black solid), and GLE (black solid).

what follows, we restrict our attention to transparent quantities, which we present in graphs and tables.

5.7 Constraints on generality

The generalizability of our findings on laterals in German Learner English is bounded in the following regards:

- *Language-external scope*: While our learners are predominantly from the south of Germany, there seem to be no dialectal features that would limit our inferences to this region (more specifically: northern Bavaria). The following findings may therefore be extended to other populations of German learners including Standard High German. Our subjects represent the type English learned at school and therefore do not permit statements about natural-setting populations of learners.
- *Language-internal scope*: The degree of generalizability to pre- and non-prevocalic laterals is limited. Half of our non-prevocalic contexts include (potentially) syllabic variants, and in the prevocalic set, 4 out of 5 tokens were in fact intervocalic. The findings for each condition are therefore a blend of two contexts. Further, we did not control for vocalic context, so we may be missing systematic patterns of variation along this dimension.
- *Speaking style*: As we relied on a reading task for data elicitation, the present findings extend to this type of speech production only.

5.8 Results

Let us now turn to the findings of the present study. We will deal with native speakers (§5.8.1) and learners (§5.8.2) in turn.

5.8.1 Native speakers

Before we consider the output of the regression models, let us take a look at our acoustic measurements. For each speaker in our data, we compute the average Δ_{Bk} for prevocalic and non-prevocalic contexts and then graph these by variety (AmE vs. BrE). Figure 5.15 shows the average readings for the $n = 26$ native speakers. Non-prevocalic contexts, which appear at the right-hand side, sit firmly at the dark end of the continuum for all speakers, irrespective of variety. AmE informants seem to produce slightly darker qualities, on average. In pre-vocalic position, the distributions are virtually non-overlapping and BrE informants produce decidedly clearer variants. There is a fair amount of between-speaker variation, which is perhaps slightly higher in prevocalic contexts.

Our regression models represent these patterns in more compact form. Figure 5.16 compares these estimates (points and error bars) to the range of scores from the survey of previous work (boxplots). Prevocalic and intervocalic position, which were reported separately in Figure 5.2, were collapsed for the purposes of this comparison. Figure 5.16 indicates that the results from this study are in good agreement with the literature. Table 5.6

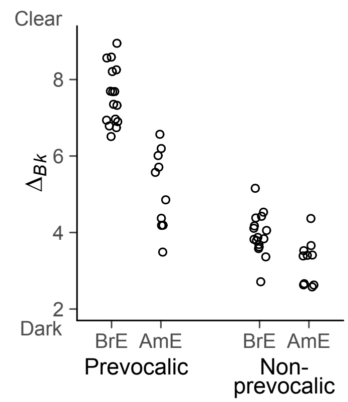


Figure 5.15: Overview of the acoustic measurements for the $n = 26$ native speakers: Average Bark difference scores by variety and phonetic context.

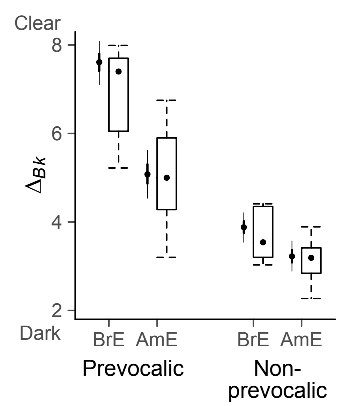


Figure 5.16: Estimates for the AmE and BrE native speakers in the present study in comparison to the literature. Boxplots show Bark difference scores for pre- and non-prevocalic position from the literature review. Error bars represent 50% and 90% posterior intervals.

lists the estimated $F_2 - F_1$ Δ_{Bk} scores for the British and American English subjects in this study. In pre- and non-prevocalic position, AmE speakers produced Δ_{Bk} scores of 3.2 and 5.1, respectively. BrE informants showed a lower degree of velarization in final position (3.9 Δ_{Bk}) and clear variants in initial position (7.6 Δ_{Bk}). In other words, as expected, both clear and dark /l/ are clearer in BrE than in AmE subjects.

		Posterior quantiles			
Position	Median	.05	.25	.75	.95
AmE					
Prevocalic	5.08	4.54	4.86	5.31	5.62
Non-prevocalic	3.22	2.89	3.08	3.36	3.57
BrE					
Prevocalic	7.61	7.11	7.41	7.81	8.09
Non-prevocalic	3.88	3.54	3.74	4.02	4.21

Table 5.6: Estimates for the AmE and BrE native speakers.

5.8.2 German learners

The acoustic measurements collected for the German learners are shown graphically in Figure 5.17, where (a) prevocalic and (b) non-prevocalic variants are juxtaposed and learners are arranged according to their FAR score (plotted on the horizontal axis). The points denote, for each speaker, the Δ_{Bk} average across all lexical items. Our continuous assessment of target variety orientation is added to this display by varying the fill color of the points: Grey denotes no specific orientation, with darker shades signaling an AmE target, and lighter values indicating orientation toward BrE. To isolate systematic aspects of the variation in lateral darkness, we now turn to a statistical model that aims to disentangle how Δ_{Bk} scores vary as a function of phonetic context (prevocalic vs. non-prevocalic), proficiency level, and target variety.

We skip technical details (see Appendix A.4) and look at a graphical

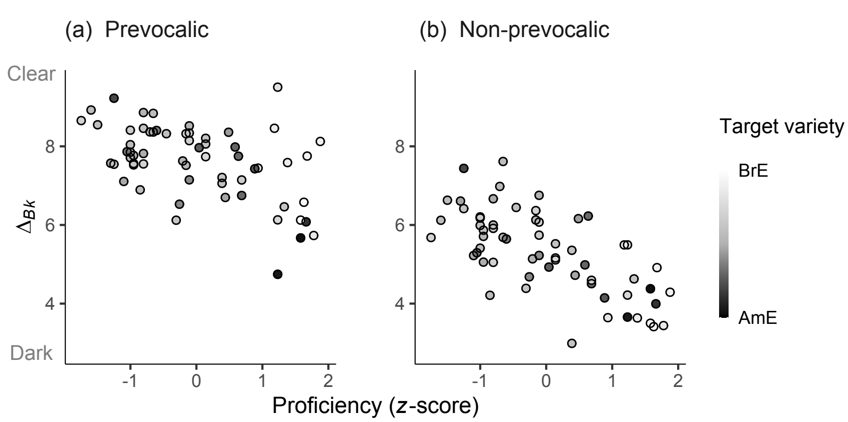


Figure 5.17: Overview of the acoustic measurements for the $n = 62$ German learners: Average Bark difference scores by proficiency levels (x-axis) and phonetic context. Grey shading is used to denote gradience in the target variety of the individual speakers. This figure was drawn with the *ggplot2* package (Wickham 2016).

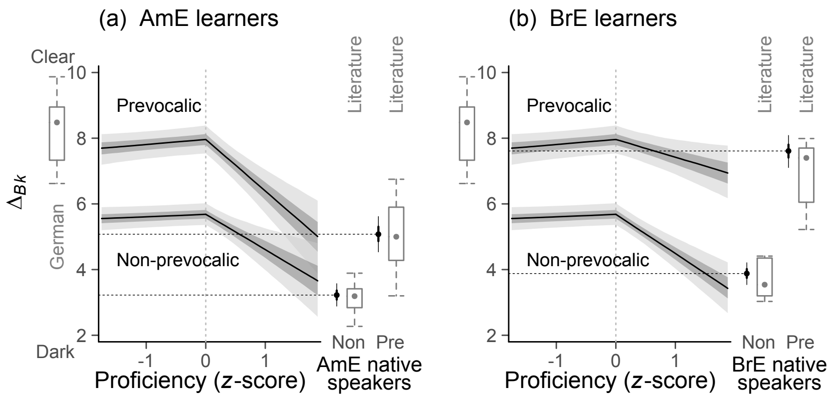


Figure 5.18: Acoustic structure of laterals in German learner productions by proficiency, target variety, and position: (a) AmE learners, (b) BrE learners. Grey boxplots at the left margin show the distribution of estimates for German native speakers from the literature. At the right margin, grey boxplots show the distribution of estimates for the (a) AmE and (b) BrE native speakers from our literature review. Estimates for the speakers recorded in the present study are denoted by black points with error bars. Error bars/bands denote 50% and 90% posterior intervals. "Non" refers to non-prevocalic position, "Pre" to prevocalic position.

summary of the results for the learners in our study. Figure 5.18 shows the findings for the two subpopulations: (a) AmE on the left, and (b) BrE on the righthand side. Up to a proficiency score of 0, these graphs are identical, since we are assuming that prior to this point, our informants resemble a single population of learners with no specific orientation in terms of target accent. Each graph shows two lines, which represent the different positions: Non-prevocalic contexts (lower trend line) and prevocalic contexts (upper trend line). The grey boxplot at the left margin shows the distribution of lateral measurements in German. At the right margin, reference values for native speakers from the respective target variety are shown, where boxplots again represent values reported in earlier work (see Figure 5.16). The following patterns are noteworthy:

- Up to intermediate levels of L2 pronunciation ability, there are virtually no changes in the quality of laterals, which remain stable at about 8 Δ_{Bk} in prevocalic position and roughly 5.5 Δ_{Bk} in non-prevocalic contexts.
- For prevocalic laterals, these values closely resemble those reported for L1 German and BrE native speakers. Non-prevocalic variants, however, fall short of BrE and AmE values by about 2 Δ_{Bk}.
- Beyond intermediate levels, BrE- and AmE-oriented learners diverge. Both groups seem to develop toward acoustic values that are characteristic of their respective target variety.
- In non-prevocalic position, both groups progress to increasingly darker renditions, shifting to about 4 Δ_{Bk}. While these values may be considered typical for native speakers of BrE, in AmE we commonly observe darker laterals of about 3 to 3.5 Δ_{Bk}.
- The key difference between the two groups is the behavior of prevocalic laterals. In AmE learners, these show a pronounced increase in darkness from 7.5 to roughly 5 Δ_{Bk}, which is in the range of scores observed for native speakers. In the BrE-oriented group, we find a slight increase in darkness to just under 7 Δ_{Bk}, which also sits well with native speaker measurements reported both in the literature and in the current study.

Proficiency	Median	Posterior quantiles .05	.25	.50	.95
Prevocalic					
Beginner	7.73	7.26	7.55	7.90	8.14
Intermediate	7.96	7.54	7.79	8.13	8.39
Advanced (AmE)	5.61	4.66	5.22	5.96	6.50
Advanced (BrE)	7.15	6.50	6.87	7.42	7.84
Non-prevocalic					
Beginner	5.57	5.23	5.43	5.70	5.91
Intermediate	5.68	5.36	5.55	5.82	6.02
Advanced (AmE)	4.07	3.19	3.71	4.44	4.99
Advanced (BrE)	3.89	3.26	3.64	4.15	4.54

Table 5.7: Estimates for learners at three proficiency levels.

5.9 Summary and discussion

This chapter dealt with the acquisition of English /l/-allophones by German learners. Our contrastive analysis of English and German laterals showed that the velarized, or dark, variant is a novel structure for L1 German learners. This allophone is produced with a secondary velarized constriction and occurs in non-prevocalic position in English. A survey of instrumental work on /l/-allophones in the standard varieties of English revealed systematic differences between BrE and AmE. In non-prevocalic position, both varieties show dark variants, but a slightly higher degree of velarization is documented for AmE. Prevocalic /l/, on the other hand, is realized differently in the two varieties: BrE shows clear variants, which are nevertheless slightly more velarized than in German. AmE shows dark variants, which are, however, less velarized than in non-prevocalic position. In contrast to English, German /l/ is clear in all positions.

Linguistic constraints are expected to interfere with the L2 acquisition of dark [ɫ]. Thus, a classification of velarized laterals as marked segments finds support from typology, L1 acquisition, and the physiology of articulation. The few indications from the literature on the perception of laterals suggest acoustic correlates of velarization to be of relatively low salience. It follows that theoretical work generally recognizes [ɫ] as a difficult structure for L1 German learners, who are expected to transfer L1 [l] in all positions and may show [w] as a transitory sound.

Empirical evidence on the acquisition of dark [ɫ] by German learners is supportive of its categorization as an intrinsically difficult structure. As such, laterals in GLE predominantly have been addressed using auditory methods, and most studies have relied on a binary classification of observed variants. A review of the quantitative literature suggests that findings may be sensitive to the number of categories chosen for the classification of allophones. Wieden & Nemser (1991), for instance, applied a three-way distinction of laterals along the clear-dark continuum and produced figures that were strikingly different from those reported in other work. Complementary instrumental techniques therefore seem warranted, as they obviate a discretization of the lateral continuum and allow for documentation of more nuanced patterns of variation.

For acoustic properties of German learners' lateral production, the literature provides less guidance. de Leeuw et al. (2012) reported on non-prevocalic /l/-production by L1 German residents in Canada with a great amount of TL experience. Findings by Baumotte (2011) on instructional-setting learners, on the other hand, are unfortunately of limited value due to methodological decisions. The objective of the present study was to extend the body of acoustic evidence on laterals in GLE by analyzing lateral production by informants across a broad range of proficiency levels. Specifically, this study aimed to shed light on the acoustic structure of pre- and non-prevocalic /l/-allophones as well as the degree to which they are contrasted in speech production.

Prior to an analysis of the data collected in this study, the literature was surveyed for instrumental records of the acoustic properties of laterals in BrE, AmE, and German. The findings for non-prevocalic /l/ in the productions of native speakers were consistent with the literature, showing that AmE speakers produce somewhat darker variants. As for prevocalic position, the results for native speakers in this study match reports in the literature, revealing a difference of comparable magnitude between the standard varieties.

The acoustic analysis of learner productions suggested that the degree of velarization in German learners' productions of non-prevocalic /l/ is gradient. While learners with low levels of L2 pronunciation ability distinguish them from prevocalic allophones, they nevertheless fall short of AmE and BrE benchmarks. Prevocalically, learners start out by producing laterals that are acoustically equivalent to both the L1 German counterpart and BrE values, at least as far as properties of the first two formants are concerned. Similar to the patterns observed for non-prevocalic /l/, production of prevocalic /l/ increases in darkness as learners advance in terms of their general L2 pronunciation ability.

From intermediate proficiency levels onwards, the acoustic structure of laterals shows increasing sensitivity to the adopted target variety. AmE- and BrE-oriented learners diverge and rearrange lateral acoustics toward native speaker values. The systematic increase in non-prevocalic lateral darkness across proficiency levels is estimated to be about 2 to 2.5 Δ_{Bk}. The scores observed in advanced learners (4 Δ_{Bk}) are consistent with those reported by de Leeuw et al. (2012), whose late German-English bilinguals produced Δ_{Bk} scores of 4 to 5.

In line with previous auditory work, the findings of this study support the characterization of dark [ɫ] as a difficult structure for German learners. Only at advanced levels do learners' productions of non-prevocalic /l/ approximate those of native speakers in terms of formant structure. While there is consensus among theories as to the difficulty of [ɫ], this verdict is reached by reference to different constraints. Consequently, there is a range of plausible explanations for the difficulty of velarized laterals:

- Meaning does not facilitate acquisition (CAH).
- Learners fail to perceive the [ɫ]-[l] contrast (SLM; OPM; MSA).

- [ɫ] is relatively more marked than [l] (MDH/SCH; OPM).
- Acquisition of [ɫ] requires suppression of natural processes that are active in the L1 (NM/NDH).
- Articulatory constraints interfere with the target-like production of [ɫ] (MSA).
- Non-prevocalic [ɫ] occurs in a *w*-marked position (LTD).

For the present, the processes underlying the acquisition of dark [ɫ] by German learners remain uncertain. Given the lack of experimental insight into German learners' perceptual sensitivity to the clear-dark contrast, statements about perceptual constraints are not warranted at this point. Nonetheless, findings of this study and de Leeuw et al. (2012) suggest that inaccuracies in speech perception are unlikely to be the only constraint on target-like production. This conclusion seems justified on the grounds that /l/-velarization comes in degrees. This seems to indicate that learners who have established a new category for [ɫ] may nonetheless fail to produce target-like laterals in non-prevocalic position.[16]

[16] However, the SLM would account for this discrepancy by stating that a new IL category may show different feature weights than TL [ɫ].

The results for non-prevocalic /l/ in the production of lower-proficiency learners did not show the expected transfer of L1 [l] in all positions. Rather, even this subpopulation of learners produced a contrast between non-prevocalic and prevocalic /l/, which amounts to about 2 Δ_{Bk}, on average. It thus appears that learners across all proficiency levels produce a systematic contrast between pre- and non-prevocalic /l/-allophones. Nevertheless, it is possible that this finding partly reflects phonetic context effects. Thus, in three of the tokens /l/ occurred in a phonetic environment favorable to velarization (*control*, *called*, *always*). Further, L1 transfer of coarticulation effects may have played a role. Apart from that, five tokens were syllabic (*people*, *animals*, *apples*, *bottles*, *vegetables*) and – as yet – little is known about the acoustic properties of this variant in German. Future work should aim to control for these confounding factors.

The interpretation of findings could also be put on a firmer basis through an analysis of laterals in both L1 and IL productions, as this would permit more valid statements about the degree to which IL variants show traces of L1 transfer. The current study relied on external evidence as reference points. The resulting interpretative difficulties should encourage future work to steer clear of such shortcomings.

Findings in the present investigation endorse the use of acoustic techniques to study /l/-variation. Thus, in non-prevocalic contexts, a large share of the L1 German subjects produced laterals at the category boundary, which may be located roughly between 5 and 6 Δ_{Bk} (see Figure 5.2). A discrete assignment of such tokens to the clear or dark category is likely to be problematic in terms of measurement reliability and does not adequately capture the phenomenon of interest: Lateral darkness in GLE comes in degrees, not in categories.

While the acoustic techniques applied have shed light on the gradience in /l/-velarization in GLE, the conclusions that are warranted on the basis of these results are nevertheless limited in several regards. For one, while Δ_{Bk} scores serve to contrast dark [ɫ] from its clear counterpart, they do not distinguish [ɫ] from vocalized allophones, as [ɫ] and [w] are similar in terms of the structure of the first two formants (Espy-Wilson 1992; Thomas 2011: 129). Another methodological caveat lies in the compounding of pre- and intervocalic /l/ for analysis. This may have masked more subtle nuances in lateral darkness, which are certainly to be expected in native speech. The selected lexical material also did not control for vocalic context, which has been shown to influence the acoustic structure of initial (AmE, BrE, German) and non-prevocalic (German) allophones. Non-prevocalically, the number of tokens elicited from each speaker was rather small, especially in prevocalic position ($n = 5$).

Future work on lateral allophony in varieties of English can build on the survey of instrumental work provided in this study. The correspondence between the patterning of Δ_{Bk} scores and auditory descriptions in the literature points to the external validity of this acoustic measure. In fact, our discussion of acoustic /l/-variation across and within varieties of English indicates that Δ_{Bk} scores offer a remarkable level of detail for the study of lateral variation. Of particular value is the fact that a quantitative reflection of /l/-velarization in target language varieties provides a frame of reference for the interpretation of learner productions. In that way, the statistical analysis is not restricted to internal comparisons but can be contextualized in a meaningful way, which allows for a more grounded interpretation of findings. For instance, consider a hypothetical investigation that reports a "statistically significant" difference between two learner groups, whose productions of non-prevocalic /l/ differ by 0.5 Δ_{Bk}. In the context of Δ_{Bk} reference values for English, it is perhaps more adequate to describe this difference as negligible. With the learners in this study representing a broad range of proficiency levels, the estimated proficiency effect in non-prevocalic lateral darkness (2 to 2.5 Δ_{Bk}) may also serve as a preliminary benchmark against which measurements gained in other populations of non-native speakers can be compared.

As discussed in the context of L2 speech theories, the implications of theoretical work for the study of IL phenomena hinge on a set of assumptions. For laterals, the literature provides fairly conclusive support of the relative markedness of dark [ɫ]. Work on the perception of lateral allophones is lacking, however. This gap in the literature needs to be addressed by future research. For the present, attempts to explain the relative level of difficulty of non-prevocalic /l/ by reference to perceptual constraints have to operate on the basis of assumptions extrapolated from studies on other populations of listeners.

Our line-up of theoretical predictions outlines further avenues for future work. Thus, competing predictions about the effect of coda complexity on the acquisition of non-prevocalic /l/ make this an area for further study. While past auditory work suggests variation in coda singletons vs. clusters to be small, this factor has not been investigated systematically to date. It also remains open for future research to determine whether syllabic [ɫ̩] develops in parallel to post-vocalic variants, and a separate treatment of syllabic [ɫ̩] may point to systematic differences. From the viewpoint of theoretical predictions, this variant appears to be more elusive as it sits at the boundary between consonants and vowels, which makes relative statements for [l ɫ ɫ̩] more difficult to derive. For instance, it is not clear whether markedness relationships for onset vs. coda and singleton vs. cluster asymmetries are relevant in the acquisition of syllabic variants.

Two further factors that deserve more attention are structural strength and speaking style. As for the latter, it is worth noting that the literature has produced contradictory evidence on the accuracy rate in more vs. less monitored speech. Non-prevocalically, the interaction between /l/-production and the immediate phonetic environment also constitutes a fruitful avenue for future work. Thus, a strong association between the acoustic properties of laterals and adjacent vowels has been reported in the literature. German /l/ is affected by the quality of the nucleus in all positions. BrE and AmE, on the other hand, only exhibit such effects in prevocalic position. For non-prevocalic /l/, on the other hand, the directionality of coarticulation is reversed, as it is the quality of the preceding vowel that shows coarticulatory variation (e.g. Wiik 1965; Cruttenden 2014). This constellation of contrasts puts forward a well-defined set of hypotheses to be addressed by future studies. In sum, the following suggestions may guide our ongoing efforts to understand the acquisition of English laterals by German learners:

- Experimental work on the perception of lateral allophones by L1 German listeners and native speakers is urgently needed.
- $F_2 - F_1$ Δ_{Bk} scores as a measure of /l/-velarization have proven useful for the investigation of both native and learner speech. One key advantage is the fact that the literature provides a firm basis for the contextualization and interpretation of results. The survey presented in Figures 5.1 and 5.2 may constitute a useful vantage point for future work on other L1 and L2 varieties of English.
- In contrast to a distinction between *clear* (prevocalic and intervocalic) and *dark* (non-prevocalic) contexts, a grouping into prevocalic, intervocalic, non-prevocalic, and syllabic allophones may provide a more nuanced account of lateral production and may yield new empirical insights into the position-sensitive realization of /l/ by German learners.
- Since AmE and BrE provide different pronunciation targets for prevocalic laterals, the adopted pronunciation model must be taken into account.
- Two further types of experimental control will strengthen the conclusions drawn from empirical work on non-native speech: (i) as coarticulation with adjacent vowels strongly influences the acoustic structure of laterals (see Figure 5.6), materials should be designed to hold context effects constant across focal conditions; (ii) the acoustic data collected for non-native speakers should include L1 lateral productions as a control condition.

6
English /r/

This chapter continues our treatment of English sonorants and turns to a cross-linguistically and dialectally heterogeneous class of speech sounds subsumed under the label *rhotics*. On a phonological level of analysis, this group is denoted by the IPA symbol /r/. What unites these sounds is their orthographic representation and the fact that, from a diachronic perspective, they derive from an alveolar tap or trill (Ladefoged & Maddieson 1996). Phonetically, however, there is no common feature; rather, they are related in terms of family resemblance, with certain members sharing articulatory and/or acoustic properties (Lindau 1985).

This chapter is concerned with the acquisition of English /r/-allophones by German learners. §6.1 starts out with a contrastive analysis, followed by a discussion of linguistic constraints underlying the L2 acquisition of English /r/ (§6.2). After an outline of theory-derived predictions about this structure in GLE (§6.3), §6.4 gives a survey of previous quantitative work on /r/ in L2 speech production. §6.5 lays down the aims of this study, followed by a description of methods and data (§6.6) and constraints-on-generality statements (§6.7). Results are presented in §6.8 and §6.9 concludes with a summary and discussion.

6.1 Contrastive analysis

German and English differ considerably with regard to the phonetic properties and distribution of rhotics. General British (GB) is a non-rhotic accent where /r/ only occurs pre-vocalically. Rhotics are thus not realized in postvocalic position, neither in singletons (*car, near*) nor in clusters (*cart, nerd*). The most common allophone in GB is the voiced post-alveolar approximant [ɹ]. It is articulated with the tip of the tongue located slightly behind the upper teeth ridge, creating a narrowing behind the alveolar ridge; air escapes freely and without friction. While the sides of the tongue touch the upper back teeth, the tongue body is lowered, with a slight retroflexion of the tongue tip. In GB, /r/ occurs word-initially in singletons (*red*, *rope*) and consonant clusters (*bring*, *true*) and word-medially in intervocalic position (*very, boring*). In partially devoiced contexts, /r/ may involve friction, e.g. in prevocalic clusters after /d/. Fully devoiced [ɹ̥] occurs after accented voiceless stops (*pray, cry*); a partially devoiced variant is typically found after unaccented voiceless stops (*battery, criterion*), voiceless fricatives (*friend, throw*) and in accented prevocalic clusters /str skr spr/ (*strong, scratch, sprint*) (Cruttenden 2014: 223–224).

In GB connected speech, /r/ may be pronounced if the following word starts with a vowel and the two words are linked (*here is* [hɪəɹɪz] vs. [hɪəʔɪz] or [hɪəɪz]). In such contexts, /r/-production has been observed to be sensitive to a number of social and linguistic factors (e.g. Bauer 1984; Hannisdal 2006; Mompeán-Gonzalez & Mompeán-Guillamón 2009). Thus, Hannisdal (2006: 164–169) reports linking-*r* to be (i) more frequent between function (vs. lexical) words (a difference[1] of $\Delta = .20$), (ii) less likely to occur before proper nouns than elsewhere ($\Delta = .50$), (iii) more likely before unstressed syllables ($\Delta = .30$), and (iv) less likely if another /r/ occurs in the immediate environment ($\Delta = .50$). In certain environments where postvocalic /r/ was lost, non-rhotic accents have developed centering diphthongs /ɪə eə ʊə/ as in *here, there, sure*. These diphthongal glides are undergoing development into monophthongal gestures. Thus, /eə/ has developed into [ɛː], /ɪə/ is increasingly heard as [ɪː] across Great Britain, /ʊə/ has an alternative pronunciation [ɔː] in many words, and /ʊə/ → [ʊː] appears to be a recent development (Cruttenden 2014: 118, 153–156).

[1] The differences quoted here and elsewhere in this chapter are absolute differences on the proportion scale.

A considerable degree of regional and social variation is observable in /r/. In GB, it may be realized as an alveolar tap [ɾ][2] in intervocalic position (*America, very*) and after a consonant, most notably /θ/ (Eckert & Barry 2005: 80). Cruttenden (2014: 225) notes that a small number of speakers produce a once-fashionable labialized variant involving strong lip rounding. The tongue tip gesture may even be omitted, resulting in a [w]-like sound. A labiodental approximant [ʋ] may also be heard and in fact appears to be spreading in southern England (Folkes & Docherty 2000; Collins & Mees 2013: 95). Uvular trills [ʀ] and fricatives [ʁ] are found in parts of northern England and Scotland.

[2] Cruttenden (2014: 224) labels this variant as conspicuous, that is, characteristic of posh or upper-class speech.

In contrast to GB, General American (GA) is a rhotic accent, where /r/ is pronounced in all positions. The distributional differences primarily concern post-vocalic contexts. GA is more conservative, retaining /r/ in words like *car, near, cart*, and *nerd*. There are also differences in realization: Retroflex [ɻ] is the more common allophone in GA. The articulation of [ɻ] may vary between speakers, however. It typically involves two constrictions, (i) narrowing in the oral cavity, which may involve the tongue tip forming a retroflex constriction or the tongue body bunching upwards, and (ii) a pharyngeal constriction (cf. Delattre & Freeman 1968; Ladefoged & Johnson 2011: 94). Vowels preceding [ɻ] may anticipate gestures of the following rhotic and thus usually show *r*-coloring (or vowel retroflexion) as a result of regressive coarticulation (especially /ɜː/, cf. *bird, fur*).

Table 6.0: (Opposite page) Major /r/-allophones in Standard High German and General British by position.

IPA symbol	Voicing	Place	Manner	Distribution
German				
Prevocalic				
[ʁ]	Voiced	Uvular	Fricative	Most common allophone; used in all regions
[ʀ]	Voiced	Uvular	Trill	Rare; more common in intervocalic position, southern varieties, and careful speech
[ʁ̞]	Voiced	Uvular	Approximant	—
[ʕ̞]	Voiced	Pharyngeal	Approximant	Word-medially; in unstressed *er*; southwestern Germany, Ostmitteldeutsch
[ɾ r]	Voiced	Alveolar	Tap/trill	Tap [ɾ] more common than trill [r]; in Bavaria
[ɹ]	Voiced			Younger speakers; in English loanwords
Postvocalic				
[ɐ̯]	Vocalic	Central-open	Non-syllabic	After long vowels [iː eː ɛː yː øː uː oː] and in the prefix *er-*; in careful speech also in *ver-*, *zer-* (where [ɐ ɛ ə] are also common) and in *her-* (*hervor*); in casual speech also before [i̯] (*Ferien*) and in unstressed *-en* (*hören*) where careful speech has one more syllable (with onset *r*)
[ɐ]	Vocalic	Central-open	Syllabic	In casual speech in the 1st Pr. Sg. and Imp. Sg. of verbs ending in *-ern* (*änd*(*e*)*re*, *zög*(*e*)*re*)
[ʁ ʁ̞]	Voiced	Uvular	Fricative/ approximant	Also [ʀ ʕ̞ ɾ r ɹ]; in careful speech after short vowels [ɪ ɛ ʏ œ a ʊ ɔ] word-finally or when followed by a consonant (*gern*, *dort*, *dürr*); however, [ɐ̯] is very common in these contexts
∅	—	—	—	Common after [aː], word-finally (*Haar*) and when followed by a consonant (*Bart*)
General British				
Prevocalic				
[ɹ]	Voiced	Post-alveolar	Approximant	Most common allophone; prevocalic (*run*) and intervocalic singletons (*very*); after voiced consonants (except /d/); in the sequence [əɹ], [ɹ] is syllabic if [ə] is deleted (*motorist*)
	Partially devoiced	Post-alveolar	Approximant	In unaccented onset clusters after [p t k] (*mattress*), across syllable boundaries in rapid speech; in onset clusters [spɹ stɹ skɹ] (*sprint*, *strong*); in accented onset clusters after voiceless consonants (*true*, *pray*), after voiceless fricatives also in unaccented position (*mushroom*)
[ɹ̥]	Devoiced	Post-alveolar	Approximant	After accented [p t k] (*true*, *pray*)
[ɹ̝]	Voiced	Post-alveolar	Fricative	In onset clusters after [d] (*dry*)
[ɾ]	Voiced	Alveolar	Tap	In conspicuous speech intervocalically (*very*) and after [θ ð] (*three*)
Postvocalic				
[ə]	Vocalic	Central	Non-syllabic	After [ɪ i e ʊ]
∅	—	—	—	Elsewhere
[ɹ]	Voiced	Post-alveolar	Approximant	In the sequence [rə], [ɹ] occurs pre-consonantally if [ə] is deleted (*parrot*, *barrel*)

In German, vocalic and consonantal realizations of /r/ are distinguished. The major SHG allophone is produced with a uvular constriction, and, depending on the amount of narrowing, is either realized as a voiced fricative [ʁ] or an approximant [ʁ̞] (Wiese 1996: 171).[3] In SHG, the uvular fricative [ʁ] is found prevocalically. In postvocalic position, /r/ is typically vocalized and pronounced as a central open non-syllabic vowel [ɐ̯] (the diacritic indicating non-syllabic status). *R*-vocalization occurs regularly after long vowels and in the prefixes *er-, ver-* and *zer-*; after short vowels and /aː/, however, it is optional (Mangold 2015: 52). Here, variation between vocalized and consonantal realization depends on style, with [ɐ̯] being more likely to occur in casual speech (Hall 1993: 88). In unstressed syllables with orthography *-er*, vocalic *r* [ɐ], a syllabic vowel slightly more open than [ə], is found. The distribution of German /r/-allophones also shows regional variation. In some southern German varieties, the consonantal variants are realized as an alveolar tap [ɾ] or trill [r].

[3] With the IPA providing no symbol for a uvular approximant, we will use a diacritic to indicate a lowered uvular fricative, a speech sound that lacks friction and resembles an approximant.

Table 6.0, which shows the positional /r/-allophones in SHG and GB[4], reveals considerable variation in the realization of rhotics. The major contrasts are the following: While German learners need to learn new allophones in all positions, the two standard varieties represent different targets. In prevocalic position, the new approximant allophones are GB post-alveolar [ɹ] or GA retroflex [ɻ], respectively. For speakers of SHG, this involves a change in place of articulation (uvular → alveolar) and manner of articulation (fricative/approximant → approximant). In postvocalic contexts, GB and GA also differ. While /r/ is typically vocalized to [ɐ] in German, it is dropped completely in GB. Centering diphthongs are an exception, since the offset may be considered a form of vocalization. However, there are differences in vowel quality, with German [ɐ] having a more open quality, and (possibly) syllabic status; GB surface realizations are never syllabic. If the target is GA, [ɻ] must be learned postvocalically, where the L1 almost always shows vocalized variants.

[4] I am unaware of similarly detailed descriptions of GA allophones, which are therefore omitted.

6.2 *L2 acquisition of /r/: Linguistic factors*

6.2.1 *Markedness*

While there is considerable cross-linguistic variation in the phonetic realization of rhotics, typological comparisons indicate a bias toward certain variants. The UPSID database includes detailed descriptions of 282 *r*-sounds (Maddieson 1984: 78–82). The most commonly attested ones are trills (occurring in .46 of the languages, almost exclusively voiced) and taps (.37, almost exclusively voiced). Continuant rhotics are rather uncommon. This group includes voiced approximants of the type found in English (.10) as well as voiced fricatives (.03), which occur in German. In terms of place of articulation, retroflex (.12) and uvular rhotics (.01) are clearly disfavored; the large majority of *r*-sounds are alveolar and/or dental. This suggests that the German allophone [ʁ] is marked for both place and manner of articulation. The dialectal variants [r] and [ɾ] are unmarked for both parameters. The GB post-alveolar approximant [ɹ] is relatively less marked for manner than retroflex and uvular variants. As regards manner of articulation, the approximants [ɹ ɻ] fall in between taps/trills and fricatives. The cross-linguistic frequency hierarchies for place and manner features can thus be stated as follows:

- Place of articulation: alveolar/dental [r ɾ ɹ] > retroflex [ɻ] > uvular [ʁ]
- Manner of articulation: trill/tap [r ɾ] > approximant [ɹ ɻ] > fricative [ʁ]

Studies on the L1 acquisition of English consonants have reported /r/ to be among the segments that are acquired late (Smit et al. 1990; Dodd et al. 2003; but see Olmsted 1971). Figure 6.1b gives an overview of accuracy rates reported in different studies (see §3.3.2), which are summarized with trend lines for word-initial (grey) and -final (black) position. Figure 6.1a shows that /r/ belongs to the group of difficult consonants. The age of acquisition is 5;0 in initial position (grey circle) and about 3;9 in final position (black circle). The rate of acquisition differs by prosodic position. Rather unexpectedly, postvocalic /r/ (Templin 1957; Prather et al. 1975) and [ɚ] (Wellman et al. 1931; Smit et al. 1990) are acquired earlier than /r/ in onset position.

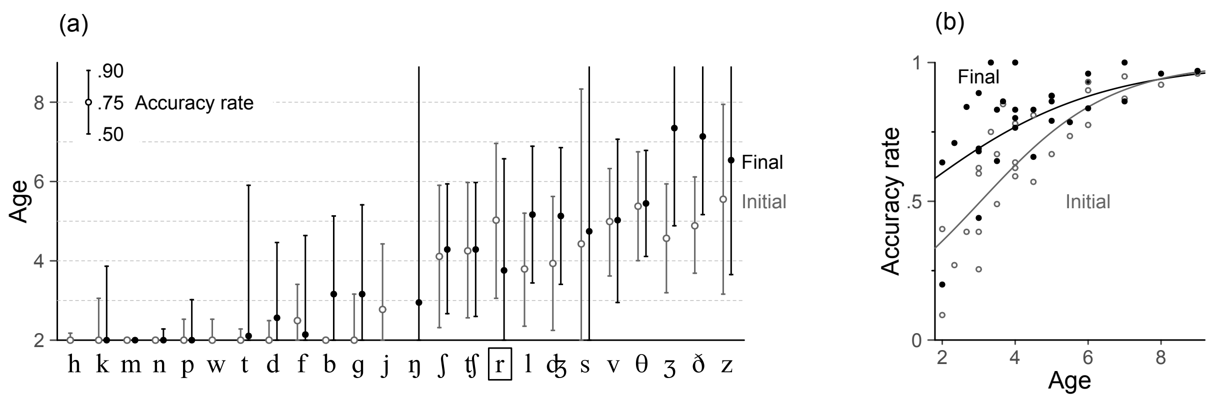


Figure 6.1: /r/ in English L1 acquisition in comparison to other consonants. Panel (a): Points mark the age of acquisition (.75), bars indicate the age of customary production (.50 accuracy, lower end), and age of mastery (.90, upper end). Estimates are based on a joint analysis of 4 studies (see §3.3.2). Panel (b): Accuracy rate in initial (grey) and final position (black) by age. The points denote estimates reported in the individual studies, the trend lines are averages based on the joint analysis.

In L1 acquisition, postvocalic /r/ typically shows vocalization (Smit 1993). Prevocalic /r/, on the other hand, is affected by gliding, a commonly observed pattern whereby the glides [w j] replace other sounds (Locke 1983: 67). Gliding of /r/ to [w] involves articulatory simplification by omission of the lingual constriction; it is a relatively persistent error pattern in L1 phonological acquisition.[5] Gliding of /r/ to [w] is also attested in speech disorders (Johns & Darley 1970).

Substitutes in L1 acquisition have been observed to vary in terms of phonetic characteristics. Cruttenden (2014: 225), for example, notes that the labiodental approximant [ʋ] (rather than the labio-velar glide [w]) frequently replaces [ɹ] in early child language. Blevins (2004: 230) reports on evidence indicating that English-learning children do not in fact equivalence-classify /r/ and /w/, but are able to distinguish instantiations of these categories in adult speech. Further, they contrast them in speech production, with *r*-approximations showing a lower level of intensity. Moreover, tongue gestures have been observed to differ between /r/- and /w/-realizations in 2 to 4-year-old English-learning children (Richtsmeier

[5] In a longitudinal study of $n = 145$ children aged 2;5 to 8 years, for instance, Roberts et al. (1990) found that while most error patterns had decreased rapidly by the age of 4, liquid gliding was among the few patterns that persisted. Similar findings were reported by Dodd et al. (2003), who analyzed $n = 684$ BrE children between 3;0 and 6;11 years of age.

2010) and perceptual studies suggest that children are able to discern these contrasts (cf. Kornfeld & Goehl 1974). As pointed out by Richtsmeier (2010), then, the label "substitution" may be misleading since in fact a covert contrast exists.

Onset-coda asymmetries in speech perception and production, which were discussed in the previous chapter on laterals (see §5.2.1), suggest postvocalic /r/ to be relatively more marked than prevocalic /r/. From the perspective of Vennemann's (1988: 21) preference laws, postvocalic singleton /r/ is disfavored due to its high sonority and resulting proximity to vowels on the sonority scale (cf. Figure 2.2). The loss of postvocalic /r/ in GB and SHG thus conforms to positional markedness asymmetries and coda preference laws. Evidence from L1 acquisition, however, lacks agreement with these principles, since postvocalic /r/ and [ɚ] emerge earlier than prevocalic /r/. This discrepancy is due to the persistence of gliding in onset position. Yet, as a comparison to the other English consonants shows, postvocalic /r/ still groups with the more difficult sounds, which may be suggestive of coda preference effects. Nevertheless, onset gliding appears to exert a stronger force in English L1 acquisition.

Post-alveolar [ɹ] is usually slightly labialized and thus involves a secondary constriction. While this makes [ɹ] a complex speech sound, the articulators are independent in the sense that they are not anatomically coupled. This, however, is not true for retroflex [ɻ]. Similar to velarized [ɫ], two lingual constrictions are involved: the alveolar gesture of the anterior tongue, and the pharyngeal constriction of the tongue root (Gick et al. 2008). The production of [ɻ] thus requires a greater degree of motor differentiation relative to other speech sounds (Boyce & Espy-Wilson 1997). This may partly explain the late emergence of [ɻ] in L1 acquisition (the majority of the studies quoted were conducted in the US). No lingual secondary constriction is involved in the GB post-alveolar variant [ɹ]. Indeed, as Wode (1977) notes, children developing toward a retroflex target typically show the progression [w] → [ɹ] → [ɻ], the retroflex variant being acquired and mastered last.

A further markedness indicator of English [ɹ ɻ] may be found in recent language change. As such, GB [ɹ] has been lost in weak positions – that is, syllable-finally. Further, tendencies in present-day BrE toward weakening of [ɹ] to [w] or [ʋ] through gestural omission parallel biases in L1 acquisition and may point to the diachronic instability of approximant rhotics. In GA, however, the rhotic has undergone gestural strengthening toward a retroflex articulation. As a result, it is resistant to coarticulatory influence (Boyce & Espy-Wilson 1997) and may therefore be less susceptible to weakening. Strengthening also led to perceptual dissimilation from neighboring vocalic segments[6], thus lending syntagmatic salience to /r/ in all positions. This parallels the behavior of /l/ in AmE, which shows liquid strengthening in laterals. As a result, laterals are generally darker and more robust to coarticulation effects (see §5.1).

[6] Except for /ɜː/ as in *nurse, merge*, which adopted *r*-coloring.

In summary, evidence from typology, L1 acquisition and the physiology of articulation points to the status of [ɹ] and [ɻ] as marked segments. Differences between the approximants pertain to articulatory complexity and L1 acquisition, which identify retroflex [ɻ] as the more marked structure.

However, unlike [ɹ], [ɻ] shows no evidence of a susceptibility of to language change.

Criterion	[ɹ]	[ɻ]
Implicational relationship	○	○
Cross-linguistic frequency	+	+
L1 acquisition	+	+
Synchronic/diachronic instability	+	○
Articulatory complexity	○	++

Table 6.1: Summary of markedness criteria for [ɹ ɻ]. Open circles ○ denote no (or inconclusive) evidence; (+)+ refers to (strong) evidence for marked status; (–)– refers to (strong) evidence for unmarked status.

6.2.2 *Frequency and orthography*

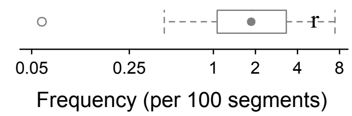


Figure 6.2: Frequency of /r/ relative to other English consonants, expressed as the number of instances per 100 segments. Rates are shown on the log scale, but tick mark labels give the absolute values of the rates. Data are pooled across 9 studies (see §3.3.3 for details).

A survey of the phoneme frequencies obtained in different studies shows that /r/ is a frequently occurring consonant in English (see §3.3.3). Estimates range from 3 to 7 instances per 100 segments, with a median rate of about 6. Figure 6.2 locates /r/ in the frequency distribution of English consonants and shows that it ranks among the segments with the highest rate of occurrence. Orthographically, /r/ is designated as <r rr wr rh>, the grapheme <r> providing an unambiguous representation. In non-rhotic accents such as GB, however, spelling provides a misleading cue to the realization of /r/ in postvocalic position.

6.2.3 *Similarity*

A number of studies have shed light on L1 German learners' perceptual sensitivity to English /r/. Weiher (1975) analyzed the perception of /r/ (spoken by a BrE native speaker) by $n = 28$ German 5th-graders.[7] Discrimination and identification tests contrasted [ɹ] with [w j l] in various positions using nonce words ($n = 94$). Overall, the learners performed well on the perception test. Correct categorization was highest in initial position, with an accuracy rate of .96 on the discrimination test and .84 on the identification test. In prevocalic singletons, [ɹ] was most likely to be confused with [w] (more than half of the substitutes in each test), but [l] and [j] were also observed. In prevocalic clusters, the error rate was slightly higher and [w] and [l] were the most likely substitutes. Triple clusters showed a bias toward [l]-substitution.

[7] On average, participants in this study were 11 years of age and had received 0.5 years of instruction.

Further insights are provided by Wieden & Nemser (1991: 159), who report on the development of speech production and perception in $n = 384$ Austrian learners of English from grades 3 to 11.[8] An X-ABC task was used to test learners' perceptual categorization of English sounds. For initial /r/, [ɹiːft] was contrasted with [ɹiːft], [wiːft] and [ʀiːft]. The results are summarized in Figure 6.3. While the L1 perceptual substitute [ʀ] only occurred at very early stages, the rate of [w]-misclassification declined from about .35 to .25. Correct identification increased rapidly and remained stable at .75 after 2 years of instruction.

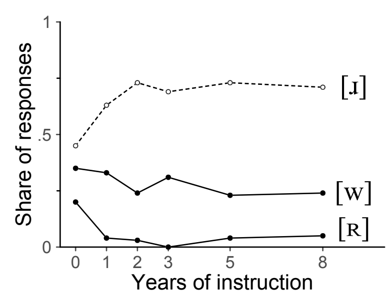


Figure 6.3: Prevocalic /r/: Perception accuracy in Austrian instructional-setting learners by years of learning. The graph shows the proportion of [ɹ]-tokens (mis-)identified as [ɹ], [w], and [ʀ]; adapted from Wieden & Nemser (1991: 159) with permission. In the original graph, it seems that the labels "correct /r/" and "correct /w/" have been confused.

[8] The study included learners at 6 proficiency levels, with $n = 64$ subjects in each group. Years of instruction was used as an indicator of developmental stage, which ranged from 0 years (grade 3; 8–9 years of age) to 8 years (grade 11; 16–17 years old).

Bohn & Best (2012) compared German and AmE listeners' perception of the AmE approximant contrasts /r/-/l/ and /r/-/w/ using a two-choice identification task and an AXB discrimination task. The L1 German group included $n = 18$ subjects, students at Kiel University with an average age of 21.[9] For the /r/-/l/ contrast, no statistical differences in the overall performance on the discrimination task were observed (German: .73; English: .78). For the /w/-/r/ contrast, German listeners gave statistically more /w/-responses near the category boundary and English listeners' rate of correct discrimination was statistically higher (.75 vs. .68). German listeners showed a bias toward /w/ in the middle of the continuum, indicating perceptual difficulties with the acoustic cues differentiating /r/ and /w/.[10]

A study by Engstrand et al. (2007) points to general difficulties in the perceptual decoding of place features in /r/-approximants. The authors investigated $n = 21$ Swedish listeners' perception of approximant realizations of /r/ with different places of articulation ranging from coronal to dorsal. An identification and a discrimination test were used. Overall, responses were predominantly non-categorical and the authors observed substantial uncertainty in the discrimination of stimuli and their identification as front or back articulations. These findings indicate considerable ambiguity in the perception of the place of articulation of approximant /r/-variants, which the authors propose as a factor underlying heterogeneity and sound change in the class of rhotics in general.

To recapitulate, German learners' perception of English /r/ shows relatively high accuracy rates, with target-like categorization accounting for the highest proportion across all stages. Perceptual interference of L1 [ʀ] is only observable at the initial stages of IL development. Across all proficiency levels, however, the distinction between /r/ and [w]-like sounds appears to cause the greatest difficulties and perceptual inaccuracies seem to persist at fairly advanced stages. None of the studies included the uvular L1 fricative [ʁ] and approximant [ʁ̞]. Especially the latter resembles the TL sound more closely than the trill [ʀ] and may thus be a more likely trigger of perception difficulties. Further, little is known about the perception of /r/ in postvocalic position.

[9] The native speaker data and the stimulus material (spoken by an AmE phonetician) were taken from Best & Strange (1992).

[10] Acoustically, the two sounds are in fact remarkably similar (Ladefoged & Johnson 2011: 203). Both show upward trajectories of the second and third formants; however, the F_2 rise is sharper in [w], whereas the F_3 rise is sharper in [ɹ]. It appears that German listeners are not sensitized to these nuances.

6.3 L2 acquisition of English /r/: Theoretical predictions

Based on the linguistic factors discussed, theoretical contributions provide guidance on the expected patterns in the acquisition of English /r/-allophones by German learners. This section gives an overview of theory-based predictions.

CAH The contrastive analysis of rhotics showed that German learners need to learn new allophones of a shared phoneme. Meaning therefore does not facilitate acquisition, and English /r/-allophones are predicted to be relatively difficult structures. In prevocalic position, the CAH anticipates transfer of uvular variants [ʁ ʁ̞]. Postvocalically, cross-linguistic influence is expected to surface in the transfer of vocalized variants [ɐ] after long vowels and in unstressed syllables with orthography *-er*. After short vowels and monophthongs similar to German /aː/, careful speech may also evidence transfer of [ʁ ʁ̞].

SLM Our literature review suggests the following auxiliary assumptions for prevocalic /r/: (i) [w] and [ʋ] are the perceptually closest L1/IL sounds; (ii) [w ʋ] are perceived as moderately similar to [ɹ ɻ]. While this implies relatively accurate perception of English /r/ by the majority of German learners, a minor population may experience perception-based difficulties in the production of English /r/. Accordingly, the SLM predicts target-like perception (and production) for the majority of German learners. Perceptual confusion is most likely to surface in [w]-like sounds in speech production. However, these predictions only apply to prevocalic /r/, since the indications in the literature are restricted to this context.

FCM Based on the FCM, prominence-driven biases in perception translate into the rank order of variants given in Table 6.2. This suggests a preference for [ʀ ʁ ʁ̞] in speech perception.

Consonant	Rank	[cons]	>	[ant]	>	[DOR]	>	[LAB]
[ʀ ʁ ʁ̞]	1	+		−		+		−
[ʋ]	2	+		+		−		+
[w]	3	−		−		+		+
[ɹ ɻ]		+		−		−		−

Table 6.2: Substitutes for [ɹ ɻ]: Predictions of the FCM.

SCH/MDH Based on the literature reviewed above, three auxiliary assumptions may be formulated: (i) [ɹ ɻ] are marked segments; (ii) [ɻ] is relatively more marked than [ɹ]; and (iii) [ɹ ɻ] are relatively less marked than [ʀ ʁ ʁ̞] in both place and manner of articulation. Acquisition thus involves replacement of a marked structure by a relatively less marked one. Hence, the MDH predicts English /r/-allophones to be acquired with relative ease. The SCH, on the other hand, arrives at different predictions. Since [ɹ ɻ] are marked structures, difficulty is expected. Further, regular error patterns observed in L1 acquisition and language change are predicted to surface in learner productions. As for substitute sounds, two tendencies appear plausible against the backdrop of the SCH. First, a universal inclination toward unmarked structures might surface in the tendency toward unmarked rhotics – specifically, the alveolar tap [ɾ] or trill [r]. Further, giving primacy to articulatory constraints, gestural simplification is expected to surface in gliding to [w ʋ] in prevocalic position and vocalization in postvocalic position. Finally, universal generalizations are also relevant at the level of syllable structure, with expected ease of acquisition in onsets relative to codas, and singletons relative to clusters. This means that postvocalic [ɹ] is expected to emerge later and production accuracy in both prevocalic and postvocalic position is expected to be higher in singletons.

NM/NDH Data from English L1 acquisition indicates that mastering approximant post-alveolar and retroflex rhotics [ɹ ɻ] requires the suppression of two lenition processes. Thus, prevocalic /r/ is affected by labialization, which involves omission of the lingual gesture to yield a [w]- or [ʋ]-like sound. Postvocalic /r/, on the other hand, is influenced by derhotacization – that is, articulation as a non-rhotic vowel or deletion. Both may be considered lenition processes as they reduce articulatory effort by gestural simplification. Predictions within the framework of natural phonology may therefore be formulated based on the following assumptions: (i) Labialization is a natural process that must be suppressed in the acquisition of prevocalic [ɹ ɻ]; and (ii) derhotacization is a natural process that must be suppressed in the acquisition of postvocalic [ɹ ɻ]. English L1 acquisition thus requires suppression of the process of labialization. Derhotacization, on the other hand, is active in BrE but has to be suppressed by AmE children. In German, the process of labialization is latent and derhotacization is active. The NDH thus makes the following predictions: (i) Prevocalic [ɹ ɻ]: With labialization being latent in the L1 and suppressed in the TL, the NDH predicts that this process will surface in IL, yielding [w]- or [ʋ]-like substitutes; (ii) Postvocalic ∅ [ə]: Since derhotacization is active in both the L1 and the TL, non-rhoticity is expected to be acquired with ease; and (iii) Postvocalic [ɹ ɻ]: The production of [ɹ ɻ] in postvocalic position requires the suppression of derhotacization, which is active in L1. Acquisition of postvocalic [ɹ ɻ] is thus expected to be difficult. Finally, the NM considers input frequency as a facilitating factor in perceptual learning. Figures in the literature suggest that /r/ is a frequent structure and therefore occurs at a high rate in learner input and output. Perceptual learning for /r/ is thus predicted to occur early, which in turn is a prerequisite for consistently correct production.

FM The FM postulates that correct perception of [ɹ] generates a new faithfulness constraint in the production grammar. Learning of the sensorimotor skills to produce [ɹ] leads to a new, undominated articulatory constraint. A phase of variable production conditional on focus on form is followed by a phase of overly faithful output. The model thus predicts an effect of speaking style at early stages of /r/-acquisition. At the final stages, the learner acquires the interaction of functional principles at the post-lexical level. Apart from style-dependent modifications of the degree to which distinctive features are implemented, this stage involves the acquisition of linking-*r* in connected speech. Finally, given the FM's focus on statistical learning, the working assumption related to frequency suggests early perceptual and sensorimotor learning of [ɹ].

MSA The MSA complements SLM-based perceptual constraints with the potential effect of articulatory difficulties. As mentioned above, the production of post-alveolar [ɹ] involves multiple constrictions; however, these gestures are performed by independent articulators, suggesting that motor constraints do not feature prominently in the acquisition of the BrE variant. AmE [ɻ], on the other hand, must be considered more complex due to the tongue forming two simultaneous constrictions. The MSA thus predicts retroflex [ɻ] to emerge late in L2 acquisition.

LTD Based on the LTD, we expect /r/-realization to vary as a function of structural strength. At the level of the syllable, this implies the following tendencies: (i) delayed acquisition in unstressed syllables; (ii) delayed acquisition of postvocalic /r/ in general; (iii) delayed acquisition in prevocalic clusters (vs. singletons), and (iv) no difference between postvocalic singletons and clusters, as /r/ is the first segment in postvocalic clusters and thus *s*-marked. The effect of strength marking on observed variants favors vocalization and deletion in *w*-marked units – that is, in postvocalic position generally, but also in prevocalic clusters. In postvocalic position, this may surface in a preference for deletion or L1-derived vocalized variants, and consequently relative difficulty of [ɹ ɻ]. In prevocalic clusters, we likewise expect deletion or gestural weakening to [w ʋ]. Finally, the assumption of increasingly systematic influence of phonetic and prosodic context suggests linking-*r* to emerge relatively late.

OPM Based on the discussion above, the acquisition of English [ɹ ɻ] may be best described in terms of the OPM's Markedness Corollary. We thus expect the acquisition of the English /r/-variants to initially show L1 influence in the form of transferred allophones – that is, [ʁ ʁ̞] in prevocalic and ∅ [ɐ ʁ ʁ̞] in postvocalic position. The influence of U may surface as outlined above (SCH/MDH). In prevocalic position, we would expect an emergence of unmarked rhotics [r ɾ] and/or articulatory simplification – that is, labialization to [w]- or [ʋ]-like sounds. Further, universal constraints disfavor the acquisition of approximant rhotics [ɹ ɻ]. These are expected to emerge last both in pre- and postvocalic position, the latter depending on the pronunciation model chosen. The OPM also makes predictions about the effect of speaking style. More formal speech is thus expected to show a higher degree of target-like production. Less formal speech, on the other hand, is predicted to be characterized by a stronger influence of L1 transfer.

Table 6.3 gives a summary of the predictions we have derived from theoretical work. We now turn to previous work on English /r/ in German learner speech.

6.4 *English /r/ in German Learner English: Previous work*

Several studies have dealt with the acquisition of English /r/ by German learners. While early contrastive treatments of L1 German/L2 English stated that German learners tend to transfer [ʀ ʁ] (Koziol 1959; Kufner 1971; Keutsch 1974), empirical evidence on /r/-production by German learners was first reported by Weiher (1975). A sample of $n = 12$ early-stage instructional-setting learners from northern Germany[11] was analyzed using three tasks: nonce words, familiar English words, and English sentences. Subjects were asked to imitate tokens produced by a BrE native speaker and therefore only produced /r/ in prevocalic position. Overall, high accuracy rates were observed both in prevocalic singletons and clusters

[11] The participants in this study were between 10 and 11 years of age, with an average of 0.5 years of instruction. In the original study, findings were reported separately for two school types. For the purpose of this study all learners represent the same population (early-stage instructional-setting learners) and Weiher's (1975) findings were therefore pooled across both groups.

Theory	Predicted patterns in L2 acquisition by German learners
CAH	Late acquisition of /r/ Prevocalic substitutes: [ʁ ʁ̞] Postvocalic substitutes: Vocalized [ɐ] in general, especially after long vowels; after short and /aː/-like vowels and also [ʁ ʁ̞] (careful speech)
SLM	Prevocalic /r/: Early acquisition by the majority of learners Substitutes in prevocalic position: [w ʋ] in a minor population of learners
MDH	Prevocalic /r/: Early acquisition
SCH	Late acquisition of /r/ Acquisition order: onset > coda; singleton > cluster Prevocalic substitutes: gliding to [w ʋ]; emergence of unmarked [r ɾ] Postvocalic substitutes: vocalization [ə ɐ], deletion ∅
NM/NDH	Frequency-driven early perceptual learning facilitates correct /r/-production Substitutes in prevocalic position: [w ʋ] as transitory sounds Postvocalic [ɹ ɻ] acquired late Postvocalic ∅ [ə] acquired early
FM	High input frequency leads to early perceptual learning High output frequency effects early sensorimotor learning Development: deviant production > stylistic variability > overly faithful production > variability reflecting connected speech processes Effect of speaking style (focus on form) at early stages of acquisition Linking-*r* acquired late
MSA	Acquisition order: [ɹ] > [ɻ]
LTD	Substitutes in onset clusters: ∅ [w ʋ] Substitutes in coda position: ∅ [ɐ] Late acquisition of linking-*r* Acquisition order: stressed > unstressed syllables; onset > coda; prevocalic /r/: singleton > cluster; postvocalic /r/: singleton = cluster
OPM	Late acquisition Prevocalic substitutes: [ʁ ʁ̞] at early stages; [w ʋ] at intermediate stages; [ɹ ɻ] at advanced stages Postvocalic substitutes: ∅ [ɐ ʁ ʁ̞] at early stages; ∅ [ə] at intermediate stages; [ɹ ɻ ɚ] / ∅ [ɐ] at advanced stages Effect of speaking style (focus on form): Higher/lower degree of TL-like production in more/less formal speech

Table 6.3: The acquisition of /r/ by German learners: Summary of predictions based on theoretical work.

(all ≥ .95), with no noticeable effect of task type on production accuracy. However, substitution patterns varied as a function of onset complexity. In the nonce word task, the predominant substitute in prevocalic singletons was [w] (.90). In clusters, [ɾ] was the most frequent substitute (around .50), followed by [w] (< .20), [l] and [ʁ] (each .15). Figures were similar in Cr- and CCr-clusters. In the English words, [w] was also the preferred substitute in prevocalic singletons (.67). For the English sentences, no information on context-sensitive substitution patterns was reported. However, [w] was again the most frequently observed realization (.50).[12]

[12] Three judges rated the acceptability of all /r/-productions. The results showed that while participants' productions were accurate in terms of place and manner of articulation, the temporal coordination of gestures as well as the transitions in sound sequences were not mastered in a NS-like fashion. The acceptability ratings were slightly higher in the sentences compared to the individual words.

Learner productions were reported to exhibit a high degree of labialization. Thus, /r/ was produced with lip rounding even in unrounded vocalic contexts.

Wieden & Nemser (1991: 165) reported on the production of /r/ by n = 384 Austrian learners of English.[13] In an imitation test, participants repeated nonce words they heard via headphones. Figure 6.3 summarizes the results for the three items [ʃɹuːd], [ɹaʊdʒd], and [θɹebz]. The correct production rate climbed from 0 to .75. [w]-type substitutes decreased sharply at early stages. Phonetically deviant responses – which included L1-derived [ʁ ʁ̞] and variants intermediate between [ɹ] and [w] – fluctuated around .25. Different patterns were observed in the production test, where German words were translated into English. The proportion of [w]-types was not as high as in the imitation test, likely due to an influence of orthography. The authors only report figures for the correct responses; these roughly follow the same trajectory as in the imitation test, rising from .15 to .60.

Wieden & Nemser (1991: 92) also report on the nature of deviant productions. They observe a higher tendency toward rounded variants in the cluster /gr/ (*green*, .20) and in *written* (.40), the latter presumably reflecting an influence of orthography (rate of rounding in other contexts: .10). The tendency toward a velar place of articulation was also strongest in *green* (.30) compared to the other contexts (.10). Finally, the highest rates of vocalization were observed in *very* (i.e. intervocalically) and *green* (.20), followed by prevocalic singletons (.13) and clusters where /r/ combined with a non-velar consonant (.07).

Insights into the naturalistic acquisition of English /r/-allophones by German learners are provided by Wode (1977), who studied n = 4 L1 German children (age: 4–9) during a 6-month stay in California. The author reports the developmental sequence [w] → [ɹ] → [ɻ]. L1-derived substitutions were also observed, but the author argues that they do not belong to what he refers to as the natural sequence of L2 acquisition. According to Wode (1977), children acquiring an L2 in a natural setting follow the same developmental path as L1 English-learning children, given that the target sound is sufficiently different from any potential L1 substitute.

Sieg (2004, quoted in Wode 2009) studied primary school immersion-setting learners from grade 1 to 4. Production accuracy was high, fluctuating between .70 and .90 (development from grade 1 to 4: .70/.90/.75/.85). The proportion of [w]-type substitutes (including [w ɹw ʋ] decreased (.90/.40/.40/.20); [ʁ] increased and then decreased (.10/.40/.10/.05), which could be due to an influence of orthography instruction during the first two years at primary school. Alveolar trills [r] were observed sporadically at early stages (.05/.15/.00/.00). The rate of zero-realizations increased steadily (.00/.05/.50/.70). Since the author provides no information on prosodic position, it is difficult to interpret these figures, however.

Langguth (2009) analyzed /r/-production in three groups of German learners (n = 7 each): 5th-graders (age: 10–11), 9th-graders (14–15) and 12th-graders (17–18). The main objective of the study was to determine the influence of the Eastern Franconian dialect on L2 English. Unlike the

[13] The study included learners at 6 proficiency levels, with n = 64 subjects in each group. Years of instruction was used as an indicator of developmental stage, which ranged from 0 years (grade 3; 8–9 years of age) to 8 years (grade 11; 16–17 years old).

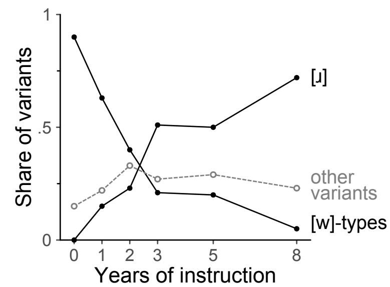


Figure 6.4: Prevocalic /r/: Production accuracy in Austrian instructional-setting learners by years of learning, ranging from 0 (grade 3) to 8 (grade 11). The graph shows the proportion of target-like [ɹ], [w]-types, and other variants. Adapted from Wieden & Nemser (1991: 165) with permission.

velar trill [ʀ], alveolar taps [ɾ] or trills [r] are characteristic for this southern German dialect. Three speaking styles were elicited: a word list, a (near-)minimal pair list, and semi-free speech prompted with questions on a text. Overall, the author noted that most errors were due to deletion (.64), resulting from a failure to produce /r/ either in initial consonant clusters or as a linking sound in connected speech. Across all task types, an increase in correct production rate in prevocalic /r/ was observed (.90/.95/.97), excluding one outlier in 9th grade. The rate of postvocalic /r/ realization varied as a function of prosodic context. Overall, learners produced .58 of postvocalic singletons, .63 of /r/ in postvocalic clusters, and .35 of postvocalic singletons in word-medial position before a consonant.

In another study on instructional-setting German learners, Langguth (2010) investigated /r/-production in 12th-graders at a grammar school in northern Bavaria (*n* = 34; age: 17–20). Subjects read a word list and a text, and answered questions on the text. Only 2 of the subjects showed consistently non-rhotic productions. While in the word list .37 of postvocalic /r/s were produced, the rate in the text was .21 and in free speech .28. For *r*-colored schwa [ɚ], the overall production rate was .33, compared to .51 after full vowels. These results indicate an effect of speaking style and structural strength on the production of postvocalic /r/.

Kautzsch (2017) reported on the production of postvocalic /r/ by *n* = 40 advanced German learners. Participants were between 20 and 25 years of age, had been taught English since the age of 10 or 11, and formed two groups in terms of target accent (BrE vs. AmE; *n* = 20 each). The study used a reading passage for data elicitation and also included *n* = 5 native speakers for each variety. Key findings are summarized in Figure 6.5. Native speakers of AmE (grey dashed lines) consistently produced high rates of approximant realizations. Linking contexts showed categorical rates, and in other contexts the prevalence of [ɹ ɹ̩] varied as a function of syllable stress (stressed .97, unstressed .82) and word class (content words .96, function words .79). In BrE native speech (black dashed lines), on the other hand, [ɹ] was categorically absent except for linking contexts, where function words (.80) and content words (.82) showed comparable rates.

The German learners in Kautzsch (2017) aligned with their target accent. However, the BrE group on average produced [ɹ ɹ̩]-rates that were slightly too high in non-linking contexts (.05 vs. 0 for native speakers) and too low in linking contexts (.20 vs. .80). The AmE group, on the other hand, produced fewer approximant realizations both in non-linking contexts (.40 vs. .90) and linking contexts (.75 vs. 1). Similar to native speech, the production of postvocalic /r/ is sensitive to the factors stress and word class: In non-linking contexts, the share of [ɹ ɹ̩] was higher in stressed contexts (Δ = .02/.09 for BrE/AmE subjects) and in content words (Δ = .02/.16 for BrE/AmE subjects). In linking contexts, the rate of [ɹ ɹ̩] was higher in function words (Δ = .23/.14 for BrE/AmE subjects).

An auditory study by Pascoe (1987)[14] explored the acquisition of /r/ by *n* = 26 learners from southern Bavaria. The informants – 8th–10th graders from different school types – had been learning English for approximately 3 to 6 years. The type of speech elicited can be described as semi-free: the retelling of a picture story and a guided telephone conversation. In

[14] See §3.3.1 for details.

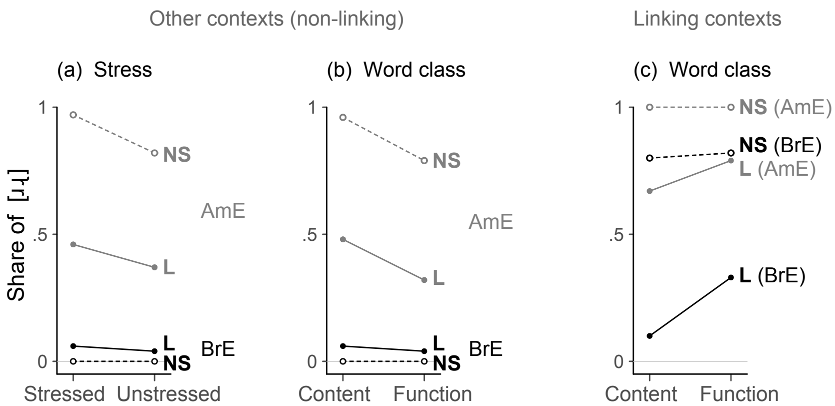


Figure 6.5: Postvocalic /r/: Rate of [ɹ ɻ] in advanced German learners with different target variety orientation in comparison to native speakers. AmE speakers (learners and native speakers) are shown in grey, BrE speakers in black; learners are represented with solid lines, native speakers with dashed lines. Panels (a) and (b) focus on non-linking contexts: (a) compares the rate in stressed vs. unstressed contexts, and (b) contrasts content and function words. Panel (a) shows data for linking contexts, broken down by word class (postvocalic /r/ in content words vs. function words). Data are from Kautzsch (2017).

prevocalic contexts, the author observed a high correct realization rate of .87. In the group of subjects with poor pronunciation ($n = 10$) the rate was .66. Pascoe (1987) also reports figures for linking-*r*, with an overall production rate of .18.

Quantitative evidence on /r/-production by instructional-setting learners from southern Bavaria was also reported by Kucharek (1988), who studied $n = 580$ learners divided into 5 groups: students in grade 6 ($n = 169$), 7 ($n = 138$), 8 ($n = 117$), and 12 ($n = 94$) and teachers ($n = 62$). Production accuracy was assessed with 6 items (*raining, wrong, tried, address, terrible, district*), which were elicited with a reading passage. Accuracy ranged between .88 and .94, with a steady increase from grade 6 to 12. Teachers produced target [ɹ] at a rate of .98.

In sum, the literature provides considerable guidance on /r/-production by German learners. Table 6.4 summarizes current knowledge about patterns of variation. Attested variants in postvocalic position are ∅ [ɹ ɐ] , and the occurrence of [ɹ] in this context appears to be sensitive to a number of factors including speaking style, linking context, position in the word and structural strength. As for prevocalic position, the set of observed substitutes includes [w ɾ l ʁ ɣ ʋ]. [w ʋ] are the predominant variants, with one study identifying [ʁ] as the most likely substitute (Pascoe 1987). The realization of prevocalic /r/, on the other hand, appears to vary conditional on a number of internal and external factors, which are summarized in Table 6.4. Target-like production has been observed to increase with years of instruction and L2 pronunciation ability. Developmental patterns have also been observed in the choice of substitute. Other factors that have been noted to underlie the varying nature of prevocalic /r/ are onset complexity, task type, and dialectal background.

6.5 Aims of this study

The objective of this study is to provide new quantitative insights into the following aspects of /r/-production by German learners:

- cross-sectional indications of developmental patterns in the realization of /r/ in prevocalic and postvocalic position,

- linguistic factors constraining the choice of variants in both positions, and
- systematic differences between BrE- and AmE-oriented learners in postvocalic position.

Table 6.4: Empirical evidence from auditory studies on prevocalic and postvocalic /r/ in German Learner English: reported production/accuracy rates and constraints on accuracy and substitutes.

Results		Notes	Study
Prevocalic r			
Accuracy rate	.95–1.0	Grade 5; age 10–11	Weiher 1975
	.90–.97	Grades 5, 9, 12; age: 10–18 years	Langguth 2009
	.88–.94	Grades 6–12; age: 11–18 years	Kucharek 1988
	.00–.75	Grades 3–11; age: 8–17 years	Wieden & Nemser 1991
	.70–.90	Grades 1–4; age: 6–10 years	Sieg 2004
	.66–.87	Grades 8–10; age 14–16 years	Pascoe 1987
Constraints on accuracy			
Complexity	$\Delta = 0$	No effect	Weiher 1975
Speaking style	$\Delta = 0$	Imitation: nonce words = English words = sentences	Weiher 1975
Constraints on substitutes			
Complexity		Singleton: [w]; cluster:[ɾ] > [w l ʁ]	Weiher 1975
		∅ in onset clusters	Langguth 2009
Familiarity		No effect	Weiher 1975
Task type		[w] more frequent in an imitation vs. translation task	Wieden & Nemser 1991
Development		[w]-like substitutes decrease rapidly	Wieden & Nemser 1991
		[w]-types decrease rapidly	Sieg 2004
		[ʁ] increases and then decreases	Sieg 2004
		[r] rare and only at early stages	Sieg 2004
		In naturalistic acquisition: [w] → [ɹ] → [ɻ]	Wode 1977
Dialect		[r] in Franconian speakers	Langguth 2009
Postvocalic r			
Production rate	.04	Advanced learners; target accent: BrE	Kautzsch 2017
	.79	Advanced learners; target accent: AmE	Kautzsch 2017
	.18	Grades 8–10; age: 14–16	Pascoe 1987
	.21–.37	Grades 5, 9, 12; age: 10–18	Langguth 2009
	.36–.63	Grade 12; age: 17–20	Langguth 2010
Constraints on production			
Speaking style	$\Delta = .16$	Word list (.37) > free speech (.28) > reading passage (.21)	Langguth 2010
Linking context	$\Delta = .20$	Semi-free speech	Pascoe 1987
	$\Delta = .35$	Advanced learners; target: AmE; linking (.75) > other (.40)	Kautzsch 2017
	$\Delta = .17$	Advanced learners; target: BrE; linking (.22) > other (.05)	Kautzsch 2017
Complexity	$\Delta = -.05$	Singletons (.58) < cluster (.63)	Langguth 2009
Position in word	$\Delta = .23$	Singleton: final (.58) > medial before consonant (.35)	Langguth 2009
Strength	$\Delta = .20$	Consonantal /r/ (.51) > *r*-colored schwa [ɚ] (.33)	Langguth 2010
	$\Delta = .09$	Advanced learners; target: AmE; stressed (.46) > unstr. (.37)	Kautzsch 2017
	$\Delta = .02$	Advanced learners; target: BrE; stressed (.06) > unstr. (.04)	Kautzsch 2017
Word class	$\Delta = .16$	Advanced learners; target: AmE; content (.48) > function (.32)	Kautzsch 2017
	$\Delta = .02$	Advanced learners; target: BrE; content (.06) > function (.04)	Kautzsch 2017

Cross-sectional patterns: Prevocalic /r/

Theory-based predictions are in disagreement regarding the degree of difficulty involved in the acquisition of [ɹ ɹ̧]. While some contributions suggest ease of acquisition (SLM; MDH; NM/NDH; FM), others expect difficulties (CAH; SCH; OPM). Predictions also diverge for the type(s) of expected substitutes. In prevocalic position, the range of predicted substitutes includes:

- [ʁ ʁ̥] (CAH); OPM: at initial stages
- [w ʋ] (SLM; SCH/MDH; NM/NDH); LTD: in clusters; OPM: at intermediate stages
- [ɾ r] (SCH/MDH)

A developmental progression of substitutes is predicted by the OPM, with L1-derived [ʁ ʁ̥] at early stages, followed by [w ʋ] as transitory sounds and ultimate emergence of [ɹ ɹ̧]. The OPM further predicts substitutes to vary with speaking style, with a tendency toward transferred [ʁ ʁ̥] in less formal speech and target-like [ɹ ɹ̧] in more formal styles. This study seeks to provide further insights into the distribution of variants at different proficiency levels.

Cross-sectional patterns: Postvocalic /r/

A number of accounts postulate [ɹ ɹ̧]-acquisition to be delayed in postvocalic relative to prevocalic position (SCH; NM/NDH; LTD; OPM). Similar to prevocalic /r/, the theoretical literature anticipates different substitutes in postvocalic position:

- [ʁ ʁ̥] CAH: in careful speech after short and /aː/-like vowels
- [ɐ] (CAH; SCH/MDH; LTD; NM/NDH)
- [ə] (SCH/MDH; NM/NDH)
- ∅ (SCH/MDH; LTD; NM/NDH)

The OPM predicts variants to depend on (i) developmental stage (with transferred [ʁ ʁ̥ ɐ] at early stages and U substitutes ∅ [ɐ ə] at intermediate stages) and (ii) style (with L1 variants in less monitored speech and target-like production in formal speech). The current investigation will focus on cross-sectional patterns, which have not been addressed previously.

The present study will also investigate a number of linguistic constraints that have been put forward in the theoretical literature. Specifically, the focus will be on the following:

Linguistic factors: Prevocalic /r/

- *Complexity*: Various accounts suggest an asymmetry between prevocalic singletons and clusters in L2 acquisition, predicting relative ease of acquisition before vowels (SCH/MDH, LTD). Empirical evidence reported by Weiher (1975), however, shows no effect of prevocalic complexity. The present study aims to shed more light on this discrepancy.

Linguistic factors: Postvocalic /r/

- *Complexity*: The theoretical literature generates opposite predictions about the acquisition of /r/ in postvocalic singletons vs. clusters. While the SCH/MDH assume relative ease of acquisition for postvocalic singletons, the LTD anticipates no differences due to the fact that /r/ occurs cluster-initially (thus being *s*-marked). Results reported by Langguth (2009) run counter to both predictions, showing a higher [ɹ ɻ]-rate in postvocalic clusters.
- *Linking context*: The LTD and the FM predict late emergence of /r/ in linking contexts. Evidence reported by Pascoe (1987) and Kautzsch (2017) seems to support this claim.
- *Structural strength*: The LTD predicts more target-like production in *s*-marked contexts, and there are indications that postvocalic [ɹ ɻ] first emerge after full and/or stressed vowels (Langguth 2010; Kautzsch 2017).
- *Speaking style*: The theoretical literature makes different claims about variation conditional on focus on form. The FM expects stylistic variation at relatively early stages of /r/-acquisition, where focus on form facilitates production accuracy. The OPM, on the other hand, predicts a bias toward target-like variants in monitored utterances and a tendency toward transferred structures in less formal situations. The literature offers some evidence of style-dependent variation in postvocalic [ɹ ɻ]-production, with the rate of production being higher in monitored speaking styles (Langguth 2010).

Postvocalic /r/ and target variety

Finally, we will follow up on the findings reported by Kautzsch (2017) and look at the realization of postvocalic /r/ depending on the adopted target accent. Our focus will be on how, and to which extent, in terms of rhoticity, German learners develop into two subpopulations over the course of L2 development.

6.6 *Method and data*

Informants, materials and procedures

The analysis of the data gathered in this study will be complemented by three data sets:[15] the *SpIL* corpus (Pascoe 1987), the recordings provided by Wunder (2012), and the data from Rank (2018).

- The *SpIL* corpus contains a total of $n = 303$ prevocalic and $n = 346$ postvocalic tokens from $n = 26$ different learners.
- The materials used by Wunder (2012) contain $n = 22$ prevocalic tokens and $n = 26$ postvocalic tokens. With $n = 30$ learners in the study, this adds to $n = 658$ prevocalic tokens and $n = 778$ postvocalic /r/-tokens ($n = 2$ missing observations each) for analysis.
- The study by Rank (2018) focused exclusively on postvocalic contexts and elicited $n = 1{,}933$ tokens from $n = 21$ subjects in total.

[15] The data sets analyzed in the present chapter are openly available via the *Tromsø Repository of Language and Linguistics* (*TROLLing*) at https://doi.org/10.18710/YDKDFG (Sönning et al. 2020a). See §3.3.1 for further information on the supplementary data sets.

- The materials for the current study included $n = 17$ prevocalic and $n = 48$ postvocalic /r/-tokens. The $n = 62$ learners produced a total of $n = 1{,}037$ prevocalic tokens ($n = 17$ missing observations) and $n = 2{,}933$ postvocalic tokens ($n = 43$ missing cases) for analysis.

The analysis of learners in the *SpIL* corpus and the data from Rank (2018) relied on the phonetic transcriptions provided by these authors. Subjects recorded by Wunder (2012) and in this study were analyzed auditorily by the present author and productions were assigned to five categories: [ɹ ɻ], [w ʋ], [ʁ ʁ̞], [ɐ] and ∅. In unstressed syllables, it was in many cases difficult to differentiate between central-open [ɐ] and central [ə]. No distinction was therefore made between the two, and all instances of unstressed vowels that showed no trace of rhoticity were coded as ∅.

Table 6.5 summarizes the distribution of variants observed in the four data sets. Due to the low number of observations for certain variants, these were aggregated for the ensuing analysis. Thus, for prevocalic variants, a binary distinction was made between *target* vs. *non-target*. In postvocalic position, [w ʋ] and [ʁ ʁ̞] were rare and therefore excluded from analysis. Accordingly, Table 6.5 distinguishes between *observed variants* and *categories used for analysis*. Overall, we see similar distributions across the data sets.

All /r/-tokens were coded for complexity – that is, whether they occurred as a singleton or constituted part of a cluster. For postvocalic

	Pascoe 1987		Wunder 2012		Rank 2018		This study	
Realization	Pr	(*n*)	Pr	(*n*)	Pr	(*n*)	Pr	(*n*)
Prevocalic /r/								
Observed variants								
[ɹ ɻ]	.86	(262)	.78	(510)	—	—	.87	(898)
[ʋ w]	.01	(3)	.22	(142)	—	—	.11	(110)
[ʁ ʁ̞]	.09	(27)	.01	(5)	—	—	.00	(0)
[ɐ]	.00	(0)	.00	(0)	—	—	.01	(6)
[ɣ]	.02	(6)	.00	(0)	—	—	.00	(0)
∅	.02	(5)	.00	(1)	—	—	.02	(23)
Categories used for analysis								
[ɹ ɻ]	.86	(262)	.78	(510)	—	—	.87	(898)
Other	.14	(41)	.22	(148)	—	—	.13	(139)
Postvocalic /r/								
Observed variants								
∅	.77	(270)	.72	(562)	.44	(841)	.68	(1992)
[ɹ ɻ]	.12	(43)	.23	(182)	.52	(1012)	.25	(738)
[ɐ]	.09	(33)	.04	(34)	.04	(80)	.07	(203)
[ʋ w]	.00	(1)	.00	(1)	.00	(0)	.00	(3)
[ʁ ʁ̞]	.01	(3)	.00	(0)	.00	(0)	.00	(6)
Categories used for analysis								
∅	.78	(270)	.72	(562)	.44	(841)	.68	(1992)
[ɹ ɻ]	.12	(43)	.23	(182)	.52	(1012)	.25	(738)
[ɐ]	.10	(33)	.04	(34)	.04	(80)	.07	(203)

Table 6.5: Distribution of variants in the three data sets. "Pr" denotes the proportional share of a (group of) variant(s).

/r/ in word-medial position, the maximal onset principle was followed. Postvocalic tokens were classified as occurring in a linking context where a word ended in <r(e)> and was followed by a word beginning with a vowel. Instances in which a speaker paused were not considered linking contexts. Finally, structural strength was coded using three categories:

- *strong* contexts, where /r/ occurred in a lexical word after a full vowel carrying lexical stress (e.g. *door, turn, chair*), including forms of *be* in main verb usage,
- *weak* contexts, i.e. unstressed syllables in polysyllabic words (e.g. *after, dinner, hour*), and
- *grammatical words*: monosyllabic function words (e.g. *for, or*) and contracted forms of *are*.

Postvocalic tokens recorded in the present study were further coded for speaking style, with productions from the word list and the short phrases classified as *isolated* (reflecting increased focus on form), and tokens from the short conversations reading task labeled *sentences*, which presumably exhibit less monitored speech.

Table 6.6 shows the distribution of observations across the levels of the linguistic factors in the data sets. While the token count gives the total number of observations per category, the type count documents the range of lexical items that served to elicit each condition. Due to differences in the tasks used for data elicitation, the data from Pascoe (1987) and

	Pascoe 1987		Wunder 2012		Rank 2016		This study	
Condition	Types	Tokens	Types	Tokens	Types	Tokens	Types	Tokens
Prevocalic /r/								
Complexity								
Cluster	23	125	7	300	—	—	9	868
Singleton	23	178	5	360	—	—	3	186
Postvocalic /r/								
Complexity								
Cluster	16	68	7	240	167	1314	15	930
Singleton	54	282	13	540	99	626	23	2046
Linking context								
Other context	65	292	15	600	249	1591	38	2901
Linking context	22	58	6	180	55	349	1	75
Structural strength								
Strong	38	172	13	480	178	1592	26	1676
Weak	35	178	7	300	86	348	14	1300
Speaking style								
Conversation	—	—	—	—	224	617	—	—
Text	—	—	—	—	41	903	18	1116
Word list	—	—	—	—	21	420	23	1860

Table 6.6: Distribution of types and tokens across the levels of the linguistic factors in the three data sets.

Rank (2018) show greater type frequencies. As a result of controlled data collection in the other two studies, the categories are represented by fewer types, which are, however, balanced across subjects. The next section gives details about the statistical procedures and may be skipped without loss of understanding.

Statistical analysis

Prevocalic /r/ was categorized as a binary outcome (target vs. non-target) and therefore analyzed with a multilevel logistic regression model with proficiency and complexity as predictors. In terms of target accuracy, singleton and cluster contexts were allowed to show different trajectories across proficiency levels and the model therefore estimates a trendline for each.[16] The variational structure of the model included varying intercepts on subject and word, and, following Barr et al.'s (2013) recommendations, by-word varying slopes for proficiency and by-subject varying slopes for complexity. The same model was fit to all three data sets, and a mathematical definition of it is given in Appendix A.5.1. As for the prior information included into the model, the insights gained from the literature review were combined into an informative prior for the population-averaged intercept of the singleton and cluster curves. This intercept denotes the average probability of observing target-like [ɹ] across learners with an intermediate level of pronunciation ability (i.e. at $z = 0$). Figure 6.6 shows the probability distribution, which was specified to align with the accuracy rates reported in the literature. The a priori expectation concentrates .90 of its mass in the interval [.22, .99]. As specified in more detail in Appendix A.5, all other parameters were given regularizing priors.

The data on postvocalic /r/, which was coded using three categories, were analyzed with a multilevel multinomial logistic regression model that included proficiency, linking context, structural strength, complexity, and (where applicable) speaking style and target variety as predictors. The random part of the model included varying intercepts on subject and word, as well as (i) varying slopes on subject for linking context, structural strength, complexity, and (where applicable) speaking style; and (ii) varying slopes on word for proficiency and (where applicable) speaking style, target variety and linking context. The analysis of the data gathered in the present study further included a breakpoint at the intermediate proficiency level of $z = 0$. Beyond this point, developmental trends were allowed to diverge based on the target variety chosen. This is to say that, for upper intermediate and advanced learners, the model returns separate estimates for AmE- and BrE-oriented learners. A detailed description of the model is given in Appendix A.6.[17] Previous studies supply prior information for the population-averaged intercept, which denotes the average probability of observing a given variant across learners with an intermediate proficiency level. As illustrated in Figure 6.7, 90 percent of the prior probability mass for [ɹ] is between .06 and .76. For ∅, this interval is [.14, .87] and for vocalized variants it is [.00, .44]. For all other parameters, regularizing priors were specified.

[16] In technical terms, the model included an interaction between complexity and proficiency.

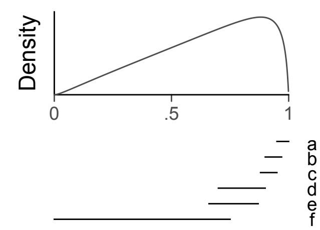


Figure 6.6: Prevocalic /r/: Prior for the intercept parameter in the model in comparison to the rates reported in the literature ([ɹ ɻ] only): (a) Weiher 1975, (b) Langguth 2009, (c) Kucharek 1988, (d) Sieg 2004, (e) Pascoe 1987, (f) Wieden & Nemser 1991.

[17] Appendix A.6 also reports convergence diagnostics and the posterior distribution of parameters.

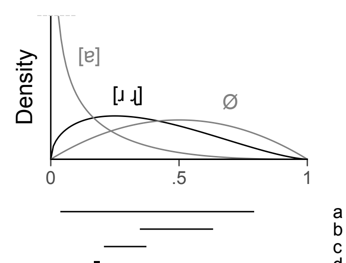


Figure 6.7: Postvocalic /r/: Prior for the intercept parameters in the model in comparison to the rates reported in the literature ([ɹ ɻ] only): (a) Kautzsch 2017, (b) Langguth 2010, (c) Langguth 2009, (d) Pascoe 1987.

6.7 Constraints on generality

Before turning to the results on pre- and postvocalic /r/, let us pause and reflect on the scope of our inferences.

- *Language-external scope*: As the learners in all four data sets are predominantly from southern Germany and northern Bavaria in particular, dialectal features of these German varietes would caution us against extending our inferences to other populations of German learners. As discussed above (see Table 6.0) the varieties traditionally spoken in this part of the country have distinct /r/-allophones (i.e. [ɾ r ʀ]). However, the distribution of variants observed in the present study shows that these dialectal variants do not feature at all in the data. While this does not imply that our subjects may be safely considered as representing a broader (rather than southern) population of learners, we would nevertheless argue that generalizations to other populations of German learners seem warranted. Further, we must keep in mind that the current pool of subjects represents instructional-setting learners, which limits inferences to this population.
- *Language-internal scope*: To comment on the degree of generalizability to English rhotics in general, we need to consider the representativeness of the lexical types investigated. As can be read from Table 6.6, the data sets differ in the amount of information they offer for language-internal generalizations. Thus, the controlled materials used in the present study and in Wunder (2012) yield a limited number of types, which constrains the generality of their insights. The data sets by Pascoe (1987) and Rank (2018), on the other hand, offer a broader inferential base. Apart from these differences in lexical coverage, there appear to be no systematic differences between the lexical types represented in the four data sets, apart from a general underrepresentation of less frequent words.
- *Speaking style*: The findings based on the present study and Wunder (2012) extend to reading tasks only and may therefore feature orthography-induced speech behavior, with the realization of rhotics possibly being influenced by their orthographic representation. This suggests that null variants ∅ may be underrepresented in these data sets. The data from Pascoe (1987) and Rank (2018), on the other hand, (also) represent more natural speech.

6.8 Results

We will first take a look at our results for prevocalic /r/ (§6.8.1) and then move on to postvocalic contexts (§6.8.2).

6.8.1 Prevocalic /r/

Let us first turn to a descriptive summary of the data. Figure 6.8 shows, for each data set, the distribution of the three variants by proficiency level. Each point denotes one learner. The top row displays the proportional

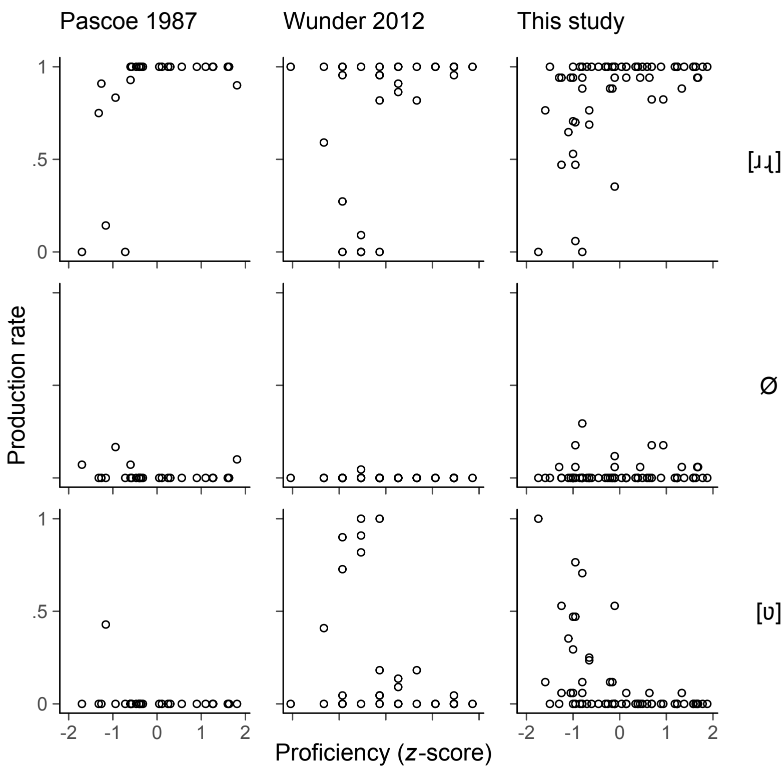


Figure 6.8: Realization of prevocalic /r/ by proficiency level in the three data sets. Production rate expressed as the proportional share of realizations.

share of [ɹ ɻ], the middle row that of ∅, and [ʋ] variants are given at the bottom. Overall, the three studies produce very similar patterns:

- [ɹ ɻ]: Toward higher proficiency levels, the rate of [ɹ ɻ] approaches 1. Learners at lower proficiency levels show considerable variation, with production rates extending across the entire scale.
- ∅: We can see that zero realizations (middle row) are consistently rare across all proficiency levels, hovering around 0, with occasional outliers with a rate of up to .25.
- [ʋ]: Labiodental substitutes (bottom row) are observed only at lower proficiency levels. There is considerable variation among lower-proficiency learners: The proportion of [ʋ] extends from 0 to 1 and individuals are spread out fairly evenly across the scale.

The results of the regression model are shown in Figure 6.9 and Table 6.7. Figure 6.9 concentrates on the distribution of target [ɹ ɻ] by onset complexity and proficiency level. The top panels show the predicted rate for [ɹ ɻ] across proficiency levels, irrespective of onset complexity (i.e. averaging across singletons and clusters). The error bands denote 50% and 90% posterior uncertainty intervals. The second row of panels shows expected rates separately for singleton contexts (solid lines) and clusters (grey dashed lines). Specific values are reported in Table 6.7. The three panels at the bottom show estimates for the difference between target [ɹ ɻ]-rates in prevocalic clusters and singletons. Values above 0 reflect

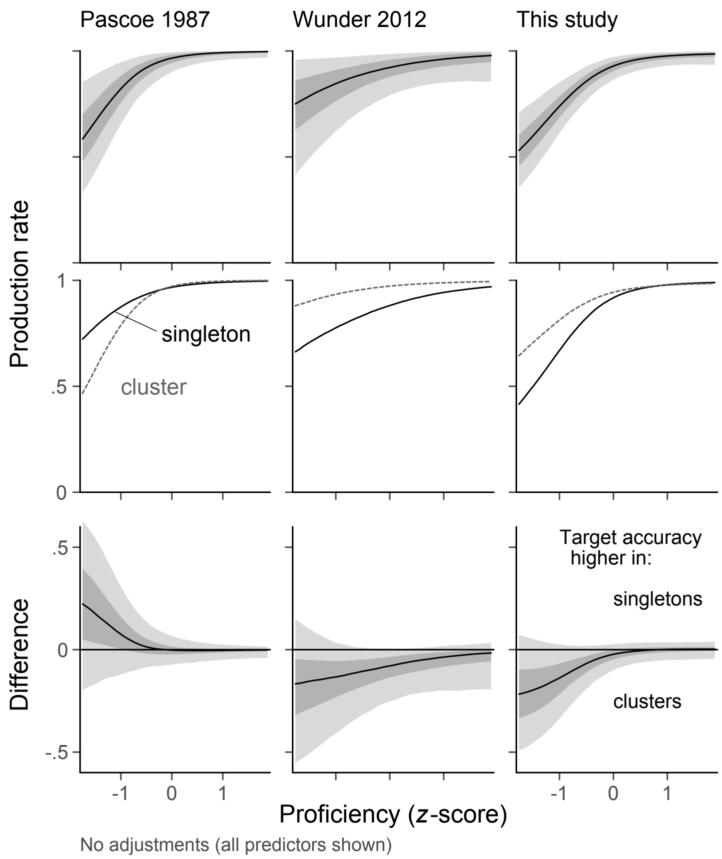


Figure 6.9: Results for prevocalic /r/: Estimated correct production rate by proficiency level and complexity. The top row shows the overall accuracy rate, i.e. the simple (unweighted) average over singleton and cluster contexts. The panels in the middle row show estimated trend lines for the two contexts. The bottom row of panels shows the observed difference in accuracy rate between singleton and cluster contexts. Error bands mark the 50% and 90% posterior intervals.

higher accuracy rates in singletons (consistent with expectation), and negative values signal more target-like production in clusters. Figure 6.9 indicates the following patterns:

- As for the overall accuracy rate across proficiency levels, the three data sources are in good agreement. The rate of target [ɹ ɻ] in prevocalic contexts varies from about .60 at lower proficiency levels to (just below) 1 at intermediate stages.
- Differences between singleton and cluster contexts only surface at lower proficiency levels. Recall that, in general, higher accuracy rates were expected in singletons. While the data from Pascoe (1987) are consistent with the expected cline, the subjects in Wunder (2012) and the current study show the opposite pattern. In absolute terms, differences are greatest at the lowest proficiency levels and amount to about .18 (Pascoe 1987), –.16 (Wunder 2012) and –.20 (the present study). Given the conflicting trends and the large margin of error, however, these data provide weak grounds for inferring differential accuracy rates in singletons and clusters.

	Overall			Singleton			Cluster			Difference		
Proficiency	*Mdn*	.05	.95	*Mdn*	.05	.95	*Mdn*	.05	.95	*Mdn*	.05	.95
Pascoe 1987												
Beginner	.66	.42	.88	.78	.45	.97	.57	.23	.89	+.18	−.17	+.55
Intermediate	.97	.90	.99	.97	.89	1.00	.97	.89	1.00	.00	−.07	+.06
Advanced	1.00	.97	1.00	.99	.95	1.00	1.00	.97	1.00	.00	−.04	+.02
Wunder 2012												
Beginner	.78	.48	.96	.70	.31	.95	.90	.54	1.00	−.16	−.51	+.11
Intermediate	.92	.79	.99	.88	.68	.98	.97	.86	1.00	−.08	−.25	.00
Advanced	.97	.86	1.00	.96	.77	1.00	.99	.91	1.00	−.02	−.19	+.03
This study												
Beginner	.59	.42	.75	.49	.25	.74	.70	.51	.85	−.20	−.46	+.06
Intermediate	.93	.87	.97	.92	.83	.97	.95	.88	.98	−.02	−.10	+.03
Advanced	.98	.94	1.00	.99	.92	1.00	.98	.94	1.00	.00	−.05	+.04

Table 6.7: Prevocalic /r/: Target accuracy estimates by study and proficiency level. *Mdn* refers to the median of the posterior distribution, and the .05 and .95 quantiles mark the 90% posterior uncertainty bounds.

6.8.2 *Postvocalic /r/*

For the findings on postvocalic /r/, we again first turn to a visual data summary. Figure 6.10 shows the distribution of variants by proficiency level and adopted target variety for the four data sets. The top row displays the share of [ɹ ɻ] in postvocalic position for each learner, the middle row that of null realizations, and the bottom row that of vocalized variants. For the data from the current study, the target variety orientation of learners is indicated via fill color: Black signals a strong inclination toward the AmE standard, white denotes BrE as a target variety. The following insights

	[ɹ ɻ]			∅			[ɐ]		
Proficiency	*Mdn*	.05	.95	*Mdn*	.05	.95	*Mdn*	.05	.95
Pascoe 1987									
Beginner	.05	.01	.15	.82	.67	.91	.12	.04	.25
Intermediate	.09	.04	.16	.83	.74	.90	.08	.03	.14
Advanced	.15	.06	.32	.78	.62	.90	.05	.01	.14
Wunder 2012									
Beginner	.07	.02	.15	.88	.77	.95	.05	.01	.13
Intermediate	.15	.09	.23	.82	.74	.89	.02	.00	.06
Advanced	.29	.15	.46	.70	.53	.84	.01	.00	.04
This study									
Beginner	.09	.03	.19	.85	.74	.93	.05	.01	.12
Intermediate	.21	.13	.30	.77	.67	.85	.02	.01	.06
Advanced (AmE)	.61	.31	.86	.38	.14	.67	.01	.00	.05
Advanced (BrE)	.06	.02	.17	.93	.83	.98	.00	.00	.02
Rank 2018									
Advanced (AmE)	.75	.61	.86	.23	.14	.37	.01	.00	.04
Advanced (BrE)	.11	.05	.20	.88	.78	.94	.01	.00	.04

Table 6.8: Postvocalic /r/: Production rate estimates by study, proficiency level, and target variety. The column labeled *Mdn* reports the median of the posterior distribution, and the .05 and .95 quantiles mark the limits of the 90% posterior uncertainty intervals. All predictors are set to zero, which denotes the following default condition: (i) non-linking context, simple (i.e. unweighted) average over (ii) strong and weak contexts, (iii) singletons and clusters, and (where applicable) (iv) all speaking styles.

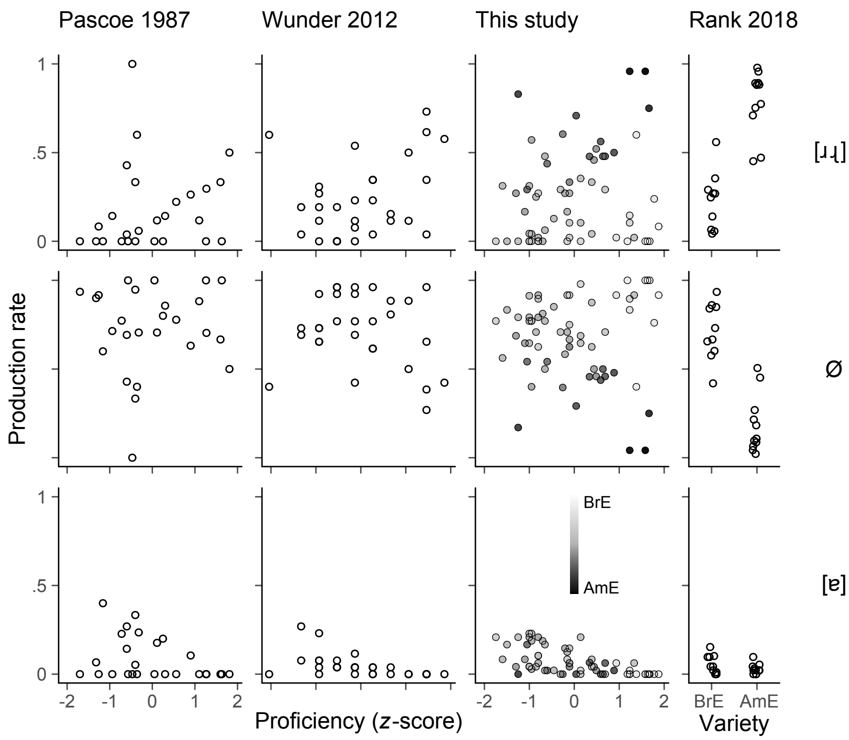


Figure 6.10: Realization of postvocalic /r/ by proficiency level for the four data sets. The production rate is expressed as the proportion of realizations. For the subjects in the present study, fill color is used to denote target variety orientation (black: AmE; white: BrE).

emerge from Figure 6.10:

- [ɹ ɻ]: The share of approximant /r/-variants is highly variable across all proficiency levels and in all data sets. Overall, their rate appears to increase slightly across proficiency levels.
- ∅: The proportion of null realizations also varies considerably across learners at all proficiency levels. On average, the prevalence of ∅ seems to decrease slightly as pronunciation ability increases.
- [ɐ]: In each study, vocalized variants play a minor role. [ɐ] tends to be observed only in lower proficiency learners, and rates below .40 seem to be typical.
- The target variety coding of the learners in the current study reveals that, as pronunciation ability increases, German learners diverge into two fairly distinct populations. In line with their adopted target varieties, BrE-oriented learners show lower rates of postvocalic [ɹ ɻ]. Subjects leaning toward the AmE standard, on the other hand, show high rates of [ɹ ɻ] in postvocalic contexts.
- The data from Rank (2018) show that upper intermediate and advanced learners align with their target variety, with AmE-oriented learners showing a much higher share of [ɹ ɻ] in postvocalic contexts. The distribution of production rates is consistent with the behavior of upper intermediate and advanced learners in the present study.

Figure 6.11 and Table 6.8 report model-based estimates. Figure 6.11 shows how the production rates for the three variants change, on average, across proficiency levels:

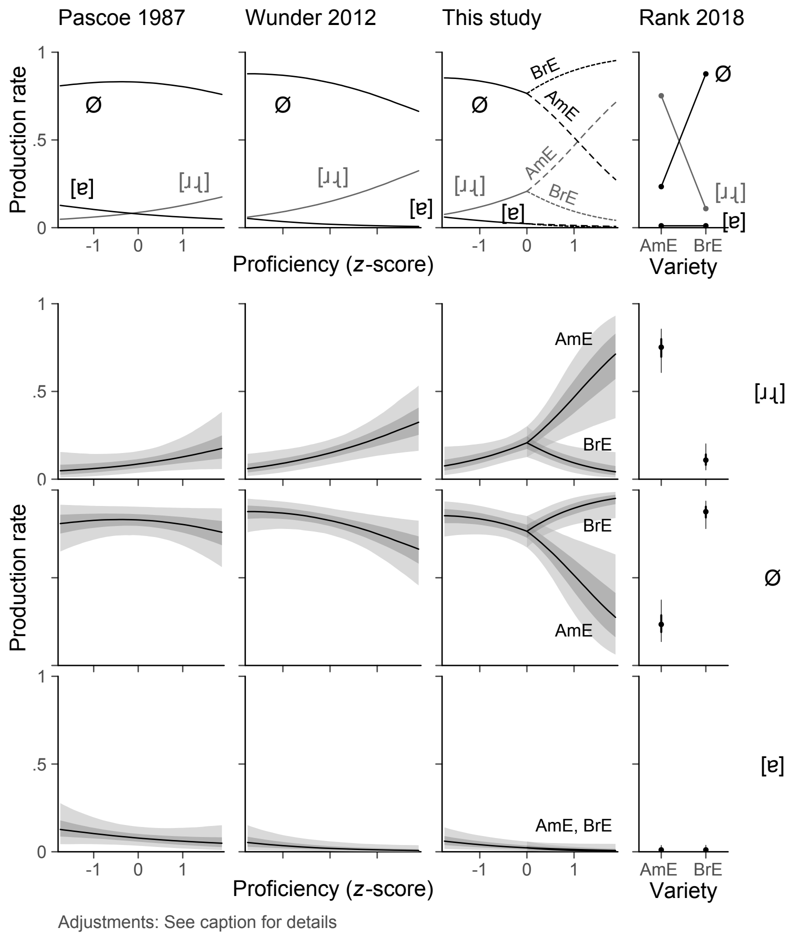


Figure 6.11: Postvocalic /r/: Comparison of the results for four data sets. The error bands mark 50% and 90% posterior uncertainty intervals. All predictors apart from the factor of interest are set to zero, which denotes the following default condition: *Linking context*: Other (i.e. non-linking) context; *Structural strength*: Simple (i.e. unweighted) average over strong and weak contexts; *Cluster context*: Simple average over singletons and clusters; *Speaking style* (only applies to the data from Rank 2018 and the present study): Simple average over all styles.

- The rate of [ɹ ɻ] increases from about .05 to .15 (Pascoe 1987) and to .30 (Wunder 2012). The data from the present study show that advanced German learners form two distinct populations: In AmE-oriented learner English, [ɹ ɻ] account for .61 (this study) and .75 (Rank 2018) of the postvocalic /r/-tokens. For learners with an inclination toward BrE, this rate drops to .06 (this study) and .11 (Rank 2018).
- Overall, null variants account for the largest share, with rates between about .70 and .90. Only in advanced learners with a rhotic target accent does this share drop to .38 (this study) and .23 (Rank 2018).
- Vocalized variants are rare, with rates of about .05 to .10 at the lower end of the proficiency scale and virtually 0 at upper intermediate stages.

We next turn to the linguistic factors constraining the production of postvocalic /r/. In the interest of simplicity, we will only be concerned with the share of approximant variants [ɹ ɻ]. Judging from Figures 6.10 and 6.11, little information is lost by this move, since two variants account for the vast majority of realizations. In what follows, we will utilize the output of our regression models to compare the rate of [ɹ ɻ] in different conditions. For linking context, for instance, we compare the rate of [ɹ ɻ] in linking vs. other contexts. Assume, for instance, that the share of [ɹ ɻ] is .50 in linking contexts and .15 in other contexts. We can express this comparison using a simple difference, i.e. +.35, signaling a higher share of [ɹ ɻ] in linking contexts. Our model allows us to draw this comparison at different proficiency levels, which means that we are essentially comparing two curves, i.e. how the share of [ɹ ɻ] in the two conditions varies across proficiency levels. It follows that we also obtain a trend of difference estimates, which might show us that the increment in the rate of [ɹ ɻ] in linking contexts may be as small as .05 at lower proficiency levels (e.g. .10 vs. .05) and increase to .40 at advanced stages (e.g. .50 vs. .10).

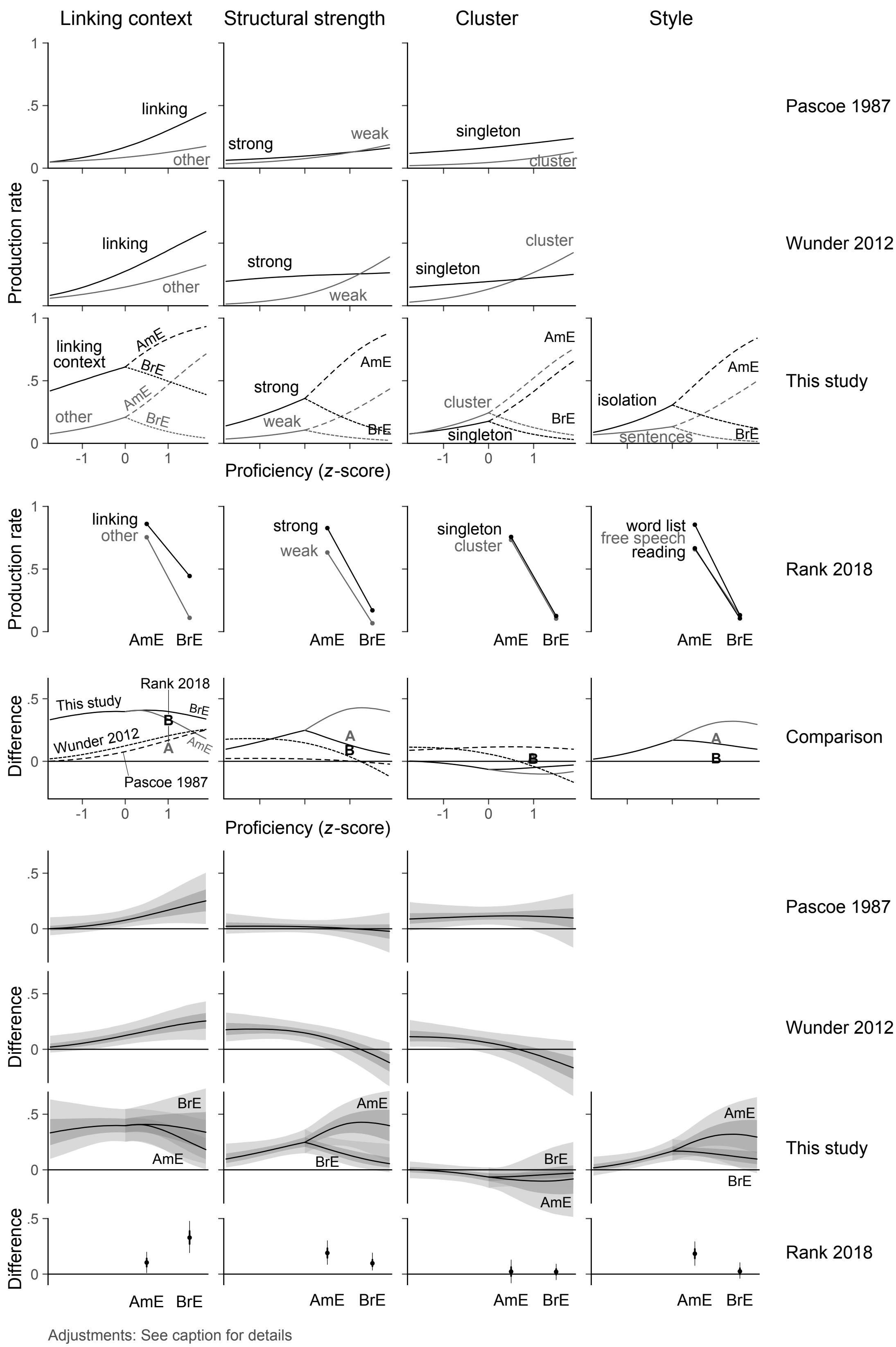
Figure 6.12 shows the relevant estimates based on the four data sets. The linguistic factors are arranged into columns (from left to right: linking context, structural strength, cluster context, speaking style). The rows represent the different data sets (from top to bottom: Pascoe 1987, Wunder 2012, this study, and Rank 2018). The upper part of the display (the top 4 rows) show how, for each condition of interest, the rate of [ɹ ɻ] varies across proficiency levels. The 5th row graphs, for each study, the difference between the curves in the panels above. Thus, the left-most panel displays three curves, which express, for each data set, the difference between the [ɹ ɻ]-rate in linking contexts and that in other contexts. This row of panels allows for a direct assessment of the link between the rate of [ɹ ɻ] and a given linguistic constraint, and whether the directionality and magnitude of this link is stable across proficiency levels and across data sets. The four rows of panels at the bootom break down these difference curves by study and add uncertainty intervals to the estimates. The following can be seen from Figure 6.12:

- Linking context: The share of [ɹ ɻ] is consistently higher in linking contexts, with difference estimates ranging from about 0 to .40 at lower proficiency levels and from roughly .20 to .40 at higher proficiency levels. Overall, the difference seems to increase slightly from lower to intermediate proficiency levels.
- Structural strength: Strong contexts seem to be more favorable to [ɹ ɻ], with differences ranging from 0 to .40. Difference estimates appear to show no consistent trend across proficiency levels.
- Cluster contexts: The comparison of singleton and cluster contexts yields inconsistent patterns, with difference estimates ranging between roughly –.10 and +.10.
- Speaking style: The rate of postvocalic [ɹ ɻ] is higher in isolated contexts compared to connected speech. Difference estimates range between 0 and +.25.

6.9 *Summary and discussion*

The objective of this chapter was to describe patterns of variation in the production of English /r/-allophones by German learners. A contrastive analysis revealed that the two languages differ markedly in the realization of rhotics. Prevocalically, English postalveolar and retroflex approximants

Figure 6.12: (Opposite page) Results for postvocalic /r/: Estimated rate of target-like /r/-production by proficiency level and various linguistic factors for each data set; from the top: Pascoe (1987); Wunder (2012); this study; Rank (2018). From left to right: Linking context, structural strength, complexity, speaking style. Error bands mark the 50% and 90% posterior intervals. All predictors apart from the factor of interest are set to zero, which denotes the following default condition: *Linking context*: Other (i.e. non-linking) context; *Structural strength*: Simple (i.e. unweighted) average over strong and weak contexts; *Cluster context*: Simple average over singletons and clusters; *Speaking style* (only applies to the data from Rank 2018 and the present study): Simple average over all styles.


Linking context
Structural strength
Cluster
Style
Pascoe 1987
Wunder 2012
This study
Rank 2018
Comparison
Pascoe 1987
Wunder 2012
This study
Rank 2018
Production rate
Difference
Proficiency (z-score)
linking
other
strong
weak
singleton
cluster
linking context
AmE
BrE
isolation
sentences
word list
free speech
reading
This study
Rank 2018
Wunder 2012
Pascoe 1987
A
B
Adjustments: See caption for details

[ɹ ɻ] contrast with German uvular fricative and approximant variants [ʁ ʁ̞]. The two standard varieties of English diverge in postvocalic contexts. While GB shows /r/-vocalization (to [ə]) after /ɪ e ʊ/ and deletion elsewhere, speakers of GA tend toward retroflex approximants in all positions. German, on the other hand, typically shows vocalization to [ɐ]. Prevocalically, then, German learners need to acquire [ɹ] or [ɻ] as new allophones. In postvocalic contexts, the two standard varieties represent different targets, either vocalization/deletion similar to L1 German (GB) or a postalveolar/retroflex approximant (GA).

From the perspective of linguistic constraints, English [ɹ ɻ] may be considered marked structures. This conclusion receives support from typology and L1 acquisition. Experimental work has shown that the majority of German learners perceive English /r/-allophones accurately. From this vantage point, theoretical contributions put forward a diverse set of suggestions. In general, there is disagreement as to the relative level of ease and expected substitutes in learner productions. Concerning expected patterns in /r/-variation, a number of linguistic factors have been suggested to show a systematic relationship with surface realizations.

The evidence offered in the literature is generally supportive of prevocalic /r/ being acquired early by L1 German learners. Onset complexity appears to exert little influence. The rate of postvocalic [ɹ], on the other hand, has been observed to be sensitive to a number of linguistic factors. The objective of this study was to make empirical contributions to the study of /r/ in GLE by exploring the distribution of variants in pre- and postvocalic contexts and across different proficiency levels. Further, this study followed up on indications in the theoretical and empirical literature, investigating variation in prevocalic /r/ as a function of onset complexity, and studying variation in postvocalic /r/ conditional on the factors linking context, structural strength, coda complexity, and speaking style.

6.9.1 Prevocalic /r/

German learners' realization of prevocalic /r/-tokens showed high accuracy rates, which approached 1 at intermediate stages. This is coherent with reports in the literature (see Table 6.4) and confirms theory-based predictions about a low degree of difficulty. It appears that the CAH and the SCH were less successful in predicting the relative level of ease for prevocalic /r/. Thus, it does not appear that lack of a functional contrast between TL [ɹ ɻ] and L1-derived substitutes impedes acquisition. As far as predictions of the SCH are concerned, it seems that markedness of English [ɹ ɻ] does not constrain acquisition in a notable manner. Contrary to suggested parallels to English L1 acquisition, German learners did not show a strong tendency toward gliding. While the occurrence of [ʋ w] in learner speech does show similarities to regular error patterns in the productions of English-learning children, only a minority of lower-proficiency subjects evidenced these substitution patterns. This indicates that this process operates much more strongly in L1 acquisition. Considering those contributions that correctly anticipate the early acquisition of prevocalic /r/, it is worth noting that these observations are explained by reference to different principles:

- The SLM attributes this to the perceptual distinctness of /r/ relative to the close L1 and IL sounds [ʋ w].
- The MDH puts forward the relatively less marked status of [ɹ ɹ̠] in comparison with [ʁ ʁ̞].
- The NM and FM point to input/output frequency to explain the early emergence of [ɹ] in onset position.
- The LTD refers to the fact that syllable onsets are *s*-marked.

Considering substitution patterns it can be stated that there was no evidence for the emergence of unmarked rhotics [r ɾ]. This is unsurprising given the salient acoustic and articulatory differences between taps and trills on the one hand, and approximants on the other. It therefore appears that, in the case of the heterogenous class of rhotics, perceptual constraints may block the emergence of unmarked /r/-sounds that do not occur in the learner's L1 (dialect) or the TL. As for the predominant substitutes in German learners, the results of the *SpIL* corpus (a clear trend toward [ʁ ʁ̞]) diverged from those obtained in the other data sets (almost categorically [ʋ w]). One explanation for this finding may be related to changes in English language teaching and the level of exposure to native speech through the media. Indeed, the gap between the more recent studies and that by Pascoe (1987)[18] amounts to almost 30 years. A less speculative account of the observed divergence, however, may be given by reference to stylistic effects. Thus, the speaking style elicited in the *SpIL* corpus differs from that of the other studies. In semi-free speech, learners have at their avail fewer cognitive capacities to focus on form. Reading passages, in contrast, allow for monitoring. The divergent patterns are thus also in agreement with the stylistic prediction made by the OPM: In casual speaking styles, transfer from L1 (in this case, L1-derived onset [ʁ ʁ̞]) is more likely to surface. A further confounding factor may have been the L1 of the rater. Given our current state of knowledge about L1-specialized speech perception, it is conceivable that raters with different L1 backgrounds arrive at different auditory categorizations of learner productions. Nevertheless, I feel that postalveolar and uvular approximant rhotics can be distinguished with a sufficient degree of reliability. Distinguishing between [ɹ] and [ʋ], on the other hand, is more difficult in many cases.

[18] The recordings were made between 1983 and 1985.

The occurrence of [ʋ w] as substitutes, which was predicted by several theories, parallels observations from English L1 acquisition. This substitution pattern may be interpreted as resulting from universal constraints, such as articulatory simplification, gestural weakening, or the natural process of labialization. Further, perception-based explanations cannot be ruled out since predictions from the perspective of the SLM are also consistent with the data.

Finally, findings on the role of onset complexity corroborate earlier work by Weiher (1975), showing little systematic difference between /r/-production in singletons and clusters, at least as far as accuracy rates are concerned (i.e. the share of [ɹ ɹ̠]). On the basis of current empirical insights, then, the role of onset complexity in the production of prevocalic /r/ remains unclear.

6.9.2 Postvocalic /r/

The realizations observed in postvocalic contexts indicate that L1-derived consonantal variants [ʁ ʁ̞] are rare. Vocalized [ɐ] appears to be slightly more likely to be transferred, but shows little variation as a function of proficiency level. Based on the CAH, [ɐ] was predicted to be more likely after long (vs. short) vowels other than /aː/-like nuclei. Vocalized variants may thus be more sensitive to linguistic (vs. learner-related) factors. Following the leads suggested by the CAH, systematic variation in the rate of L1-derived vocalized [ɐ] as a function of the preceding vowel appears to be a worthwhile avenue for future work.

Overall, the productions of German learners were predominantly non-rhotic, with a clear bias toward deletion across all proficiency levels. This is in agreement with predictions made by the SCH, NM/NDH, and LTD. While generating the same predicted patterns, these theoretical contributions arrive at different explanations. The SCH and NM/NDH refer to the natural process of de-rhotacization in coda position. The LTD, on the other hand, points to the fact that codas are structurally weak environments and favor weakened realizations, with deletion evidencing a form of maximal lenition. Thus, while the occurrence of deletion in postvocalic contexts cannot be ascribed to L1 transfer, it remains unclear whether it is attributable to the influence of universal tendencies (as suggested by the SCH, NM/NDH, and LTD) or whether these are instances of target language (GB) attainment. For instructional-setting learners who are exposed to rhotic and non-rhotic model accents, it is not possible to differentiate between these two structural components.

A further consistent pattern was the low rate of [ɹ]-variants in postvocalic contexts, which ranged from 0 to .20 among beginners, but increased, on average, at higher proficiency levels. The late emergence of postvocalic [ɹ] is indeed consistent with the prediction derived from the theoretical literature (MDH/SCH; NM/NDH; LTD). While these findings may be interpreted as indicating a general developmental increase in rhoticity in GLE, this interpretation is clouded by the fact that GB and and GA (or, more generally, rhotic and non-rhotic accents) provide different targets for the production of postvocalic /r/. The patterns revealed by the advanced learners in the present study and Rank (2018) indicate the emergence of two different populations of learners, one showing development toward consistently non-rhotic productions, the other toward a realization of /r/ in all positions. The proportional share of variants in these two studies is remarkably consistent, with BrE-oriented subjects showing [ɹ]-rates of about .10, compared to .70 for rhotic learners. These figures also align with those reported by Kautzsch (2017), who observed shares of .05 and .40 among BrE and AmE learners, respectively.

The present results are suggestive of linguistic constraints underlying the emergence of rhoticity. Thus, [ɹ] was observed to be favored in more monitored speaking styles and certain environments, namely in linking contexts and structurally strong positions. The latter finding is consistent with LTD predictions. Evidence for stylistic effects was observed in the data obtained in the present study and Rank (2018). As predicted by both the

FM and the OPM, the rate of postvocalic [ɹ] was higher in more monitored speech. However, the FM and the OPM made different predictions about the developmental stage at which focus on form is expected to surface in learner productions. While the OPM assumes stylistic constraints to apply across all stages, the FM stipulates what may be termed a window of monitor-induced variation at early stages of acquisition. It is difficult to empirically differentiate between these claims since, due to the late emergence of coda [ɹ], learners who are classified as proficient overall may be in the early stages of coda [ɹ] acquisition. The level of granularity offered by the foregoing analyses therefore does not license comments on the timing of stylistic variability.

In order for future studies on /r/-acquisition to build more confidently on theoretical work, efforts must be directed toward understanding perceptual constraints on /r/-production. As yet, little is known about the perception of /r/ in postvocalic position, which makes it difficult to hypothesize about perceptual biases in the acquisition of rhoticity. Further, it is not clear to what extent acquisition of retroflex [ɻ] vs. post-alveolar [ɹ] may be constrained perceptually.

While an auditory analysis may be successful in capturing categorical patterns in /r/-production, complementary acoustic analyses may provide insights at a finer level of detail. Thus, in the present investigation no distinction was made between post-alveolar [ɹ] and retroflex [ɻ] variants. With L2 phonological theories offering testable suggestions about the acquisition order of these two approximants, future work in this domain may profit from the application of acoustic techniques to gauge the degree of retroflexion in pre- and postvocalic position. Furthermore, the present investigation did not distinguish between GB target [ə] and L1-derived syllabic [ɐ] in unstressed syllables, despite the fact that these variants resemble different structural IL components (L1 vs. L2, in terms of the OPM). Instrumental techniques might be of value for advancing our understanding of more nuanced patterns of variation exemplified by these two distinctions ([ɹ] vs. [ɻ]; [ə] vs. [ɐ]).

To sum up, let us recapitulate some points that may inspire future work on /r/-allophones in GLE:

- Given that /r/-production varies systematically with rhoticity and theory-derived statements predict variation conditional on prosodic position, findings for onset and coda position need to be analyzed, interpreted, and reported separately.
- Studies concerned with linguistic constraints on the acquisition of prevocalic /r/ should focus on learners with low levels of L2 pronunciation ability, as /r/-production in this position is virtually categorical at more advanced stages.
- Research on systematic patterns in the acquisition of postvocalic /r/, on the other hand, should target advanced learners, who are more likely to show variable production rates that may reflect the operation of underlying constraints.
- Research on postvocalic /r/ should control for the pronunciation model chosen by the subjects (rhotic or non-rhotic variety), especially when

targeting more advanced learner speech.

- To provide a sounder basis for perception-based predictions about the acquisition of English /r/-allophones, experimental insights into the perceptual processing of postvocalic approximant rhotics and the [ɹ]-[ɻ] contrast are needed.

7 The labio-velar glide /w/

This chapter turns to the English labio-velar glide /w/ and completes our treatment of sonorants in GLE. After a contrastive analysis (§7.1), we discuss linguistic constraints underlying /w/-acquisition by German learners (§7.2). §7.3 derives predictions from theory and §7.4 reviews existing empirical accounts of /w/-production in this learner variety. The aims and method of this study are laid out in Sections 7.5 and 7.6, followed by the limits of the external validity of our results (§7.7). Findings are presented in §7.8 and §7.9 concludes with a summary and discussion.

7.1 Contrastive analysis

A comparison of the English and German consonant inventories reveals that the labio-velar glide /w/ is absent from German. Like other approximants, /w/ lacks a noise component and is therefore phonetically similar to vowels. Its articulation involves two strictures – rounded lips and a tongue gesture similar to that of a high back vowel. On phonological grounds, /w/ is classed as a consonant, as it occurs at the margins rather than the center of syllables. In English, /w/ is only found in onset position, where it may surface as a singleton or constitute part of a cluster.

The realization of /w/ is sensitive to phonetic and prosodic context. The vocalic allophone [w] occurs in initial position (*winter*), intervocalically (*away*), and after voiced consonants (*dwell*). /w/ is slightly devoiced after /sk/ (*square*), after an accented voiceless fricative (*switch*), and in an unaccented syllable after /p t k/ (*slipway, footwork, awkward*). The voiceless bilabial fricative [ʍ] is found after accented /t k/ (*twenty, quick*). Phonetic voicing in /w/ is thus sensitive to the voicing of the preceding segment. The degree of lip rounding varies with the following vowel. A higher degree of labialization is found before back, rounded vowels as opposed to front and more open vowels (Cruttenden 2014: 233f.).

While German lacks /w/, it has the phonetically similar vowel /uː/. Phonologically, the closest sound is /v/, which is commonly realized as a labiodental approximant [ʋ] in word-initial position (see §8.1). Apart from differences in place of articulation, English [w] and German [ʋ] are equivalent in manner of articulation and distribution, as both occur in onset position. In some southern German dialects, /v/ is realized as a bilabial fricative [β] or approximant [β̞] (Mangold 2015: 71).

In sum, the phonetically closest German sound to English /w/ is the close back vowel /uː/. Phonologically, however, the labiodental consonant /v/ is closest. In word-initial position, its common allophone [ʋ] also resembles English [w] in manner of articulation. The learning task is thus not a phonetic one, since German has the articulatorily similar vowel /uː/. Rather, it is phonological, as a new distribution must be learned for a familiar speech sound.

7.2 *L2 acquisition of English /w/: Linguistic factors*

7.2.1 *Markedness*

Cross-linguistically, the labio-velar glide is a relatively frequent speech sound: It occurs in .74 of the languages in the UPSID database (Maddieson 1984). It is also acquired early in L1 acquisition (Wellman et al. 1931; Templin 1957; Smit et al. 1990; Dodd et al. 2003). Figure 7.1b, which summarizes accuracy rates reported in the literature (see §3.3.2), shows that mastery of /w/ is attained early in L1 acquisition. A comparison of its developmental trajectory to those observed for other consonants (see 7.1b) reveals that /w/ is acquired before the age of 2, and therefore among the consonants that emerge early in L1 acquisition. Olmsted (1971) noted that [w] is a frequent substitute for other phones in L1 acquisition, ranking second after [d]. This is mostly due to the fact that [w] is the preferred replacement for [l] (.86 of the *l*-errors). Usage of /w/ as a substitute for other sounds is an instance of gliding, a regular error pattern observed in L1 acquisition, whereby sounds resembling the glides /w/ or /j/ replace other segments, typically /r/ and /l/.[1]

[1] Substitution of [w] for labiodental fricatives in the acquisition of L1 German was observed in a case study by Elsen (1991). While Olmsted (1971) observed cases of [w]-substitution for around 2/3 of all consonants, there was no instance in which it replaced /v/. However, Smith's (1973) case study on an English-learning child did report [w]-substitution for labiodental fricatives.

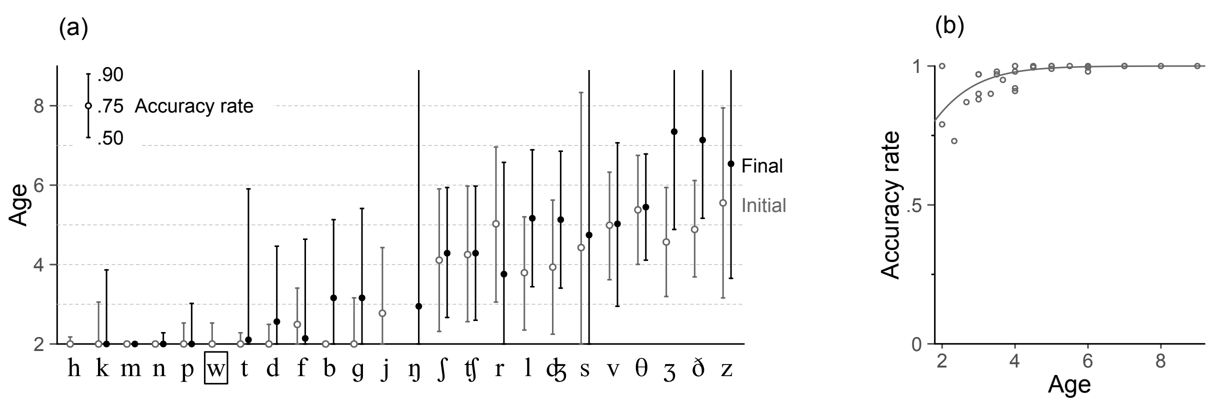


Figure 7.1: /w/ in English L1 acquisition in comparison to other consonants. Panel (a): Points mark the age of acquisition (.75), bars indicate the age of customary production (.50 accuracy, lower end), and age of mastery (.90, upper end). Estimates are based on a joint analysis of 4 studies (see §3.3.2). Panel (b): Accuracy rate by age. The points denote estimates reported in the individual studies, the trend lines are averages based on the joint analysis.

The early acquisition of /w/ by English-learning children may in part be due to the fact that the labio-velar glide only occurs in onset position. As has been discussed above (see §5.2.1), onsets are perceptually and articulatorily more robust than codas and therefore the unmarked prosodic environment. Another factor accounting for the patterns observed in L1 acquisition may be articulatory ease. While [w] involves two constrictions, these are formed by independent articulators. The typological distribution of vowel quality further indicates that high back vowels are strongly associated with lip rounding (cf. Figure 4.10, p. 60), suggesting a natural link between these strictures.

To recapitulate, typological records, L1 acquisition, and the physiology of articulation indicate that /w/ may be considered a natural, unmarked sound. Table 7.1 summarizes support for this classification.

Criterion	[w]
Implicational relationship	○
Cross-linguistic frequency	–
L1 acquisition	– –
Synchronic/diachronic instability	○
Articulatory complexity	–

Table 7.1: Summary of markedness criteria for [w]. Open circles ○ denote no (or inconclusive) evidence; (+)+ refers to (strong) evidence for marked status; (–)– refers to (strong) evidence for unmarked status.

7.2.2 *Frequency and orthography*

As Figure 7.2 shows, our survey of English phoneme frequencies (see §3.3.3) positions /w/ above average relative to other English consonants, with an estimated median frequency of 2.4 instances per 100 segments. Orthographically, /w/ is denoted by <w> (*water*) and <qu> (*query*), but also <wh> (*when*) and <u> (*linguist*). Among the exceptions are the high-frequency words *one* and *once* (Cruttenden 2014: 232).

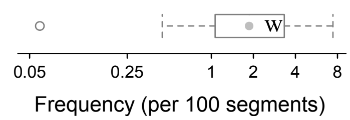


Figure 7.2: Frequency of /w/ relative to other English consonants, expressed as the number of instances per 100 segments. Rates are shown on the log scale, but tick mark labels give the absolute values of the rates. Data are pooled across 9 studies (see §3.3.3 for details).

7.2.3 *Similarity*

The perception of English /w/ by German learners has received some attention in the literature. Bohn & Best (2012) studied L1 German listeners' perceptual sensitivity to the English approximant contrasts /w/-/r/ and /w/-/v/ in onset position. A two-choice identification and an AXB discrimination task were used. A total of $n = 18$ students from the University of Kiel in northern Germany participated in the study ($M_{Age} = 21$). For the /w/-/r/ category boundary location and slope, the perception of German learners was equally categorical to that of native speakers. German listeners, however, gave statistically more *w*-responses near the category boundary. As a result of this bias and German listeners' lower sensitivity at the /w/-end of the continuum, English listeners' rate of correct discrimination was statistically higher (.75 vs. .68 for German subjects). In the AXB discrimination tests for the /w/-/j/ contrast, however, it was German listeners who showed a higher rate of correct discrimination (.90 vs. .75), even at the /w/-end of the continuum. The authors attribute this to German listeners' oversensitivity to lip rounding as a distinctive feature. L1 experience may thus result in a bias toward acoustic cues reflecting rounded-unrounded oppositions, with possible overshadowing effects of the acoustic correlates of [± round].

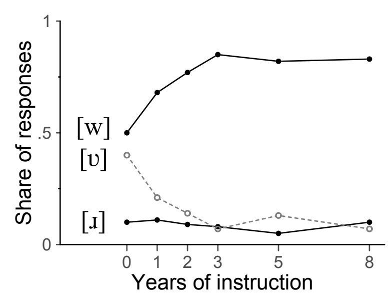


Figure 7.3: Perception accuracy of [w] against English [ɹ] and German [ʋ] in Austrian instructional-setting learners by years of learning, ranging from 0 (grade 3) to 8 (grade 11). The graph shows the share of [w]-tokens (mis-)identified as [w], [ɹ] and [ʋ]; adapted from Wieden & Nemser (1991: 159) with permission. In the original graph, it seems that the labels "correct /w/" and "correct /r/" have been confused; this, however, does not alter our conclusions. Note that the category [ʋ] is omitted in the original chart.

Wieden & Nemser (1991: 159) investigated cross-sectional patterns in the perception of English [w] by $n = 384$ Austrian learners.[2] An X-ABC task contrasted the nonce words [wiːft] vs. [wiːft], [ʋiːft], and [ɹiːft] spoken by a BrE native speaker. The perceptual matchings at different proficiency

[2] The study included learners at 6 proficiency levels, with $n = 64$ subjects in each group. Years of instruction was used as an indicator of developmental stage, which ranged from 0 years (grade 3; 8–9 years of age) to 8 years (grade 11; 16–17 years old).

levels are reproduced in Figure 7.3. Across all stages, [w] is most likely to be perceived correctly. At the initial stages, [w] is confused with [ʋ] at a rate of .40. Correct matching increased rapidly and plateaued at about .80 at later stages. Perceptual confusion with [ʋ] was observed to drop sharply and level out at .10. Confusion with [ɹ] remained constant at a rate of approximately .10.

Iverson et al. (2008) studied the perception of English /w/ and /v/ by $n = 22$ German learners ($M_{Age} = 23$, Range [19; 46]). In an identification task, subjects heard natural stimuli embedded in VCV sequences spoken by native speakers. While the median level of accuracy was .94, considerable variation was observed among learners, with some performing at chance level. The authors divided the informants into two groups based on their score in this task. The groups showed drastically different performance in subsequent perception experiments with synthesized stimuli. Tasks were designed to measure reliance on place and manner cues to distinguish English /w/ and /v/ and locate the listeners' best exemplar for each category. While the more proficient listeners showed perceptual matchings and best exemplar locations similar to native speakers, the low-proficiency group exhibited no systematic discrimination between /w/ and /v/ in terms of place and manner features.

In brief, L1 German learners show relatively high accuracy rates in the perception of English [w]. On average, the rate of target-like categorization is consistently higher than the misidentification rate across all developmental stages. When contrasted with the approximants /r j/, German learners show categorical discrimination. However, they appear to be less sensitive at the /w/-end of the respective continua if the contrast does not involve lip rounding as a distinctive feature. At the initial stages, perceptual difficulties have been documented for the contrast between English [w] and German [ʋ]. Nevertheless, sensitivity to this contrast appears to improve, on average, during the course of IL development. For the English [w]–[v] contrast, perceptual learning may yield cue reliance and weighting similar to that of native listeners. However, substantial differences may be observable among adult learners in terms of their ability to discriminate English [w] and [v].

7.3 L2 acquisition of English /w/: Theoretical predictions

This section considers /w/-acquisition by German learners from the theoretical viewpoints discussed in Chapter 2.

CAH As a familiar allophone of a new phoneme, the labio-velar glide /w/ is expected to cause no difficulties for German learners. While phonologically grounded L1 transfer is predicted to surface in substitution by [ʋ v], distributional reassignment of the German high back vowel /uː/ would allow positive transfer, resulting in a phonetically similar segment in speech production.

SLM Based on the results of perception studies reviewed above, we may assume that (i) [ʋ] is the perceptually closest L1 sound, and (ii) [ʋ] is perceived as moderately similar to [w]. This points to relatively accurate perception of English /w/ by the majority of German learners. A minor population, however, may experience perception-based difficulties in /w/-production. Accordingly, the SLM predicts target-like perception (and production) by the majority of German learners. Perceptual confusion is most likely to surface in [ʋ]-like sounds in speech production.

FCM The phonological features contrasting [w] from [uː ʋ v] are: consonantal (0.95) > anterior (0.50) > sonorant (0.27) > DORSAL (0.23). Table 7.2 shows the prominence-based ranking of substitutes. The FCM predicts [uː] as the perception-based substitute in speech production.

Consonant	Rank	[cons]	>	[ant]	>	[son]	>	[DOR]
[uː]	1	−		−		+		+
[ʋ]	2	+		+		+		−
[v]	3	+		+		−		−
[w]		+		−		−		+

Table 7.2: Substitutes for [w]: Predictions of the FCM.

SCH/MDH The evidence cited above suggests that [w] is an unmarked segment. The SCH and MDH therefore predict /w/ to be acquired with ease. Further, /w/ only occurs in onset position, which may be considered a further facilitating factor. Due to the relatively more marked status of clusters, singleton /w/ is expected to be acquired earlier.

NM/NDH Evidence from L1 acquisition is not suggestive of any natural processes that require suppression in the acquisition of [w]. Rather, the labio-velar glide appears to be the outcome of natural lenition processes such as liquid gliding and may thus be considered a naturally emerging structure. Accordingly, it should pose no difficulties in L2 acquisition. A mildly facilitating effect of input frequency may be assumed. This claim rests on the insight that /w/ is a moderately frequent structure.

FM The FM anticipates perceptual and sensorimotor learning to initiate a phase of variable production sensitive to focus on form, which is followed by a phase of overly faithful production. At the final stages, learner speech is predicted to reflect functional principles that operate at the post-lexical level. This includes style-dependent modification of articulatory effort, which is expected to surface in different degrees of lip rounding and/or velarization. Further, similar to native speech, phonetic context is expected to affect [w]-production in two ways: (i) voicing may vary depending on stress and the sonority of the preceding segment; and (ii) labialization may vary with the place of articulation of the following vowel (see §7.1). In addition, a slightly facilitative frequency effect is expected to surface in /w/-perception and production.

MSA With no known articulatory constraints on [w]-production, the MSA reduces to the SLM.

LTD The LTD posits [w]-accuracy to be sensitive to structural strength. At the level of the syllable, we expect the following tendencies: (i) delayed acquisition in unstressed syllables; and (ii) delayed acquisition in onset clusters relative to onset singletons. Strength marking is further assumed to be reflected in the type of variants observed. Weakened variants in *w*-marked positions may thus show lower degrees of voicing, velarization and/or labialization.

OPM In OPM terminology, /w/ is a normal structure. The model thus predicts initial L1 transfer of [ʋ v]. A moderate influence of universal constraints may in fact be hypothesized to promote acquisition of [w] since it is an unmarked segment. Given that target [w] may be motivated by both L2 and U, it is expected to emerge early in GLE.

The set of hypotheses about expected patterns in GLE is summarized in Table 7.3. Next, we take a look at previous empirical work on /w/-production by German learners of English.

Theory	Predicted patterns in L2 acquisition by German learners
CAH	Early acquisition Phonological substitute: [ʋ v] Phonetic substitute: [uː]
SLM	No difficulty for the majority of learners Substitution by [ʋ] in a minor population of learners
FCM	Perception-based substitute in speech production: [uː]
MDH/SCH	Early acquisition Acquisition order: singleton > cluster
NM/NDH	Mildly facilitating effect of frequency Early acquisition
FM	Mildly facilitating effect of frequency Development: deviant production > stylistic variability > overly faithful production > variability reflecting connected speech processes Late acquisition of context-dependent variation in labialization and voicing
LTD	*s*-marked positions: early acquisition; more canonical realizations of [+voiced, +velar, +labial] *w*-marked positions: delayed acquisition; less canonical realizations of [+voiced, +velar, +labial] Acquisition order: stressed > unstressed syllables; singleton > cluster
OPM	Early acquisition Substitutes at early stages: [ʋ v]

Table 7.3: The acquisition of /w/ by German learners: Summary of predictions based on theoretical work.

7.4 *English /w/ in German Learner English: Previous work*

Contrastive accounts of German and English phonology typically make some mention of /w/ as it is a new consonant for German learners. Thus, Mair (1995: 15) considers the labio-velar glide an easy sound and states that it is acquired early by German learners. Arnold & Hansen (1974: 77) and Kufner (1971: 45) consider German /v/-allophones the most likely substitutes for English /w/. Keutsch (1974: 67) notes that German learners tend to anticipate features of the following vowel in their productions of English /w/ and therefore may produce insufficient lip rounding before unrounded vowels and insufficient velarization before front and central vowels. Kufner (1971: 45), on the other hand, remarks that German learners experience particular difficulty in pronouncing /w/ intervocalically and after voiced consonants. In other words, /w/-accuracy is claimed to be sensitive to the sonority of the phonetic environment.

Empirical evidence on the acquisition of /w/ by German learners has been reported in a number of studies. Sieg (2004, quoted in Wode 2009) investigated /w/-production in L1 German primary school immersion-setting learners from grade 1 to 4. The majority of the subjects had attended a pre-school immersion program, with only few learners having had no prior knowledge of English. The correct production rate for /w/ was high across all age groups (.93/.95/.99/.99). The predominant substitute at all stages was [v], followed by [β]. No distinction appears to have been made between [v] and [ʋ].

A study by Langguth (2009) investigated /w/-production in three groups of German learners with different amounts of instruction (n = 7 each): 5th-graders (age: 10–11), 9th-graders (14–15) and 12th-graders (17–18). Production accuracy increased with age and was consistently high (.92/.96/.99), the most frequent substitutes being [v] and [β]. This study also did not distinguish between [v] and [ʋ].

Iverson et al. (2008) studied the production and perception of English /v/ and /w/ by n = 22 German learners. The subjects were between 19 and 46 years of age (*Mdn* = 23) and had started learning English between the age of 6 and 44 (*Mdn* = 11). Four trained phoneticians (English native speakers) rated the accentedness of both consonants produced in a single sentence (*The heavy wind swept away Valerie's velvet scarf*) on a scale from 1 to 7, the highest score indicating native-like production. The ratings showed that /w/ was produced in a near native-like fashion by the majority of learners (*Mdn* = 6).

Pascoe (1987)[3] studied n = 26 instructional-setting learners from southern Bavaria (age: 14–16, grades 8–10). An overall correct production rate of .54 for /w/ was noted, with a large gap between students with poor and good pronunciation (.19 vs. .93). The most frequently observed substitute was [ʋ], with a considerable amount of inter- and intra-speaker variation. These figures contrast with those reported by Sieg (2004, quoted in Wode 2009) and Langguth (2009), who observed correct production rates exceeding .90. This may in part be due to the L1 background of the researchers. While the learner productions in Sieg (2004) and Langguth (2009) were rated by native speakers of German, Pascoe's subjects were assessed by BrE speakers. It is likely that listeners with different L1 backgrounds may rate learner productions differently and empirical accounts in the literature may therefore not be directly comparable.

Further evidence on /w/-production by instructional-setting learners

[3] See §3.3.1 for details.

from southern Bavaria is reported by Kucharek (1988: 131, 144), who studied a total of n = 580 learners divided into 5 groups: students in grade 6 (n = 169), 7 (n = 138), 8 (n = 117), and 12 (n = 94), and teachers (n = 62). Production accuracy was assessed with 4 items (*when, watched, walked, we*), which were elicited with a reading passage. As shown in Figure 7.4, accuracy rates ranged between .70 and .80, with a difference of about .10 in absolute terms between 7th- and 12th-graders. Teachers showed a production accuracy of .98.

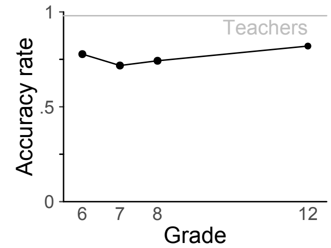


Figure 7.4: /w/-accuracy in the productions of German instructional-setting learners in grade 6, 7, 8, and 12 in comparison to teachers; data from Kucharek (1988).

Table 7.4 summarizes insights gained in empirical work on /w/-production by German learners. Reported accuracy rates range from .20 to 1, with a tendency toward high rates. Observed substitutes are [ʋ v β ɹ]. Only little is known about factors underlying variability in production accuracy.

Table 7.4: Empirical evidence from auditory studies on /w/ in German Learner English: reported accuracy rates.

Results		Notes	Study
Accuracy rate	.93–.99	Grades 1–4; age: 6–10 years	Sieg 2004
	.92–.99	Grades 5, 9, 12; age: 10–18 years	Langguth 2009
	.71–.82	Grades 6–12; age: 11–18 years	Kucharek 1988
	.19–.93	Grades 8–10; age 14–16 years	Pascoe 1987

7.5 *Aims of this study*

This investigation aims to make contributions to the study of /w/ in German Learner English by focusing on two aspects:

- cross-sectional patterns in the realization of /w/ and
- linguistic factors constraining the realization of /w/.

Cross-sectional patterns in the realization of /w/

Predictions derived from theoretical work unanimously consider /w/ an easy structure that is expected to be acquired early by the majority of German learners (CAH, SLM, MDH/SCH, NM/NDH, OPM). Predicted L1-derived substitutes are [v] and [ʋ] (CAH, OPM), with perceptual interference causing bias toward the latter (SLM). Learners may also transfer L1 [uː] to produce phonetically target-like variants (CAH; FCM). Quantitative evidence in the literature suggests a developmental increase in target-like production and this study aims to contribute to the empirical record.

Linguistic factors

Linguistic factors influencing /w/-production are an understudied area. Constraints put forward by theoretical accounts include onset complexity (SCH/MSH; LTD) and structural strength (LTD). This study will focus on phonetic context effects that have been reported to influence /w/-production in native and non-native speech:

- *Preceding segment*: Based on the literature, /w/-accuracy is hypothesized to be sensitive to characteristics of the preceding segment. In native speech, the degree of voicing in /w/ increases with the sonority of the preceding segment (see §7.1). For GLE, Kufner (1971: 45) claims that production accuracy is lower after more sonorous sounds. Specifically, difficulties in /w/-production are stated for intervocalic contexts and after voiced segments. This study aims to offer insights into /w/-variation as a function of the sonority of the preceding context.
- *Following vowel*: It was noted above that /w/-realization in native speech varies conditional on properties of the following vowel (see §7.1), with /w/-tokens preceding back rounded (vs. more front and more open) vowels exhibiting greater lip rounding. Keutsch (1974: 67) reports the same tendency for GLE, but offers no data to support this claim. Due to the status of /w/ as a double articulation, the following vowel may influence /w/-accuracy in GLE in two ways. The constriction location of the vowel may affect the degree of velarization, with more canonical instantiations of the feature [+velar] before close back vowels. The degree of lip rounding, on the other hand, may be sensitive to the degree of vowel labialization, with more rounded variants in the vicinity of rounded vowels. This study will explore possible carry-over effects of articulatory characteristics of the nucleus. Specifically, /w/-accuracy is expected to vary with (i) the constriction location and (ii) lip rounding of the following vowel.

7.6 Method and data

Informants, materials and procedures

The following analyses rely on four data sets, three of which are taken from previous work on GLE:[4]

- The *SpIL* corpus (Pascoe 1987) includes a total of n = 323 instances of /w/.
- The database provided by Wunder (2012) contains n = 14 labio-velar glides per subject, adding to n = 420 tokens for analysis.
- The third data set was provided by Rank (2016) and contains n = 46 /w/-items per speaker (word list: n = 16; reading passage: n = 30). This sums to n = 1,196 tokens in total.
- In the present study, each learner produced n = 19 /w/-tokens (n = 5 in the word list; n = 14 in the short conversations), amounting to a total of n = 1,152 tokens for analysis (n = 26 missing observations).

The recordings of Wunder (2012) and this study were analyzed auditorily by the present author. Learner productions were assigned to three categories: [w], [ʋ β], and [v]. No distinction was made between the bilabial approximant [β] and the labiodental approximant [ʋ] as it was often difficult to differentiate between them. The auditory classifications

[4] The data sets analyzed in the present chapter are openly available via the *Tromsø Repository of Language and Linguistics* (*TROLLing*) at https://doi.org/10.18710/F1A34O (Sönning et al. 2020b). See §3.3.1 for further information on the supplementary data sets.

	Pascoe 1987		Wunder 2012		Rank 2016		This study	
Realization	Pr	(*n*)	Pr	(*n*)	Pr	(*n*)	Pr	(*n*)
Observed variants								
[w]	.54	(174)	.71	(299)	.80	(957)	.78	(898)
[ʋ]	.46	(149)	.21	(90)	.08	(96)	.22	(254)
[v]	.00	(0)	.07	(31)	.12	(143)	.00	(0)
Categories used for analysis								
Target	.54	(174)	.71	(299)	.80	(957)	.78	(898)
Non-target	.46	(149)	.29	(121)	.20	(239)	.22	(254)

Table 7.5: Distribution of variants in the three data sets. "Pr" denotes the proportional share of a variant.

for Pascoe (1987) and Rank (2016) are the ones offered by these authors. Table 7.5 summarizes the distribution of variants in the data sets. For the ensuing analyses, [v] and [ʋ] were collapsed into a single category, resulting in a binary classification of learner productions into *target* and *non-target*. Accordingly, Table 7.5 distinguishes between *observed variants* and *categories used for analysis*. The aggregation of [v] and [ʋ] allowed for better comparability (i) between the four data sets considered and (ii) with earlier studies that appear to have made no distinction between these variants (Sieg 2004; Langguth 2009). Overall, target accuracy ranged between .50 and .80 in the four data sets considered.

The categories chosen for the preceding segment aimed to reflect sonority using six classes. First, three broad contexts were distinguished, depending on whether /w/ followed a pause, a consonant, or a vowel. Consonants were further subdivided into voiced and voiceless. Vowels were further broken down into spread [iː ɪ e], neutral [ɜː ə ɑː ʌ æ] and rounded [ɒ ɔː ʊ uː]. The following vowel, on the other hand, was assigned to one of 6 categories, which represent major regions of the vowel space: *i* [iː ɪ], *e* [e eɪ æ], *a* [a ʌ ɑː], *o* [ɒ ɔ ɔː], *u* [ʊ uː], and *central* [ɜː ə].

Table 7.6 shows the distribution of the predictor variables in the data sets. While the token frequency gives the total number of /w/-observations in each study, the type frequency reflects the range of different lexical items on which the total token count is based. In the balanced data sets by Wunder (2012), Rank (2016) and this study, is it roughly the number of tokens divided by the number of subjects (minus missing cases). The next paragraph provides some details about the statistical analysis and may be bypassed with no loss of continuity.

Statistical analysis

Since /w/-production was categorized as a binary outcome (target vs. non-target), all data sets were analyzed with multilevel logistic regression models. These models followed Barr et al.'s (2013) recommendation and included random intercepts on subject and word, by-subject random slopes for the linguistic factors (preceding segment, following vowel), and by-word random slopes for proficiency. In order to make the estimates for the linguistic predictors more comparable across the data sets, custom contrast

	Pascoe 1987		Wunder 2012		Rank 2016		This study	
Condition	Types	Tokens	Types	Tokens	Types	Tokens	Types	Tokens
Preceding segment								
Vowel								
Spread	11	68	3	90	6	182	3	186
Neutral	10	30	—	—	5	130	3	186
Round	7	14	4	120	1	26	2	124
Consonant								
Voiced	14	79	—	—	2	52	3	186
Voiceless	14	40	3	120	11	390	1	62
Pause	12	92	2	90	12	416	6	434
Following vowel								
a [a ʌ ɑː]	3	36	—	—	5	156	—	—
e [e eɪ æ]	9	55	2	90	9	338	4	372
i [iː ɪ]	7	129	4	150	7	364	3	248
o [ɒ ɔ ɔː]	6	89	4	120	3	156	5	558
u [ʊ uː]	1	4	—	—	2	104	—	—
central [ɜː ə]	5	10	1	60	3	78	—	—

Table 7.6: Distribution of types and tokens across the levels of the linguistic factors in the four data sets.

were used to represent these categorical predictors in the analysis (see Schad et al. 2020). The coding, which is listed in Appendix A.7, streamlines the intercept across the analyses to represent the average accuracy rate across shared levels of these predictors: For the following vowel, this is the weighted average *e*, *i*, and *o*; for the preceding segment, that of neutral vowels and voiced consonants. Since all conditions are then contrasted with this shared set of contexts, the coefficients can be compared somewhat more accurately. The structure of the regression models is detailed in Appendix A.7. Regularizing priors were attached to all predictors, and a weakly informative prior was specified for the intercept to add external knowledge about the expected accuracy rate at a proficiency level of 0. Figure 7.5 shows the correspondence between the indications in the literature and this distribution, which assigns 90% probability to the interval [.39, .99].

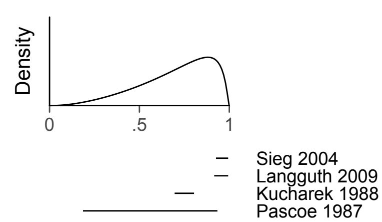


Figure 7.5: Prior for the intercept parameter in the model in comparison to the rates reported in the literature.

7.7 *Constraints on generality*

Before we come to the results, we should specify to which extent they allow us to generalize beyond the data at our disposal.

- *Language-external scope*: To the best of my knowledge, there are no dialectal features that would restrict the generalizability of the present findings to other populations of German instructional-setting learners.
- *Language-internal scope*: It seems that there are no limitations due to the choice of lexical items in the materials across the studies. However, the set of words is likely biased toward higher-frequency items.
- *Speaking style*: The findings based on the present study, Wunder (2012)

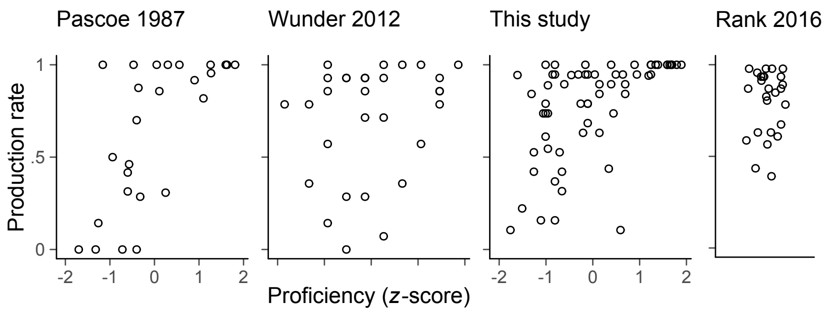


Figure 7.6: Accuracy rate for /w/ by proficiency level in the different data sets. The graphs show the proportion of target realizations.

and Rank (2016) extend to reading tasks only. The data from Pascoe (1987), on the other hand, represent more natural speech.

7.8 Results

Let us start with a visual examination of the data. Figure 7.6 shows, for the four data sets, how the accuracy rate for /w/ varies among speakers. Except for the speakers in Rank (2016), subjects are arranged by proficiency level. Each point shows the rate of target-like [w] for a particular learner. The dispersion of individuals in the four panels is very similar:

- At lower levels of pronunciation ability, there is considerable variation between learners in terms of /w/-accuracy. Target rates extend across the entire range and many individuals appear to alternate frequently between target and non-target renditions.
- Between-learner variation decreases at higher proficiency levels, as the share of [w] in German learners' productions approaches 1, on average.
- Categorically target-like renditions appear to be only observable among high-proficiency learners.

We sidestep technical details about the regression models, which we defer to Appendix A.7[5], and directly turn to a visual inspection of relevant quantities. Figure 7.7 gives a graphical summary of the results for /w/-accuracy by proficiency. The regression lines give a condensed summary of the descriptive patterns above. Error bands indicate 50% and 90% uncertainty intervals. Estimates of the [w]-rate at representative levels of pronunciation proficiency are given in Table 7.7. Figure 7.7 offers the following insights:

[5] Appendix A.7 reports convergence diagnostics for each analysis and the posterior distribution of the model parameters.

- While the three data sets yield similar estimates for advanced learners, there is notable divergence at the lower end of the proficiency scale.
- Pascoe's (1987) data suggest an increase from about .05 to almost 1 across proficiency levels.
- The systematic variability in the productions of Wunder's (2012) subjects is much smaller, extending from about .70 to .90.
- The learners in the present study show an increase from about .40 to .95, on average.

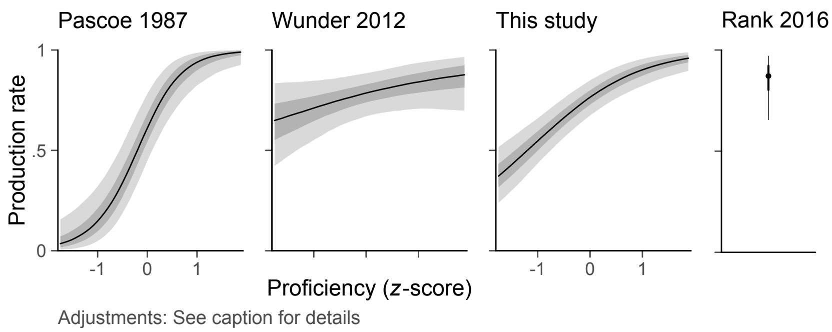


Figure 7.7: Results for /w/: Estimated correct production rate by proficiency level. Error bands mark the 50% and 90% posterior intervals. The predictor *preceding segment* is held at the simple (i.e. unweighted) average of neutral vowels and voiced consonants; for *following vowel*, values are set to the simple average over the categories *e*, *i*, and *o*.

	Pascoe 1987			Wunder 2012			This study			Rank 2016		
Proficiency	*Mdn*	.05	.95	*Mdn*	.05	.95	*Mdn*	.05	.95	*Mdn*	.05	.95
Beginner	.05	.01	.19	.67	.46	.84	.42	.29	.56	—	—	—
Intermediate	.63	.46	.79	.79	.67	.89	.77	.68	.85	—	—	—
Advanced	.98	.90	1.00	.86	.70	.95	.94	.87	.98	.87	.66	.97

Table 7.7: Accuracy rate estimates for /w/ by study and proficiency level: Posterior medians and 5% and 95% quantiles, which mark the limits of 90% uncertainty intervals.

Findings for the effect of the following vowel are summarized in Figure 7.8. In the left panel, the vowel classes are ordered along the horizontal axis, from rounded to neutral to spread. This arrangement aligns with the hypothesized association between the degree of velarization in /w/ and the place of constriction of the following vowel – that is, vowels are ordered according to their approximate distance from cardinal [u]. The vertical axis in Figure 7.8 signals variation in target accuracy: Points in the upper part of the display indicate higher accuracy rates. Letters are used to denote the source of an estimate (e.g. ”P” for Pascoe 1987). What we are looking for in this graph is whether there are vowel classes that are consistently, or at least on average (across studies), associated with higher (or lower) accuracy rates. The estimates are on the logit scale. For instance, the category *a* [a ʌ ɑː] appears to boost the accuracy rate by +2 logits, compared to the reference condition, which is the average across the categories *e*, *i*, and *o*. The correspondence between differences in logits and the more familiar differences in proportions depends on where along the scale from 0 to 1 comparisons are made. The right-hand panel establishes a link between the scales. To this end, the horizontal axis shows anchor values, that is, potential accuracy rates for the reference set (i.e. the average across *e*, *i*, and *o*). Thus, for lower proficiency learners (with a reference accuracy rate of about .50 to .60), an increment of +2 on the logit scale corresponds to a difference of about +.35 on the proportion scale. For intermediate learners (reference rate: .70), the [w]-share climbs by about +.25, and for advanced learners (reference rate: .90), the ceiling effect shrinks the increment to roughly +.05. Thus, the magnitude of a facilitative or impeding ”effect” of certain vowel categories on the proportion scale depends on the reference rate. Figure 7.8 hints at the following patterns:

- There is no discernible trend when vowel groups are arranged by their

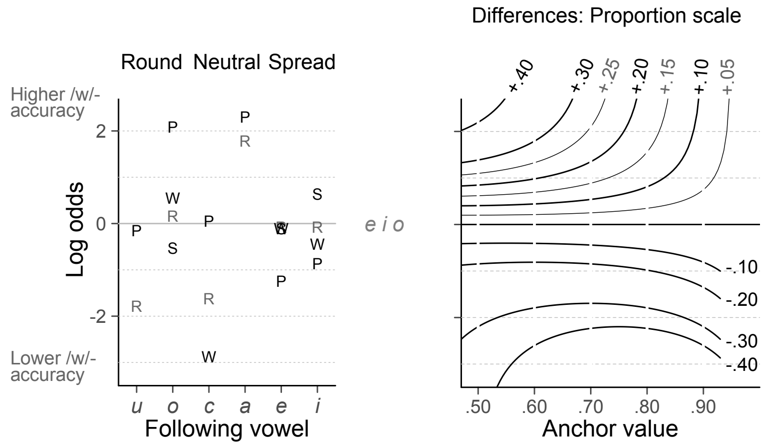


Figure 7.8: Left panel: Variation in /w/-accuracy before different groups of vowels, which are shown on the horizontal scale. The plotted symbols denote, on the logit scale, the deviation of a vowel group compared to the average accuracy rate before [*e i o*]. Letters indicate the data set from which the estimate was computed ([P]ascoe 1987, [W]under 2012, [R]ank 2016, this [S]tudy). Higher values denote higher /w/-accuracy. Right panel: Translation of the logit differences to differences on the proportion scale for different anchor values, i.e. average accuracy rates before [*e i o*].

distance from cardinal [u]. That is, there is no consistent cline across vowel categories from left to right.

- As for vowel labialization, there appears to be no systematic link with accuracy rates: Round (left), neutral (center), and spread vowels (right) appear to show similar rates, on average.
- High back and central vowels (*u, c*) seem to be associated with lower levels of /w/-accuracy.
- On balance, mid and open back, and open central vowels (*o, a*) appear to favor correct production.
- The indications for mid and high front vowels (*e, i*) are weak and inconsistent.

Figure 7.9 is constructed in the same way and shows the patterns observed for the preceding segment. The contexts are ordered by sonority, in decreasing order from left to right. The following points are noteworthy:

- Higher accuracy rates are found after a pause and neutral vowels. Voiced consonants appear to be linked with lower accuracy rates.
- Sonority: Apart from a tendency toward higher accuracy after voiceless vs. voiced consonants, there is no across-the-board trend for sonority since vowels do not differ systematically from consonants.

To allow for comparison of the variation by preceding segment and following vowel, Figures 7.8 and 7.9 were drawn with the same vertical axis limits. Given the classification scheme applied in the present study, it appears that variation as a function of the following vowel is greater in magnitude, indicating /w/-accuracy to be more sensitive to the following nucleus than to characteristics of the preceding context.

7.9 *Summary and discussion*

This chapter was concerned with German learners' acquisition of the English labio-velar glide /w/. A contrastive analysis revealed the absence of this consonant from German, whose high back vowel /uː/ is the phonetically closest sound. Phonologically, the closest L1 sound is /v/, which is

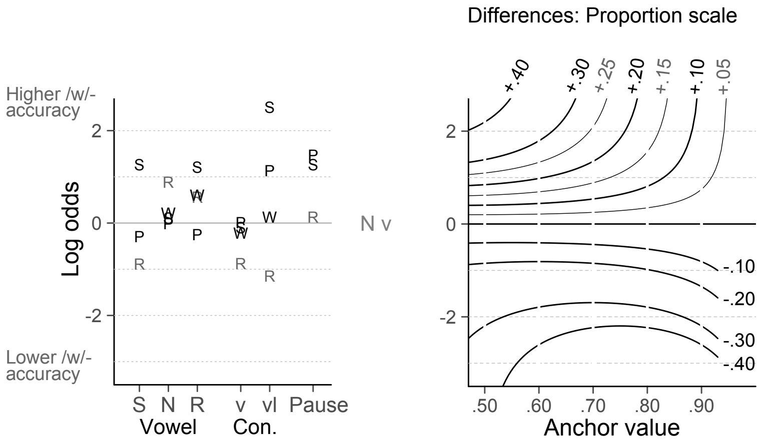


Figure 7.9: Left panel: Variation in /w/-accuracy after different segment groups, arranged on the horizontal scale by sonority. The plotted symbols denote, on the logit scale, the deviation of a segment group, or sonority level, compared to the average accuracy rate before neutral vowels and voiced consonants. Letters indicate the data set from which the estimate was computed ([P]ascoe 1987, [W]under 2012, [R]ank 2016, this [S]tudy). Higher values denote higher /w/-accuracy. Right panel: Translation of the logit differences to differences on the proportion scale for different anchor values, i.e. average accuracy rates before neutral vowels and voiced consonants.

typically realized as a labiodental approximant [ʋ]. From the viewpoint of linguistic constraints, English [w] was classed as an unmarked structure, based on insights gained from typology and L1 acquisition. Perceptually, English [w] appears to pose no difficulties for most German listeners. Consequently, a theory-guided assessment of /w/ unanimously predicted relative ease of acquisition for this structure. Empirical evidence is broadly consistent with this claim, with most studies reporting accuracy rates between .70 and 1. Linguistic factors underlying /w/-accuracy in GLE have so far received no attention in the literature.

The objective of this study was to provide further quantitative insights into the rate of target-like production across different levels of L2 pronunciation ability by means of a parallel analysis of four data sets. Guided by anecdotal evidence in the literature, an exploratory analysis of phonetic context effects was conducted, in expectation of lower accuracy rates after sonorous segments. /w/-accuracy was also hypothesized to be sensitive to lip rounding in the preceding vowel, with more target-like production after rounded vowels. As for the following vowel, it was assumed that shared features trigger higher accuracy rates. Specifically, /w/ was expected to be produced more accurately before rounded (vs. neutral and spread) vowels and high back (vs. more front and more open) vowels.

Findings for variation in target accuracy across proficiency levels were coherent with those in the literature but revealed variable accuracy rate estimates. The largest discrepancies were observed between lower proficiency learners, whose estimated correct production rate ranged from .05 to .70. For intermediate-stage learners, these figures varied between .60 and .90 and for advanced learners the three data sets consistently suggest accuracy rates of .90 or higher.

Two factors may help explain the variation in accuracy rate that was observed between the lower proficiency learners in the three studies. First, the type of speech elicited by Pascoe (1987) can be considered the most natural. A less monitored speaking style may account for the low accuracy rates observed among the lower proficiency learners in the *SpIL* corpus. This explanation sits well with the OPM, which posits an increase in L1-derived [v ʋ] and a decrease in target [w] in less monitored speech. The fact that variation is primarily observed in less proficient speakers is unsurprising from the viewpoint of the FM, which predicts stylistic variability at early stages of /w/-acquisition. A further indication of the relevance of stylistic factors is the consistency of figures derived from more monitored speech (i.e. Wunder 2012, Rank 2016, and the current study) with those reported in earlier work, which also (predominantly) resorted to reading tasks for data elicitation (Kucharek 1988; Langguth 2009). A second source of variation may be the L1 of the rater. Thus, the study by Pascoe (1987) was the only one in which a native speaker categorized the variants produced by learners. It is possible that as a result of L1 experience, raters with different native language backgrounds arrive at different categorizations of productions that vary along a continuum. The category boundary between [w] and [ʋ] may assume slightly different locations in the perception grammars of L1 English and L1 German listeners. In particular, while English listeners have established a robust category for [w], it is a new sound for German listeners. Even with considerable phonetic training, English listeners may apply different criteria for an assignment of learner productions to the /w/-category.

With all data sets revealing considerable variation up to intermediate levels of L2 pronunciation ability, the strong prediction about general ease of acquisition is not supported by the present investigation. Rather, the patterning of subjects is consistent with earlier empirical work on the perception of [w] by German listeners, which pointed to deficient perceptual parsing by a minor population of German learners (Wieden & Nemser 1991; Iverson et al. 2008).

The association between the observed variation in target accuracy and characteristics of the following vowel did not lend support to observations by Keutsch (1974: 67), who noted that high and mid front unrounded vowels (*e*, *i*) disfavor target-like /w/-production. The findings for these vowel types were mixed, and the predicted pattern was only borne out by Pascoe's (1987) data (cf. Figure 7.8). For back vowels, the analyses yielded unexpected results. Thus, there was no boost in accuracy before *u*-vowels, despite the shared features [+round] and [+back]. Back rounded *o*-vowels, on the other hand, are associated with more target-like production. As far as these results go, it seems that the association between /w/-accuracy and the following vowel cannot be pinned down to tongue configuration and lip rounding. Rather, this association appears to be mostly idiosyncratic with certain vowels favoring /w/-accuracy. The picture is far from clear-cut, however, and more work is needed to elucidate the link between /w/-accuracy and the nucleus of the syllable.

Results for the preceding phonetic context showed consistently higher accuracy rates after a pause. Little systematic variation was observed for sonority and characteristics of the vowel, and lip rounding and place of articulation of the following vowel did not show the expected association with production accuracy. These findings thus only partially support Kufner's (1971: 45) observations on the relationship between /w/-accuracy and sonority. With target-like production being consistently higher after a pause, it appears that utterance-initial position is a preferred environment for [w] in GLE. A consistent effect of the sonority of the preceding segment, on the other hand, could not be detected, as there was no discernible cline between voiceless and voiced consonants or between consonants and vowels. As for place features of the preceding vowel, lip rounding appears to play a negligible role.

From the viewpoint of L2 phonological theory, knowledge about phonetic context effects of the kind investigated in the present study remains scarce. It seems that the immediate phonetic environment has so far been neglected in theoretical work. Empirical research on within-structure variation that reflects adjacency effects therefore lacks principled guidance on expected patterns of variation. While hypotheses may be scaffolded by recourse to patterns of variability observed in the TL and other populations of (non-)native speakers, this approach lacks formal grounding. From an empirical perspective, suggestions for future efforts toward an understanding of the mechanisms underlying this form of within-structure variation can be made. Thus, to provide a sound basis for quantitative work, it appears that three aspects of phonetic context effects should be taken into consideration:

- First, whether phonetic features in [w] interact with the immediate environment may depend on the prosodic proximity of focal segments, i.e. the level of prosodic representation at which they are conjoined. Thus, adjacency effects may be more likely between elements in the same syllable (e.g. in consonant clusters or between onset and nucleus or nucleus and coda) and decrease in likelihood at higher-level prosodic boundaries. This could explain why /w/-accuracy was more strongly associated with the following nucleus than with the preceding context, as the left-ward interaction crossed a syllable boundary.
- Second, phonetic context effects can operate from two directions, i.e. forwards or backwards. Little is known about the likelihood of directionality in learner speech. Other things being equal, focal segments may be more likely to preserve or anticipate features, reflecting preferential sensitivity to the preceding or following context.
- Third, adjacency effects may lead to feature assimilation and/or dissimilation. Thus, identical features in neighboring segments may occur as a result of coarticualtion, or they may be actively disfavored, leading to dissimilation.

It appears that efforts toward filling this niche in L2 phonological theorizing may profit from close links between empirical and theoretical work. While formalized accounts may be guided by quantitative evidence, empirical work on within-structure variation would benefit from theoretical advances in this domain, which may offer deeper explanatory accounts of observable patterns.

A limitation of the present study are the narrow conclusions that are warranted on the basis of a binary distinction between *target* and *non-target* variants in learner speech. This is especially true in the case of /w/. Due to its status as a double articulation, two strictures contribute to production accuracy. Binary approaches fail to provide insight into the acquisition of the labialization and velarization of target /w/. While it is desirable for future studies on GLE to contrast [v] and [ʋ], it must be conceded that a distinction between these variants can in many instances not be reliably made. Complementary instrumental techniques seem to offer promising opportunities for further research. In particular, acoustic measurements may provide a more nuanced reflection of variation along the [w]–[ʋ]–[v] continuum.[6] While the formant structure of these sounds provides acoustic cues corresponding to the degree of labialization and velarization, measures such as the harmonicity median and the center of gravity may serve to distinguish between [v] and [ʋ w] (cf. Hamann & Sennema 2005a; see §8.1). Acoustic information may thus differentiate between the three major /w/-variants observed in GLE and sidestep valid concerns about the possibility of L1 bias in the auditory categorization of variants.

[6] In a sociolinguistic study on allophonic approximant variation in Mandarin, for instance, Wiener & Shih (2013) successfully applied F_2 as a measure for distinguishing between labio-velar [w] and labiodental [ʋ].

Several linguistic factors suggested in the theoretical literature are also worthy of attention in future work. These are structural strength, onset complexity, and speaking style. While the latter was invoked to offer an explanation for the divergence between Pascoe's (1987) data and the other data sets (as well as the literature), this explanation remains speculative due to lack of direct comparability. More work is needed to provide insights into stylistic variation in learner productions.

In sum, the following suggestions can be made for future studies on /w/-production by German learners:

- If the focus of a study is on linguistic factors underlying variation in /w/-production, subjects should be sampled from lower to intermediate proficiency levels, as this subpopulation of German learners has been consistently observed to exhibit the greatest variation.
- The few indications offered by the present analysis suggest phonetic context effects to operate more strongly in natural speech, and to exhibit the closest correspondence to predicted patterns in this speaking style. This may serve as an impetus for future work to turn attention to less monitored learner productions.
- Care should be taken when comparing word list data with connected speech, as initial position has been documented to consistently favor target-like production. Better comparability may then be achieved by controlling for preceding phonetic context.
- The description of variants in learner productions should strive to distinguish between [v] and [ʋ]. Given the status of /w/ as a double articulation, a comprehensive analysis may venture to disentangle labial and velar components of target accuracy.
- As mentioned above, the application of acoustic methods bears potential for the study of /w/-production by German learners. This unchartered territory remains to be explored in future work.

8 The labiodental fricative /v/

Following our treatment of English vowels and sonorants, the next three chapters turn to the production of obstruents by German learners. In contrast to sonorants, obstruents differ categorically from vowels both phonetically and phonologically and can be considered prototypical consonants. The two major classes of obstruents are fricatives and stops. While fricatives are the focus of this and the next chapter, Chapter 10 deals with obstruents in coda position.

The noise component characteristic of fricatives is produced by bringing two articulators sufficiently close together to cause turbulence of the escaping air. This turbulence may be accompanied by vocal fold vibration. From a typological perspective, English and German have complex sets of fricatives. With only .06 of the $n = 314$ languages in the UPSID database (Maddieson 1984: 47) having more than 8 fricatives (see Figure 8.1), the inventories of English ($n = 9$) and German ($n = 10$) are relatively large.

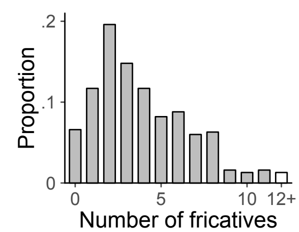


Figure 8.1: Distribution of the number of fricatives in the world's languages; data from Maddieson (1984).

Based on articulatory and acoustic characteristics, fricatives fall into two classes. Sibilants (e.g. /s z ʃ ʒ/) are produced with the airstream hitting an obstacle, giving rise to a high-frequency noise component. Non-sibilants such as /f v θ ð/, on the other hand, are characterized by airflow parallel to the obstacle and lower-frequency noise. Phonological descriptions capture this difference with the feature [±strident].

This chapter deals with the acquisition of the English labiodental fricative /v/ in onset position. §8.1 contrasts English and German in terms of their realization of this shared phoneme. After a consideration of linguistic constraints involved in the acquisition of /v/ (§8.2), §8.3 consults theories of L2 phonology to formulate expected patterns in GLE. §8.4 offers a review of previous empirical work on /v/-acquisition by German learners. This is followed by an outline of the aims (§8.5) and methods (§8.6) of this study. §8.7 delimits the generality of our results, which are presented in §8.8. §8.9 concludes with a summary and discussion.

8.1 Contrastive analysis

English and German share a pair of labiodental fricatives /f v/. These are articulated with contact between the inner surface of the lower lips and the edge of the upper teeth, causing turbulence in the escaping air stream. While this friction is voiceless for /f/, /v/ may involve vocal fold vibration.

English /v/ occurs in initial (*very, visit*), medial (*ever, oven*) and final position (*leave, cave*). In the onset of a syllable, it may combine with /j/ (*view, Vienna*). In coda position, more cluster types exist – however, most of these include a morpheme or syllable boundary, as /v/ may combine with inflectional allomorphs (/vz/ *leaves, caves*; /vd/ *loved, saved*) or a syllabic consonant as the second cluster constituent (/vn/ *even, oven*; /vl/ *evil, grovel*). The sequence /lv/ (*solve, twelve*) is an exception as it occurs in monomorphemic and monosyllabic words (Cruttenden 2014: 196–198).

In connected speech, word-final /v/ is prone to coarticulatory influence and may be realized as /f/ if the following word begins with a voiceless consonant. This is usually observed in common sequences such as *have to* [ˈhæftə] and *of course* [əfˈkɔːs], but also surfaces in other environments. /v/ may be elided in casual speech, especially before consonants (Roach 2009: 114). In particular, the high-frequency function words *of* and *have* may show elision of /v/. In *of*, it is especially likely to be dropped when followed by /ð/ (Collins & Mees 2013: 126).

While a phonemic comparison of English and German reveals no contrasts, divergence occurs at the allophonic level. Due to the devoicing rule of obstruents in coda position (see Chapter 10), German /v/ is only licensed in syllable onsets. It usually occurs in initial position (*Wasser*, /ˈvasɐ/; *weil* /vaɪl/) and less commonly in medial position (*Möwe* /ˈmøːvə/) (Kohler 1995: 154). In syllable-final position it is devoiced both as a singleton (*aktiv* /akˈtiːf/) and in consonant clusters (*aktivster* /akˈtiːfstɐ/). Moreover, syllable-initial /v/ is realized differently in the two languages. English /v/ is generally described as being produced with more energy and stronger contact between the articulators (Koziol 1959; Arnold & Hansen 1974: 84; Scherer & Wollmann 1986; Eckert & Barry 2005). Specifically, the lower lips are pressed harder against the upper teeth and its duration is longer (Koziol 1959: 80). Pascoe (1987) describes the typical German realization of /v/ as a weak labiodental approximant [ʋ], which may involve contact between the teeth and the back of the lower lip, as opposed to the teeth sitting on top of the lower lips. Phonetically, German /v/ thus shares features with English [v] and [w]. Most Germans produce i a weak labiodental approximant [ʋ], a tendency that is most pronounced in initial position. While [ʋ] and [v] involve the same articulators, [ʋ] and [w] are equivalent in manner of articulation. Mangold (2015: 71) notes that weak realizations are particularly characteristic of southern German dialects, where /v/ may also be realized as a voiced bilabial fricative [β] or approximant [β̞].

Acoustic evidence reported by Hamann & Sennema (2005a) is supportive of the characterization of German initial /v/ as an approximant-like sound. Productions of adult female speakers of Dutch and German (n = 5 each) were compared. In contrast to German, Dutch has the three-way labiodental phonemic contrast /f v ʋ/. Three acoustic parameters were used to compare the realization of German /f v/ and Dutch /f v ʋ/ in onset position (monosyllabic context: [_a]): consonant duration, the harmonicity median[1], and the center of gravity.[2] The authors observed that the realization of German /v/ was acoustically close to Dutch /ʋ/ across all three measures. The results, which are reproduced in Figure 8.2, show that, from

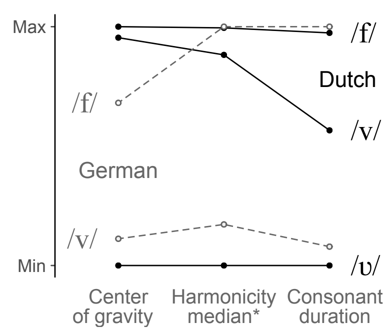


Figure 8.2: Acoustic properties of the realization of German /f v/ and Dutch /f v ʋ/ in initial position; data from Hamann & Sennema (2005a). *Note: To yield a regular visual arrangement, the values for the harmonicity median are shown in reversed order (i.e. the maximum at the bottom).

[1] The harmonicity median expresses the amount of friction relative to the periodicity in the speech sound and is thus highest for vowels and sonorants and lowest for voiceless fricatives.

[2] This measure gives a general summary of the spectral characteristics of a speech sound.

a phonetic point of view, German /v/ is closer to an approximant than to a fricative in initial position.

To summarize, the phoneme /v/ is realized differently in English and German. English has a voiced labiodental fricative [v] in all positions; coda /v/ may in certain contexts be realized as voiceless [f]. In German, initial /v/ is usually lenited to [ʋ], an approximant-like sound that involves no friction. In coda position, /v/ is regularly devoiced. The acquisition of English /v/ thus requires German learners to produce the new allophone [v] both in onset and coda position. In syllable onsets, this involves fortition of [ʋ] to [v]; in coda position, [f] needs to be lenited to [v].

8.2 *L2 acquisition of /v/: Linguistic factors*

8.2.1 *Markedness*

A comparison of obstruent inventories in languages around the world shows that the distribution of fricatives is more restricted compared to stops. While stops appear to be an absolute universal – that is, they are present in all natural languages – some consonant inventories lack fricatives (Maddieson 1984). As for the types of fricatives attested cross-linguistically, there are discernible patterns. In general, we find a preference for voiceless ones and fricatives are more likely to be contrasted in place of articulation than voicing. The most common type are voiceless *s*-sounds (dental and alveolar sibilants, .84 of the languages in the UPSID database), followed by /ʃ/ (.46) and /f/ (.43). The respective voiced counterparts occur at a lower rate, with .31 for *z*-sounds, .21 for /v/ and .16 for /ʒ/ (Maddieson 1984: 45). Statistical trends therefore suggest a frequency bias toward voiceless (vs. voiced) and alveolar/dental (vs. labiodental/postalveolar) fricatives. From a typological perspective, /v/ thus appears to be disfavored due to the following characteristics: It is (i) a fricative (rather than a stop), (ii) voiced (rather than voiceless), and (iii) labiodental (rather than alveolar/dental).

Cross-linguistic patterns are consistent with evidence from English L1 acquisition. English-learning children acquire stops before fricatives and they master voicing contrasts in stops earlier than in fricatives. Figure 8.3 summarizes the age of acquisition for English stops and fricatives averaged over 4 studies (see §3.3.2). Two consistent patterns emerge (i) stops are acquired before fricatives and (ii) voiceless obstruents are acquired before their voiced counterparts.[3] On average, the lag between voiceless and

[3] This pattern is also observed in other L1s (cf. Locke 1983: 72–80). Furthermore, stop-like production of fricatives is a regular error pattern in L1 acquisition (Ingram 1989: 371).

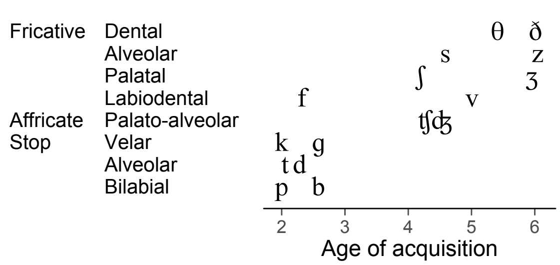


Figure 8.3: Age of acquisition of English obstruents: Comparison of voiced and voiceless stops and fricatives. Estimates are averages across initial and final position and based on a survey of 4 studies (see §3.3.2).

voiced pairs is larger for fricatives than for stops. This is coherent with the typological rareness of fricative voicing contrasts.[4] Note that the gap is particularly pronounced for /f v/.

[4] Evidence from aphasia is also supportive of this asymmetry, as voicing oppositions are less likely to be lost in stops than in fricatives (see Jessen 1998: 181f.).

Figure 8.4b shows the accuracy rates reported in the literature for different age groups. Word-initial and -final contexts are distinguished and the curves give a summary of the trends. Figure 8.4a reveals that /v/ is among the sounds that are acquired late by English-learning children. The age of acquisition in both initial and final contexts is 5;0. Voiced fricatives thus appear to constitute an area of difficulty in L1 acquisition. Counter to the delay that is commonly observed in coda obstruents, initial and final /v/ show very similar developmental trajectories. This underscores the status of onset /v/ as a difficult sound in L1 acquisition. The most typical substitute reported for onset /v/ is [b], followed by [f] (see Olmsted 1971; Ingram et al. 1980 and references cited therein).

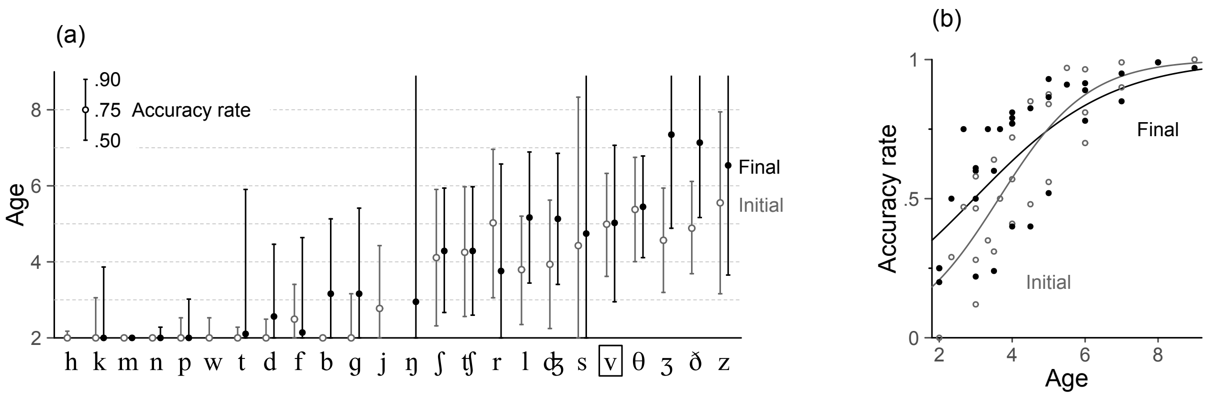


Figure 8.4: /v/ in English L1 acquisition in comparison to other consonants. Panel (a): Points mark the age of acquisition (.75), bars indicate the age of customary production (.50 accuracy, lower end), and age of mastery (.90, upper end). Estimates are based on a joint analysis of 4 studies. Panel (b): Accuracy rate in initial (grey) and final position (black) by age. The points denote estimates reported in the individual studies, the trend lines are averages based on the joint analysis.

Evidence for the susceptibility of /v/ to sound change may be found in Creole English, where it appears to be receptive to substitutions and substrate transfer. Thus, /v/-stopping – that is, production as /b/ ('V-B Confusion', Wells 1982: 568) – is observable across varieties such as West Indian and Caribbean creole and resembles error patterns in L1 acquisition. /v/-lenition toward an approximant articulation ('V-W Confusion', ibid.: 568) is also attested in varieties such as Gullah, Bahamian, Bermudan, and Vincentian creole (ibid.: 568; Holm 2000: 163).

The patterns observed in typology and L1 acquisition may be partly rooted in the physiology of articulation. In general, the production of fricatives involves greater precision than that of stops (Ladefoged & Maddieson 1996: 137). Friction only arises with the right amount of narrowing, which requires fine motor control over the active articulators. Articulatory constraints may also account for the rarity of fricative voicing. As Ohala (1993: 201–202) suggests, conflicting aerodynamic requirements may disfavor voiced fricatives. For the vocal folds to vibrate, oral pressure has to be sufficiently low; at the same time, it has to be high enough to effect turbulence in the escaping air stream. It is argued that meeting both requirements simultaneously may be difficult.

To summarize, evidence from typology and L1 acquisition identifies English /v/ as a marked speech sound. It is cross-linguistically rare relative to other obstruents and acquired late by English-learning children. Physiological factors may at least partly explain these patterns, with articulatory constraints disfavoring the production of voiced fricatives.

Criterion	[v]
Implicational relationship	○
Cross-linguistic frequency	+
L1 acquisition	++
Synchronic/diachronic instability	(+)
Articulatory complexity	+

Table 8.1: Summary of markedness criteria for [v]. Open circles ○ denote no (or inconclusive) evidence; (+)+ refers to (strong) evidence for marked status; (–)– refers to (strong) evidence for unmarked status.

8.2.2 *Frequency and orthography*

Studies investigating the frequency of English phonemes (see §3.3.3) have shown that /v/ occurs at a moderate rate compared to other consonants (see Figure 8.5). Its median frequency of 2 per 100 segments, however, is strongly influenced by the high-frequency function words *of* and *have*, where /v/ occurs in coda position. While the studies in our survey do not break down frequency estimates by prosodic position, it may be assumed that the rate of onset /v/ is certainly much lower.

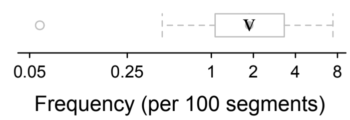


Figure 8.5: Frequency of /v/ relative to other English consonants, expressed as the number of instances per 100 segments. Rates are shown on the log scale, but tick mark labels give the original values of the rates. Data are pooled across 9 studies (see §3.3.3 for details).

English /v/ is denoted by orthography <v> in the great majority of instances (Cruttenden 2014: 196), an important exception being the function word *of*. As shown in Figure 8.6, German /v/ is represented by the graphemes <v w>; <v> in turn may denote /f/ or /v/. Thus, while English spelling (apart from *of*) provides an unambiguous cue to the /f v w/-distinction, German orthography is ambiguous, with /f/ represented by <f v> and /v/ by <v w>.

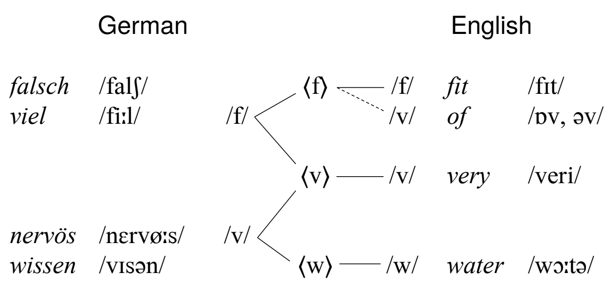


Figure 8.6: Orthographical representation of English and German /v/. While English spelling provides a straightforward symbol-to-sound correspondence (except for *of*), German spelling is ambiguous.

8.2.3 *Similarity*

A number of studies have shed light on German learners' perception of English /v/. Iverson et al. (2008) investigated the perception of English /v/ and /w/ by n = 22 German learners. Subjects were 19 to 46 years old (*Mdn* = 23) and had started learning English between 6 and 44 years of age (*Mdn* = 11). Several tests were administered to assess perceptual separation between the two sounds and the relative weight of place and manner cues to perceptually distinguish between /v/ and /w/. In a natural stimulus identification task, subjects heard [v] and [w] embedded in V-C-V sequences (spoken by native speakers) and were asked to indicate which consonant they heard. While the overall level of accuracy was high (*Mdn* = .94), there was a substantial amount of variation among German subjects. Thus, two proficiency groups were formed based on the identification accuracy obtained in this task (cut-off: .75; a quarter of the subjects scored lower). The authors provide a detailed analysis of the subjects' perceptual sensitivity to the English /v/-/w/ contrast. To this end, a set of synthetic stimuli (medial context: [ɑː_ɑː]) was designed to cover the perceptual space between the categories. Tokens were varied along two dimensions reflecting manner of articulation (4 steps from approximant to voiced fricative) and place of articulation (4 steps from labiovelar to labiodental). Upon hearing these stimuli, subjects had to indicate whether they heard /v/ or /w/ and whether the stimulus was a good exemplar of the respective category.

The patterns observed in the two proficiency groups differed. Learners who performed well on the natural stimulus identification task also showed good category separation in the two-dimensional space. Similarly to native speakers, both place and manner cues were used. However, differences were noted in the goodness ratings for English /v/ on the place dimension. While English subjects judged step 4, the most labiodental stimulus, as the best exemplar, German subjects' best exemplar was located at step 3, representing a less purely labiodental articulation. The results for the less proficient learners showed poor category identification, which points to perceptual distortions in this population of listeners.

Further indications on German learners' perception of the /v/-/w/ contrast were reported by Ankerstein & Morschett (2013), who used a forced-choice discrimination task to study a similar population of *n* = 21 learners (age: 18–28). The /v/-/w/ distinction was tested in the [_ɑː]) context. The authors observed a mean error rate of .41 and considerable variation among subjects (*SD* = .21). This corroborates findings by Iverson et al. (2008), as sensitivity to this contrast appears to be variable in adult German listeners.

Hamann & Sennema (2005b) provide insights into German learners' perception of labiodental speech sounds. An identification task was used to test the perception of the Dutch three-way contrast /f v ʋ/ by *n* = 21 German learners of Dutch and a control group of *n* = 6 German listeners with no prior knowledge of Dutch (age: 18–40). The authors noted near-perfect identification rates for Dutch /ʋ/ in both groups. /v/, on the other hand, was more likely to be misheard by German learners as /ʋ/ than vice versa. This indicates a robust perceptual representation for [ʋ] and perceptual bias toward [ʋ] rather than [v]. These results suggest that German /v/ is perceptually more similar to Dutch /ʋ/ than it is to Dutch /v/, corroborating acoustic evidence on the approximant-like status of German /v/ (cf. Figure 8.2 above).

Hamann et al. (2010) found differences in the perceptual cues used by German and English listeners to distinguish voiced and voiceless labiodental fricatives. Reliance on two signals was investigated: the harmonics-to-noise ratio and the duration of the fricative. The harmonics-to-noise ratio quantifies the balance of periodicity and friction in the speech signal (Yumoto et al. 1982; Boersma 1993). Segments that are produced without friction (i.e. vowels and sonorants) have a high ratio (or, vice versa, a low noise-to-harmonics fraction); voiced fricatives are intermediate and voiceless fricatives are characterized by a low ratio. The harmonics-to-noise ratio and fricative duration have been identified as reliable acoustic cues for the distinction between labiodental speech sounds (Hamann & Sennema 2005a,b). Hamann et al. (2010) studied $n = 94$ listeners from 4 different L1 backgrounds. These L1s differ in the size of their labial sound inventories. While German and Croatian have 4 labial consonants /p b f v/, English and Polish have an additional category /w/. Stimuli were manipulated along the two acoustic dimensions and presented in a forced-choice identification task. It was observed that English and German differ in the perceptual boundary between /f/ and /v/. The English threshold is located at a higher noise fraction, indicating more friction in English /v/. The harmonics-to-noise ratio appears to be the primary cue for speakers of all languages.

These observations point to contrasts between English and German in the organization of the perceptual space. Figure 8.7 schematically compares the phonological segmentation of the [w]–[ʋ]–[v]–[f] continuum in the two languages. This is clearly a simplification, as it reduces a multidimensional space to a one-dimensional representation and neglects L1-specific differences in cue weighting. Nevertheless, it serves to highlight relevant contrasts. Thus, presence of /w/ in English requires native listeners to establish a /w/-/v/ category boundary, which the German perception grammar lacks. Presence of /w/ apparently shifts the category boundary of /v/ further toward the [f]-end of the continuum (as noted by Hamann et al. 2010). The findings reported by Hamann & Sennema (2005b) suggest that – relative to English /v/ – German /v/ is located further toward the [w]-end of the continuum. This appears to result from the fact that the acoustic space is not occupied by another (consonantal) category. Consequently, German /v/ covers a larger portion of the acoustic space, extending further toward the [w]-end.

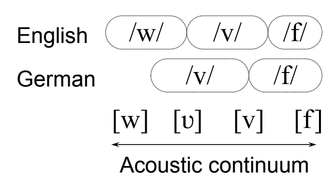


Figure 8.7: Organization of the English and German perceptual space along the [w]-[f] continuum: Schematic representation.

In sum, the literature suggests heterogeneity among German learners in the perceptual sensitivity to the English /v/-/w/ contrast in initial and intervocalic position. While perceptually weak learners show no sensitivity to place and manner cues, those with established categories achieve excellent discrimination but with slightly different best exemplar locations for /v/. Empirical evidence indicates that perceptual problems may arise from contrasts between German and English with regard to the organization of the perceptual space. Presence of /w/ in English appears to shift category boundaries. As yet, no study has explicitly addressed German learners perceptual sensitivity to the [v]–[ʋ] contrast. However, findings by Iverson et al. (2008) and Hamann & Sennema (2005b) may be interpreted as providing indirect evidence for a lower sensitivity of German learners to the [ʋ]–[v] contrast compared to the [ʋ]–[w] contrast.

8.3 L2 acquisition of English onset /v/: Theoretical predictions

The linguistic factors discussed in the preceding section allow us to derive predictions about the acquisition of English onset /v/ by German learners. This section outlines the types of patterns anticipated by different theoretical strands.

CAH English onset [v] is a new allophone of a shared phoneme and therefore a difficult structure. L1 transfer is predicted to yield substitution by [ʋ].

SLM The literature review above puts forward the following assumptions regarding German learners' perception of English onset /v/: (i) [ʋ] is the perceptually closest L1 sound; and (ii) the degree of perceived similarity between [ʋ] and [v] is moderate to high. This suggests perception-based transfer of L1 [ʋ] by the majority of German learners. Empirical evidence indicates that even learners with established perceptual categories still differ from native speakers in cue weighting, with an influence of features of their L1 [ʋ] – that is, a less labiodental place of articulation. This suggests that L1 influence in perception persists throughout IL development. Consequently, productions of English /v/ in onset position are expected to show varying degrees of L1 transfer, which may surface in less purely labiodental and/or (slightly) more approximant-like productions.

FCM The phonological features contrasting English [v] with German /v/-allophones are: voice (0.55) > sonorant (0.27). As shown in Table 8.2, the most likely perceptual substitute is [ʋ].

Consonant	Rank	[voice]	>	[sonorant]
[ʋ]	1	+		+
[f]	2	–		–
[v]		+		–

Table 8.2: Substitutes for [v]: Predictions of the FCM.

SCH/MDH Based on our literature survey, we may assume that [v] is a marked structure and consequently a difficult segment in L2 acquisition. The range of clusters in onset position is limited to /vj/, in which /v/ is expected to show delayed acquisition.

NM/NDH Evidence from English L1 acquisition showed onset /v/ to be susceptible to substitution by [b], reflecting the fortition process of stopping. We may therefore assume that stopping is a natural process that has to be suppressed in the acquisition of onset [v]. In language change, on the other hand, fortitions are less likely to be observed compared to lenitions (see Lass 1984: 178). This is also reflected in a number of gestural reduction patterns that are frequently observed cross-linguistically

(Blevins 2004: 145). The fricative /v/ may therefore also be affected by opening and undergo approximantization to [ʋ] (cf. V-W confusion, §8.2.1). This suggests that approximantization is a natural process that has to be suppressed in the acquisition of [v]. From the perspective of German learners, stopping of onset /v/ may be disfavored or blocked due to the presence of the phonemic category /b/, which also occurs in their L1. Approximantization, on the other hand, is active in German, yielding the typical realization as a labiodental approximant [ʋ]. As a result, German learners need to suppress a process that is active in their L1, which is expected to be difficult. As Lass (1984: 182) points out, the preferred environment for lenitions is between vowels V_V. Pre-consonantal contexts, on the other hand, are protected environments. We thus expect /v/-lenition to prevail in onset and intervocalic position. Finally, the NM also takes into account input frequency as a facilitating factor in perceptual learning. The frequency estimates in the literature suggest that /v/ is a moderately frequent structure but occurs at a low rate in onset position. Onset /v/ is thus a relatively rare structure in learner input and output. Frequency may thus be assumed to restrain the L2 acquisition of onset /v/.

FM Upon perceptual and sensorimotor learning of [v], an initial phase of deviant production is followed by three consecutive stages: variation reflecting focus on form, overly faithful production, and variation sensitive to functional principles. With learning being driven stochastically, the above-formulated working assumption suggests an impeding effect of frequency.

MSA The MSA complements the prediction of the SLM by incorporating articulatory constraints. Acquisition of English [v] in onset position is thus expected to be further delayed due to aerodynamic difficulties in the production of voiced fricatives.

LTD The LTD predicts sensitivity to *s*/*w*-marking, yielding the following tendencies: (i) delayed acquisition in unstressed syllables; and (ii) no difference between onset singletons and the cluster /vj/ (/v/ is *s*-marked in both cases). As for the variants observed in speech production, *s*-marked environments are expected to show a tendency toward strengthened renditions. Specifically, [ʋ] may show fricativization to [v], devoicing to [f], and/or stopping to [b̥]. Weak environments, on the other hand, are more likely to show lenited [ʋ], which ties in with the general expectation of higher accuracy rates in structurally strong positions.

OPM Since onset [v] may be considered a marked and similar structure, the OPM predicts IL development in terms of both the Similarity and Markedness Corollary.[5] Figure 8.8 blends these corollaries to generate the following predictions: An initial phase of persistent L1 transfer yields substitution by [ʋ]. Acquisition is expected to be further delayed by a phase of U-dominance. Based on the discussion above, we expect U to surface in the tendency toward lenition – that is, opening to [ʋ]. Since L1 transfer

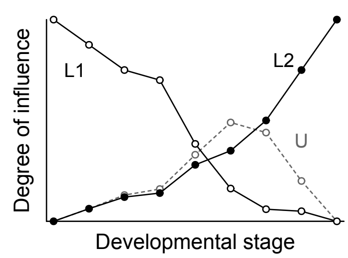


Figure 8.8: Extrapolation of the OPM to structures that are both similar and marked. This graph should be compared to Figure 2.6.

[5] See §4.3 (Figure 4.18, p. 66) for a discussion of the rationale underlying the fusion of these corollaries.

predicts the same variant, the effects of L1 and U cannot be distinguished, but target-like [v] is expected to emerge very late.

Table 8.3 gives a summary of the predictions we have been able to formulate. Next, we summarize previous empirical work on onset /v/ in German Learner English.

8.4 Onset /v/ in German Learner English: Previous work

This section surveys existing empirical work on onset /v/-production by German learners. Contrastive treatments of English and German have pointed out potential difficulties involved in the acquisition of English /v/ by German learners. Thus, Scherer & Wollmann (1986: 93) note that L1 German speakers tend to produce insufficient contact between the lower lips and the upper teeth, yielding a frictionless approximant sound. Dretzke (1998: 54) also points to the tendency of German learners to produce a sound that is too weak.

A developmental error that is often noted in connection with the acquisition of English /v/ is "over-correction" (Barry & Gutknecht 1973: 40) or "hypercorrection" (Dretzke 1998: 54). It describes a developmental error pattern in L2 acquisition reflecting the following stages:

- Stage 1: Initial transfer of L1-derived [ʋ] for both English /v/ and /w/

Table 8.3: The acquisition of /v/ by German learners: Summary of predictions based on theoretical work.

Theory	Predicted patterns in L2 acquisition by German learners
CAH	Late acquisition Substitute: [ʋ]
SLM	Late acquisition by the majority of learners Persistent substitution by [ʋ] L1 influence of place and manner features also likely in advanced learners
FCM	Perception-based substitute in speech production: [ʋ]
MDH/SCH	Late acquisition Acquisition order: singleton > cluster
NM/NDH	No effect of frequency Late acquisition Substitution patterns: persistence of lenited variants [ʋ] in pre- and intervocalic position; stopping to [b b̥]
FM	No effect of frequency Development: deviant production > stylistic variability > overly faithful production > variability reflecting connected speech processes
MSA	Late acquisition
LTD	*s*-marked positions: early acquisition; strengthenend substitutes [v f b̥] *w*-marked positions: delayed acquisition; lenited variant [ʋ] Acquisition order: stressed > unstressed syllables; in onsets: singleton = cluster
OPM	Late acquisition Very late acquisition by a minor population of learners (cf. SLM) Substitutes at early and intermediate stages: [ʋ]

leads to productions such as *weather* [ˈʋɛðə], and *very* [ˈʋɛ.ɹi].

- Stage 2: The new sound [w] is learned and the learner makes the change [ʋ] → [w] in speech production. In words with an underlying /w/, this yields a closer approximation to the target sound. However, the learner over-corrects and also makes the change [ʋ] → [w] in words with an underlying /v/, which leads to less target-like production. This results in developmental errors such as *very* [ˈwɛ.ɹi].
- Stage 3: The learner corrects these errors, successfully rendering /v/ as [v]: *very* [ˈvɛ.ɹi].

In the error analysis approach to learner speech, phenomena of this kind are classed as intra-lingual errors and covered under the more general label 'overgeneralization' (Richards 1971; Selinker 1971, 1972). This term refers to phenomena in speech production where a certain TL regularity is extended to contexts where it is not appropriate. While this error is not explicitly anticipated in the light of theoretical contributions, developmental patterns of this kind are usually attributed to general cognitive processes and thus classify as a universal feature of learner speech (cf. Major 2001: 47–48).

A number of studies have reported on the production of /v/ by German learners. Wieden & Nemser's (1991) analysis of n = 384 Austrian learners[6] found that correct production rates for /v/ varied by context. In syllable onsets, the authors observed an accuracy of .18 in initial position (*very*) and .13 in medial position (*television*). Deviant productions were largely the result of lenition – that is, approximantization.

Sieg (2004, quoted in Wode 2009) reports on speech production development in German immersion-setting learners. Correct production rates for initial /v/ were relatively low (.50/.35/.45/.95). The most frequently observed substitute was [w]; [f] was only observed in 1st-graders.

Langguth (2009) investigated /v/-production by German learners in grades 5, 9, and 12 (n = 7 each). The author notes that only phonemically deviant productions were counted as an error, thus excluding [ʋ] – that is, approximant-like realizations that were not velarized. Overall, onset /v/ was produced with an accuracy rate of .60 word-initially and .90 in intervocalic position.

Pascoe (1987: 178) studied /v/-production by n = 26 instructional-setting learners from southern Bavaria (grade 8–10; age 14–16) and observed an overall correct production rate of .70 for /v/. A considerable difference was noted between subjects with good (.90) and poor pronunciation (.50). However, as the author does not distinguish between onset and coda position, these results are difficult to interpret.

In a further study on instructional-setting learners from southern Bavaria, Kucharek (1988) investigated onset /v/-production in 5 learner groups: students in grade 6 (n = 169), 7 (n = 138), 8 (n = 117), and 12 (n = 94) and teachers (n = 62), adding to a total of n = 580 participants. Production accuracy for onset /v/ was assessed with only one item (*very*), which was elicited with a reading passage. The results showed accuracy rates between .70 in grade 6 and .80 in grade 12. Among teachers, production accuracy was .92.

In sum, the empirical evidence on onset /v/-production in GLE docu-

[6] The study included learners at 6 proficiency levels, with n = 64 subjects in each group. Years of instruction was used as an indicator of developmental stage, which ranged from 0 years (grade 3; 8–9 years of age) to 8 years (grade 11; 16–17 years old). Figures for the individual stages were not reported.

ments variable levels of production accuracy. Table 8.4 gives a summary of accuracy rates and the factors that have been reported to influence /v/-production. These include proficiency level and years of instruction as well as position in the word. As for the latter, reported effects on the accuracy rate vary. This may be attributable to differences concerning which types of variants were rated as deviant. Observed variants for onset /v/ are [ʋ w f v].

Table 8.4: Empirical evidence from auditory studies on /v/ in German Learner English: reported production/accuracy rates and constraints on accuracy.

Results		Notes	Study
Accuracy rate	.35–.95	Grades 1–4; age: 6–10 years	Sieg 2004
	.70–.80	Grades 6–12; age: 11–18 years	Kucharek 1988
	.60–.90	Grades 5, 9, 12; age: 10–18 years	Langguth 2009
	.13–.18	Grades 3–11; age: 8–17 years	Wieden & Nemser 1991
	.50–.90	Grades 8–10; age 14–16 years	Pascoe 1987
Constraints on accuracy			
Position in word	Δ = –.30	Initial (.60) < intervocalic (.90); [ʋ] was not counted as an error	Langguth 2009
	Δ = .05	Initial (.18) > intervocalic (.13); [ʋ] was counted as an error	Wieden & Nemser 1991

8.5 Aims of this study

This study aims to shed light on three aspects of onset /v/-production by German learners:

- developmental patterns reflected in cross-sectional data on the realization of onset /v/ and
- linguistic constraints on the realization of onset /v/.

Cross-sectional patterns in the realization of onset /v/

The L2 phonological theories under consideration unanimously consider English /v/ a difficult structure that is expected to be acquired late (CAH; SLM; MDH/SCH; NM/NDH; MSA; OPM). The labiodental approximant [ʋ] is broadly considered the most likely substitute in onset position (CAH; SLM; FCM; NM/NDH; MSA; OPM). However, on the basis of the assumptions formulated, the NM/NDH also suggests [b b̥] as likely variants, and the LTD posits strengthened segments [f v b b̥] in *s*-marked positions. While the empirical literature documents increases in /v/-accuracy as a function of IL development, little is known about the likelihood of different variants across proficiency levels. Earlier work either appears to make no distinction between [ʋ] and [v] (Langguth 2009; Sieg 2004) or does not report on the distribution of variants in different populations of learners (Kucharek 1988). This study extends the empirical record by shifting focus to cross-sectional patterns in German learner productions. Specifically, over-correction errors have not previously been addressed empirically and this investigation aims to address this gap. Considering the three hypothetical developmental stages described above, the expected patterning of /v/-variants in GLE is shown in Figure 8.9. Thus, learners are expected to show initial transfer of L1 [ʋ], followed by a tendency toward [w]-over-correction. Target-like production of [v] is expected to occur at a relatively late stage.

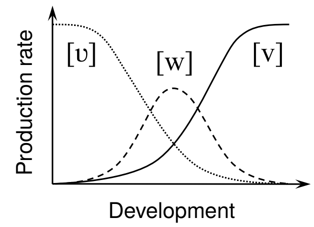


Figure 8.9: Expected developmental pattern for onset /v/ production by German learners: Distribution of variants over the course of IL development.

Linguistic factors

Position in the word is the only linguistic constraint that has been considered in the empirical literature (Langguth 2009). The present study seeks to broaden the scope by focusing on the following constraints:

- *Structural strength*: According to the LTD, the effect of structural strength should surface in two ways. First, a higher accuracy rate is expected in *s*-marked positions. Second, deviant renditions in structurally strong environments are predicted to show a tendency toward strengthened variants [f v b̪].
- *Speaking style*: The FM and OPM expect stylistic variation as a function of focus on form. While the former predicts an effect at relatively early stages of /v/-acquisition, the OPM expects style-dependent variation across all stages.

8.6 *Method and data*

Informants, materials and procedures

The following analyses are based on three data sets:[7]

- The *SpIL* corpus (Pascoe 1987) contains $n = 30$ occurrences of onset /v/, which are produced by 11 (out of 26) subjects. Due to the limited number of observations, these data will only be discussed descriptively.
- The second supplementary data set is from a study by Rank (2016). A total of $n = 44$ items were elicited with two tasks: a word list ($n = 29$) and a reading passage ($n = 15$). This sums to a total of $n = 1{,}144$ tokens for analysis.
- In the current study, each learner produced $n = 7$ /v/-tokens, summing to a total of $n = 429$ renditions for analysis ($n = 5$ missing cases).

The analyses of the supplementary data sets relied on the transcriptions provided by the respective authors. In contrast to other studies (e.g. Sieg 2004; Langguth 2009), both Pascoe (1987) and Rank (2016) distinguished between the labiodental fricative [v] and the approximant [ʋ]. /v/-realizations in the recordings collected for the current study were assigned to three categories [v ʋ w]. Productions that may be described as weak labiodental stops [b̪] with audible contact between the lips and the upper teeth occurred sporadically and were assigned to the category [v]. It should be noted that the distinction between [w] and [ʋ] is in many cases difficult to make, since velarization is a gradient feature and approximant productions may in fact show a simultaneous labiodental and bilabial constriction. In uncertain cases, classification resorted to the degree of

[7] The data sets analyzed in the present chapter are openly available via the *Tromsø Repository of Language and Linguistics* (*TROLLing*) at https://doi.org/10.18710/B276ZX (Sönning & Rank 2020). See §3.3.1 for details on the supplementary data sets.

velarization, with a token being assigned to [ʋ] if it exhibited no back vowel resonance.

Table 8.5 presents the distribution of variants. There is considerable discrepancy between the auditory classifications in the three data sets. Thus, while Pascoe (1987) reports no instances of [w], comparable rates of [w]-variants are noted in Rank (2016) and this study. The proportion of target-like [v] was highest in Rank (2016), which may be due to differences in proficiency, as the study was restricted to intermediate to advanced learners.

	Pascoe 1987		Rank 2016		This study	
Realization	Pr	(*n*)	Pr	(*n*)	Pr	(*n*)
[v]	.60	(18)	.73	(838)	.35	(151)
[ʋ]	.37	(11)	.11	(129)	.45	(194)
[w]	.00	(0)	.15	(177)	.20	(84)
[f]	.03	(1)	.00	(0)	.00	(0)

Table 8.5: Distribution of variants in the three data sets. "Pr" denotes the proportional share of a variant.

The factors structural strength and speaking style were only analyzed in the data set provided by Rank (2016). Speaking style reflects the task types used – that is, isolated words read from a list and a reading passage. Structural strength was coded using two categories:

- strong contexts: stressed syllables in lexical words (primary or secondary stress)
- weak contexts: unstressed syllables in lexical words

Table 8.6 shows the type and token frequency for the linguistic factors in Rank's (2016) data.

Condition	Types	Tokens
Structural strength		
Weak	5	182
Strong	27	962
Speaking style		
Word list	14	390
Sentences	29	754

Table 8.6: Distribution of types and tokens across the levels of the linguistic factors in the data by Rank (2016).

Statistical analysis

As learner productions were described using the three categories [v ʋ w], the data were analyzed with multilevel multinomial regression models. While the factors speaking style and structural strength were explored using the data provided by Rank (2016), systematic variation across proficiency levels was assessed using the data collected in this study. As a bell-shaped pattern was predicted for the over-correct variant [w], a quadratic term for proficiency was added to the model. This adds flexibility to the fitted regression lines and makes them more sensitive to non-linear patterns. Following the maximalist approach outlined by (Barr et al. 2013), both models included random intercepts on subject and word; the model for the data from the present study further includes random slopes on word for proficiency (simple and quadratic term), and the model for Rank's (2016) data includes random slopes on subject for the two linguistic factors, structural strength and speaking style. The model structure is outlined in Appendix A.8. Vague priors were assigned to the intercepts, which denote the proportion of each variant while holding (i) proficiency at 0 (the present study) and (ii) averaging over the linguistic factors (Rank 2016). The priors specified over the three variants are shown in Figure 8.10: The 90% prior probability interval for [v] is [.01, .71], that for [ʋ] [.05; .83] and for [w], it is [.03, .78]. For all predictors, regularizing priors were specified and weakly informative priors were assigned to the remaining ancillary parameters.

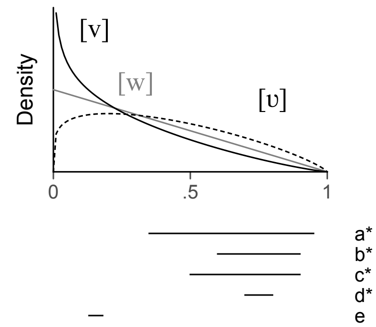


Figure 8.10: Prior for the intercept parameters in the model in comparison to the rates reported in the literature. The asterisk (*) marks studies that do not differentiate between [ʋ] and [v] and essentially estimate the share of non-[w] renditions. Key to the studies: (a) Sieg 2004, (b) Langguth 2009, (c) Pascoe 1987, (d) Kucharek 1988, (e) Wieden & Nemser 1991.

8.7 Constraints on generality

Prior to a discussion of our findings, we should clarify the extent to which these generalize to wider contexts.

- *Language-external scope*: According to Mangold (2015: 71), accents spoken in the south of Germany show weaker /v/-variants in onset position. For speakers of these dialects, the perceptual space illustrated in Figure 8.7 may then be structured differently, with category boundaries located further to the left. This would mean that the acoustic space covered by L1 /v/ would extend further into the phonetic domain of English /w/. This would be expected to increase perceptual difficulties. At the same time, L1-derived variants may also be weaker for speakers of such dialects. It follows that the generalizability of the present findings is somewhat limited, as the share of [ʋ]- and [w]-variants may be slightly higher compared to other populations of German instructional-setting learners.
- *Language-internal scope*: We are not aware of any systematic ways in which our choice of lexical items may compromise the generality of our findings.
- *Speaking style*: Since English orthography provides an unambiguous grapheme-to-phoneme mapping for onset /v/, the data elicited with reading tasks should not be directly extended to free speech. We would expect target-like [v]-renditions to be higher in read speech; the overcorrection rate (i.e. the share of [w]), on the other hand, may be lower than in spontaneously produced material.

8.8 Results

Let us first take a look at the distribution of variants across proficiency levels. Figure 8.11 shows how the proportions of [ʋ w v] vary across proficiency levels. The share of [ʋ], shown in the left panel, is highly

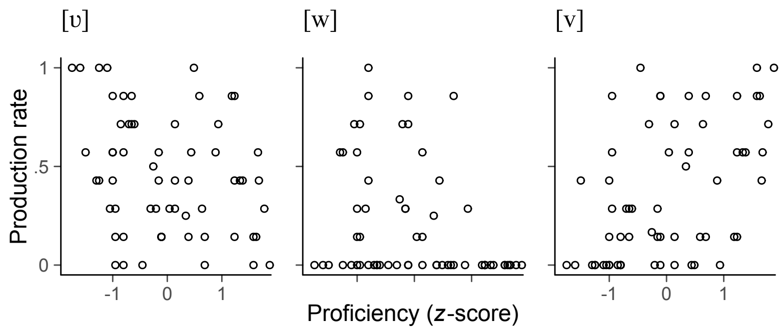


Figure 8.11: Realization of onset /v/ by proficiency level: The production rates for [ʋ w v], expressed as the share of realizations for each learner in the present study.

variable: Across almost all proficiency levels, learners span the entire scale from 0 to 1. There appears to be a mild decrease in the share of [ʋ] at advanced proficiency levels. The rate of [w] is shown in the middle panel. Variation among learners is particularly pronounced at intermediate levels of pronunciation ability. In the productions of advanced learners, on the other hand, this variant appears to be rare. Finally, the right panel shows that the share of target [v] appears to increase, on average, with proficiency. Nevertheless, substantial variation between learners is evident at all proficiency levels.

Figure 8.12 shows the distribution of variants in the data provided by Rank (2016). Each line shows, for a particular speaker, the proportional profile of variants observed (i.e. $n = 26$ profiles). The typical distribution appears to be a high proportion of [v], at roughly .80, and minor shares of roughly .10 for each [ʋ] and [w]. Four learners deviate from this pattern and show a relatively high rate of [w] at about .50.

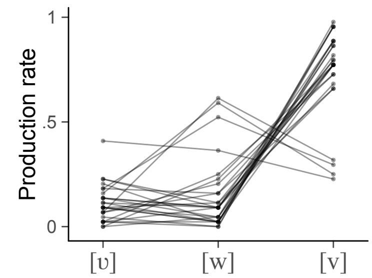


Figure 8.12: Production rates for [ʋ w v] in the data from Rank (2016). The lines show a profile for each learner, connecting the observed proportion of each variant from left to right.

The technical output of the models is deferred to Appendix A.8, where posterior estimates for parameters are listed. Here, we directly turn to a visual summary of the indications offered by the models. Let us first take the data from the present study, to see whether the realization of /v/ varied conditional on proficiency level. Figure 8.14, which graphs estimated production rates by proficiency, reveals the following trends:

- The rate of L1-derived [ʋ] decreases steadily from about .85 to .40.
- The rate of target-like [v] increased from .10 to .60, on average.
- On average, the rate of [w] shows a mild increase to about .20 at lower intermediate stages and decreases to zero at upper intermediate stages.

If we compare these model-based trends to the dispersion of learners in Figure 8.11, it is clear that the estimated rates for different proficiency levels do not offer a representative description of the informants in this study. The estimated production rates that are reported in Table 8.7 therefore mask the extensive amount of variation that is observed around them.

The results for structural strength and speaking style, which are based on an analysis of the data provided by Rank (2016) are shown in Figure 8.13. The following points are noteworthy:

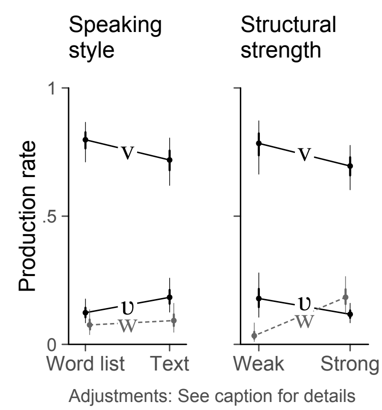


Figure 8.13: The distribution of variants by speaking style and structural strength. Error bars reflect 50% and 90% uncertainty intervals; data from Rank (2016). The following adjustments were made: In the left panel, estimates are adjusted to the simple (i.e. unweighted) average over weak and strong contexts; in the right panel, the predictor speaking style is set to the simple average over word list and reading style.

- Production accuracy is .80 for the word list and .72 for the reading passage (a difference of .08).

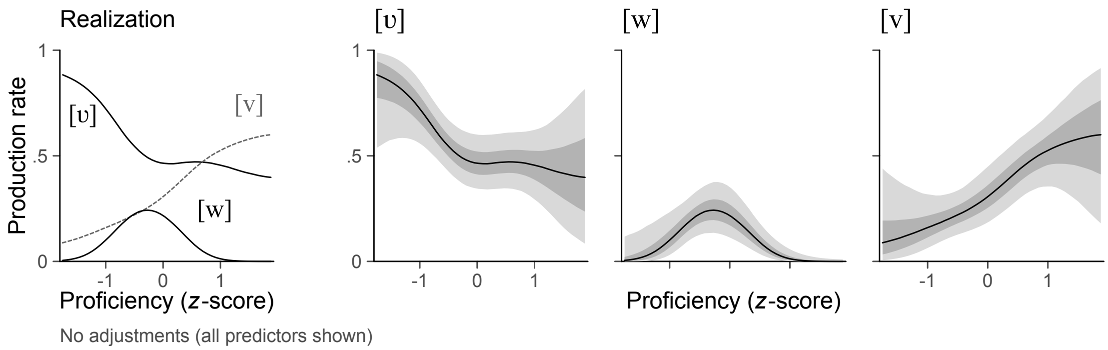


Figure 8.14: Rate of onset /v/ variants by proficiency level. Error bands reflect 50% and 90% uncertainty intervals.

- Production accuracy is higher in strong (.78) compared to weak contexts (.70) (a difference of .08).
- The rate of [w]-variants did not vary with speaking style (.08 in the word list, .09 in the reading passage).
- The rate of [w]-variants was higher in strong (.18) compared to weak contexts (.03) (a difference of .15).

	[ʋ]			[w]			[v]		
Proficiency	*Mdn*	.05	.95	*Mdn*	.05	.95	*Mdn*	.05	.95
Beginner	.85	.57	.97	.01	.00	.14	.11	.01	.39
Lower Intermediate ($z = -0.5$)	.56	.44	.69	.21	.12	.33	.21	.12	.33
Upper Intermediate ($z = +0.5$)	.47	.35	.61	.10	.04	.19	.42	.28	.57
Advanced	.42	.16	.71	.00	.00	.02	.58	.28	.84

Table 8.7: Estimates for the production rate of variants at representative proficiency levels (data from this study): Posterior medians (*Mdn*) and the upper and lower limits of a 90% uncertainty interval.

8.9 Summary and discussion

This chapter was concerned with German learners' acquisition of the English labiodental fricative /v/ in onset position. While English and German share a phonemic category for /v/, a comparison of its realization in onset position revealed that the typical English allophone [v] is produced with greater friction than the typical German realization [ʋ], which may be described as a weak labiodental approximant. Concerning linguistic constraints on the acquisition of onset /v/, evidence from typology, L1 acquisition and the physiology of articulation suggests the classification of /v/ as marked. As for perception, previous studies have found German listeners' perceptual sensitivity to the /v/–/w/ contrast in non-final position to be variable. Further, there is indirect evidence for a relatively greater level of difficulty in distinguishing labiodental approximants [ʋ] from fricatives [v]. Judging from research on best exemplar ratings, it appears that a lenited representation for /v/ prevails in listeners who otherwise show excellent perceptual categorization of English /v/ and /w/. Theory-informed predictions unanimously consider /v/ a difficult

	[ʋ]			[w]			[v]		
Condition	*Mdn*	.05	.95	*Mdn*	.05	.95	*Mdn*	.05	.95
Structural strength									
Strong	.12	.08	.16	.18	.12	.26	.70	.60	.78
Weak	.18	.11	.28	.03	.01	.08	.78	.66	.87
Speaking style									
Word list	.12	.08	.18	.08	.04	.14	.80	.71	.87
Text	.18	.13	.26	.09	.05	.16	.72	.62	.80

Table 8.8: Estimates for the production rate of variants in the different conditions; data from Rank (2016).

structure, expecting [ʋ] but also [f b̥] as IL variants. Insights gained from empirical observations further identify [w] as a likely surface realization, resulting from an over-correction of [ʋ] to [w] in /v/-contexts. This error pattern, which is generally described as reflecting a three-stage process, is not anticipated by theoretical accounts.

While broadly consistent with theory-based predictions, a comparative evaluation of empirical work on /v/-production by German learners is complicated by the fact that some studies either do not differentiate between [v] and [ʋ] or provide no information about prosodic position. The latter neglect is particularly unfortunate as the processes and outcomes in onset and coda position are predicted to differ profoundly. The latter position is expected to show strong tendencies toward devoiced [v̥] or voiceless [f] as a result of transfer of final devoicing, a regular phonological process in L1 German (see Chapter 10). Further, not much attention has been devoted to linguistic constraints on /v/-accuracy and the interaction between /w/ and /v/ in the course of acquisition. From an empirical perspective, there is little documentation of the process of over-correction in /v/-acquisition. The objective of this study was to investigate the distribution of variants across different levels of L2 pronunciation ability and test predictions about the sensitivity of /v/-realization to speaking style and structural strength.

Findings in the present investigation provide further evidence on the status of onset /v/ as a difficult structure for German learners. Even highly proficient speakers show an accuracy rate of only .60, on average. The results pointed to a high level of between-learner variation, with all variants showing great variability between subjects at all proficiency levels. There was also evidence of a bell-shaped emergence and disappearance of over-corrected [w]-variants at lower intermediate levels. The rate of [w]-variants was also highly variable across the informants in this study. The patterns observed for the linguistic factors, on the other hand, lend only partial support to theory-based expectations. Findings for /v/-production in different speaking styles expectedly showed higher accuracy rates in more monitored speech – that is, in the word list task as compared to the reading passage. LTD predictions about a boost in target-like production in *s*-marked positions were not borne out by the data, however. In fact, lower accuracy rates were observed in the onset of syllables carrying primary or secondary stress, which conversely showed a higher rate of overgeneralized

[w]-renditions. This further challenges LTD predictions, which posit a tendency toward fortition in structurally strong positions.

The high level of variability between subjects points to the importance of learner-related factors in accounting for /v/-realization by L1 German speakers. This finding is consistent with insights offered by perception studies, which indicates heterogeneity among German learners in terms of their perceptual sensitivity along the [v]–[ʋ]–[w] continuum. While the findings for speaking style are consistent with expectations, it is critical that future work focus on more natural speech. The investigation of less monitored productions seems warranted on the grounds that the average over-correction rate (0 to .20) seems rather low.[8] As indicated by the low prevalence of onset /v/ in the *SpIL* corpus (Pascoe 1987), one challenge that studies investigating free speech will face is the low rate at which onset /v/ occurs.

[8] This judgement is based on my personal impression of the actual over-correction rate in German learner speech.

As for the relationship between /v/-realization and structural strength, the rate of [w] showed the greatest sensitivity to this factor. In fact, its share in unstressed contexts is close to zero, which points to the potential status of structural strength as a (near-)categorical constraint. An explanation for this finding may be found in parallels to /w/-contexts. Thus, the distribution of over-correct variants may follow the distribution of [w]-realizations for /w/, which could provide a phonological template for over-correction. In fact, an inspection of the top 1,000 words in the COCA (Davies 2008-) reveals that 49 out of the 50 items with /w/ occur in the onset of a stressed syllable (the only exception being *always*). Given this highly skewed distribution of /w/ in terms of structural strength, it is likely that /v/ → [w] over-correction is favored in (or perhaps restricted to) the onset of syllables carrying primary or secondary stress. While more work is required to shed further light on this constraint on over-correction, the following predictions follow from these considerations:

- Over-correction is more likely in syllables carrying primary or secondary stress.
- Over-correction is disfavored in unstressed syllables, before schwa, and before syllabic consonants.

Concerning L2 speech theories, future efforts must also address how over-correction errors can be predicted without reliance on prior empirical observation. While an explanation in terms of general cognitive learning processes is certainly feasible, it appears that intra-lingual errors remain outside the predictive scope of our current state of theory. Efforts should therefore be directed toward establishing testable constraints on over-correction. These include the conditions under which it occurs, the TL categories that are affected and the directionality of influence. The insights provided by the /v/-/w/ setting for over-correction may be suggestive of necessary conditions for this type of intra-lingual error: If, perceptually, an L1 sound is located between two TL categories for which there is no other L1-derived substitute, then this L1 sound is initially used to realize both TL sounds. If category formation occurs earlier for one of the TL sounds, then it is this newly established IL category that triggers over-correction errors in the other. Predictions about over-correction can thus be formulated on the basis of insights into speech perception, with the degree of perceived dissimilarity suggesting the order of category formation and therefore the directionality of influence.[9]

As far as linguistic constraints on the acquisition of onset /v/ by German learners are concerned, our current understanding of perceptual distortions in German listeners is still limited. Specifically, knowledge about cue reliance in initial position remains scarce, and researchers currently have to extrapolate from findings on the perception of intervocalic position /v/ to formulate predictions about other contexts. Further, applied research would profit from experimental insights into the degree of perceived dissimilarity between German [ʋ] and English [v] and [w].

Empirical work on onset /v/ in GLE may profit from the application of acoustic techniques, as these yield a more nuanced description of learner productions and do not require a categorization of tokens along the [v]–[ʋ]–[w] continuum. Instrumental measurements may serve to document more subtle traces of cross-linguistic influence in advanced learners, who have been observed to exhibit such L1-induced biases in speech perception.[10] Finally, instrumental studies should also report on the acoustic properties of learners' L1 /v/-productions to allow for more reliable assessments of the nature and degree of cross-linguistic influence.

Let us conclude with suggestions for future work on onset /v/ in GLE:

- Onset and coda variants should be clearly distinguished in the analysis, interpretation, and presentation of results. Studies failing to make this distinction are of little comparative value.
- Auditory studies should strive to distinguish between [ʋ] and [v], as these reflect different structural IL components – that is, L1-derived and TL variants, respectively. Failure to distinguish between the two may miss developmental changes in substitution patterns.
- As a distinction between [v] and [ʋ] in learner productions may be difficult in certain cases, it appears worthwhile to consider applying acoustic methods for the investigation of variation in onset /v/ production by German learners. Instrumental studies should not fail to report on the acoustic properties of the learners' L1 /v/-variants.
- At this point, it appears that research on more natural speaking styles is clearly warranted. It should be kept in mind, however, that onset /v/ may by rare in spontaneous speech.
- In light of the empirical evidence reported in this study, structural strength appears to be a promising factor for the investigation of over-correction.

[9] In retrospect, we encountered very similar settings when dealing with the TL categories TRAP vs. DRESS (see Chapter 4) and clear [l] vs. dark [ɫ] (see Chapter 5). The acquisition of these structures by German learners may therefore offer further insights into the intra-lingual mechanisms giving rise to over-correction.

[10] Hamann & Sennema (2005a) successfully applied acoustic measures to distinguish between [v] and [ʋ].

9 Dental fricatives

Our survey of segmental features in German learner phonology continues with a pair of obstruents that is absent from the German phoneme inventory: the dental fricatives /θ ð/. After a contrastive analysis of English and German in §9.1, we discuss linguistic constraints relevant for the L2 acquisition of /θ ð/ (§9.2). Theoretical frameworks are consulted in §9.3 to derive suggestions about systematic patterns in GLE. A review of quantitative work on /θ ð/ in L2 speech production (§9.4) is followed by a statement of the aims (§9.5) and method (§9.6) of this study, which focusses on the voiced variant /ð/. §9.7 identifies the circumstances to which the present findings may generalize, followed by a presentation of results in §9.8. §9.9 closes with a summary and discussion.

9.1 Contrastive analysis

In contrast to German, the English phoneme inventory includes a pair of dental fricatives. These are denoted by the IPA symbols /θ ð/ for the voiceless and voiced variant, respectively. While /θ ð/ are sometimes described as being complementarily distributed allophones, their status in present-day English is clearly phonemic, as they contrast a number of minimal pairs such as *mouth* (noun vs. verb) or *teeth* vs. *teethe*.[1] The place of articulation for /θ ð/ may vary between *post*-dental (the tip of the tongue touches the edge and back side of the upper teeth) and *inter*-dental (the tongue protrudes the upper and lower teeth). In terms of place of articulation, post-dental variants are more common in BrE and interdental ones in AmE (Ladefoged & Maddieson 1996: 144; Ladefoged & Johnson 2011: 12). The term *dental* fricative will hereafter subsume both variants.

[1] This opposition, however, carries little functional load and diachronically /θ/ and /ð/ in fact derive from a single phoneme.

In comparison to other languages, English has a dense system of fricatives. In particular, the three-way contrast between dental [θ ð], alveolar [s z] and palato-alveolar [ʃ ʒ] requires fine control of the articulators. Besides constriction location, these sounds differ in tongue configuration[2]: for [s z], air escapes along a groove, for [θ ð] along a slit. The articulation of [θ ð] thus involves no contact between tongue and palate.

[2] Cruttenden (2014: 193) stresses that tongue configuration is at least as relevant as place of articulation in the production of fricative sounds.

/θ/ and /ð/ differ in distribution. Both occur in onset and coda position, but only /θ/ is licensed in word-initial clusters. Singleton /θ/ occurs in initial (*thick, thought*), medial (*author, method*), and final position (*teeth, bath*). In onset clusters, it may combine with an approximant, most frequently /r/ (*three, throw*) and less commonly /w/ (*thwart*) or /j/ (*enthusiasm*). In coda position, it may constitute part of a wide range of clusters consisting of up to four segments (*lengths*) (Cruttenden 2014: 198). Singleton /ð/ is likewise found in initial (*these, though*), medial (*mother, weather*) and final position (*breathe, teethe*). While /ð/ does not occur in onset clusters, it combines with a number of segments in final position to form dyadic clusters (*rhythm*).

In present-day BrE and AmE, the production of dental fricatives varies conditional on segmental context and speaking style. In general, assimilation of [θ ð] to adjacent segments is likely to be observed in unstressed (vs. stressed) positions, informal (vs. formal) styles, and rapid (vs. slow) speech (Collins & Mees 2013: 125). Adjacent sibilants, for example, cause articulatory difficulty and may trigger gestural modifications. It is thus not uncommon for dental fricatives to be elided in clusters (e.g. *months* [mʌns], *clothes* [kləʊz]) or assimilated in connected speech (e.g. *Pass the salt* [ˈpɑːszəˈsɔːlt], *Is this it* [ɪˈzɪˈsɪt]). Sequences involving /s z/ appear to be particularly susceptible to coarticulatory influence (Cruttenden 2014: 199).

Phonetic context effects primarily affect /ð/, a phoneme that permeates connected speech due to its occurrence in several high-frequency function words (*the, this, that, these, those, they, them, there*). As Roach (2009: 45) argues, since /ð/ lacks friction in natural speech, this consonant may – on phonetic grounds – be more adequately classified as a weak dental stop. In casual speech it may also be elided (*give them* [ˈɡɪvəm]) or undergo assimilation, such as loss of friction as a result of perseverative manner assimilation when preceded by an alveolar nasal (*in the* [ɪn̪n̪ə]) or stop (*and the* [ənd̪d̪ə]). The definite article *the* is particularly susceptible to such processes when preceded by /n l s z/; nevertheless, a contrast to the indefinite article *a* is still maintained (Collins & Mees 2013: 126). When preceded by an alveolar sibilant, /ð/ may undergo place assimilation in casual speech as in *Is there any* [ˈɪz zəɹˈeni] or *What's the time* [ˈwɒts zə ˈtaɪm] (Cruttenden 2014: 199).

Table 9.1 provides an overview of attested patterns of perseverative coarticulation in word-initial /ð/. In terms of place of articulation, the literature suggests that only alveolar consonants trigger place assimilation. However, no such effect is reported after /t d/, which conversely show anticipatory place assimilation to /ð/. Preceding /r/ also does not seem to affect place features in /ð/. Further, initial /ð/ has been observed to accommodate manner features of preceding consonants. It can thus be produced as a (dental) lateral following /l/, as a (dental or alveolar) nasal following /n/, and as a dental plosive following /t d/. Finally, voicing features have also been observed to spread to word-initial /ð/.

In brief, due to the absence of /θ ð/ in their native language, German learners need to acquire new allophones of two new phonemes. These involve a novel place of articulation, which has a meaning-distinguishing function in English (e.g., *sink* vs. *think*, *three* vs. *free*; *there* vs. *dare*). /θ ð/ are subject to variation in English. This is especially true for /ð/, which occurs in high-frequency function words and is susceptible to connected speech processes including elision and assimilation to the preceding segment.

Parameter	Trigger	Result	Example	Reference
Place				
Alveolar	[s]	[s]	*What's the time* [wɒts sə taɪm]	Collins & Mees 2013: 126
	[z]	[z]	*Is there any* [ɪz zə ɹeni]	Cruttenden 2014: 199
	[n]	[n]	*in the morning* [ɪn nə mɔːnɪŋ]	Cruttenden 2014: 199
	[l]	[l]	*all the time* [ɔːl lə taɪm]	Collins & Mees 2013: 126
Manner				
Plosive	[t]	[t̪]	*Get them* [get̪ t̪əm]	Roach 2009: 113
	[d]	[d̪]	*Read these* [ɹiːd̪d̪iːz]	Roach 2009: 113
Nasal	[n]	[n̪]	*in the* [ɪn̪n̪ə]	Roach 2009: 113
		[n]	*on the shelves* [ɒn nə ʃelvz]	Collins & Mees 2013: 126
Lateral	[l]	[l]	*all the way* [ɔːl lə weɪ]	Collins & Mees 2013: 126
Voicing				
Voiceless	[t]	[n̪]	*Get them* [get̪ t̪əm]	Roach 2009: 113
	[s]	[s]	*What's the time* [wɒts sə taɪm]	Collins & Mees 2013: 126

Table 9.1: Effect of preceding consonants on /ð/: Attested patterns of perseverative coarticulation of place, manner and voicing in native speech.

9.2 L2 acquisition of dental fricatives: Linguistic factors

9.2.1 Markedness

Dental non-sibilant fricatives are typologically rare; according to Maddieson's (1984) figures, voiced and voiceless dental fricatives occur at a rate of only .05 in the languages in the UPSID database. It is common for languages to have an *s*-sound, a sibilant with either an alveolar or (less commonly) a dental place of articulation (Fromkin & Rodman 1978; Maddieson 1984: 44).

Dental fricatives have consistently been reported to be among the last speech sounds acquired by English-learning children (Wellman et al. 1931; Templin 1957; Olmsted 1971; Prather et al. 1975; Dodd et al. 2003). Figure 9.1 presents the results of a pooled analysis of 4 studies (cf. §3.3.2). Panel b shows estimates for /ð/-accuracy reported in different studies. Word-initial and -final position are distinguished and the curves give a summary of of the cross-sectional patterns. Figure 9.1a, which shows dental fricatives in the context of other consonants, indicates that /θ/ and /ð/ are among the last sounds to emerge in L1 acquisition. For /ð/, the age of acquisition is 4;11 in initial position and 7;2 in final contexts. The predominant substitutes for initial /θ/ observed in English-learning children are [f] > [s] > [t] (.40/.24/.10, Olmsted 1971; .62/.16/.04, Ingram et al. 1980). The preferred substitute for initial /ð/, on the other hand, is almost categorically [d] (.90, Olmsted 1971). For /ð/, the acquisition in coda position shows a considerable delay of about 2 years. No effect of prosodic position is noted for /θ/.

The status of dental fricatives as marked structures is further substantiated by their synchronic and diachronic instability. Their susceptibility to language change is reflected in the patterns of variation observed across varieties of English. Based on Schneider et al. (2004), Blevins (2006: 11)

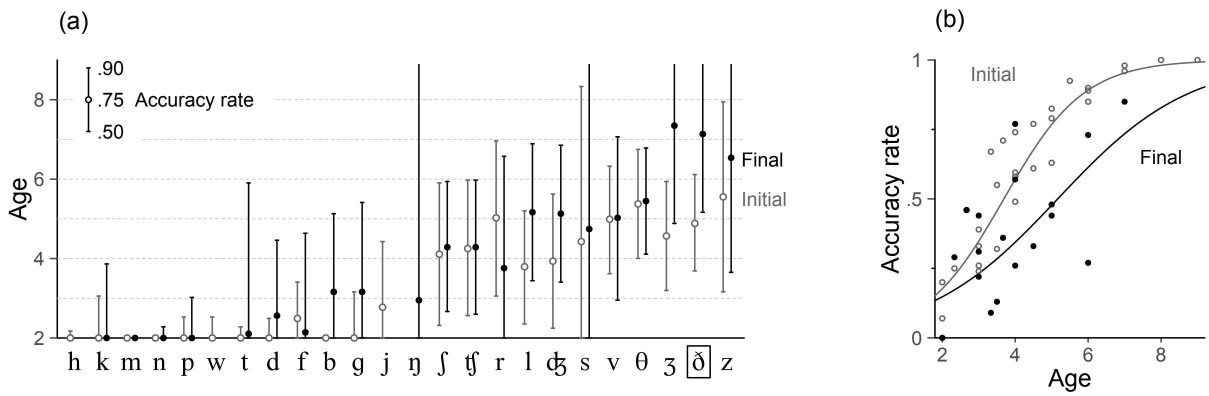


Figure 9.1: /ð/ in English L1 acquisition in comparison to other consonants. Panel (a): Points mark the age of acquisition (.75), bars indicate the age of customary production (.50 accuracy, lower end), and age of mastery (.90, upper end). Estimates are based on a joint analysis of 4 studies (see §3.3.2). Panel (b): Accuracy rate in initial (grey) and final position (black) by age. The points denote estimates reported in the individual studies, the trend lines are averages based on the joint analysis.

and Jekiel (2012: 61) summarize evidence of two widespread tendencies:

- *th*-stopping, the stop-like production of /θ ð/ (e.g. Shetland, West Ireland, Newfoundland, Maori English, Fiji English, New York English, African American Vernacular) and
- *th*-fronting, the use of labiodental substitutes (Southeastern England, New Zealand, Australia, African American Vernacular).

Substitution by [f v] is rooted in speech perception (see below), and both changes are motivated by articulatory simplification. While dental fricatives are not intrinsically complex gestures, articulatory difficulties arise in sequences – that is, in combination with other consonants, most notably sibilants (cf. Cruttenden 2014: 199).

Criterion	[θ]	[ð]
Implicational relationship	○	○
Cross-linguistic frequency	++	++
L1 acquisition	++	++
Synchronic/diachronic instability	+	+
Articulatory complexity	(+)	(+)

Table 9.2: Summary of markedness criteria for [θ ð]. Open circles ○ denote no (or inconclusive) evidence; (+)+ refers to (strong) evidence for marked status; (–)– refers to (strong) evidence for unmarked status.

Table 9.2 summarizes aspects that identify /θ ð/ as marked sounds. From a typological perspective, they are disfavored speech sounds. The same is true for L1 acquisition, as they are among the last segments acquired by English-learning children. They also appear to be an easy target for language change and show synchronic instability across varieties of English. One aspect that may contribute to these facts is the articulatory complexity arising in combination with other sounds.

9.2.2 *Frequency and orthography*

The dental fricatives /θ ð/ differ substantially in frequency, which is due to the fact that /ð/ features in high-frequency function words. Based on the frequencies reported in previous work (see §3.3.3), estimates for /ð/ range between 2 and 4, with a median rate of 3 per 100 segments. Voiceless /θ/, on the other hand, is one of the least frequent consonants, with estimates extending from .04 to .07 (*Mdn* = .04). As is evident from Figure 9.2, the rate of /ð/ is above average and /θ/ is among the rarest consonants. Orthographically, the English dental fricatives are unambiguously denoted by <th>. However, spelling provides no cue to the voicing contrast.

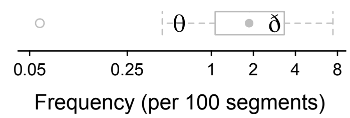


Figure 9.2: Frequency of /θ ð/ relative to other English consonants, expressed as the number of instances per 100 segments. Rates are shown on the log scale, but tick mark labels give the original values of the rates. Data are pooled across 9 studies (see §3.3.3 for details).

9.2.3 Similarity

One explanation that has been put forward for the typological rarity of /θ ð/ is their lack of perceptual salience (e.g. Maddieson 1984: 51). Acoustically, they are similar to other non-sibilant fricatives, most notably labiodental [f v]. Not surprisingly, therefore, [f v] are the speech sounds dental fricatives are most likely to be misheard as, both by native and non-native listeners. Miller & Nicely (1955) studied AmE listeners' confusion patterns for 16 English consonants and observed that [θ] was most likely to be misheard as [f] (.72); [ð] was most often confused with [v] (.65). These results have been corroborated by other studies (Tabain 1998; Cutler et al. 2004; Brannen 2011).

Similar confusion patterns have been noted for German learners at different proficiency levels. Weiher (1975) investigated the perception of dental fricatives by *n* = 12 northern German learners in grade 5 (age: 10 to 11), with an average of 0.5 years of instruction. A discrimination and an identification task with nonce words were used (*n* = 176; spoken by a BrE native speaker).[3] Target segments were embedded in different phonetic contexts. Overall, perception accuracy was similar for [θ] and [ð], with rates of .72/.75 in the identification task, and .82/.83 in the discrimination task, respectively. Perception accuracy varied by context. On average, dental fricatives were perceived more accurately as singletons than in clusters. For singleton [θ], the accuracy rate was higher in initial position, for singleton [ð] in final position. Across all contexts, the preferred perceptual substitutes were [f] for [θ] and [v] for [ð], matching the patterns observed for native speakers.

[3] The identification task employed an X-ABC format, where [θ] and [ð] were contrasted with two alternatives from the set [t s f] and [d z v], respectively. In the discrimination task, an X-AB format was used.

Similar results are reported by Hancin-Bhatt (1994) for *n* = 10 advanced German learners.[4] An identification task with nonce words (*n* = 168) included [ð θ t d s z f v] in initial, intervocalic and final position. Accuracy rates ranged from .30 to .50 and were somewhat higher for [θ] (initial .47, medial .45, final .53) than for [ð] (initial .41, medial .27, final .53). Figures for the categorization as labiodental fricatives [f v] were not reported, but the author notes that "there was a strong tendency toward this type of misperception, even among native speakers" (Hancin-Bhatt 1994: 258).

[4] The subjects in this study were, on average, 24 years of age, had 7.6 years of instruction, and started learning English at the age of 13.

Further evidence is reported by Hanulíková & Weber (2012), who employed an AXB discrimination task to test *n* = 12 German university students' discrimination between [θ] and [f s t]. In line with the results of earlier studies, the error rate was highest for the [θ]-[f] contrast (.26) compared to that for [θ]-[s] (.11) and [θ]-[t] (.10).

Cross-sectional patterns in the perception of word-initial [θ ð] were reported by Wieden & Nemser (1991: 158, 162). A total of *n* = 384 Austrian learners at different proficiency levels took part in the study.[5] Nonce

[5] The study included learners at 6 proficiency levels, with *n* = 64 subjects in each group. Years of instruction was used as an indicator of developmental stage, which ranged from 0 years (grade 3; 8–9 years of age) to 8 years (grade 11; 16–17 years old).

words were used in an X-ABC task, where [ðiːft] was contrasted with [ðiːft]-[ziːft]-[ʋiːft], and [θiːft] with [θiːft]-[ðiːft]-[fiːft]. Figure 9.3a shows that correct perceptual matching increased steadily for [ð]. There was a preference for [ʋ] as a substitute at early stages (around .30). Different results were obtained for word-initial [θ]: Perceptual problems persisted across all stages with virtually no developmental progress. Correct perception rates were slightly above .50, with labiodental [f] as the preferred substitute.

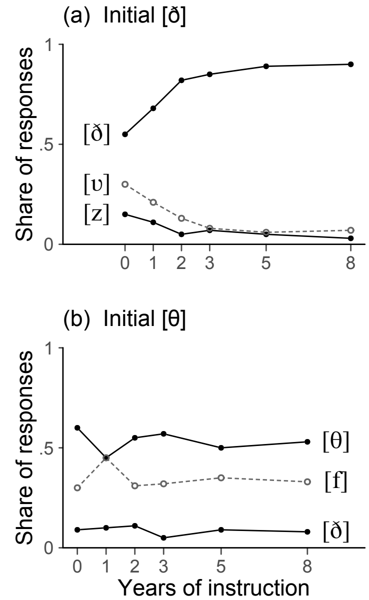


Figure 9.3: Perception accuracy of (a) [ð] against [f] and [ð], and (b) [θ] against [ʋ] and [z] in Austrian instructional-setting learners by years of learning, ranging from 0 (grade 3) to 8 (grade 11). Adapted from Wieden & Nemser (1991: 159, 162) with permission.

To recapitulate, the literature shows that, not unlike native speakers, German learners have difficulties with the perception of dental fricatives in nonce words. Accuracy rates for [θ] range from .45 to .80, those for [ð] from .30 to .90. These figures vary across studies and are influenced by task design. In particular, the type of contrasts included among the experimental stimuli affects performance. Dental fricatives are most likely to be confused with labiodental sounds, especially the fricatives [f v], but also the approximant [ʋ]. The same tendency is observed in native listeners. Further, perception accuracy has been found to vary as a function of phonetic context, prosodic position and developmental stage. Thus, [θ ð] are perceived more accurately as singletons than in clusters. While accuracy rates for singleton [ð] were higher in final position, results for the effect of prosodic position on [θ] are mixed. Finally, perception accuracy for initial [ð] appears to increase with IL development, but no such progress has been documented for [θ].

9.3 Acquisition of dental fricatives: Theoretical predictions

Building on the linguistic factors just outlined, L2 phonological theories generate a range of predictions about dental fricative production by German learners. This section considers the acquisition of /θ ð/ in the light of the theoretical literature.

CAH Both /θ/ and /ð/ are new allophones of new phonemes and the CAH therefore predicts ease of acquisition. From the perspective of the CAH, L1 transfer may motivate /f v s z t d/. The process of final devoicing favors production of voiceless fricatives in coda position.

SLM The literature on German learners' perception of dental fricatives can be condensed into the following assumptions: (i) [v ʋ] are the L1 sounds perceptually closest to [ð] and the degree of perceived similarity between [v ʋ] and [ð] is moderate to high; (ii) [f] is the L1 sound perceptually closest to [θ] and the degree of perceived similarity between [f] and [θ] is high. Accordingly, the SLM predicts initial transfer of L1 [v ʋ] for /ð/ and [f] for /θ/, the latter being expected to be more persistent.

FCM The L1 sounds [f v s z t d] differ from /θ ð/ in three features: continuant (0.55) > CORONAL (0.45) > LABIAL (0.23). As Table 9.3 shows, the FCM predicts [s z] as the predominant perceptual substitutes and expects them to surface in speech production.

Consonant	Rank	[cont]	>	[COR]	>	[LAB]
[s z]	1	−		−		+
[f v]	2	+		+		+
[t d]	3	+		+		−
[θ ð]		+		−		−

Table 9.3: Substitutes for [θ ð]: Predictions of the FCM.

SCH/MDH The following markedness assumptions follow from the literature review: (i) /θ ð/ are marked segments; (ii) /θ/ is relatively more marked than /f s t/; and (iii) /ð/ is relatively more marked than /v z d/. The acquisition of dental fricatives therefore involves mastering a relatively more marked structure, which identifies /θ ð/ as difficult segments. Further, delayed acquisition is expected in codas and in clusters. Patterns that are observable in natural languages include substitution of /θ/ by [f s] and /ð/ by [d] in L1 acquisition, and substitution of /θ/ by [t t̪ f] and /ð/ by [d d̪ v] in varieties of English. A further regularity that may be expected in the production of /ð/ is the assimilation of variants to adjacent segments in connected speech (see Table 9.1).

NM/NDH Our markedness considerations strongly suggest the presence of natural constraints affecting /θ ð/-production. It is important to distinguish between constraints that are grounded in perception vs. production. In fact, tendencies toward *th*-fronting – that is, production as a labiodental fricative – are clearly rooted in speech perception, both in L1 acquisition and in language variation and change (cf. Blevins 2004: 134–135). *Th*-stopping, on the other hand, is more likely to be the result of production constraints. As such, stopping of fricatives is a common, regular sound change that is also observable in L1 acquisition (ibid.: 228–229). We may thus assume that stopping is a natural process that has to be suppressed in the acquisition of /θ ð/. In GB and GA, the process of *th*-stopping is suppressed. Since German does not have dental fricatives, this process is latent for /θ ð/. The NDH therefore predicts that this process will surface in IL, yielding [t d]-like substitutes in learner productions. However, in L1 acquisition stopping rates for /θ/ (.10) are much lower than for /ð/ (.90). This is presumably due to the fact that /ð/-stopping is also motivated perceptually – as noted above, /ð/ usually lacks friction in natural speech (Roach 2009: 45). The NM further considers frequency in the input as a factor in L2 acquisition and the literature review suggests that /θ/ is a rare structure and /ð/ a frequent structure. Frequency may thus be assumed to facilitate the acquisition of /ð/ but not /θ/.

FM The FM predicts the interplay of newly established articulatory and faithfulness constraints to surface in four developmental stages: (i) deviant production, (ii) variation reflecting focus on form, (iii) overly faithful production, (iv) variation sensitive to functional principles. The last stage may be characterized by style-dependent implementation of distinctive features and connected speech processes that arise from an articulatory interaction with adjacent segments. The effect of functional principles is expected to be particularly notable in /ð/-initial function words.

MSA Taking the predictions of the SLM as a basis, the MSA further incorporates physical constraints operating within the talker. These may arise from intrinsic attributes, as is the case in *th*-stopping, or extrinsically – that is, when [θ ð] occur in articulatorily complex sequences. In native speech, this is observable in combination with sibilants. The MSA thus adds to SLM-based predictions a further delay of the acquisition of dental fricatives in general (due to intrinsic constraints) and in articulatorily demanding contexts (as a result of extrinsic constraints).

LTD According to the rationale of the LTD, dental fricatives will first emerge in *s*-marked positions. This suggests the following tendencies: (i) delayed acquisition in unstressed syllables; (ii) delayed acquisition in coda position; (iii) no difference between onset singletons and clusters, as /θ/ is the first segment in onset clusters and thus *s*-marked; and (iv) delayed acquisition of non-initial coda cluster elements. The effect of structural strength on the observed substitutes is predicted to surface in a tendency toward strengthened variants in *s*-marked positions, and weakened variants in *w*-marked positions. It follows that *s*-marking is associated with a tendency toward *th*-stopping and/or devoicing. Weakened variants, on the other hand, may show approximantization – that is, loss of friction. The LTD's view of learner productions showing an increasingly systematic influence of phonetic and prosodic context suggests consistent coarticulation patterns to emerge relatively late.

OPM The discussion above indicates that /θ ð/ classify as both marked and similar structures, suggesting a consideration in terms of the combined corollary[6] shown in Figure 4.18. Thus, perception-driven transfer of similar L1 structures [f v ʋ] is expected to occur at initial stages and persist for L1-derived [f]. This is followed by the influence of universal constraints, which may surface in the tendency toward *th*-stopping in general and toward articulatory simplification in the vicinity of sibilants. Target-like production is expected to emerge late.

Predictions are summarized in Table 9.4. In the next section, we take a look at the insights provided by empirical investigations into /θ ð/ in German learner speech.

[6] See §4.3 (Figure 4.18, p. 194) for a discussion of the rationale underlying the fusion of these corollaries.

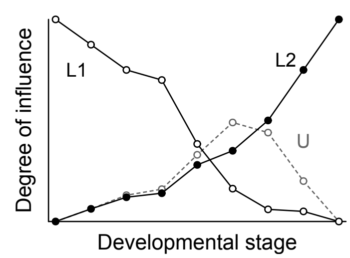


Figure 9.4: Extrapolation of the OPM to structures that are both similar and marked. This graph should be compared to Figure 2.6.

9.4 *Dental fricatives in German Learner English: Previous work*

The production of dental fricatives by L1 German learners has been addressed by several studies. Contrastive accounts with a pedagogical focus differ in the type of substitutes listed for /θ ð/. As Table 9.5 shows, there is consensus that [s] is the preferred reflex for /θ/; however, different variants are reported for /ð/. Further, it has been noted that the combination of /θ ð/ with other fricatives, most notably /s z/, causes difficulties for German learners (Dretzke 1998: 55; Eckert & Barry 2005: 88).

Theory	Predicted patterns in L2 acquisition by German learners
CAH	Early acquisition Substitutes for /ð/: [z v ʋ d] in onset position, [s f t] in coda position Substitutes for /θ/: [s f t]
SLM	Late acquisition by the majority of learners Substitutes: /ð/ [v ʋ]; /θ/ [f] New category formation: /ð/ > /θ/
FCM	Perception-based substitutes in speech production: [s z]
MDH/SCH	Late acquisition Substitutes: /θ/ [s f t t̪] /ð/ [d d̪ v] and coarticulation with adjacent segments Acquisition order: onset > coda; singleton > cluster
NM/NDH	Positive frequency effect for /ð/ Negative frequency effect for /θ/ Substitutes: /θ/ [t t̪]; /ð/ [d d̪]
FM	No effect of frequency Development: deviant production > stylistic variability > overly faithful production > variability reflecting connected speech processes
MSA	Late acquisition by the majority of learners (SLM) Substitutes (cf. SLM): /ð/ [v ʋ] then stopping [d d̪] or coarticulation with adjacent segments; /θ/ [f] then stopping [t t̪] or coarticulation with adjacent segments Late acquisition in contexts involving syntagmatic articulatory difficulty
LTD	*s*-marked positions: Early acquisition Strengthened substitutes: /ð/ [d d̪ t t̪ θ]; /θ/ [t t̪] *w*-marked positions: Delayed acquisition Lenited substitutes: /ð/ [d̞ ʋ]; /θ/ [ð v z d] Acquisition order: stressed > unstressed syllables; onset: singleton = cluster; coda: initial > non-initial cluster element
OPM	Late acquisition Very late acquisition by a minor population of learners (cf. SLM) Late acquisition in contexts involving syntagmatic articulatory difficulty Substitutes (cf. SLM): /ð/ [v ʋ] then [d d̪]; /θ/ [f] then [t t̪]

Table 9.4: The acquisition of /θ ð/ by German learners: Summary of predictions based on theoretical work.

Weiher (1975) investigated the production of dental fricatives by German 5th-graders ($n = 12$) using three tasks: (i) the repetition of nonce words, (ii) the repetition of English words, and (iii) the reading and repetition of English sentences. The findings show an effect of elicitation method. For both /θ/ and /ð/ the correct production rate was lower in nonce words (.47 and .45, respectively) than in the English words (.93 and .82) and English sentences (.83 and .86). Due to the 5th-graders' limited vocabulary, only a small number of English words with dental fricatives

Reference	/θ/	/ð/
Scherer & Wollmann 1986	[s]	[z]
Mair 1995	[s f t]	[d z v]
Arnold & Hansen 1974	[s t]	[z d]
Koziol 1959	[s]	[z d]
Keutsch 1974	[s]	[z]
Kufner 1971	[s f]	[θ d z]

Table 9.5: Substitutes for /θ ð/ in German Learner English: Preferences reported in contrastive analyses.

could be tested. A detailed analysis of the substitution patterns was thus only possible for nonce words. Results were similar to those obtained in the perception tasks: Labiodental fricatives were the most frequent substitutes (.79 for both /θ/ and /ð/). Different patterns were observed for English words, however, where [f] and [s] had roughly equal shares, and also for the English sentences, where .82 of the substitutions were [s] or *s*-like. These findings show that while the imitation of unfamiliar items is subject to perceptual constraints, familiar items are affected by constraints on speech production.[7]

Sieg (2004: quoted in Wode 2009) investigated phonological development in immersion-setting learners from grade 1 to 4. Accuracy rates for both dental fricatives increased steadily (/ð/: .30/.45/.70/.80; /θ/: .70/.80/.80/.85). For /θ/, [s] was the predominant substitute (1/.75/1/.95); marginal error patterns involved [t] (.00/.25/.00/.00) and [f] (.00/.00/.00/.05). The predominant substitute for /ð/ was [d] (.85/.60/.65/.80), followed by [z] (.10/.35/.30/.15).

In another study on instructional-setting learners, Langguth (2009) reported on the production of dental fricatives by German learners from a grammar school in northern Bavaria (Bamberg). The cross-sectional design included three groups ($n = 7$ each) from grade 5 (age: 10–11), 9 (14–15) and 12 (17–18). Three speaking styles were elicited: (i) a word list, (ii) a (near-)minimal pair list, and (iii) semi-free speech – that is, answers to follow-up questions on a short text. The effect of proficiency level was not as pronounced as in other studies. Overall, /θ/ and /ð/ were produced at similar accuracy rates (.80 and .75, respectively). While there was an upward trend for /ð/ (.65/.65/.90), the pattern for /θ/ was U-shaped (.80/.60/.95). The effect of speaking style was rather small with a correct production rate of .75 in free speech and .80 in single words for /θ/ and the same pattern for /ð/ (.72 vs. .78, respectively). Across all subjects, the most frequent substitute for /θ/ was [s], followed by [ð] .21, [f] .17, [t] .07, [d] .07, and [v] .03. Developmental substitution patterns were also observed: [s] occurred most frequently in grade 5, [f] and [t] in grade 9. The distribution of substitutes across proficiency levels was not reported. For /ð/, the predominant substitute was [d] (.48), followed by [s] .33, [θ] .12, and [z] .06. Again, [s] was dominant among younger learners, [d] at later stages. Higher accuracy rates were observed in intervocalic position (.83) than in initial position (.73).

Further quantitative evidence is reported by Wieden & Nemser (1991: 188, 189), who studied the production of /θ ð/ by $n = 384$ Austrian learners

[7] This suggests fundamentally different mechanisms underpinning the imitation of familiar vs. unfamiliar items. An effect of word familiarity on the correct production of [θ ð] was also reported by Nemser (1971b), who studied L1 Hungarian learners' production and perception of [θ ð]. In the perception task, the preferred reflex was a labiodental fricative: [θ] was predominantly perceived as [f] (.82 vs. [s] .08 and [t] .10) and [ð] as [v] (.59 vs. [d] .36 and [z] .04). This is in line with the results for German learners discussed above. In the production task, which involved a translation of familiar words from Hungarian to English, the dominant substitute for [θ] was [t] (.92 vs. [s] .08) and the preferred reflex for [ð] was [d] (.96 vs. [z] .04).

of English.[8] As Figure 9.5 shows, correct production rates for /ð/ varied by proficiency level and prosodic position. Overall, higher accuracy rates were found in initial (*this*, *the*, *there*) and medial position (*mother*) compared to final position (*clothes*, *with*), where devoiced [θ] featured as the most frequent substitute. The rate of target-like realizations in initial and medial position climbed from 0 to .40 in initial position and from .10 to .60 in medial position. For /θ/, the authors reported higher accuracy rates in singletons than in clusters. Thus, production accuracy increased from close to 0 to .40 in clusters and .65 in singletons.

[8] The study included learners at 6 proficiency levels, with n = 64 subjects in each group. Years of instruction was used as an indicator of developmental stage, which ranged from 0 years (grade 3; 8–9 years of age) to 8 years (grade 11; 16–17 years old).

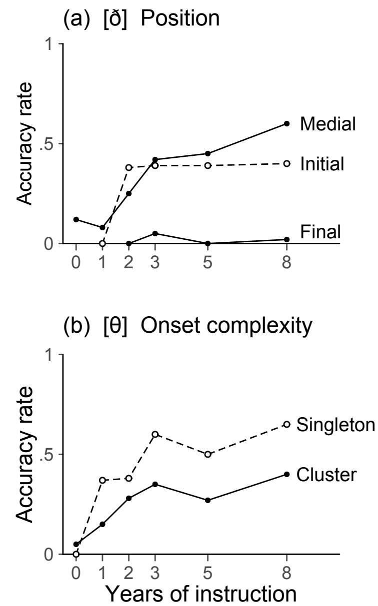


Figure 9.5: Production accuracy of /θ ð/ in Austrian instructional-setting learners by years of learning, ranging from 0 (grade 3) to 8 (grade 11): Correct production rate (a) for /ð/ by position, and (b) for /θ/ by onset complexity. Adapted from Wieden & Nemser (1991: 188, 189) with permission.

Hanulíková & Weber (2010) analyzed the production of word-initial /θ/ by n = 37 German learners (mean age: 23 years). Half of the items were produced correctly. Among the substitutes for /θ/, [s] was observed most frequently (.60), followed by [t] .14, and [f] .11. The authors also analyzed acoustic properties of the word-initial sounds and observed that the [s]-substitute typically differed from intended [s]-realizations in duration, amplitude, and spectral properties. This indicates that, rather than substituting [s] for [θ], German learners produce a transitory sound acoustically intermediate between [s] and [θ]. The label "*s*-type" substitute may therefore be more appropriate, unless an auditory analysis distinguishes between [s] and [s̪].

Pascoe (1987) reports on dental fricative production by n = 26 learners from southern Bavaria (see §3.3.1 for details). Overall accuracy rates were .55 for /ð/ and .68 for /θ/. While subjects with good pronunciation (n = 7) produced .85 of both /θ ð/ correctly, the group of students with poor pronunciation (n = 10) showed lower accuracy rates (.25 for /ð/ and .45 for /θ/).

Further work on dental fricative production by instructional-setting learners from southern Bavaria was carried out by Kucharek (1988), who studied n = 580 learners divided into 5 groups: students in grade 6 (n = 169), 7 (n = 138), 8 (n = 117), and 12 (n = 94) and teachers (n = 62). Production accuracy was assessed with a reading passage, which served to elicit 3 items, one for /ð/ (*father*) and two for /θ/ (*something*, *thirty*). Accuracy rates for /θ/ ranged between .70 and .90, those for /ð/ between .75 and .90. The accuracy of both structures increased steadily from grade 6 to 12; teachers showed a target-like production rate of .97.

In summary, quantitative work on dental fricatives in GLE has identified several patterns of variation, which are summarized in Table 9.6. /θ/-accuracy has been observed to increase with proficiency and years of instruction and appears to be sensitive to word familiarity (Δ = .46), complexity (Δ = 0 to .25), and speaking style (Δ = .05 to .10). The substitutes that have been reported are [s s̪ f t ð d v], with *s*-types occurring most frequently. The choice of variant has been observed to be linked to proficiency level, word familiarity, and speaking style. Varying estimates have also been reported for /ð/-accuracy, which increases with proficiency and years of instruction and appears to be sensitive to word familiarity and position in the word. Results for stylistic variation are inconclusive. /ð/-substitutes include [d d̪ z z̪ v f t] and the type of variant produced has been observed to vary as a function of pronunciation proficiency, word familiarity, and speaking style.

Table 9.6: Empirical evidence from auditory studies on /θ ð/ in German Learner English: reported production/accuracy rates and constraints on accuracy and substitutes.

Results		Notes	Study
/θ/			
Accuracy rate	.80–.95	Grades 5, 9, 12; age: 10–18 years	Langguth 2009
	.70–.90	Grades 6–12; age: 11–18 years	Kucharek 1988
	.70–.85	Grades 1–4; age: 6–10 years	Sieg 2004
	.47–.93	Grade 5; age: 10–11 years; lowest rate: nonce words	Weiher 1975
	.45–.80	Grades 8–10; age: 14–16 years	Pascoe 1987
	.00–.65	Grades 3–11; age: 8–17 years	Wieden & Nemser 1991
Constraints on accuracy			
Familiarity	Δ = .46	Familiar (.93) > nonce words (.47); imitation task	Weiher 1975
Complexity	Δ = [0; .25]	Singleton [0; .65] > cluster [0; .40]; effect greater at later stages	Wieden & Nemser 1991
Speaking style	Δ = .10	Single words (.93) > sentences (.83)	Weiher 1975
	Δ = .06	Single words (.78) > free speech (.72)	Langguth 2009
Constraints on substitutes			
Familiarity		[f v] most common in nonce words (imitation task)	Weiher 1975
		[f s] in familiar English words	Weiher 1975
Speaking style		[f s] in single words	Weiher 1975
		[s] in English sentences	Weiher 1975
Proficiency		*s*-type substitutes more frequent at early stages	Langguth 2009
		[f t] more frequent at later stages	Langguth 2009
/ð/			
Accuracy rate	.75–.90	Grades 6–12; age: 11–18 years	Kucharek 1988
	.65–.90	Grades 5, 9, 12; age: 10–18 years	Langguth 2009
	.30–.80	Grades 1–4; age: 6–10 years	Sieg 2004
	.45–.86	Grade 5; age: 10–11 years; lowest rate: nonce words	Weiher 1975
	.25–.80	Grades 8–10; age: 14–16 years	Pascoe 1987
	.00–.60	Grades 3–11; age: 8–17 years	Wieden & Nemser 1991
Constraints on accuracy			
Familiarity	Δ = .37	Familiar (.82) > nonce words (.45); imitation task	Weiher 1975
Position	Δ = [0; .50]	Medial, initial [0; .50] > final [0; .05], effect greater at later stages	Wieden & Nemser 1991
	Δ = [0; .20]	Medial [.10; .60] > initial [0; .40], effect greater at later stages	Wieden & Nemser 1991
	Δ = .10	Medial (.83) > initial (.73)	Langguth 2009
Speaking style	Δ = .05	Single words (.80) > free speech (.75)	Langguth 2009
	Δ = –.04	Single words (.82) < sentences (.86)	Weiher 1975
Constraints on substitutes			
Familiarity		[f v] most common in nonce words (imitation task)	Weiher 1975
		[f s] in familiar English words	Weiher 1975
Speaking style		[f s] in single words	Weiher 1975
		[s] in English sentences	Weiher 1975
Proficiency		*s*-type substitutes more frequent at early stages	Langguth 2009
		d-type substitutes more frequent at later stages	Langguth 2009

9.5 *Aims of this study*

This investigation aims to make empirical contributions to the study of dental fricatives in GLE by focusing on the following aspects:

- cross-sectional patterns in the realization of onset /ð/ and
- linguistic factors constraining the realization of onset /ð/.

Cross-sectional patterns in the realization of onset /ð/

Most theoretical contributions arrive at a classification of dental fricatives as difficult structures (SLM, MDH/SCH, MSA, OPM); only the CAH predicts ease of acquisition. As for expected substitutes, a wide range of variants appears plausible from a theoretical viewpoint. Some frameworks make predictions conditional on phonetic and prosodic context. Thus, the LTD considers dental fricative substitution to be sensitive to structural strength, and the MSA and SCH suggest coarticulation effects induced by adjacent segments. The following set of substitutes for onset /ð/ may be postulated on the basis of theoretical work:

- [v ʋ] CAH; SLM; SCH; MSA; OPM; LTD: [ʋ] in *w*-marked positions
- [z] CAH; FCM; MSA: after [s z] MSA
- [d] CAH; SCH; NM; OPM; LTD: in *s*-marked positions
- [d d̪] SCH; NM; OPM; MSA: after [d]; LTD: in *s*-marked positions
- [n n̪] MSA, SCH: after [n]
- [l] MSA, SCH: after [l]
- [t t̪ θ] LTD: in *s*-marked positions
- [d̥] LTD: in *w*-marked positions

There is some evidence in the literature suggesting that substitution patterns vary with IL development. Thus, Langguth (2009) reported a tendency toward *s*-types at earlier stages, followed by a tendency toward *d*-types. The current study shall throw further light on systematic variation in the realization of onset /ð/ at different proficiency levels.

Linguistic factors

The effect of phonetic context on the production of onset /ð/ has been approached differently in the literature, as studies have foregrounded different phonetic features in the surrounding context.[9] Quite surprisingly, place of articulation has so far been largely neglected. Given what is known about /ð/-variation in native speech (cf. §9.1), however, this feature seems to be clearly motivated. In non-native speech, [s z] have been observed to be a source of difficulty (e.g. Wenk 1979), and it has been pointed out that this is also the case in GLE (Dretzke 1998: 55; Eckert & Barry 2005: 88). The aim of the current study is to shed light on phonetic context effects by considering how onset /ð/ production varies after [s z] compared to other sounds.

[9] For instance, researchers have turned to the sonority of adjacent segments (Gatbonton 1978), manner of articulation, voicing, perceived cross-linguistic similarity and lexical frequency (Trovimovich et al. 2007).

9.6 Method and data

Informants, materials and procedures

The following analyses are based on the data gathered in this study and two additional data sets:[10]

- The *SpIL* corpus (Pascoe 1987) includes *n* = 304 instances of onset /ð/, with the definite article *the* accounting for the majority of occurrences (77%). Demonstrative pronouns and determiners (*this, that, these, those*) made up 11% of the tokens.
- Wunder's (2012) data contain *n* = 863 relevant tokens (48% *the*, 17% demonstratives, 17% personal pronouns).
- The recordings from this study yielded *n* = 1,531 instances of onset /ð/ (74% *the*, 4% demonstratives).

[10] The data sets analyzed in the present chapter are openly available via the *Tromsø Repository of Language and Linguistics* (*TROLLing*) at https://doi.org/10.18710/DYAGZG (Sönning et al. 2020c). See §3.3.1 for further information on the supplementary data sets.

For the analysis of Pascoe's (1987) data, we relied on the phonetic transcriptions provided in the *SpIL* corpus. The data from Wunder (2012) and this study were analyzed auditorily by the author of the current study. Given the fact that /ð/ is typically produced as a weak dental stop in native speech (Roach 2009), stop-like renditions [d̪] were classified as target-like and assigned to the same category as [ð]. Table 9.7 shows the distribution of variants in the three data sets. A distinction is made between *observed variants* and *categories used for analysis*. The latter aggregates variants into 4 major groups, reflecting target-like production [ð d̪], *d*-type substitutes [d d̥], *z*-type substitutes [z z̪ s s̪][11] and other variants [θ f v] and ∅. These four groups were used for the statistical analysis of the data.

[11] In the auditory classification of the tokens, some features were difficult to pin down – especially the distinction between the *z*-type fricatives [z̪ s̪] and dental [ð]. I based the classification on the perceived frequency of the noise component, which, of course, is a continuous attribute. The distinction between target-like dental [d̪] and alveolar [d] was also subtle in some cases, but I felt that, in most cases, the burst of noise allowed me to identify the place of articulation with a satisfactory level of confidence.

Table 9.7: Distribution of variants in the three data sets. "Pr" denotes the proportional share of a (group of) variant(s).

	Pascoe 1987		Wunder 2012		This study	
Realization	Pr	(*n*)	Pr	(*n*)	Pr	(*n*)
Observed variants						
[ð d̪]	.56	(169)	.69	(597)	.65	(990)
[d]	.24	(73)	.17	(148)	.18	(281)
[z z̪]	.20	(62)	.01	(10)	.02	(26)
[s s̪]	.00	(0)	.07	(57)	.04	(60)
[f v]	.00	(0)	.02	(21)	.02	(23)
[θ]	.00	(0)	.02	(17)	.08	(120)
[d̥]	.00	(0)	.00	(0)	.01	(12)
∅	.00	(0)	.02	(13)	.01	(18)
Categories used for analysis						
[ð d̪]	.56	(169)	.70	(990)	.69	(597)
d-type [d d̥]	.24	(73)	.21	(294)	.17	(148)
z-type [z z̪ s s̪]	.20	(62)	.05	(76)	.08	(67)
other [θ f v] ∅	.00	(0)	.03	(49)	.06	(51)

As for the linguistic factor, Table 9.8 shows how preceding [s z] and other contexts are represented in the data sets. While the token frequency gives the total number of observations per factor level, the type frequency indicates the number of different words with onset /ð/ that occurred in each condition. Next, we outline key features of the statistical analysis. Readers less interested in technical details may safely skip ahead to §9.7.

	Pascoe 1987		Wunder 2012		This study	
Condition	Types	Tokens	Types	Tokens	Types	Tokens
Other	9	279	7	746	6	1364
Sibilant	4	25	3	121	1	62

Table 9.8: Distribution of types and tokens across the levels of the linguistic factors in the three data sets.

Statistical analysis

Since the realization of onset /ð/ was coded using four categories, a multilevel multinomial regression model was fit to the data. The structure of the model is defined in Appendix A.9. To account for the hierarchical data structure, and following Barr et al.'s (2013) recommendations, the model included random intercepts on subject and word, as well as by-word random slopes for proficiency, and by-subject random slopes for phonetic context. Weakly regularizing priors were specified for all predictor slopes, as well as the secondary parameters, that is, all correlation and standard deviation parameters that define the variational structure. For the intercept, the information from the literature review was used to define diffuse informative priors. These denote the a priori expectation about the rate of each variant in non-[s z] contexts, at an intermediate proficiency level ($z = 0$). The connection between these these distributions and the values reported in the literature is shown in Figure 9.6. These priors assign 90% probability to the following intervals: [.15, .88] for [ð], [.03, .69] for *d*-types, [.00, .44] for *z*-types, and [.00, .26] for other variants.

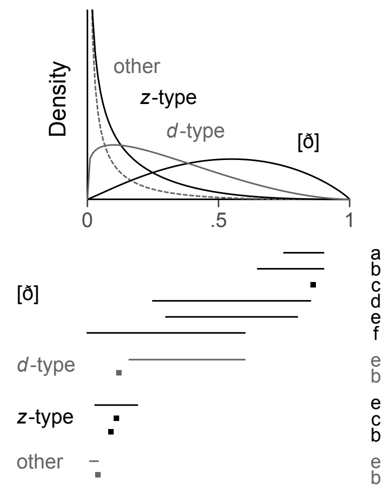


Figure 9.6: Prior for the intercept parameters in the model in comparison to the rates reported in the literature: (a) Kucharek 1988, (b) Langguth 2009, (c) Weiher 1975, (d) Pascoe 1987, (e) Sieg 2004, (f) Wieden & Nemser 1991.

9.7 Constraints on generality

Before we proceed to the results of our investigation, let us delimit the circumstances to which these apply.

- *Language-external scope*: We are not aware of any speaker-related factors that would compromise the applicability of our findings to other populations of German instructional-setting learners.
- *Language-internal scope*: In the materials used for data elicitation, onset /ð/ almost exclusively occurred in function words (at least 80% of the tokens in all data sets), with *the* accounting for the largest share. While this approximates distributional patterns in natural language use, the findings presented in the following may therefore only apply to these contexts, that is, unstressed function words.
- *Speaking style*: The findings based on the present study and Wunder (2012) reflect reading style. Since dental fricatives are unambiguously represented by <th>, orthography may have a facilitating effect in these data sets, suggesting that the rate of target-like variants may be higher than in free speech.

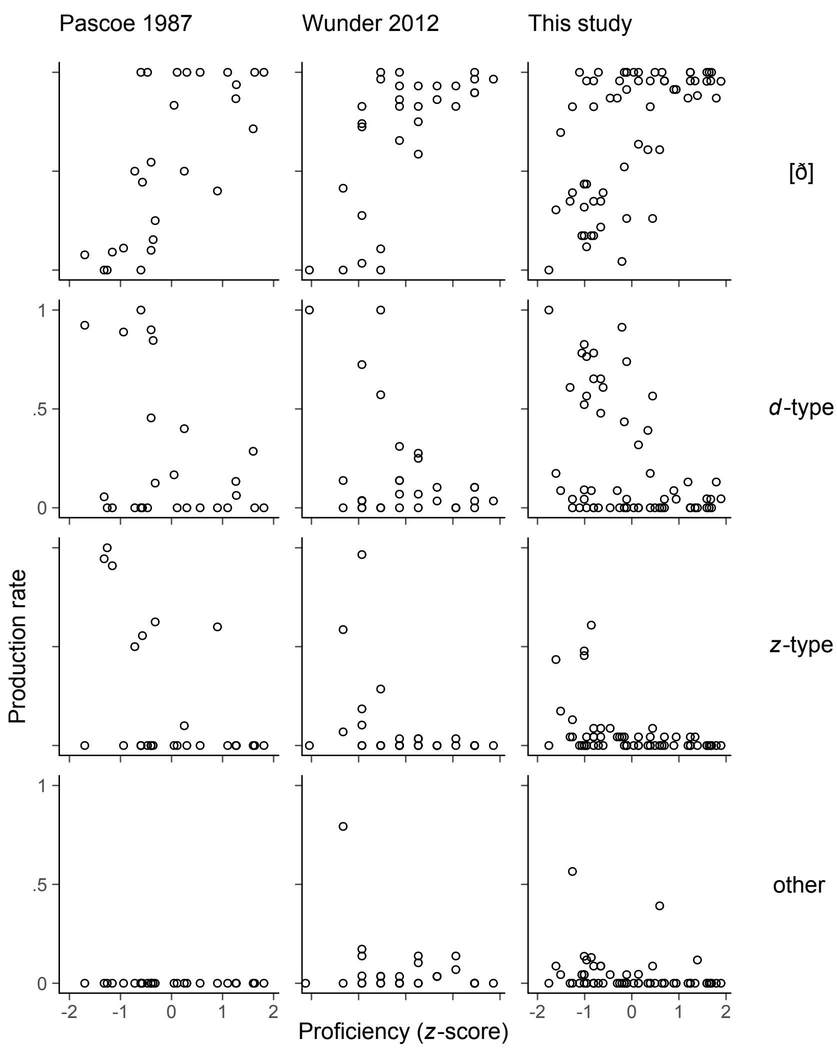


Figure 9.7: Onset /ð/: Production rates for [ð], *d*-types, *z*-types, and other variants by proficiency level. The graphs shows, for each learner, the proportion of the variants.

9.8 Results

We begin with a descriptive look at the data. The distribution of the four variants across different proficiency levels is shown in Figure 9.7, where each point denotes, for a particular learner, the proportion of a given variant among all onset /ð/ productions. Overall, the patterns are in good agreement. While the share of target [ð] increases with pronunciation ability, at lower and intermediate levels we observe substantial variation among learners. It is only at advanced stages that learners, on average, begin to approach a categorically target-like realization rate. Accordingly, the proportion of non-target variants shows decreasing trends. In general, the rate of *d*-types [d d̪] appears to be higher than that of *z*-types [z z̪ s s̪], at least in the data from Wunder (2012) and the present study. Other variants appear to play a minor role and there are only few individuals who show an appreciable share.

For technical details about the regression models, please refer to Ap-

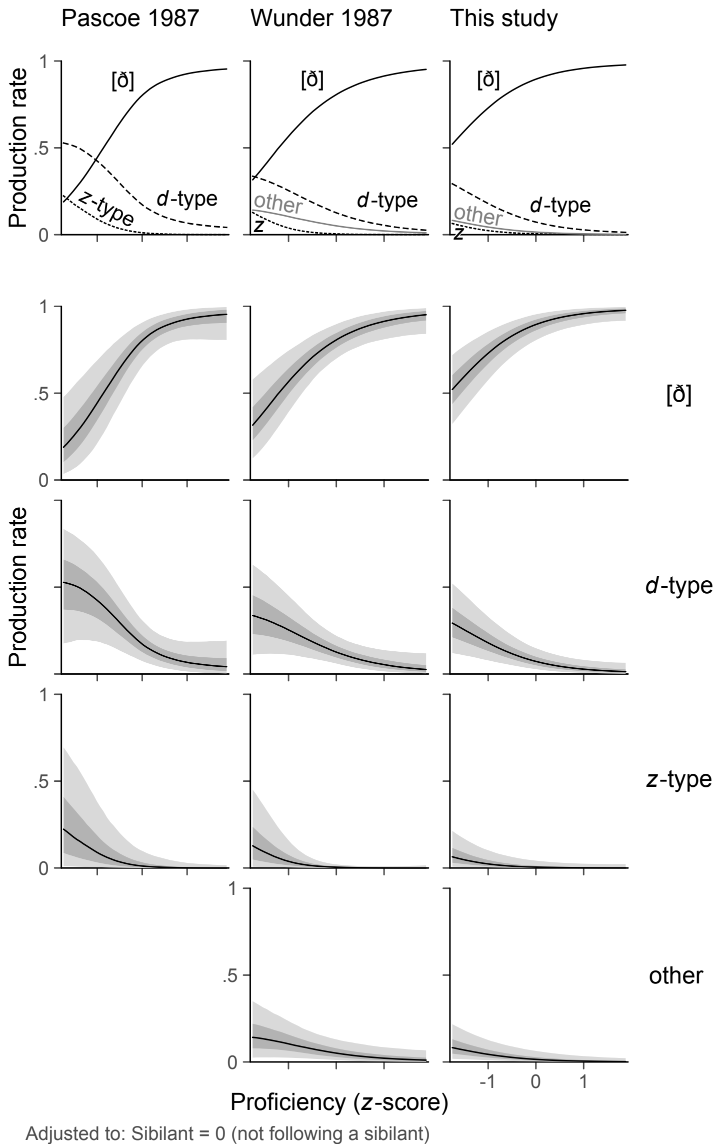


Figure 9.8: Realization of onset /ð/ in phonetic contexts other than after [s z]: Production rate of variants across proficiency levels, in the three data sets. Lines show estimated regression curves. The top row compares the proportion of variants and the rows below show the statistical uncertainty associated with the estimates. Error bands reflect 50% and 90% uncertainty intervals.

pendix A.9. We will immediately turn to a visualization of the estimates offered by our analyses. Figure 9.8 shows the average rate of each variant across proficiency levels. Corresponding estimates for representative proficiency levels are listed in Table 9.9. The figures describe the realization of onset /ð/ in contexts other than after [s z]. The average trends may therefore deviate from those in the descriptive diagrams above, where realization rates are pooled across phonetic contexts. The top row compares the four variants, and in the rows below, the statistical uncertainty surrounding these estimates is indicated. Figure 9.8 offers the following insights:

- In contexts other than after [s z], accuracy rates for onset /ð/ increase rather quickly and approach 1 at intermediate levels.
- In these contexts, the rate of *d*-type substitutes seems to be higher, on

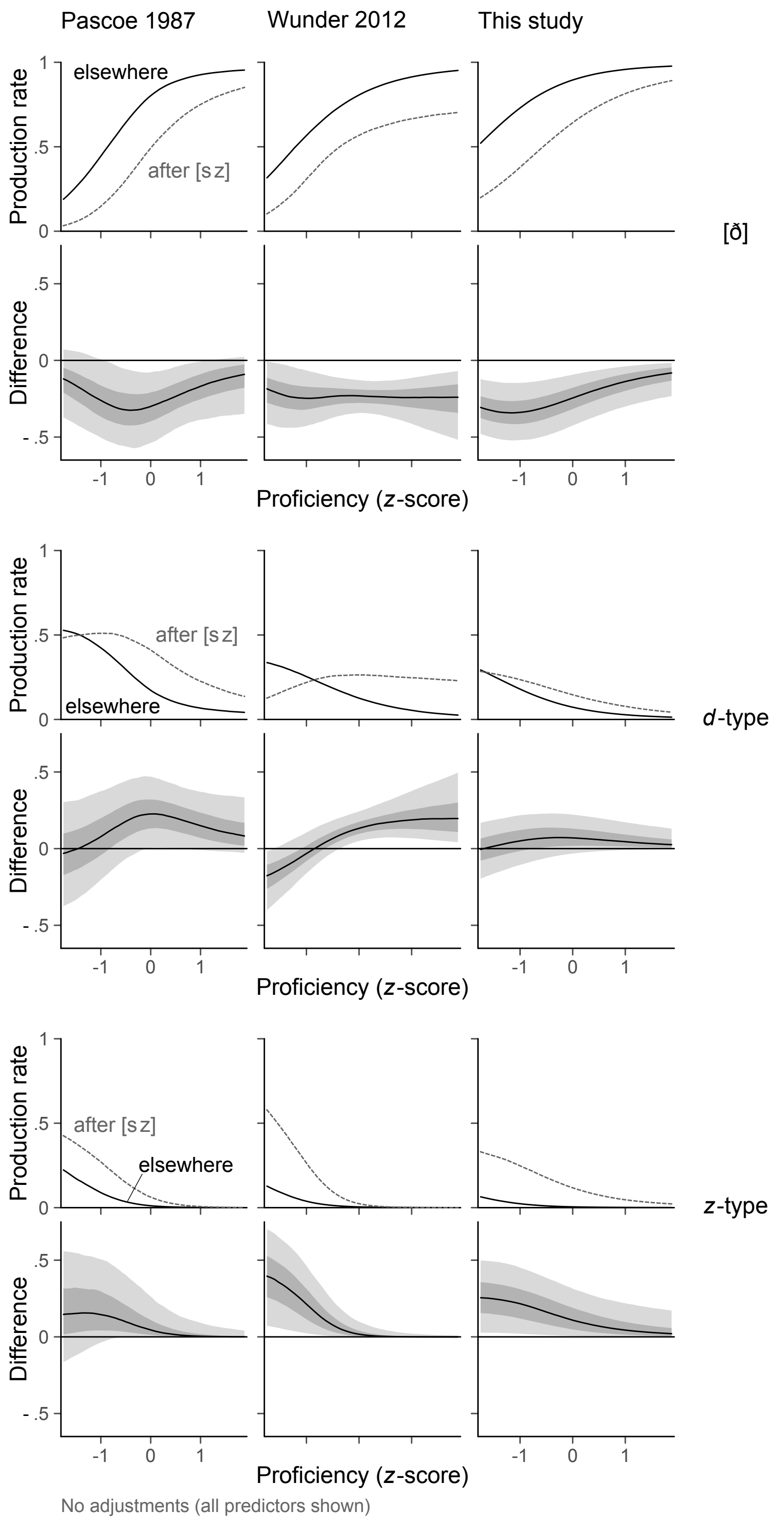


Figure 9.9: Realization of onset /ð/ conditional on the preceding segment. The figures show how the rate of [ð], *d*-types, and *z*-types varies after [s z] (grey dashed lines) compared to other contexts. For each variant, the first row of panels shows the estimated rate, and the second row of panels shows the estimated difference in the rate of this variant. Values below the curve indicate a lower rate after [s z]. Error bands indicate 50% and 90% uncertainty intervals.

average, than that of *z*-types and other variants. However, the statistical uncertainty surrounding these estimates is quite large.

- The rate of *d*-type substitutes decreases, on average from about .50 (Pascoe 1987), .30 (Wunder 2012) and .25 (this study) at the lowest levels of pronunciation ability, to below .15 at intermediate levels.
- The rate of *z*-types drops from beginner rates of roughly .20 (Pascoe 1987), .10 (Wunder 2012) and .05 (this study), to below .01 at intermediate levels.

Findings for the preceding phonetic context are summarized in Figure 9.9, where estimates are shown for three variants: [ð] (rows 1 and 2), *d*-types (rows 3 and 4), and *z*-types (rows 5 and 6). The upper panels in each pair of rows shows two estimated curves for a variant: the average rate after [s z] (grey dashed line) and in other contexts (solid black line). Note that the solid black lines are those shown in Figure 9.8 above. The question of interest here is how the distribution of variants changes when onset /ð/ occurs after [s z]. The lower panels in each pair of rows shows the difference between the two curves. If the difference is below the horizontal line marking zero, this means that the share of the variant is lower after [s z], if it is above the zero line, its prevalence is higher in these contexts. The error bands mark the 50% and 90% uncertainty intervals for these difference estimates. The following observations may be made:

- The rate of target-like [ð] is lower after [s z]. The magnitude of this drop varies across proficiency levels and seems to be highest, in absolute terms, at beginner and lower intermediate levels, where it amounts to about $\Delta = .15$ to $\Delta = .30$. These estimates are subject to a fair amount of statistical uncertainty.
- The rate of *d*-type substitutes shows a less regular pattern: Among beginners, these variants appear to be equally (or somewhat less) likely after [s z] compared to other contexts. While this pattern is broadly consistent across the three data sets, the magnitude of these differences varies: In general, *d*-types appear to be about .03 to .20 more likely after [s z] among intermediate and advanced learners.
- As for *z*-type variants, we observe consistent trends: After [s z], *z*-types are more likely, especially among beginners, where the difference amounts to .14 to .40, in absolute terms.

9.9 Summary and discussion

This chapter was concerned with the acquisition of onset /ð/ by German learners. A contrastive analysis showed that dental fricatives are new sounds for German learners, whose L1 lacks a dental place of articulation. /ð/ is common in English and occurs in initial position in a number of high-frequency function words, where it is subject to connected speech processes including deletion and assimilation to the preceding segment. From the perspective of linguistic constraints, dental fricatives must be considered marked structures, as they are rare in languages around the world and among the last sounds acquired and mastered by English-learning

Table 9.9: Estimates for the production rate of variants, in phonetic contexts other than after [s z], at representative proficiency levels.

	[ð]			*d*-types			*z*-types			other		
Proficiency	*Mdn*	.05	.95	*Mdn*	.05	.95	*Mdn*	.05	.95	*Mdn*	.05	.95
Other contexts												
Pascoe 1987												
Beginner	.25	.06	.54	.51	.19	.80	.18	.01	.61	—	—	—
Intermediate	.81	.67	.91	.17	.07	.30	.01	.00	.10	—	—	—
Advanced	.95	.81	.99	.05	.01	.19	.00	.00	.02	—	—	—
Wunder 2012												
Beginner	.39	.19	.63	.32	.12	.58	.10	.01	.36	.13	.03	.31
Intermediate	.81	.70	.90	.12	.06	.22	.00	.00	.02	.05	.01	.13
Advanced	.94	.83	.98	.03	.01	.12	.00	.00	.01	.01	.00	.07
This study												
Beginner	.59	.40	.77	.26	.11	.47	.05	.01	.17	.07	.02	.19
Intermediate	.90	.81	.96	.07	.03	.15	.00	.00	.04	.01	.00	.06
Advanced	.97	.91	.99	.02	.00	.07	.00	.00	.02	.00	.00	.03
After [s z]												
Pascoe 1987												
Beginner	.05	.00	.38	.50	.10	.88	.39	.03	.86	—	—	—
Intermediate	.50	.24	.76	.41	.13	.67	.06	.00	.31	—	—	—
Advanced	.82	.53	.97	.17	.01	.46	.00	.00	.07	—	—	—
Wunder 2012												
Beginner	.15	.03	.42	.15	.03	.44	.49	.10	.84	.13	.02	.39
Intermediate	.57	.42	.73	.26	.14	.42	.02	.00	.11	.11	.03	.25
Advanced	.69	.41	.88	.24	.07	.54	.00	.00	.02	.05	.00	.19
This study												
Beginner	.25	.09	.48	.27	.07	.52	.31	.05	.61	.12	.02	.34
Intermediate	.65	.45	.83	.14	.03	.32	.11	.01	.34	.05	.01	.17
Advanced	.87	.68	.96	.05	.00	.19	.03	.00	.21	.01	.00	.08

children. This is partly attributable to contiguous motor constraints, as the production of /θ ð/ in combination with sibilants requires fine motor differentiation. Another relevant factor is their lack of perceptual salience. Both native and non-native listeners frequently mishear dental [θ ð] as labiodental [f v] and perceptual difficulties have also been identified in the population of L1 German listeners.

Resting on these insights into linguistic influences on the acquisition of /θ ð/, most theories arrive at a classification of dental fricatives as difficult segments in L2 acquisition. As for their realization in GLE, a large set of variants is motivated from the viewpoint of L2 speech theories. These include L1-derived segments, assimilations resulting from regressive coarticulation, and strengthened or weakened variants.

Claims of dental fricatives being difficult structures have received support from quantitative research on GLE, with accuracy rates ranging between 0 and .90. Several factors have been linked to variation in accuracy rates, including proficiency level, word familiarity, and position in the syllable and word. As yet, there has been no systematic investigation of phonetic context effects on the production of /ð/ by German learners. The objective of the present study was to explore cross-sectional patterns in the choice of variant for onset /ð/. A study by Langguth (2009) reported on differential substitution conditional on developmental stage, with a stronger tendency toward *s*-type substitutes at earlier stages. The current investigation further set out to shed light on the effect of preceding phonetic context. In this domain, empirical work on GLE is lacking.

Findings on the accuracy rate of onset /ð/ were consistent with evidence in the literature, with estimates of around .25 to .60 at lower proficiency levels and close to 1 at higher proficiency levels. With high-proficiency learners showing almost categorical target-like production, onset /ð/ may be considered a structure of moderate difficulty. As for deviant realizations, *d*-type and *z*-type substitutes accounted for the greatest share of non-target variants. *D*-types consistently occurred at a higher rate across all proficiency levels. A closer inspection of the association between preceding sibilant and the distribution of variants revealed that a drop in target variants after sibilants (between .15 and .30 in absolute terms) was compensated by a rise in *d*-type and *z*-type realizations. While a sibilant-induced boost in *z*-types appeared to be restricted to the lower proficiency levels, there was a tendency for *d*-type substitutes to show a sibilance-related increase at intermediate and advanced stages.

While most theoretical perspectives concur in predicting difficulties for the acquisition of onset /ð/ by German learners, these claims rest on different assumptions about underlying processes. These include the relative markedness of /θ ð/ (MDH), the operation of natural processes (NM/NDH), syntagmatic motor complexity (MSA) and speech perception (SLM; MSA; OPM). Similar to earlier work on dental fricatives in GLE, labiodental substitutes were rarely observed. This runs counter to predictions made by the SLM and models that incorporate its tenets (MSA, OPM). Concerning the dispreference for labiodental variants, three factors may play a role:

- Dental fricatives have a robust orthographic representation, which provides no graphemic cue to a labiodental place of articulation. Rather, it points to an alveolar constriction location, the place of articulation most frequently observed in substitutes.
- Speech perception has been observed to be closely related to visual information, a phenomenon that is known as the McGurk effect (McGurk & MacDonald 1976). In the case of dental fricatives, the external articulators provide relevant visual information in face-to-face settings. Observation of the lips reveals that /θ ð/ are not labiodental, but involve internal articulators.
- A third factor that may play a role is markedness, with observed substitution patterns reflecting the emergence of unmarked feature values. Thus, Rice (2007) summarizes arguments for the coronal place of articulation to be considered unmarked in comparison to labial and dorsal consonants. Coronal substitutes may thus be generally favored as substitutes for dental fricatives.

A comparison of substitution patterns by learners from different L1 backgrounds certainly shows that a preference for dorsal substitutes is not categorical. Nevertheless, these factors may partly account for the predominance of dorsal substitutes in German Learner English despite the fact that labiodental fricatives are perceptually much closer. It is worth noting that, contrary to the SLM, the FCM predicted *s*/*z*-type substitutes on the grounds that /θ ð/ and /s z/ share the features continuant and coronal. It should be noted, however, that predictions of the FCM primarily concern speech perception and thus relate to speech production only indirectly. The expectation of *s*/*z*-type substitutes in the perception of /θ ð/ is at odds with the established empirical observation of /f v/ being the preferential substitutes in speech perception. Thus, while the FCM correctly predicts *s*/*z*-type substitutes in production, the fact that this prediction is not borne out in speech perception interferes with its line of reasoning. It can thus be stated that the processes assumed by the FCM do not provide an adequate account of the empirical observations, at least as far as the set of phonological features employed in this study is concerned.

While the factors just outlined may contribute to an explanation of the preference for dorsal variants, they do not differentiate between stops (*d*-type) and fricatives (*z*-type). Nevertheless, the observed tendency toward *d*-types is consistent with the following theory-derived claims:

- MDH: Fricatives are relatively more marked than stops. A higher rate of *d*-type substitutes may therefore reflect a preference for unmarked segments.
- SCH: Turning to substitution patterns observable in other natural languages, the SCH predicted a greater likelihood of *d*-type substitutes. German learner speech thus seems to structurally conform to the universal tendencies observed in other natural languages.
- NM/NDH: The bias toward *d*-types may result from the universal phonological process of stopping, which is also observed in L1 acquisition.
- LTD: The emergence of stop-like variants is also explainable in terms of onset strengthening – that is, the tendency toward fortition of *s*-marked segments.

Thus, when accounting for the observed patterns in the realization of onset /ð/ by German learners, proper consideration ought to be given to alternative views. This concerns (i) the clarification of the empirically substantiated dispreference for perceptually motivated substitutes and (ii) the explanation of the distribution of major variants. For the present, empirical patterns lend themselves to multiple explanations.

Findings on the distribution of deviant variants are in agreement with the studies by Sieg (2004: quoted in Wode 2009) and Langguth (2009), who also noted a preference for *d*-type substitutes. To the extent that these empirical findings can be considered representative of /ð/-production by German learners, it appears that lists of substitutes reported in contrastive analyses (see 9.5) may need to be modified. Only the order given by Mair (1995) reflects patterns that have been documented for GLE. Nevertheless, judgement on the descriptive adequacy of any set of preferential substitutes must be suspended, as no study to date has provided an extensive survey of /ð/-production by German learners. Previous research is suggestive of differential substitution patterns (Wieden & Nemser 1991; Langguth 2009), which are also motivated on theoretical grounds. In particular, two distinctions deserve closer attention in future work: prosodic position (onset vs. coda) and, for onset /ð/, position in the word. The present study was restricted in scope to onset position. With coda obstruents being regularly devoiced in L1 German (see Chapter 10), the substitution patterns in final position are governed by different processes yielding a different set of variants. As for onset /ð/, the few indications in the literature suggest further variation conditional on position in the word. As far as accuracy rates are concerned, onset /ð/ in medial, intervocalic position appears to be more favorable to target-like renditions ($\Delta = 0$ to .20, Wieden & Nemser 1991; $\Delta = .10$, Langguth 2009). A separate analysis of initial and medial position appears warranted for two reasons. First, for intervocalic /ð/-tokens (*father*, *weather*), the phonetic context is invariable. Thus, word-medial intervocalic position is a protected environment in the sense that variation conditional on connected speech processes of the type attested for initial /ð/ is blocked. For onset /ð/, we therefore expect a statistical interaction between speaking style and position in the word. This is to say that only word-initial /ð/ is likely to be sensitive to connected speech processes and is thus expected to be realized differently in isolation vs. connected speech. Second, another factor relevant for the description and prediction of substitution patterns is the favorability of intervocalic segments to lenition processes (Lass 1984: 182). This is expected to surface in a smaller share of *d*-type substitutes, which may be considered the outcome of *th*-stopping, a fortition process.

Given the variability of /ð/ in place, manner and voicing features, one challenge involved in the investigation of this structure in IL phonology is the interpretation of variants that deviate from a canonical voiced dental fricative. As highlighted by the FM, two sources of within-learner variation must be distinguished: that induced by monitoring effects and that related to functional principles. Thus, variants that deviate from [ð] in place, manner, or voicing features may be the outcome of different mechanisms – that is, deviations may be due to incomplete acquisition or reflect connected speech processes that are also observed in native speech. The same surface realization may therefore be interpreted in fundamentally different ways depending on which type of variation it is assumed to represent. For instance, a German learner's production of *Is this it* as [ɪˈzɪˈsɪt] may reflect a transfer-induced deviation from target /ð/ or a target-like connected speech process (see §9.1). Guidance provided by theory suggests that interpretative difficulties may be partly resolved by controlling for speaking style and proficiency level.

To conclude, the following points may be considered by future studies on dental fricative production by German learners:

- Variation in /ð/-production by German learners is subject to different constraints, which in turn vary by prosodic and phonetic context. Empirical documentation should therefore (minimally) distinguish between three contexts: onset, coda, and word-medial (intervocalic) position.

- Studies investigating sources of /ð/-variation should consider sampling from lower proficiency levels, as variation in target-like production has been observed to be highest in this population of learners.
- Researchers should be aware of issues involved in the interpretation of IL variants that deviate from /ð/ in place, manner, or voicing features, as these may be the outcome of different mechanisms and therefore represent different structural IL components.
- Given that adjacency effects play a role in /ð/-production by native and non-native speakers, studies should not fail to record information about phonetic context. As different attributes of contiguous segments may be of interest in future work, contextual information is best provided in raw form (i.e. using IPA symbols).
- From an explanatory perspective, two questions remain open for future work: Why do German learners favor dorsal over labiodental variants? Why do learners prefer *d*-type over *z*-type substitutes?

10
Final voiced obstruents

This chapter closes our treatment of segmental structures in German Learner English with an analysis of the consonantal feature [±voice]. As such, the phonetic description of consonants rests on three parameters: place of articulation, manner of articulation, and voicing. In the obstruent inventories of English and German, voicing has a meaning-distinguishing function, contrasting pairs such as /t d/, /s z/, and /p b/. The phonetic realization of the phonological feature [±voice] varies across languages and different articulatory and acoustic features may be employed to cue obstruent voicing contrasts. Variation is also observable language-internally, conditional on prosodic context. For instance, the implementation of [±voice] differs systematically in onset vs. coda position. In fact, the labels *voiced* and *voiceless* may not be adequate from a phonetic point of view, as vocal fold vibration – that is, phonetic voicing – need not be the primary distinctive cue. Different terms have therefore been suggested for the phonological feature that distinguishes pairs such as *back* vs. *pack* or *ice* vs. *eyes*. Jakobson & Halle (1956) use the terms *tense* and *lax*, focusing on the muscular effort involved; by the same token, Kohler (1984) suggested the labels *fortis* and *lenis*. In the following, the terms *voiced* and *voiceless* will be used in a phonological sense, with the understanding that these labels need not correspond to what is observed at the phonetic level.

This chapter is concerned with the production of final voiced obstruents (FVOs) by German learners. Following a discussion of contrasts between English and German, §10.2 turns to inherent properties of FVOs that constrain their L2 acquisition. Theoretical positions on the L2 acquisition of final obstruent voicing contrasts are presented in §10.3, which derives hypotheses from formalized accounts. A survey of previous empirical work on FVOs in German learner speech (§10.4) is followed by an outline of the aims (§10.5) and method (§10.6) of this study. After a consideration of the constraints on generality (§10.7), results are presented in §10.8. §10.9 then closes with a summary and discussion.

10.1 Contrastive analysis

While both English and German use [±voice] as a contrastive feature, differences emerge in terms of its distribution. Thus, the two languages signal obstruent voicing contrasts in syllable onsets (e.g. *Pass* - *Bass*

/pas/-/bas/; *Seite - Seide* /ˈzaɪ.də/ - /ˈzaɪ.tə/). In German, however, the contrast is neutralized in syllable-final position (e.g. *Rad - Rat* [ʁaːt]). This regular process is referred to as *final devoicing* and affects obstruents in coda singletons and clusters. As a consequence, voiced obstruents are only licensed in onset position. The fact that final devoicing neutralizes an underlying voicing opposition is apparent from the occurrence of regular morphophonemic alternations: Upon suffixation, coda obstruents may be resyllabified to the onset of a vowel-initial suffix and consequently reflect the underlying voiced phoneme (e.g. *Rad* [ʁaːt], *Rades* [ˈʁaːdəs] and *Lied* [liːt], *Lieder* [ˈliːdɐ]). In contrast to German, English upholds a contrast in all positions, yielding minimal pairs such as *bed* [bed] vs. *bet* [bet].

Acoustically, obstruent voicing is signaled by a combination of cues. These may be *intrinsic* (i.e. found in the consonantal segment itself[1]) or *extrinsic* (i.e. surfacing in adjacent segments, most notably vowels[2]). In English, these acoustic parameters differ in reliability and weight. In syllable-final position, where voiced stops and fricatives are partially or completely devoiced phonetically, intrinsic features are backgrounded. A robust and reliable cue to the voiced-voiceless distinction in codas is the duration of the preceding vowel, which is longer before voiced obstruents. The perceptual relevance of this signal is well-documented. In fact, it has relatively strong weight and may even serve as a sufficient cue to the perceptual categorization of coda obstruents (Denes 1955; Raphael 1972; Crowther & Mann 1992; Broersma 2005).

The duration of the preceding vowel as an extrinsic voicing cue is attested across languages and has received different labels in the literature.[3] Here, the term *preceding vowel duration* (Tauberer 2010) will be used (hereafter PVD). The degree to which vowel duration varies as a function of the following obstruent will be referred to as the *PVD effect*.

While it has been suggested that the PVD effect operates universally in natural languages (e.g. Chen 1970), its magnitude has been observed to vary considerably cross-linguistically. From a methodological point of view, the PVD effect may be expressed in absolute terms – that is, as a difference in duration (e.g. in ms) – or in relative terms, as a ratio. Since vowel duration varies intrinsically (e.g. /ɪ/ vs. /iː/) and as a function of other factors such as speech rate, a relative measure is more suitable for comparisons between studies and speakers. The *PVD ratio* will therefore be used as a measure to compare findings reported in the literature.[4]

Figure 10.1 summarizes empirical evidence on the PVD effect in English and German.[5] As the studies listed differ in methodology, these figures should be interpreted as giving a quantitative reflection of the heterogeneity of estimates reported for each language or accent. As such, the PVD effect varies as a function of linguistic factors, which will be further discussed below. Position in the word has been identified as a particularly strong predictor (Sharf 1962; Klatt 1973; Tauberer 2010) and the following discussion therefore distinguishes two (necessarily broad) word-level contexts:

- final vowels, which occur in monosyllabic words (*bad*) or the final syllable of a (usually disyllabic) word (e.g. *believe*), and

[1] Examples are the timing and intensity of the glottal signal during the hold stage of a stop, or the period of friction in a fricative.

[2] Acoustically, this may be reflected in the duration and formant structure of the preceding vowel, or F_0 in the following vowel.

[3] Among these are *vowel length variation* (Chen 1970), *vowel-length effect* (Kluender et al. 1988), *voicing-dependent vowel duration* (Hussein 1994), *pre-lenis lengthening* (Chomsky & Halle 1968), and *pre-fortis clipping* (Wells 1990).

[4] See §9.6 for methodological considerations.

[5] For details, see the online supplement: https://osf.io/fsnjz/. For a comprehensive meta-analysis on the PVD effect in German, see Nicenboim et al. (2018).

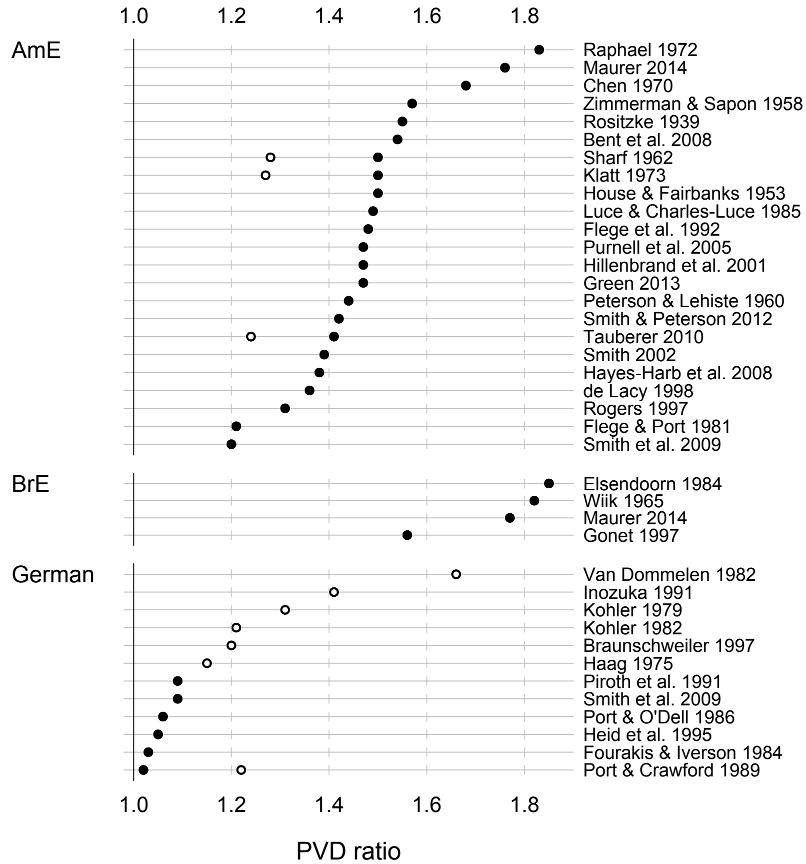


Figure 10.1: The PVD effect in AmE, BrE, and German: Duration ratio of vowels preceding voiced vs. voiceless obstruents in final (•) and non-final (◦) position (details can be found in the OSF repository: https://osf.io/fsnjz/).

- non-final vowels, which (mostly) occur in the penultimate syllable of a trochaic disyllabic word (e.g. *Seide, mieden*).

Studies on the German PVD effect report ratios between 1.0 and 1.1 in final position. This indicates little systematic PVD variation in word-final position. However, there is evidence that German exhibits a notable PVD effect in non-final position, with values in the literature ranging from 1.2 to 1.6. A comparison between German, BrE, and AmE unambiguously shows that the PVD effect in final position is much larger in English than in German. To facilitate comparisons, Figure 10.2 summarizes the distribution of measurements in the literature, including values reported for other languages[6] in final position[7] and non-final position[8].

Cross-linguistic differences are evident in final position, where AmE and BrE show the highest scores. The PVD effect has been observed to be strongest in BrE, with a median ratio of 1.76. In AmE, it is smaller (1.47), but still of considerable size compared to other languages. In German, the PVD effect in final position is small, with a median ratio of 1.07 across studies. As for non-final position, the magnitude of the PVD effect appears to be similar in AmE (*Mdn* = 1.27) and German (*Mdn* = 1.22); no data are available for BrE. Figure 10.2 shows that the word-final PVD effect appears to be distinctly large in English.[9]

[6] The data are deposited to https://osf.io/fsnjz/.

[7] References: French: Chen 1970; Kohler 1979; Laeufer 1992; Arabic: Flege & Port 1981; Czech and Polish: Keating 1980; Russian: Chen 1970; Dimitrieva et al. 2010; Italian: Esposito 2002

[8] References: Russian and Korean: Chen 1970; Spanish: Zimmerman & Sapon 1958; Norwegian: Fintoft 1962; Finnish: Wiik 1965; Dutch: van Dommelen 1982

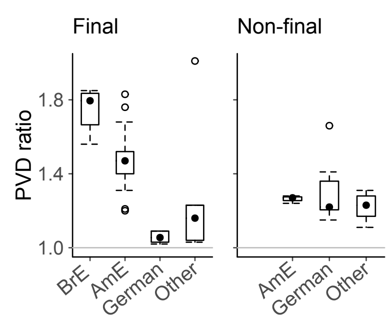


Figure 10.2: Comparison of the PVD effect in final and non-final position in English, German and other languages.

[9] These cross-linguistic contrasts are corroborated by studies that directly compare English to other languages using the same design and methodology (Zimmerman & Sapon 1958; Chen 1970).

As noted above, the PVD effect varies systematically as a function of prosodic and segmental factors. Besides position in the word, speech rate has been found to play a role, with higher tempo characteristically showing smaller ratios. Tauberer (2010) reported ratios of 1.4 in slow speech, 1.3 in normal speech, and 1.2 in fast speech. Similar effects were noted by Smith (2002), who observed PVD ratios of 1.4 (normal) and 1.3 (fast), and Port (1981), who found ratios of 1.2 (slow) and 1.1 (fast). In connected speech, further variation may be observed conditional on prosodic context. Umeda (1975, quoted in Tauberer 2010) reported the largest values for pre-pausal position (1.5), compared to word-final and word-medial position (1.4 and 1.1, respectively). In the measurements reported by Luce & Charles-Luce (1985), however, position (phrase-final vs. other) did not have an effect on the PVD ratio in monosyllabic words.[10]

The magnitude of the PVD effect also seems to vary depending on features of the vowel. First, vowel length seems to play a role. Across all studies that permitted this comparison, the PVD ratio was larger in long vowels compared to short vowels. The absolute difference between the ratios ranged from +0.06 to +0.38, with an average of about +0.20.[11] Vowel height also seems to be linked to the magnitude of the PVD effect - it is consistently larger in close compared to open vowels. The absolute difference between the ratios ranged from +0.05 to +0.47, with an average of about +0.23.[12] As for consonant class (stops vs. fricatives), there is variation in the directionality of the moderating effect.[13]

To summarize, English and German differ in the realization of coda obstruents. While English retains a voicing contrast in final position, this opposition is neutralized in German. In English, a prominent cue that distinguishes voiced and voiceless coda obstruents is the duration of the preceding vowel, which is longer before voiced consonants. Compared to other languages including German, the PVD effect is large in BrE and AmE. The literature further provides guidance on variation in the magnitude of the PVD effect in English. In general, it appears to be *larger*

- in BrE (vs. AmE),
- in final and pre-pausal environments (vs. non-final position),
- in long vowels and diphthongs (vs. short vowels), and
- in close/high vowels (vs. open/low vowels).

The PVD effect thus seems to vary systematically conditional on the factors variety, prosodic position, vowel length and vowel height. While German does not exhibit a PVD effect in final position, it shows the same voicing-dependent variation in vowel duration in non-final position as (American) English and other languages.

10.2 *L2 acquisition of FVOs: Linguistic factors*

10.2.1 *Markedness*

Two attributes of final voiced obstruents are relevant for assessing their markedness status: prosodic position and voicing. Onset-coda asymmetries were discussed in §5.2.1. In brief, the literature suggests codas to be

[10] While the absolute PVD difference (expressed in ms) reported by the authors differed statistically, the PVD ratio reflects no position effect (I calculated a PVD ratio of 1.47 for both contexts in Experiment 1 and 1.53 (final) vs. 1.54 (non-final) in Experiment 2). This shows that the conclusions drawn from a study may depend on how the PVD effect is operationalized – that is, whether it is assessed in absolute (PVD difference) or relative terms (PVD ratio).

[11] The following values are documented in the literature:

- BrE: +0.38 (Elsendoorn 1984), +0.33 (Wiik 1965), and +0.20 (Gonet & Stadnicka 2006)
- AmE: +0.17 (Rositzke 1939), +0.15 (House 1961), +0.15 (Luce & Charles-Luce 1985), +0.14 (Peterson & Lehiste 1960), and +0.06 (Hillenbrand et al. 2001).

[12] The values in the literature are:

- BrE: +0.47 (Wiik 1965) and +0.19 (Gonet & Stadnicka 2006)
- AmE: +0.33 (House 1961), +0.23 (Luce & Charles-Luce 1985), +0.09 (Hillenbrand et al. 2001), and +0.05 (Rositzke 1939).

[13] These are the values in the literature:

- BrE: –0.11 (Elsendoorn 1984) and –0.03 (Wiik 1965)
- AmE: +0.16 (House 1961) and +0.18 (Peterson & Lehiste 1960).

marked relative to onsets, which identifies FVOs as marked in terms of prosodic position. As for voicing, a universal preference for voiceless (over voiced) fricatives was established in §8.2.1. The same tendency can be observed for stops, with voiceless variants being more common cross-linguistically (Maddieson 1984: 27). Voiced obstruents are further constrained phonotactically – that is, they are licensed in a smaller number of clusters. FVOs are thus also marked with respect to voicing.

Dinnsen & Eckman (1975) established a universal hierarchy to account for the cross-linguistic distribution of obstruent voicing contrasts. They stipulate an implicational relationship between medial and final fricatives and stops: A voicing contrast in word-final fricatives implies a contrast in medial fricatives and final stops, both of which imply a contrast in medial stops. This ranking compounds the effect of position (final more marked than medial) and type of obstruent (fricatives more marked than stops). Eckman (1977) included initial position into the implicational hierarchy and postulated the following ranking for prosodic position: A voicing contrast in final position implies the presence of a voicing contrast in medial position, which in turn implies the presence of a voicing contrast in initial position. Obstruent voicing contrasts are thus disfavored in final (vs. initial and medial) position and also in fricatives (vs. stops).

Observations from L1 acquisition support this ranking. In general, voiceless obstruents are acquired earlier than their voiced counterparts by English-learning children (cf. Figure 8.3). Further, a preference for voiceless obstruents in final position appears to be a universal tendency in L1 acquisition. Within the framework of natural phonology, for example, final devoicing is classified as a natural process that must be suppressed in the acquisition of word-final voicing contrasts (Stampe 1979). Figure 10.3 compares the age of acquisition of voiced obstruents in initial vs. final position.[14] Two trends are easily discerned:

[14] The data are pooled across 4 studies (see §3.3.2 for details).

- Position: In general, voiced obstruents are acquired earlier in initial position (I) than in final position (F). The labiodental fricative /v/ is the only exception to this pattern.
- Manner of articulation: There is a consistent cline, for both voiced and voiceless variants, with stops being acquired first, followed by affricates and then fricatives.

Developmental differences have also been observed in acoustic studies. Smith (1979) compared 2- and 4-year-old children and adult speakers of

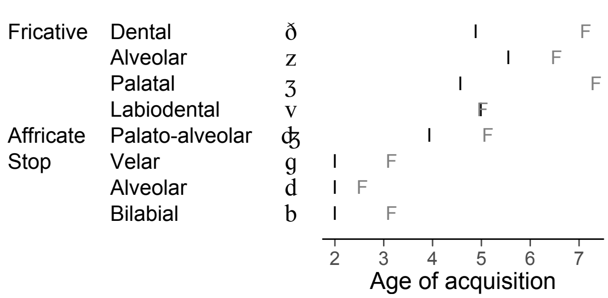


Figure 10.3: Age of acquisition of English voiced obstruents in initial (I) vs. final (F) position. Data points drawn at the left margin indicate acquisition by the age of 2. Estimates are based on a survey of 4 studies (see §3.3.2).

AmE (n = 5 each). Children showed a higher rate of devoicing in final obstruents (.98 and .92 for 2- and 4-year-olds) than adults (.50).[15] Smith & Stoel-Gammon (1983) reported developmental patterns for n = 4 AmE children between 1;6 and 3;0. The error rate of word-final stops decreased from 1 (1;6 to 2;0) to .84 (2;0 to 2;6) and .65 (2;6 to 3;0). Naeser (1970) investigated final voiced stops in n = 9 AmE children aged 1;8 to 2;10 months. The error rate decreased from .50 (1;8) to .02 (2;10).

[15] Differences in the acoustic properties of the devoiced segments were also noted. Thus, the degree of devoicing was higher in children, whose closure duration showed an average of .30 voicing, compared to .64 for adults.

An explanation for the typological preference and earlier acquisition of voiceless obstruents may be partly found in articulatory constraints. As pointed out by Ohala (1993: 194), vocal fold vibration requires appropriate positioning of the vocal cords and sufficient air flow. In §8.2.1, we pointed to conflicting aerodynamic requirements in the production of voiced fricatives. As for stops, the fact that they involve a closure of the vocal tract is at odds with the requirement of sufficient air flow. Concerning the universal preference for voiceless obstruents in final position, Blevins (2004: 103–106) suggests perceptual and articulatory factors to play a role. It is argued that a trigger may be the regular, cross-linguistically observed process of phrase-final (or pre-pausal) lengthening. Syllables preceding a pause or an intonation phrase boundary are lengthened, including the coda. The articulatory effect of coda consonant lengthening is the inhibition of voicing due to a prolonged hold phase, the acoustic result being convergence of voiced and voiceless obstruent duration. Both factors favor the perception of final lengthened consonants as voiceless, since absence of vocal fold vibration and increased duration are intrinsic attributes of voiceless obstruents.

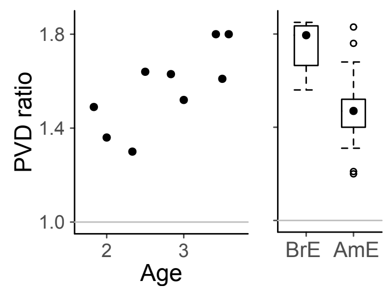


Figure 10.4: PVD effect in L1 acquisition: PVD ratio by age in English-learning children compared to the values reported for adult speakers of BrE and AmE.

While intrinsic characteristics of FVOs appear to be relatively marked, different conclusions must be drawn for PVD variation. In the study by Naeser (1970), for example, all age groups produced adult-like PVD ratios of 1.49 (at 1;8) and 1.63 (at 2;10). Indeed, studies investigating the PVD effect in the acquisition of L1 English have produced consistent evidence for a systematic variation in preceding vowel duration in infant speech. The findings suggest that children make this durational distinction before the age of 2;0 (Naeser 1970; Buder & Stoel-Gammon 2002; Ko 2007). Similarly, Raphael et al. (1980) noted that 3- and 4-year-old AmE children showed the same PVD effect as adults. Figure 10.4 and Table 10.1 give an overview of the values reported in previous work. The ratios produced by children resemble those found in adult speech, suggesting that the PVD effect is mastered early in L1 acquisition.

Buder & Stoel-Gammon (2002) suggest that the production of longer vowels before voiced consonants is the default, unmarked setting in L1 acquisition. They argue that children start out by producing PVD differences and subsequently develop the specific phonetic patterns of their L1. Empirical evidence reported by the authors supports this claim. They investigated AmE and Swedish children at the age of 2;0 and 2;6.[16] In contrast to AmE, Swedish has been reported to exhibit no PVD effect.[17] In 2-year-old children, PVD ratios of 1.64 (Swedish) and 1.40 (American) were observed.[18] The two groups showed different developmental trends. At the age of 2;6, the PVD ratio for the Swedish children had decreased markedly from 1.70 to 1.13 while that for the AmE children increased

[16] The sample sizes were, for AmE, n = 8 (2;0), n = 8 (2;6), and for Swedish n = 5 (2;0) and n = 10 (2;6)

[17] A ratio of 1.03 is reported by Elert (1964; as quoted in Buder & Stoel-Gammon 2002: 1855.).

[18] These values were calculated from the Table IV (p. 1860), using the harmonic mean, i.e. the exponentiated mean of the logged ratios.

Study	Age	Ratio	N	Position
Naeser 1970 [1]	2;10	1.63	6	Final
	1;10	1.49	3	Final
Buder & Stoel-Gammon 2002[2]	2;0	1.40	8	Final
	2;6	1.65	8	Final
Tauberer 2010[3]	2;4	1.30	2	Mixed
Greenlee 1978[1]	3;0	1.52	4	—
Raphael et al. 1980[4]	3;0–4;0	1.61	20	Final
DiSimoni 1974[5]	3;0–4;0	1.80	—	—
Krause 1982[5]	3;0–4;0	1.80	—	—

Table 10.1: PVD ratios observed in English L1 acquisition: Overview of studies. Notes: [1]quoted in Buder & Stoel-Gammon (2002: 1856); [2]averaged over extrinsic values in Tab. IV, p. 1860; [3]corpus data; [4]averaged over pairs in Tab. I, p. 338; [5]quoted in Tauberer (2010: 29).

from 1.40 to 1.65. It appears that Swedish children started with a default setting, which they subsequently had to unlearn. Further, the discussion above has shown that the PVD effect occurs in a number of languages, with Chen (1970) suggesting this phenomenon to be universal. In the absence of quantitative evidence on its cross-linguistic frequency, however, this claim needs to be substantiated by future studies. In terms of articulation, however, there appear to be no constraints on variation in vowel duration.

To summarize, the evidence in the literature, which is summarized in Table 10.2, suggests different markedness statements for instrinsic and extrinsic cues to voicing in coda obstruents. Voiced obstruents in final position appear to be marked structures. This is evidenced by implicational relationships between final and medial/initial obstruents, typological frequency distributions and observations from L1 acquisition. This markedness asymmetry may partly be grounded in physiological facts, which have been argued to disfavor voiced obstruents in general, and particularly in final position. As for extrinsic cues to the voicing distinction in final position, it can be stated that the PVD effect emerges early in L1 acquisition and may even be the default setting in children. Further, vowel duration is not subject to any known articulatory limitations. The PVD as a cue to voicing contrasts may therefore be considered an unmarked feature.

Criterion	FVOs	PVD effect
Implicational relationship	++	○
Cross-linguistic frequency	++	(–)
L1 acquisition	+	– –
Synchronic/diachronic instability	○	○
Articulatory complexity	+	– –

Table 10.2: Summary of markedness criteria for FVOs and the PVD effect. Open circles ○ denote no (or inconclusive) evidence; (+)+ refers to (strong) evidence for marked status; (–)– refers to (strong) evidence for unmarked status.

10.2.2 Similarity

German learners' perception of voicing in English final obstruents has been addressed by two studies. Weiher (1975) investigated $n = 12$ learners from northern Germany aged 10 to 11, with an average of 0.5 years of instruction. A discrimination and an identification task with nonce words ($n = 98$) were used. In the discrimination task, subjects were asked to identify the deviant form in a sequence of four stimuli (e.g. [klenz] - [klens] - [klenz] - [klenz]). In the identification task, items had to be matched with one out of three that followed (e.g. [zɒlvz] - [zɒlfs] - [zɒlfs] - [zɒlvz]). Overall, perception accuracy was .82 in the discrimination task and .88 in the identification task. The learners' performance on both tasks varied depending on the characteristics of the preceding vowel, with a drop in performance if the preceding vowel was a long monophthong. The accuracy rate in the discrimination task was .73 (vs. .87 for short monophthongs and .55 for diphthongs), in the identification task it was .83 (vs. .87 for short monophthongs and .92 for diphthongs). As noted by Weiher (1975), this finding is unexpected since the PVD effect is larger in intrinsically long vowels. Reliance on PVD is thus expected to yield a different pattern in perception accuracy. These findings may indicate that early-stage German learners may not systematically exploit PVD as a perceptual cue.

Smith et al. (2009) analyzed the perception of word-final obstruents by $n = 15$ L1 German speakers and $n = 15$ native speakers of AmE. The German subjects were between 18 and 40 years old and mostly students at a US university. The perceptual experiment employed a forced-choice identification task with tokens produced by native speakers and German learners. Overall, L1 German listeners' scores were lower than those of the native speakers (.76 vs. .87 correct identification). Native listeners correctly identified .95 of the tokens produced by native speakers, but only .78 of those produced by German speakers. German listeners correctly identified .78 of the tokens produced by native speakers and .74 of those produced by German speakers. The fact that German listeners performed only slightly better on stimuli spoken by native speakers ($\Delta = .04$) shows that they may not be sensitized to the cues available in this material.

To the best of my knowledge, there are no studies on German learners' reliance on the preceding vowel duration for the categorization of word-final voiced and voiceless obstruents. However, sensitivity to this cue has been investigated in English learners from other L1 backgrounds. Two types of L1s may be distinguished: (i) languages that do not have obstruents in syllable-final position at all and (ii) those that have only voiceless coda obstruents. Crowther & Mann (1992) studied the use of PVD as a perceptual cue to final obstruent voicing in native speakers of English and L1 Japanese and L1 Mandarin Chinese learners, whose L1s do not permit stops in coda position. It was observed that English listeners were most sensitive and L1 Mandarin listeners least sensitive to this information in the acoustic signal; the L1 Japanese listeners fell in between. The authors suggest that experience with phonemic vowel length in their L1 accounts for the relative advantage of Japanese listeners, as the Mandarin vowel system does not include length as a distinctive feature. The results indicate that L1 experience with vowel length as a contrastive feature may aid the perception of voicing contrasts in English final obstruents.

Several studies have shed light on the perception of voicing contrasts in final position by learners whose L1 does have final obstruents but no voicing contrast. Rojczyk (2010) focused on Polish learners' reliance on vowel duration as a perceptual cue to coda voicing contrasts. The study included $n = 26$ beginners and $n = 24$ advanced learners as well as a control group of $n = 11$ native speakers. Vowel duration was varied systematically in the nonce word *theep* in steps of 30 ms, ranging between 142 and 292 ms. While the native speakers showed a clear cline from identifying the stimuli as /θiːp/ vs. /θiːb/, no such pattern was evident for either learner group. While advanced learners were somewhat more likely to classify tokens across the whole continuum as /θiːb/, no systematic variation conditional on the duration of the adjacent vowel was observed; .75 of the stimuli with the longest vowel duration were heard as /θiːp/. These results indicate that even advanced learners (3^{rd}-year university students) do not systematically exploit PVD as a cue to the voicing contrast; it appears that other cues, most likely internal to the consonant, overshadowed the duration of the preceding vowel.

The contribution of different types of information in the acoustic signal was the subject of a study by Broersma (2005), who ran two experiments to investigate reliance on a number of cues in the perception of final consonant voicing by BrE and advanced L1 Dutch listeners. Like German, Dutch regularly devoices syllable-final obstruents. The first experiment included two groups ($n = 20$ each) and two listening conditions: one with unedited tokens and one with the release burst truncated. All items were nonce words. Dutch listeners' performance did not differ from that of the native speakers. They categorized final consonants with the same degree of accuracy, despite the fact that Dutch has no voiced obstruents in final position. A second experiment with two new groups ($n = 28$ each) used manipulated items (with final fricatives) to determine whether Dutch listeners relied on the duration of the preceding vowel. Vowel duration was edited to mismatch other cues in the signal. The author observed that native speakers relied more strongly on vowel duration, even in cases of discrepancy with competing cues. Dutch listeners, on the other hand, were able to rely on other signals for the distinction of word-final fricatives and performed better than native speakers in terms of categorization accuracy. The results of this study indicate that native speakers of (British) English have established the duration of the preceding vowel as a robust and automated cue in their perception grammar, which overshadows competing signals. Dutch learners, on the other hand, appear to be more flexible in cue selection. While they can rely on duration in the absence of other information, their perceptual processing appears to be less entrenched. If necessary, they may switch to other cues for the distinction between voiced and voiceless coda fricatives.

In sum, research on German learners' perception of FVOs has shown that this distinction is weakly defined in their perception grammar. While there is indirect evidence that German listeners do not make use of the PVD cue in a systematic way, no study has explicitly addressed this issue. Research on listeners from other L1s suggests that L1 experience with vowel duration as a meaning-distinguishing cue sensitizes learners to this type of information in the signal. This may explain Polish learners' lack of sensitivity, since their L1 does not contrast vowels in length (Jassem 2003). L1 Dutch listeners, on the other hand, are familiar with length differences in vowels (Gusshoven 1992), which may explain why advanced learners systematically exploit this cue in perception. Extrapolating from these findings, German learners should be able to transfer reliance on temporal properties from their L1 and eventually exploit the duration of the preceding vowel. However, as yet there is no empirical evidence to corroborate this claim.

10.3 L2 acquisition of FVOs: Theoretical predictions

The assessment of FVOs and the PVD effect from the perspective of linguistic constraints forms a basis for theory-based predictions about their acquisition by German learners. This section formulates predictions about FVOs and the PVD effect in GLE.

CAH For L1 German learners, FVOs constitute shared allophones of (in most cases) shared phonemes, which differ in distribution. This type of contrast is expected to be difficult to learn, with learner productions showing L1-derived voiceless obstruents in coda position. As for the PVD effect, German learners are expected to transfer L1 duration patterns, yielding target-like ratios in non-final position and insufficient temporal contrast in final position.

SLM In the absence of empirical insights into German learners' employment of PVD as a cue to the perception of final obstruent voicing, SLM predictions about the PVD effect cannot be formulated. It is also not clear whether extrinsic cues of this kind are within the scope of the model, as its focus is traditionally restricted to intrinsic attributes of position-sensitive allophones. An extension of the SLM to include extrinsic cues in adjacent segments certainly appears feasible, however. This would require learners to extend their scope of perception to neighboring sounds. It is thus conceivable that extrinsic cues are subject to different constraints compared to intrinsic ones. To speculate, perception may primarily focus on intrinsic attributes. Other things being equal, equivalence classification may thus be considered more likely in the absence of salient intrinsic signals. This tendency may be particularly pronounced if learners are desensitized to extrinsic cues as a result of L1 experience. In the absence of experimental insights, however, these conjectures remain speculative. In general, the studies by Weiher (1975) and Smith et al. (2009) suggest perceptual difficulties with FVOs by German listeners, but no conclusions are currently warranted about cue weighting. The following assumptions may be stated: (i) the perceptually closest L1 sounds to FVOs are their voiceless counterparts; and (ii) the degree of perceived similarity between final voiced and voiceless obstruents is moderate to high. This suggests perception-based difficulties in the majority of learners.

FCM Assuming that the range of L1 candidates for FVO substitution is limited to their voiceless counterparts, the most important prerequisite for the FCM, a set of alternatives, is not met.

SCH/MDH Distinguishing between the markedness status of FVOs on the one hand, and the PVD effect as an extrinsic cue on the other, the following assumptions will be made: (i) FVOs are marked structures; (ii) FVOs are relatively more marked than final voiceless obstruents; (iii) final voiced fricatives are relatively more marked than final voiced stops; and (iv) PVD is a an unmarked feature. Intrinsic attributes of FVOs are therefore expected to be difficult to acquire. L2 acquisition is predicted to show a bias toward unmarked voiceless obstruents in final position. Further, we expect delayed acquisition in fricatives relative to stops. The PVD effect, on the other hand, should be easy to learn.

NM/NDH The literature on the natural phonology of FVOs suggests that devoicing is a natural process that must be suppressed in the acquisition of FVOs. Similar to the considerations just outlined, this process affects intrinsic attributes of FVOs. Presumably, however, there are no natural processes constraining vowel length variation. As for the effect of frequency, no evidence-based figures are available that quantify rate of occurrence. It seems reasonable to assume FVOs to be rather frequent. Considering the PVD effect, however, the tokens most relevant for perceptual learning are likely those contexts where the effect is particularly pronounced, e.g. in utterance-final and pre-pausal environments. Frequency estimates should take variation in the magnitude of the PVD effect into consideration. In the absence of such figures, it remains unclear in which way input frequency may be expected to bear on L2 acquisition.

FM Adopting a constraint-based approach to the acquisition of FVOs and the PVD effect necessitates making a distinction between two types of articulatory and faithfulness constraints. As for the latter, perceptual learning may occur for (i) intrinsic features of FVOs and/or (ii) the PVD effect as an extrinsic cue. The stochastic emergence of these faithfulness constraints as well as subsequent sensorimotor learning need not coincide. Accordingly, the FM predicts a systematic interplay of newly established articulatory and faithfulness constraints for both structures, which gives rise to four developmental stages: (i) deviant production, (ii) variation reflecting focus on form, (iii) overly faithful production, (iv) variation sensitive to functional principles. We expect the last stage to be characterized by style-dependent implementation of distinctive features and connected speech processes that arise from speech motor interaction with adjacent segments. Thus, systematic PVD variation as a function of prosodic and segmental factors is expected to surface at advanced stages. Given the markedness constraints on features inherent to FVOs, learner productions at this stage may also show increased use of extrinsic cues to signal final obstruent voicing contrasts. A further post-lexical phenomenon that may interact with the acquisition of [±voice] in FVOs is resyllabification as a result of connected speech processes. Thus, FVOs followed by a vowel in connected speech will, upon linking, be resyllabified to the onset of the following syllable, thereby assuming a prosodic position that is more favorable to intrinsic voicing features.

MSA With no known articulatory constraints on vowel duration, the MSA reduces to the SLM for the acquisition of the PVD effect. For intrinsic features of FVOs, however, constraints outlined above are assumed to further delay the acquisition of voiced obstruents in final position by favoring physiologically less demanding voiceless segments. Further, coarticulatory constraints may generally impede the production of voiced obstruents in low-sonority contexts. As a consequence, voicing may be inhibited when the sound is followed by a voiceless segment. In contrast, contexts in which a voiced obstruent is followed by a vowel or sonorant may be more favorable to intrinsic voicing cues.

LTD Since FVOs by definition occur in *w*-marked coda position, their acquisition is expected to be delayed from the viewpoint of structural strength. Other things being equal, we expect relative ease of acquisition in *s*-marked environments, such as stressed syllables. As for final obstruent clusters, the LTD's system of strength marking suggests relative ease of acquisition of cluster-initial obstruents. The model's predictions about strength effects on the observed variants contradict some of these tendencies, however. Thus, a bias toward strengthened – that is, voiceless – variants is expected in *s*-marked positions, which in fact delays acquisition in stressed syllables and cluster-initial segments. In *w*-marked units, on the other hand, a bias toward weakened variants indeed favors realizations that more closely resemble FVOs. As for the PVD effect, no such divergence unfolds: Earlier acquisition is expected in stressed positions, where strengthened – that is, longer – variants do not impede acquisition.

OPM The classification of FVOs as both marked and similar from the perspective of German learners suggests delayed acquisition, with an underlying systematic interplay of L1, L2, and U structures as outlined in Figure 10.5.[19] Perceptual constraints are thus expected to yield a prolonged phase of L1 transfer, surfacing in the prevalence of L1-derived voiceless variants in coda position. Subsequently, the marked status of FVOs implies a period of U-dominance. In accordance with the tendencies observed cross-linguistically and in L1 acquisition, we expect a persistent bias toward voiceless coda obstruents. The effect of perception-driven L1 transfer and U constraints is therefore indistinguishable in the acquisition of coda obstruents by German learners. Finally, once U decreases, L2 structures are acquired. Based on the OPM, then, the acquisition of FVOs should show a considerable delay. For the acquisition of the PVD effect, the lack of empirical evidence on the perception of this cue by German learners precludes OPM predictions. However, since there are no plausible universal constraints on the production of vowel length variation, a phase of U-dominance seems unlikely. The OPM then predicts earlier acquisition of the PVD effect relative to intrinsic features of FVOs. Finally, the Stylistic Corollary suggests more target-like production in monitored and formal speech, and a higher rate of L1 transfer in unmonitored and casual utterances.

Table 10.3 offers a summary of the theoretical predictions covered in

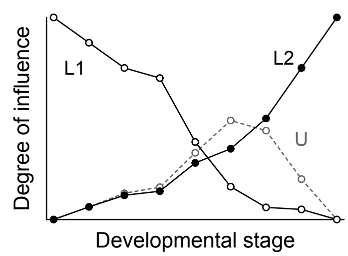


Figure 10.5: Extrapolation of the OPM to structures that are both similar and marked. This graph should be compared to Figure 2.6.

[19] See §4.3 (Figure 4.18, p. 66) for a discussion of the rationale underlying the fusion of these corollaries.

Table 10.3: The acquisition of FVOs by German learners: Summary of predictions based on theoretical work.

Theory	Predicted patterns in L2 acquisition by German learners
CAH	Late acquisition of FVOs Substitutes: voiceless counterparts
SLM	Late acquisition by the majority of learners Substitutes: voiceless counterparts
MDH/SCH	Late acquisition of intrinsic features of FVOs Substitutes: intrinsic attributes of voiceless counterparts Acquisition order: stops > fricatives Early acquisition of PVD effect
NM/NDH	Late acquisition of intrinsic features of FVOs Early acquisition of PVD effect
FM	(Independent) development for intrinsic features of FVOs and the PVD effect: deviant production > stylistic variability > overly faithful production > variability reflecting connected speech processes, resyllabification of FVOs as a result of linking Effect of speaking style (focus on form) at early stages Late stages: style-dependent feature backgrounding, systematic PVD variation as a function of prosodic and segmental factors, possible foregrounding of PVD effect (vs. intrinsic features),
MSA	Further delay in the acquisition of intrinsic features (cf. SLM) Delayed acquisition of fricatives relative to stops Delayed acquisition in low sonority contexts No further delay in the acquisition of PVD effect
LTD	Delayed acquisition of FVOs Acquisition order of FVOs and PVD effect: stressed > unstressed syllables; cluster-initial obstruent > non-initial obstruent Substitutes for FVOs: *s*-marked positions: strengthened (voiceless) substitutes *w*-marked positions: weakened (voiced/devoiced) variants
OPM	Late acquisition of intrinsic features of FVOs Earlier acquisition of PVD effect Substitutes: Intrinsic attributes of voiceless counterparts Effect of speaking style (focus on form): more formal speech: Higher degree of TL-like production; less formal speech: Higher degree of L1 transfer

this section. We now take a look at previous empirical work on FVOs in German Learner English.

10.4 *FVOs in German Learner English: Previous work*

The literature provides auditory and instrumental accounts of FVO production by German learners.

Auditory studies

Wieden & Nemser (1991: 75, 83, 191) studied the production of FVOs by *n* = 384 Austrian learners of English across different developmental stages.[20] In the production test, low accuracy rates were observed for final voiced stops (.03) and fricatives (.01). Two levels of devoicing were distinguished: voiceless (classified as a fortis consonant by native speaker judges) and devoiced (–lenis, but still classified as a lenis consonant). The degree of devoicing varied as a function of phonetic context. In singleton stops, the fortition rate was .46 (ibid.: 77). In coda clusters the fortition rate varied depending on the position of the stop in question. It was lower if the stop preceded a voiced obstruent (.33) than when it occurred in cluster-final position (.61). Voiced fricatives showed similar patterns. The fortition rate was higher in coda singletons (*five, with*) than when a voiced obstruent followed (/v/ .91 vs. .81; /z/ .74 vs. .71; /ð/ .65 vs. .41, ibid.: 86).

[20] The study included learners at 6 proficiency levels, with *n* = 64 subjects in each group. Years of instruction was used as an indicator of developmental stage, which ranged from 0 years (grade 3; 8–9 years of age) to 8 years (grade 11; 16–17 years old).

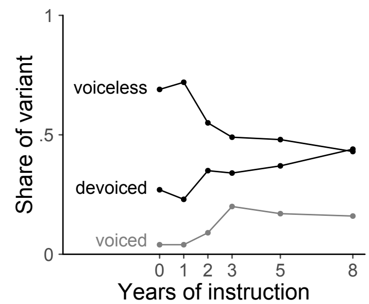


Figure 10.6: Realization of FVOs in the imitation test: rate of voiceless (fortis), devoiced (semi-fortis), and voiced (lenis) realizations. Adapted from Wieden & Nemser (1991: 167) with permission.

Wieden & Nemser (1991: 167) also report developmental trends for the realization of FVOs in the imitation test, where subjects were asked to repeat nonce words they heard via headphones. Figure 10.6 shows the trajectories for voiceless, devoiced, and voiced realizations. Developmental change occurs between stages 0 and 3, where the share of voiceless variants decreases, while the rate of voiced renditions and devoiced realizations (perceived as a voiced obstruent by native speakers) increases slightly. Little progress is evident at later stages.

The authors also observed an effect of task type. While the imitation test required subjects to imitate nonce words solely based on perception, the production test elicited English words using various techniques. Figure 10.7 contrasts the results for final /z/ in the imitation task (panel a) with those obtained in a German-to-English translation task (panel b). The patterns differ drastically. In the imitation task, the rate of voiced [z] increases steadily to about .70. Voiceless [s] decreases quickly, while devoiced [z̥] as an intermediate variant increases and then decreases. In contrast, no change was observed in the translation task, where subjects showed a clear preference for [s] across all stages.

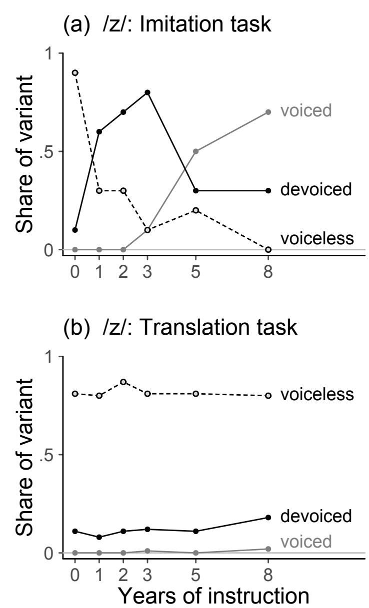


Figure 10.7: Realization of final /z/: Rate of voiceless (fortis), devoiced (semi-fortis), and voiced variants by years of instruction in an (a) imitation task and (b) a translation task. Adapted from Wieden & Nemser (1991: 191) with permission.

Weiher (1975) analyzed the production of FVOs by *n* = 28 German learners in grade 5 (age: 10–11) using three tasks: the imitation of nonce words, the imitation of English words, and the reading and imitation of English sentences. Overall, the author observed a voicing rate of .74. Little variation occurs across task types (Δ = .07; nonce words .71, English words .73, English sentences .78) and preceding vowel length (Δ = .07; short monophthongs .75, long monophthongs .70, diphthongs .77).[21]

[21] These values are based on Tables 54, 56, and 57 in Weiher (1975).

Langguth (2009) studied the production of FVOs by German learners in grades 5, 9, and 12 (*n* = 7 each; age: 10–11, 14–15, and 17–18, respectively). Three tasks were used for data elicitation: a minimal pair reading task, a word list, and semi-free speech prompted by questions on a text. Overall, the share of voiced renditions was .15, with a difference between stops (.25) and fricatives (.04). Final obstruent voicing was observed to vary with the place of articulation, both in stops (/b/ .40, /d/ .30, /g/ .05) and fricatives (/v/ .07, /z/ .01).

Wunder (2012) investigated the production of FVOs by German learners from northern Bavaria (Bamberg) in grade 6, 9, and 12 (n = 10 each). Two speaking styles were elicited: a reading passage and a minimal pair imitation task, in which each couplet was presented both on paper and auditorily, spoken by an AmE native speaker. The stimuli, which were designed to allow for analysis of several linguistic factors, were analyzed auditorily into fortis and lenis production, with an extra category for doubtful cases. The overall voicing rate was .15, with considerable differences between stops (.28), fricatives (.05) and affricates (.02). The share of voiced stops further varied depending on place of articulation, with appreciable differences between bilabial (.44), alveolar (.30), and velar (.11) segments. The production of fricatives was also sensitive to place of articulation, with rates of .08 for labiodental, .04 for alveolar, and .01 for dental strictures. No noticeable effect of speaking style was observed, with a voicing rate of .16 in the minimal pair imitation task and .14 in the reading passage. Similarly, little variation was noted between age groups, with rates of .08, .11, and .14 in grades 6, 9, and 12, respectively.

Table 10.4 summarizes auditory work on FVO production by German learners. Most studies report high devoicing rates, a curious exception being Weiher (1975). Properties of the obstruents have been observed to play a role, with a higher share of voiced variants (i) in stops vs. fricatives and (ii) – in stop consonants – for more anterior places of articulation. Further, the phonetic context seems to play a role: In clusters, higher voicing rates have been reported for FVOs closer to the nucleus. Speaking style appears to play a minor role. The effect of task type, on the other hand, may account for a considerable boost in FVO accuracy, which is higher when learners imitate model utterances by native speakers.

Table 10.4: Empirical evidence from auditory studies on FVOs in German Learner English: Reported accuracy rates (i.e. the share of voiced variants) and constraints on accuracy.

Results		Notes	Study
Accuracy rate	.74	Grade 5, age: 10–11 years	Weiher 1975
	.15	Grades 5, 9, 12; age: 10–18 years	Langguth 2009
	.01–.03	Grades 3–11; age: 8–17 years	Wieden & Nemser 1991
	.15	Grades 6, 9, 12; age: 11–17 years	Wunder 2012
Constraints on accuracy rate			
Place	Δ = .33	Stops: bilabial (.44) > alveolar (.30) > velar (.11)	Wunder 2012
	Δ = .35	Stops: bilabial (.40) > alveolar (.30) > velar (.05)	Langguth 2009
	Δ = .07	Fricatives: labiodental (.08) > alveolar (.04) > dental (.01)	Wunder 2012
	Δ = .06	Fricatives: labiodental (.07) > alveolar (.01)	Langguth 2009
Manner	Δ = .23	Stops (.28) > fricatives (.05) > affricates (.02)	Wunder 2012
	Δ = .21	Stops (.25) > fricatives (.04)	Langguth 2009
Speaking style	Δ = .07	Isolated words > sentences	Weiher 1975
	Δ = .02	Minimal pair imitation task (.16) > reading task (.14)	Wunder 2012
Cluster position	Δ = .28	Stops: Before voiced obstruent (.67) > final (.39)	Wieden & Nemser 1991
	Δ = .24	/ð/: Before voiced obstruent (.59) > final (.35)	Wieden & Nemser 1991
	Δ = .10	/v/: Before voiced obstruent (.19) > final (.09)	Wieden & Nemser 1991
	Δ = .03	/z/: Before voiced obstruent (.29) > final (.26)	Wieden & Nemser 1991
Vowel	Δ = .07	Diphthongs (.77) > short/long monophtongs (.75/.70)	Weiher 1975
Task type	Δ = 0–.70	Imitation task > translation task; /z/ only	Wieden & Nemser 1991

Acoustic studies

Empirical work has also reported on the acoustic correlates of final obstruent voicing in GLE. Green (2013) studied vowel length variation in n = 13 L1 German learners and n = 9 AmE native speakers. The German learners were students at a university in California, aged 21 to 36 with length of residence ranging from 5 months to 10 years. The learners produced PVD ratios substantially lower than the native speakers (1.16 vs. 1.47).

A study by Maurer (2014) investigated the production of FVOs by German students of English at university level. Subjects were divided into an advanced group (n = 60) who had completed a stay abroad in an English-speaking country (minimum: 4 months) and a control group (n = 6) of university students without experience abroad as well as a group of BrE and AmE native speakers (n = 10 each). Devoicing was measured using n = 24 tokens from a reading task consisting of a list of minimal pairs that differed in the voicing of the final consonant.[22] The effect of consonant voicing on the duration of the preceding vowel was measured with token pairs elicited in four reading tasks (minimal pairs, word list, sentences and a text passage). Each speaker produced a total of n = 119 token pairs, which were matched for vowel and final consonant (/k g p b t d/); 4 vowels were included in the analysis: /ɪ e æ ɒ/. The PVD ratio showed a difference between the control group (1.22) and the advanced learners (1.53); native speakers produced a ratio of 1.77.

[22] The degree to which final voiced and voiceless obstruents were contrasted was measured using Pillai scores calculated from two durational measurements: closure phase and aspiration. This score quantifies category overlap and thus expresses how well speakers differentiate final obstruents in terms of these parameters (0 = complete overlap; 1 = no overlap). The results show that both learner groups differed substantially from the native speakers in terms of their separation of final voiced and devoiced obstruents (control group: .35; advanced learners: .48; native speakers: .80).

Barry (1977) studied acoustic correlates of final obstruent voicing in n = 11 learners from northern Germany (age: 17–19) using the minimal pairs *bed–bet* and *dog–dock*. These items were elicited with reading, imitation, and naming tasks; preceding vowel and closure duration were measured. The learners exhibited a clear PVD effect, producing ratios of 1.69 and 1.61 for *bed–bet* and *dog–dock*, respectively. These figures cannot be attributed to the effect of imitation, as the subjects also showed systematic variation in vowel duration in the reading task, which started the experiment. Notable differences between learners and native speakers were observed in closure duration, however. These findings suggest that German learners are capable of attaining a native-speaker level of PVD variation in final stops. Further, the PVD effect may be easier to acquire than cues internal to the obstruent, at least as far as closure duration is concerned.

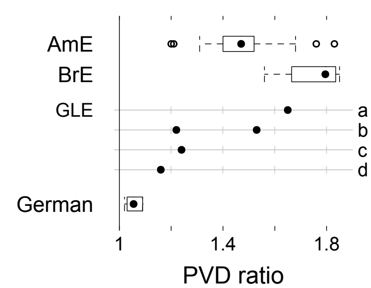


Figure 10.8: Summary of the instrumental evidence on the PVD effect in GLE in comparison to properties of the L1 (bottom) and the target language varieties (top). Key to studies: (a) Barry 1977, (b) Maurer 2014, (c) Smith et al. 2009, (d) Green 2013.

Similar results were reported by Smith et al. (2009), who analyzed several acoustic parameters in word-final obstruents produced by n = 13 L1 German speakers and n = 13 native speakers of AmE. The German subjects, with an average age of 26, had been living in the US for at least 7 months (*Mdn* = 1.7 years). The study elicited German and English monosyllabic words embedded in carrier sentences. The results showed that while German learners achieve native-like vowel length variation, other acoustic correlates such as consonant closure duration and voicing during consonant closure have values intermediate between their L1 German and those of the native speakers. Similar to the findings reported by Barry (1977), these results suggest that the PVD effect may be acquired earlier than cues intrinsic to the obstruent.

In sum, instrumental evidence on the PVD effect in GLE has revealed that learners are able to reach a native speaker level of vowel duration variation. Table 10.5 and Figure 10.8 summarize the measurements reported in the literature. Three studies observed ratios of around 1.2, which are on a par with L1 timing patterns in non-final position. Two studies record PVD ratios near 1.6, which are in the range reported for native speakers.

Study	PVD ratio	*N*
Green 2013	1.16	13
Maurer 2014[1]	1.22	6
	1.53	60
Smith et al. 2009	1.24	13
Barry 1977	1.65	11

Table 10.5: Acoustic evidence for the PVD effect in GLE: Ratios reported in previous studies. *N* refers to the number of speakers recorded. Note: [1]Averages based on Tables 4.10 and 4.11 (pp. 130, 134).

10.5 Aims of this study

Our objective is to make empirical contributions to the study of FVO production in GLE by focusing on the following aspects:

- cross-sectional patterns in the realization of FVOs,
- linguistic factors constraining the realization of FVOs, and
- developmental patterns in the magnitude of the PVD effect.

Cross-sectional patterns in the realization of FVOs

The theoretical frameworks surveyed above unanimously consider FVOs as difficult structures, expecting late acquisition by German learners (CAH, SLM, MDH/SCH, NM/NDH, MSA, LTD, OPM). The empirical literature is broadly coherent with this claim, reporting accuracy rates below .20 (Wieden & Nemser 1991; Langguth 2009; Wunder 2012). An exception is Weiher (1975), who observed a rate of .75. The substitutes expected in learner speech are devoiced obstruents and the respective voiceless counterparts. This study seeks to shed further light on the realization of FVOs by studying cross-sectional patterns in the rate of variants as manifested in learners at different levels of L2 pronunciation ability.

Realization of FVOs: Linguistic factors

The empirical literature documents various systematic sources of variation in FVO production.

- *Manner of articulation*: In line with the fricative-stop asymmetry posited by the MDH/SCH and MSA, previous work has shown that the acquisition of final voiced fricatives is delayed relative to that of stops (Langguth 2009; Wunder 2012), and it has also been observed that the fortition rate is higher in fricatives (Wieden & Nemser 1991). This study seeks to offer further insights into the link between manner of articulation and FVO production.

- *Place of articulation*: Place of articulation has been observed to play a role in the acquisition of final voiced stops: bilabial obstruents typically show the highest accuracy rates, followed by alveolar and then velar segments (Langguth 2009; Wunder 2012). We will follow up on these indications.
- *Sonority*: Coarticulation effects induced by the sonority of adjacent sounds were considered in the light of the MSA.[23] Findings by Wieden & Nemser (1991) indicate a sensitivity of FVO production to the following context. However, in GLE this effect has not been investigated systematically and will be addressed here.

[23] A consistent effect of the sonority of the right-ward phonetic context was reported by Edge (1991), who studied L1 Cantonese and L1 Japanse learners of English, and observed the highest devoicing rate before a pause and the lowest rate before vowels.

Cross-sectional patterns in the magnitude of the PVD effect

From a theoretical perspective, the PVD effect is predicted to be acquired earlier than intrinsic features of FVOs (MDH/SCH; NM/NDH; MSA; OPM). Experimental studies substantiate this prediction (Barry 1977; Smith et al. 2009). PVD ratios observed in German learners range from near-L1 figures to a native speaker level of accuracy. This study aims to trace cross-sectional patterns in the acquisition of PVD variation across a wider range of proficiency levels, to provide further quantitative insights into the emergence of this extrinsic cue in GLE.

10.6 Method and data

Informants, materials and procedures

FVO production was analyzed with (i) auditory data extracted from the *SpIL* corpus (Pascoe 1987) and (ii) acoustic measurements using the data collected in the present study.[24] The *SpIL* corpus contains $n = 647$ FVO tokens in total. We relied on the phonetic transcriptions provided by Pascoe (1987), where three types of variants are distinguished: voiced, devoiced, and voiceless. In total, four different types of FVOs occurred in the corpus (/d g v z/) and it was decided to retain these categories, as they give a transparent reflection of place and manner of articulation. The sonority of the adjacent context to the right of the focal obstruent was coded using 5 categories: vowel, voiced consonant, voiceless consonant, glottal stop, and pause. As the corpus provides detailed information including short and long pauses, assignment of observations to these categories was straightforward on the basis of Pascoe's (1987) transcriptions. Table 10.6 shows the distribution of the factor levels in the data.

[24] The data sets analyzed in the present chapter are openly available via the *Tromsø Repository of Language and Linguistics* (*TROLLing*) at https://doi.org/10.18710/DKIGE5 (Sönning & Pascoe 2020). See §3.3.1 for further information on the data provided by Pascoe (1987).

As for the acoustic data gathered from the recordings in the current study, the analysis of the PVD effect relied on temporal measurements of minimal pairs elicited with the word list and short phrases tasks. For the segmentation criteria applied, see §4.6. Each subject produced $n = 11$ pairs of words, which contrasted in final obstruent voicing. The order of items in the production tasks ensured that pairs did not occur in succession. The following minimal pairs were elicited:

- Word list: *plays–place*, *bag–back*, *played–plate*, *save–safe*
- Phrases: *heed–heat*, *god–got*, *hod–hot*, *had–hat*, *hard–heart*, *head–het*, *hid–hit*

Condition	Types	Tokens
Obstruent		
/d/	56	332
/g/	1	27
/v/	9	62
/z/	44	223
Sonority		
Pause	46	134
[ʔ]	34	93
Voiceless consonant	36	152
Voiced consonant	49	204
Vowel	28	61

Table 10.6: Distribution of the linguistic factors in the *SpIL* corpus (Pascoe 1987).

For each pair, the PVD ratio was determined by dividing the duration of the vowel preceding the voiced obstruent by that before the voiceless counterpart. By using a ratio measure, we can, at least to a certain degree, sidestep differences in the duration of vocalic intervals that may be expected to occur between learners and native speakers. Thus, absolute differences in vowel duration would have been more strongly confounded by speech rate and potential differences in intrinsic vowel duration. Further, the analysis thereby focuses directly on the measure that is most informative in the context of this study, the PVD ratio. The statistical analysis rested on $n = 11$ measurements per speaker, summing to a total of $n = 619$ data points for the learners ($n = 23$ missing cases) and $n = 284$ for the native speakers ($n = 1$ missing).

The next section describes aspects of the statistical analysis and may be omitted without loss of continuity.

Statistical analysis

The acoustic PVD measurements were analyzed with a multilevel regression model with varying intercepts on subject and word pair. The native speaker model includes variety as a predictor, which is also represented with varying slopes on words. For the learner data, proficiency was the focal predictor, which also received varying slopes across word. For the analysis of the acoustic data, the literature survey provides guidance on the magnitude of the PVD effect in native speech. This state of knowledge was included into the model in the form of informative prior distributions. Figure 10.9 shows the correspondence between estimates in the literature and the derived probability distributions. The prior for the PVD ratio in AmE concentrates 90% of its mass in the interval [1.13, 1.97] and that for BrE has 90% of its mass in the range [1.19, 2.29]. For the analysis of the learner productions, the literature suggests estimates of about 1.2 and 1.6 to be representative of low and high proficiency levels, respectively. A weakly informative prior was centered at 1.4 and the prior density assigns probability .90 to the interval [0.91, 2.44]. More information about the structure of the models can be found in Appendix A.10.

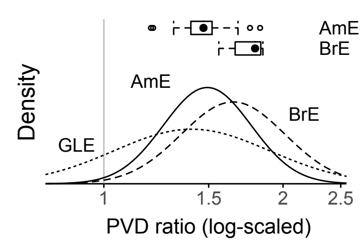


Figure 10.9: Prior distributions of the PVD ratio in AmE and BrE derived from the empirical evidence in the literature. The solid curve is the prior for AmE, the dashed curve that for BrE. The boxplots at the top show the distribution of measurements in the literature (see Figures 10.1 and 10.2).

The auditory data based on Pascoe (1987) were analyzed with a multilevel multinomial regression model, with varying intercepts on subject and word. The predictors of interest were backed by varying slopes: Proficiency was allowed to vary across words (by-word varying slopes), and the association of FVO production with the obstruent type and the following context was also given leeway to differ in strength across speakers (by-speaker varying slopes). As for priors, the intercept parameters receive a diffuse but sensible specification, which is shown graphically in Figure 10.10. It has 90% of its mass in the intervals [.01, .57] (voiced), [.03, .69] (devoiced), and [.13, .87] (voiceless). More details can be found in Appendix A.11.

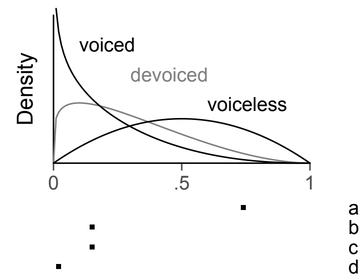


Figure 10.10: Priors for the intercept parameters in the model in comparison to the rates reported in the literature (voiced variants only): (a) Weiher 1975, (b) Langguth 2009, (c) Wunder 2012, and (d) Wieden & Nemser 1991.

10.7 Constraints on generality

When extrapolating the following findings to other speakers and circumstances, we need to bear in mind a few limitations:

- *Language-external scope*: It does not seem that there are any speaker-related variables that would restrict the generalizability to other populations of German instructional-setting learners.
- *Language-internal scope*: For the acoustic analysis of the PVD ratio, all items carrying FVOs occurred either in isolation or in a contrastive context. The findings therefore only extend to prosodically prominent environments. Further, long vowels and diphthongs are overrepresented in the set of lexical items (about 70% of the tokens, counting /æ/ as a long vowel.). Estimates of the PVD ratio, especially those for native speakers, may therefore be higher than in a more balanced selection of carrier words. The auditory data, on the other hand, may be used to generalize to a fairly broad set of contexts.
- *Speaking style*: The acoustic data only apply to closely monitored speaking styles. The auditory data are representative of more natural speech.

10.8 Results

We will first consider the acoustic data collected in the context of the current study (§10.8.1). §10.8.2 then moves on to the auditory analysis based on the *SpIL* corpus (Pascoe 1987).

10.8.1 Acoustic analysis

The distribution of PVD ratios for the learners in the current study is shown in Figure 10.11, where each point denotes, for a given speaker, the average PVD ratio over the $n = 11$ word pairs. The vertical axis is log scaled, but the original ratio values can be read from the tick mark labels. Overall, we see a steady increase in the magnitude of the PVD ratio from left to right. For low-proficiency learners, values around 1.1 seem to be typical. Intermediate learners average at about 1.2 to 1.3, and advanced learners roughly at 1.4.

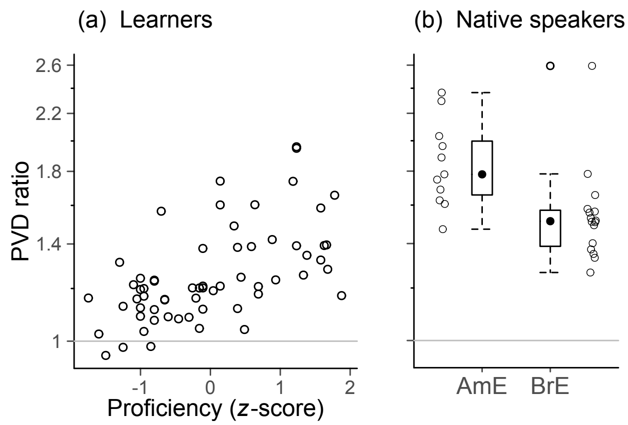


Figure 10.11: Distribution of PVD ratios for (a) learners and (b) native speakers. The y-axis is log-scaled, with tick mark labels giving the original values.

The variation between learners also appears to increase slightly from left to right.

The scores obtained for the native speakers are shown in Figure 10.11. The AmE subjects produced considerably higher PVD ratios, with a median of about 1.8, compared to 1.5 for BrE subjects. The variation among speakers is somewhat larger in the AmE group. Barring one outlier in the BrE group (average ratio: 2.6), we find two quite distinct distributions of measurements.

Technical details about the regression models are deferred to Appendix A.10.[25] Figure 10.12 compares model-based estimates for the learners and the native speakers. These are also listed in Table 10.7, which includes uncertainty estimates. The patterns offer a condensed summary of the descriptive insights provided by Figure 10.11:

- The average PVD ratio for AmE native speakers is 1.78, that for BrE subjects 1.50.
- German learners at lower proficiency levels show little systematic variation, with ratios of around 1.10, which are similar to those observed in German.
- Intermediate-level learners typically produce PVD ratios of around 1.25.
- Advanced learners average around 1.40, which is almost in the range of values for native speakers.

[25] The appendix lists convergence diagnostics for the models and reports posterior estimates for the model parameters.

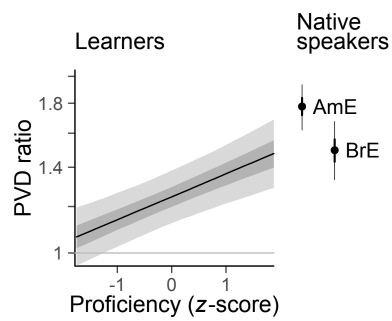


Figure 10.12: Developmental patterns in the PVD ratio: Model-based linear trend line in comparison with estimates for the native speakers groups in the present study. Error bars/bands show 50% and 90% uncertainty intervals.

		Posterior quantiles			
Group	*Median*	.05	.25	.75	.95
German learners					
Beginner	1.09	0.98	1.04	1.14	1.22
Intermediate	1.25	1.13	1.20	1.30	1.39
Advanced	1.43	1.26	1.36	1.50	1.62
Native speakers					
AmE	1.78	1.62	1.72	1.84	1.94
BrE	1.50	1.33	1.43	1.56	1.68

Table 10.7: PVD ratio estimates for learners at representative proficiency levels and the AmE and BrE native speakers (data from this study).

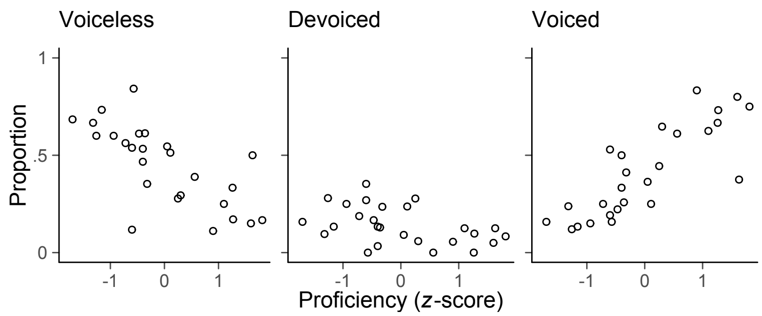


Figure 10.13: FVO production by German learners: Production rates for voiceless, devoiced, and voiced variants by proficiency level. The graphs shows, for each learner, the proportion of the variants. Data from Pascoe (1987).

10.8.2 Auditory analysis

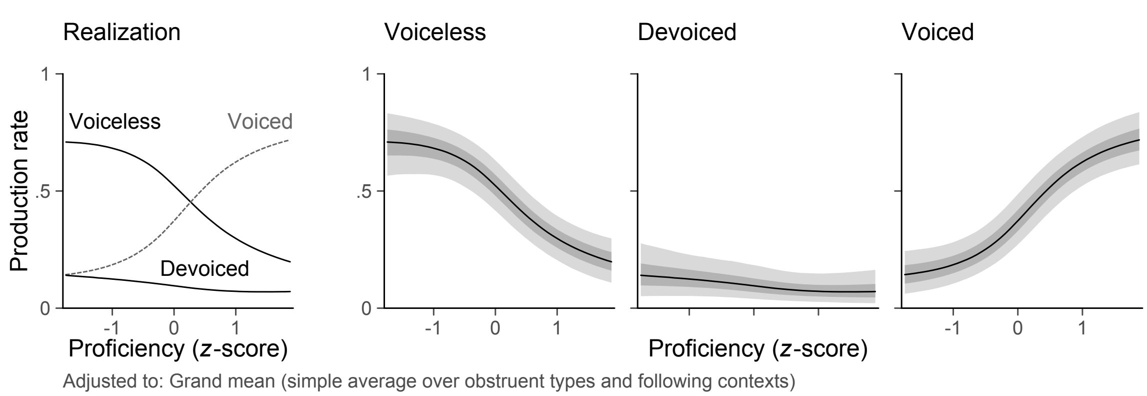
For the auditory data based on the *SpIL* corpus, we also set out with a visual inspection of the production rates observed across the $n = 26$ learners. Figure 10.13 shows, from left to right, the rate at which each learner produced voiceless, devoiced, and voiced variants of obstruents in final position. The share of voiceless renditions varies between about .10 and .80 and decreases steadily across proficiency levels. Devoiced variants appear to play a negligible role, with very few learners showing rates higher than .30; no clear trend across proficiency levels is discernible. Voiced variants, then, show the mirror image of voiceless productions, with rates between .10 and .75 and a steady increase from low to high levels of pronunciation ability.

Let us now turn to our regression analysis for a more compact summary of these trends. Details about the model are relegated to Appendix A.11.[26] We translate the output of our model back to the proportion scale. Figure 10.14 shows how the estimated production rates for the three variants change across proficiency levels. It it in close agreement with the patterns we just noted in Figure 10.13 (estimates are reported in Table 10.8):

[26] The appendix documents convergence diagnostics for the MCMC algorithm and reports posterior estimates for the model parameters.

- Voiceless variants decrease, on average, from about .70 at the lower end of the proficiency scale to about .25 among advanced learners.
- The share of voiced variants increases from about .15 to .70.
- Devoiced renditions remain at a low rate across all proficiency levels, decreasing from about .15 to .05.

Figure 10.14: FVO production by German learners: Estimated rate of variants by proficiency level. Error bands reflect 50% and 90% uncertainty intervals. Data from Pascoe (1987).



	Voiceless			Devoiced			Voiced		
Proficiency	*Mdn*	.05	.95	*Mdn*	.05	.95	*Mdn*	.05	.95
Beginner	.71	.57	.82	.14	.05	.26	.15	.07	.25
Intermediate	.52	.41	.62	.09	.04	.18	.38	.28	.49
Advanced	.23	.14	.33	.07	.02	.15	.69	.59	.80

Table 10.8: Estimates for the production rate of variants at representative proficiency levels (data from this study).

Next, we turn to the following context, which we classified into 5 ordered categories based on the sonority of the adjacent environment. For expository clarity, we restrict our attention to the production rate of voiced variants. Thus, we are interested in how the share of voiced obstruents varies with the sonority of the rightward context. To this end, the top left panel in Figure 10.15 shows how, for each of the five contexts, the share of voiced variants changes across proficiency levels. The dotted line represents the average across all conditions and offers a baseline of comparison. The panels to the right zoom in to the specific contexts, one by one, and add statistical uncertainty estimates. The dotted reference line remains as a benchmark in these panels.

Since curves are notoriously difficulty to compare visually (see Cleveland & McGill 1984), the second row of panels in Figure 10.15 displays differences between the curves shown in the upper row. To this end, the dotted reference profile, which signals the grand mean, serves as a standard of comparison, and the extent to which a given context shows higher- or lower-than-average accuracy rates is signaled by vertical deviations from the horizontal line. The accuracy-rate boost in prevocalic contexts (Vowel), for instance, varies between about +.10 and +.25 in relation to the overall average. Note how parallel curves in the top row need not correspond to stable proportional differences.

Figure 10.15 reaveals the following points:

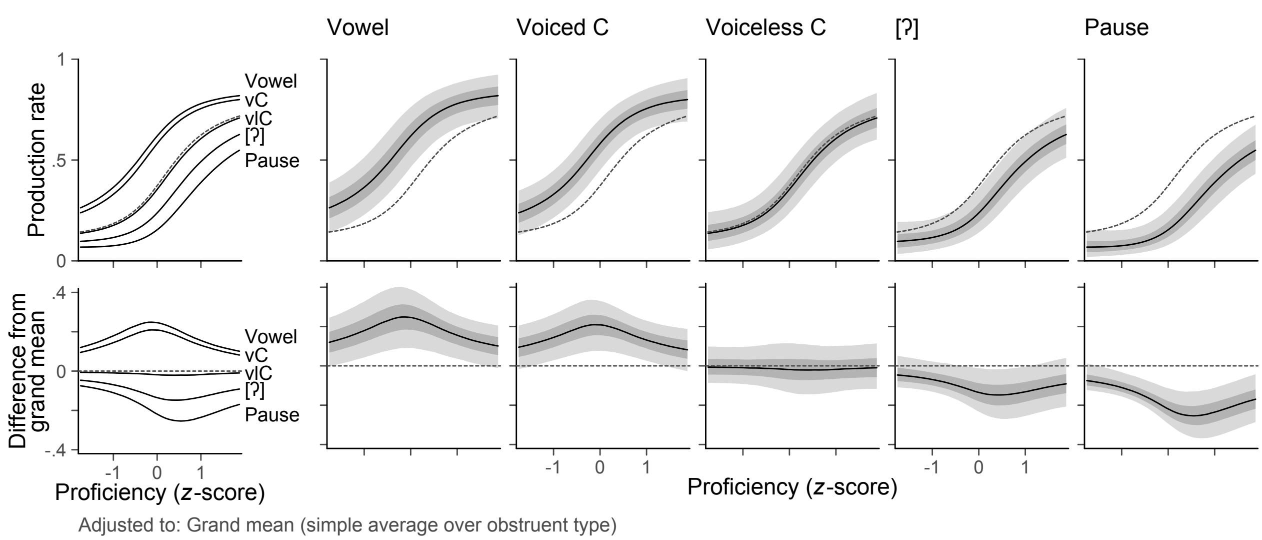


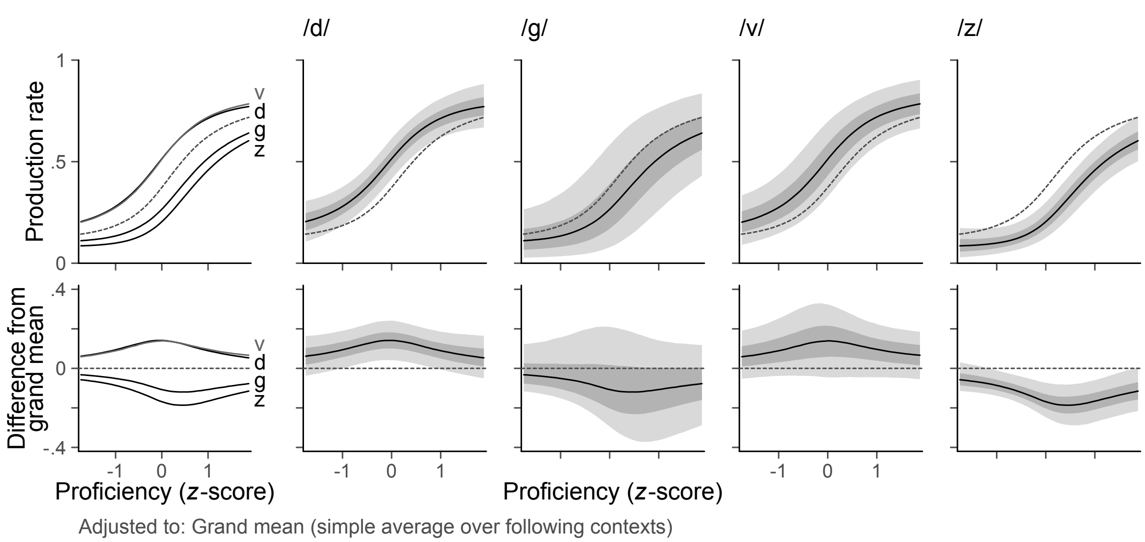
Figure 10.15: The share of voiced FVO renditions as a function of the following context. The top row shows the rate of voiced variants by proficiency in each of the 5 different contexts and the overall average (dotted line). The bottom row displays the deviation of each context from the overall mean across the five contexts (dotted line). Error bands reflect 50% and 90% uncertainty intervals. Data from Pascoe (1987).

- The rank-order of contexts is consistent with the cline in sonority: The higher the sonority of the adjacent segment, the higher the rate of voiced variants.
- Comparing the endpoints along the scale (Vowel and Pause), the net difference ranges between roughly .20 and .45 on the proportion scale.
- Before vowels and voiced consonants, the share of voiced variants is consistently above average. Among low- and high-proficiency learners, it amounts to about +.10, for intermediate learners it is roughly at +.20.
- Before glottal stops and a pause, on the other hand, the share of voiced renditions is below average, by about –.05 ([ʔ]) and –.10 (pause) among beginners, to about –.15 ([ʔ]) and –.20 (pause) at intermediate stages, to roughly –.10 ([ʔ]) and –.20 (pause) at the upper end of the proficiency scale.

Finally, we inspect how the realization of FVOs varied across obstruent classes. Figure 10.16 is constructed the same way as Figure 10.15 and we can turn our attention directly to the patterns in the graphs:

- The amount of vertical displacement among the curves shows that, in comparison to the following context, changes in accuracy rate across obstruent types is somewhat smaller.
- For voiced alveolar /d/, the proportional share of voiced renditions appears to be consistently above average. The difference amounts to about +.05 among beginners and advanced learners to roughly +.15 in the intermediate proficiency ranges.
- For voiced alveolar /z/, on the other hand, the data suggest below-average accuracy rates, with a downward deflection of about –.05 (beginners), –.20 (intermediate), and –.10 (advanced).
- For velar /g/ and labiodental /v/, the error bands indicate considerable uncertainty in the model-based estimates.

Figure 10.16: The share of voiced FVO renditions as a function of obstruent type. The top row shows the rate of voiced variants by proficiency for each of the four obstruent types and the overall average (dotted line). The bottom row displays the deviation of each type from the overall mean across the four obstruent categories (dotted line). Error bands denote 50% and 90% uncertainty intervals. Data from Pascoe (1987).



10.9 Summary and discussion

This chapter was concerned with intrinsic and extrinsic cues to coda obstruent voicing in the production of German learners. A contrastive analysis of English and German revealed differences in the distribution of voiced obstruents. In contrast to English, which upholds a voicing contrast in all positions, German regularly devoices coda obstruents. In English, the most salient and robust cue to final obstruent voicing is the duration of the preceding vowel, which is longer before voiced segments. This systematic variation in vowel duration, which is referred to as the PVD effect, is of considerable magnitude in English compared to other languages including German. A consideration of FVOs with respect to linguistic constraints clearly identified them as marked structures, a classification that finds support from typology, L1 acquisition and physiological facts. PVD as an extrinsic cue, on the other hand, appears to be unconstrained in this regard. Perception studies have shown a general lack of sensitivity to the voicing contrast in coda obstruents. While evidence on German learners' perceptual reliance on the PVD effect for FVO categorization is currently lacking, knowledge gained about learners from other L1s is suggestive of L1 experience with phonemic vowel length contrasts being facilitative for perceptual learning. As for the guidance provided by theoretical contributions, there is consensus about the assumed level of relative difficulty – FVOs are expected to emerge late in L2 acquisition. The theoretical literature also suggests target accuracy to be sensitive to linguistic factors, including obstruent type, sonority of the immediate phonetic context, and structural strength. The PVD effect, on the other hand, is considered to be generally easier to acquire than features internal to the obstruent.

Previous work on the realization of FVOs by L1 German learners is supportive of their classification as difficult structures. Apart from reporting expectedly low accuracy rates, two studies have provided insight into the relevance of linguistic constraints, documenting asymmetries between stops and fricatives as well as place of articulation in stops and, to a lesser extent, also in fricatives (Langguth 2009; Wunder 2012). It was the aim of this study to provide comparable documentation of place and manner effects, and to broaden the scope by further examining the contribution of sonority to the variability in FVO production by German learners. Another objective of this study was to provide insights into the magnitude of the PVD effect across different levels of L2 pronunciation ability and extend the empirical record to early-stage instructional-setting learners.

The analysis of the realization of FVOs revealed sensitivity of learner productions to proficiency, obstruent type, and the sonority of adjacent segments. The rate of target-like productions varied between .15 at lower proficiency levels and about .70 at the highest levels. While these estimates are broadly consistent with the literature, most of the previously reported accuracy rates appear rather low compared to these data (see Table 10.4). This may be partly attributable to differences in speaking style, as it has been revealed by previous studies that more natural styles exhibit higher accuracy rates (e.g. Edge 1991; Major & Faudree 1996). In GLE, stylistic variation has been reported to be negligible, however (Langguth 2009; Wunder 2012). Quantitative discrepancies in research on FVO production may also arise from differences in the auditory classification of variants, as raters may vary in the categorization of learner productions. Given exisiting evidence on perceptual specialization as a result of native language experience, differences between ratings offered by researchers from different L1 backgrounds are likely. This cautions against drawing strong conclusions from the comparison of studies. Stop-fricative asymmetries for alveolar /d z/ amounted to a difference of about .15 to .35 in production accuracy. This cline is consistent with figures reported in the literature (.25 in Wunder 2012; .20 in Langguth 2009). Due to considerably clouded estimates for /g/, the data analyzed here only offer weak indications on place-of-articulation constraints in stops; the direction and magnitude, however, are consistent with previous reports (alveolar > velar, $\Delta = [.10, .25]$; cf. $\Delta = .35$ in Langguth 2009 and $\Delta = .33$ in Wunder 2012). Likewise, only cautious inferences are warranted for the contribution of constriction location in fricatives (/v z/). While coherent with previous work in direction (labiodental > alveolar) the magnitude of the difference estimated here ($\Delta = [.15, .35]$ vs. .05 in and Langguth 2009 and Wunder 2012) could easily be an overestimate considering the wide range of uncertainty surrounding estimated rates for /v/. Finally, sonority had the expected effect, with the order of estimates lining up with the sonority hierarchy. This constraint also appeared to exceed place and manner of articulation in magnitude, with accuracy rates varying by .20 to .45 (in absolute terms) conditional on the sonority of the following segment.

To provide context for an analysis of the PVD ratio by German learners, this study presented a survey of the empirical literature on the systematic variation in preceding vowel variation in BrE, AmE, and German. This review offers a frame of reference for the interpretation of new findings and may prove useful for future studies on the PVD effect in varieties of English. The findings for the PVD ratio for BrE (1.6) and AmE (1.8) native speakers are in accordance with the range of estimates reported in earlier studies. Based on the literature survey, however, the varieties were expected to pattern differently. Nevertheless, this variation in reported estimates is to be expected given the limited amount of quantitative work on BrE (see Figure 10.1) and the small number of speakers recorded in the present study. As for the German learners, PVD ratios were found to increase systematically with pronunciation proficiency, growing from about 1.10 to 1.45. These figures coincide with previously reported ratios, which range from 1.15 to 1.65.

Findings in this investigation are coherent with previous experimental work, lending support to the categorization of FVOs as difficult structures in L2 acquisition. While theory-based predictions concur as to the expected quantitative patterns in learner productions, they offer different explanations for these observations. These include German learners' lack of perceptual sensitivity to final obstruent voicing contrasts (SLM), markedness asymmetries between onsets and codas, voiceless and voiced obstruents, as well as stops and fricatives (MDH/SCH), natural devoicing processes

(NM/NDH), articulatory constraints disfavoring the co-occurrence of phonetic voicing and airflow obstruction (MSA), and the structural weakness of syllable codas (LTD). Thus, findings on the overall accuracy rate and its sensitivity to obstruent type are consistent with multiple process models and consequently lend themselves to different explanations. Importantly, both perception and production constraints can be stipulated to account for these observations. This calls attention to interpretative difficulties that arise in the analysis of speech production data, as these may not be able to discriminate between different hypothesized mechanisms. This is not to suggest that empirical reports on FVO production in non-native speech are of limited value. On the contrary, quantitative work is essential for theoretical insight to be broadened and refined. However, current knowledge about FVO production is lagging behind L2 theory in the sense that there are a number of theoretically motivated constraints that have not received proper attention in applied research. Among these are structural strength, speaking style, frequency, properties of the preceding vowel and cluster complexity.

Nevertheless, the type of variation observed in this study and the literature does suggest FVO production to be constrained by principles independent of speech perception. One reliable indicator for production constraints is the evidence found for adjacent sonority. Phonetic context effects of this kind can only be accounted for by reference to the immediate phonetic environment. This type of within-structure variation cannot be attributed to explanations hinging on intrinsic obstruent attributes.

The findings for German learners' use of the PVD ratio to signal a final obstruent voicing contrast only partially support the expected relative ease of acquisition. On average, only proficient subjects produced ratios exceeding 1.4. It should be noted, however, that the hypothesized ease of acquisition was stated relative to the production of intrinsic FVO features. The insights provided in this study do not permit comparative assessments of the PVD effect and the production of FVOs, however, as these outcomes were investigated in different speaking styles produced by different speakers and analyzed with different methods. Nevertheless, it can be stated that target-like production of the PVD effect (i.e. values greater than 1.3) appears to emerge at upper intermediate levels, with learners at lower proficiency levels almost categorically patterning below this threshold.

For the acquisition of the PVD effect by German learners, guidance provided by theory is limited due to the absence of empirical insights into the perception of the preceding vowel duration as a cue to final obstruent voicing. In order for perception-based predictions to be put on a firmer empirical basis, it is essential for future work to further our understanding of likely perceptual distortions. In the absence of such insights, hypotheses generated by the SLM and other models that build on it (MSA, OPM) remain speculative.

To shed further light on cue weighting in speech production and thereby pursue the theory-guided expectations on the relative accessibility of the PVD effect, future studies may resort to instrumental techniques to study multiple acoustic cues to the final obstruent voicing contrast. The few indications in the literature are suggestive of a bias toward the PVD effect

(Barry 1977), but more work is needed to draw firmer conclusions. Efforts should also be directed toward extending the empirical base on the PVD effect in BrE.

In summary, the following suggestions may serve as a point of departure for future work on FVOs in GLE:

- A central concern should be the elucidation of German learners' reliance on the PVD effect as a perceptual cue to final obstruent voicing.
- Current knowledge about the PVD effect in BrE must be considered unsatisfactory; further empirical contributions are needed.
- The literature survey showed that estimates of the PVD effect vary across studies on the same population of speakers. This heterogeneity should be taken into consideration when passing judgement on the native-likeness of learner productions.
- The PVD effect has been observed to be sensitive to a number of linguistic variables. This suggests the necessity of controlling for factors such as vowel length and height, position in the word or utterance, and speech rate. This may be a concern for studies relying on more natural productions (e.g. corpus data).
- At the same time, systematic variation in the magnitude of the PVD effect provides new impulses for future L2 work, which may set out to investigate similar patterns in learner productions. This line of research should consider focusing on high-proficiency informants, as this population of learners has been documented to exhibit sufficient temporal contrasts that may consequently evidence internal (within-structure) variation.
- Instrumental investigations of the production of final voiced obstruents may contribute to our knowledge about the relative accessibility of the PVD effect by also documenting intrinsic acoustic properties of FVOs.

11
L2 phonological theory and data

Let us now take a step back and consider some broader issues that have emerged in the course of this study. This final chapter brings us back to L2 phonological theory. We first recapitulate the added value of organizing theoretical work from the perspective of IL variation, which aims to directly connect theory and empirical work (§11.1). §11.2 then reflects on the critical role of auxiliary assumptions when deriving empirical expectations from theory. We then change perspective and discuss issues involved in going from data back to theory to make sense of observed patterns (§11.3). §11.4 offers some concluding remarks.

11.1 Interlanguage variation and L2 phonological theory

In line with earlier overviews of theoretical work on IL phonology (e.g. Leather 1999; Gut 2009; Eckman 2012; Pickering 2012), Chapter 2 grouped contributions thematically. We distinguished between ideas related to (i) cross-linguistic influence, (ii) language universals, and (ii) development and variation. In the empirical part of this study, our survey of L2 phonological theories allowed us to formulate expectations about patterns of variation in GLE. It became clear that theories direct their attention to different dimensions of interlanguage variation. For instance, some theories are geared towards explaining why certain structures are more difficult than others, and some theories are concerned with variation in accuracy rates for a single structure. These are two fundamentally different dimensions of variation, which broadly align with intrinsic explanations, that is, explanations referring to properties of the structure, and extrinsic explanations, referring to properties external to the structure. This suggests an alternative, empirically motivated scheme for the organization of theoretical work – one that groups accounts by the type of variation that is described, predicted and explained. To recapitulate, four types of variation were distinguished:

- Variation *between learners*: In this study, between-learner variation was captured with a foreign accent rating, which allowed us to distinguish between different levels of L2 pronunciation ability. Accounting for this type of variation is the objective of research on individual differences in L2 acquisition.

- Variation *within learners*: L2 learners' productions show variation at a single point in time, an important factor being stylistic variation related to the degree of monitoring and focus on form.
- Variation *between structures*: Variation also occurs between structures, as certain sounds may be easier to learn than others. This is attributable to linguistic constraints inherent in the structure.
- Variation *within structures*: Similarly, the production of a single structure may show variation, which may arise from extrinsic constraints, such as phonetic and prosodic context effects.

This four-way distinction establishes links between theoretical work empirical practice. The frameworks discussed in Chapter 2 are summarized in Table 11.1, where they are classified according to theme (as was done in Chapter 2) and the focal type of variation. Given the present study's focus on linguistic constraints, only three kinds of variation are distinguished: between-structure (B), within-structure (W), and stylistic variation (S), the latter corresponding to constraints accounting for within-learner variability. The overview reveals that some contributions are restricted to differences between structures and some are wider in scope. The SCH, FM, SM and OPM offer the most comprehensive accounts of IL variation.

Table 11.1: Overview of theoretical contributions: Major theme and type of variation accounted for. Key to symbols: (L1) cross-linguistic influence, (U) language universals, (DV) development and variation; (B) between-structure variation, (W) within-structure variation, and (S) stylistic variation.

			Theme			Variation		
Contribution		Reference	L1	U	DV	B	W	S
Contrastive Analysis Hypothesis	CAH	Lado 1957	•			•		
Articulatory Settings	AS	Honikman 1964	•			•		
Speech Learning Model	SLM	Flege 1995	•			•		
Similarity Differential Rate Hypothesis	SDRH	Major & Kim 1996	•			•		
Perceptual Assimilation Model	PAM	Best 1995	•			•		
Desensitization Hypothesis	DH	Bohn 1995	•	•		•		
Phonological Interference Model	PIM	Brown 1998	•			•		
Feature Competition Model	FCM	Hancin-Bhatt 1994	•			•		
Markedness Differential Hypothesis	MDH	Eckman 1977		•		•	•	
Structural Conformity Hypothesis	SCH	Eckman 1991		•		•	•	•
Natural Model of L2 Phonological Acquisition	NM	Dziubalska-Kołaczyk 1990		•		•	•	
Naturalness Differential Hypothesis	NDH	Schmid 1997		•		•	•	
Functional Model of Phonological Acquisition	FM	Boersma 1998		•		•	•	•
Model of Segmental Acquisition	MSA	Colantoni & Steele 2008		•		•	•	
Gradual Diffusion Model	GDM	Gatbonton 1978			•		•	
Model of Sociolinguistic Variation	SM	Fasold & Preston 2007			•	•	•	•
Linguistic Theory of L2 Phonologial Development	LTD	James 1988			•		•	
Ontogeny Phylogeny Model	OPM	Major 2001			•	•	•	•
Dynamic Systems Theory	DST	De Bot et al. 2007			•	•	•	

This overview may guide future studies by identifying theoretical work that (i) produces predictions for a particular research problem and (ii) offers potential explanations for observed patterns. For instance, if interest lies in the difficulty of structure A relative to structure B, the set of relevant theories can be narrowed down to those accounting for *between-structure* variation. For studies dealing with the acquisition of a single structure, on the other hand, contributions focusing on *within-structure* variation may provide insights. We can also see how theories may "compete" in the context of a particular phenomenon and offer different explanations for empirical observations. We will return to this point shortly.

Our delineation of dimensions of variation suggests that a clear distinction can be made between *within-structure* and *between-structure* variation. Certain comparisons between phonological structures, however, should also be informed by our knowledge about within-structure variability. To illustrate, consider two new sounds for German learners: dark [ɫ] and /w/. We might be interested in predicting (and understanding) the relative difficulty of these structures for German learners. We could consult theories dealing with between-structure variation and evaluate inherent properties of these segments. In this case, however, the comparison between [ɫ] and /w/ does not offer a pure reflection of intrinsic properties. Rather, we are comparing a segment that is only licensed in onset position (/w/) to one that, in GB at least, only occurs in coda position ([ɫ]). Our between-structure comparison, then, is confounded by onset-coda asymmetries in speech perception and production.

If we take a closer look at the factors underlying the variation between and within structures, we can distinguish between perception and production constraints. Thus, *between-structure* variation may arise from differences in *intrinsic* perceptual salience or similarity effects or *intrinsic* production constraints related to articulatory complexity. Constraints that account for *within-structure* variation bring into focus *extrinsic* influences which are to a certain degree independent of the structure under investigation. Within-structure variation may likewise be rooted in perception or production. For instance, perceptual sensitivity has been observed to vary systematically with prosodic position and onset/coda complexity. Production constraints, on the other hand, may also operate differently in onset vs. coda position, or in singletons vs. clusters.

Table 11.2 provides an overview of intrinsic and extrinsic perception and production constraints, with references to relevant theoretical work. While this inventory does not claim to be exhaustive, it may serve as a point of departure for future empirical and theoretical work. The added value of this overview is that it brings together factors that could offer plausible explanations for IL patterns based on our current state of knowledge. The way in which these constraints shape learner speech varies depending on the type of structure and the L1-TL combination:

- Intrinsic perceptual constraints are tied to the learner's L1, since the "sieve" of the L1 phonological system determines sensitivity to contrasts between TL and L1 structures.
- Extrinsic perception constraints, on the other hand, are to a certain degree independent of the listener's L1 background and may classify as a universal factor underlying L2 speech.
- Intrinsic and extrinsic production constraints can also be assumed to influence non-native speech in largely similar ways and therefore likewise count as universal influences in non-native speech. They may therefore be of broader relevance for the study of L2 speech.

Constraint	Theory
Perception	
Intrinsic constraints	
Perceived phonetic dissimilarity	SLM
Presence of phonological features in the L1	PIM
Prominence of L1 phonological features	FCM
Extrinsic constraints	
Prosodic position: Onset vs. coda	LTD
Complexity: Singleton vs. cluster	LTD
Production	
Intrinsic constraints	
Articulatory complexity	SCH, MSA
L1 articulatory settings	AS
Aerodynamic constraints	SCH, MSA
Extrinsic constraints	
Prosodic position: Onset vs. coda	MDH, SCH
Complexity: Singleton vs. cluster	SCH
Phonetic context	SCH, GDM
Coarticulation	SCH
Structural strength	LTD
Speaking style	FM, OPM

Table 11.2: Intrinsic and extrinsic production and perception constraints.

11.2 From theory to data...

For each of the structures investigated in the present study, we started by consulting theoretical work to formulate empirical expectations. We therefore proceeded deductively, that is, we derived a set of hypotheses from theoretical statements. This brought to our attention two issues to which we now turn: (i) the need to make auxiliary assumptions and (ii) the convergence of predictions.

To distill theory into concrete predictions, we often had to make additional assumptions. For instance, to invoke the concept of markedness, we needed to determine the markedness status of a given structure. For FCM-based predictions, on the other hand, we had to settle on a scheme (and inventory) of phonological features. And expectations from the viewpoint of the MSA required an assessment of the articulatory complexity of a given sound (sequence). We call these assumptions "auxiliary" since they play a secondary role in our reasoning. This is to say that they are not of interest in themselves, but only support the bridge from theory to prediction. They nevertheless assume a critical role since they are essential links in this mental exercise.

Some auxiliary assumptions can be formulated based on empirical observation. This applies to perceptual constraints, for instance, which can be assessed by means of experimental measurements. For other constraints, we rely on the literature as a source of information. This is the case for markedness and naturalness considerations, for instance. Importantly, auxiliary assumptions relating to underlying constraints must often be considered provisional, since they are stated on the basis of limited knowledge; they may have to be refined in the light of new insights. It follows that deductively derived hypotheses may be valid in the sense that they follow logically from a given set of premises. The uncertainty surrounding our auxiliary assumptions, however, affects the degree of assurance we can have in the status of a specific hypothesis, or empirical prediction, as an instantiation of a theory. We should acknowledge assumption uncertainty and its consequences for claims relating empirical findings to L2 phonological theory.

Given the critical role of auxiliary assumptions for communicating between theory and data, they should be stated explicitly. This makes transparent the links in a deductive line of argumentation and allows others to see how we arrive at our claims. This seems to be particularly important when invoking vague frameworks such as the SCH (see §2.3.1 for discussion), whose consequences hinge critically on background assumptions. A change in the set of premises may result in a shift of expected patterns. Consider, for instance, our assumptions about the perceived dissimilarity between clear [l] and dark [ɫ]. In the absence of empirical work on the perception of lateral allophones by German learners, we speculated that the degree of perceived similarity between these sounds is high, which led to the expectation of a prolonged phase of L1 transfer of [l] in non-prevocalic contexts. Future work on L2 German learners' perception of laterals may prompt us to revise this assumption and adjust our empirical expectations.

The second issue we would like to address is the convergence of predictions. As illustrated at several points throughout this study, theoretical contributions may generate identical predictions. When going from theory to expectations about data, this is not a particular concern. Thus, if our sole aim is to predict, say, the degree of difficulty of a particular structure, convergence of predictions would in fact increase our confidence in a given prediction. If, on the other hand, our goal is to understand and explain observations in L2 speech, we face a dilemma. In the next section, we discuss difficulties that arise when drawing conclusions about theory based on data. Before we go further, however, let us summarize our main points so far in the form of a diagram.

The left part of Figure 11.1 shows an illustrative set of L2 phonological theories. These invoke different constraints (which are listed at the left margin) and sketch different processes or mechanisms to account for the sound systems acquired by non-native speakers. As we have illustrated in the preceding chapters, these theoretical notions can be translated into predictions. By prediction we mean a statistical expectation, that is, a quantitative description of the expected patterns in the data.[1] When going from theory to prediction, we have to cross the grey area, which denotes the set of auxiliary assumptions we must make. For some theories, our state of knowledge about these auxiliary premises allows us to navigate unambiguously to a single predicted pattern. In other cases, the uncertainty surrounding our background assumptions may

[1] Predictions can differ in the level of detail they provide. Consider two structures, or conditions, A and B. Quantitative predictions can range in specificity from $A \neq B$ to $A > B$ to $A{:}B = 2{:}1$.

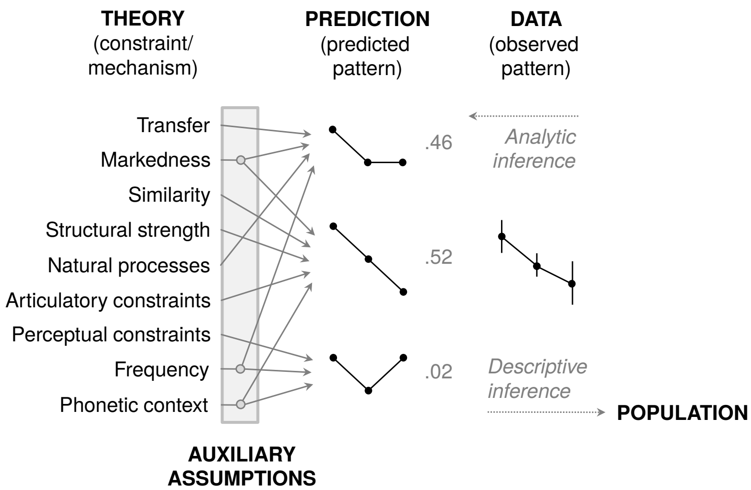


Figure 11.1: L2 phonological theory and data: Schematic illustration.

yield branching arrows to different predicted patterns. Open circles inside the grey rectangle indicate ambiguity as a result of assumption uncertainty. Consider, for instance, markedness statements about final voiced obstruents, which we discussed in §10.2.1. We concluded that intrinsic properties of FVOs must be considered marked. This was in contrast to the preceding vowel duration as an extrinsic cue to the voicing contrast, which arguably classifies as unmarked. This maps onto different expectations about the acquisition of this structure. As a further example, consider phonetic context effects in the production of word-initial /ð/. The feature we focused on was whether /ð/ occurred after a sibilant or not. This was grounded in judgements about syntagmatic articulatory difficulty. However, other properties of the phonetic context may be viewed as relevant. Thus, Trovimovich et al. (2007) examined instead the sonority of the preceding segment, which leads to different predictions.

Figure 11.1 also illustrates the many-to-one mapping of theories and predicted patterns. This situation was observed on numerous occasions throughout this study. This brings us to the next issue we would like to address: the question of what can be inferred from a set of data.

11.3 *... and back again*

When doing an empirical study, we generate a set of data summaries. Figure 11.1 illustrates a pattern we might end up with. Often, we are interested in making inferences based on data. Our inferences can extend into two directions. Let us first consider the arrow pointing rightward. Seeing that our study is based on a sample of limited size, we may wonder whether the patterns we found can be generalized to the broader population of entities represented by our sample. This is a *descriptive* (sometimes called enumerative) task, since the goal of inference is an adequate description of a certain measurable trait in the population of interest. Descriptive inference aims at increasing our *empirical knowledge* about a phenomenon, i.e. the way(s) in which it surfaces in observable statistical patterns.[2]

[2] Statistical tools that may be of help in such inference tasks include estimation (the construction of uncertainty intervals) and hypothesis testing.

The second type of inference points into the other direction, back to theory, or the data-generating mechanism. Such inferences are often labeled *analytic* (or scientific). We are interested in the how the patterns we observe came to be – that is, what process or mechanism generated the data. The goal of analytic inference is usually to increase our *understanding* of a phenomenon and to offer an *explanation* for empirical patterns. We are not content with stating that we found a certain pattern and that it is likely to hold in the underlying population (a descriptive inference). Rather, we would like to say something about why this pattern looks the way it does, why A is greater than B, and so on: We are trying to offer an explanation for our data.

If we look leftward, back to theory, we can compare our observed patterns to the set of predicted patterns we established earlier. There are tools that can help us asses the match between observation and predictions, which may be helpful in deciding which prediction comes closest to the observed facts. In our illustration, we have added exemplary weights, which can be interpreted as the probability that a given predicted pattern, relative to the other predictions in our set, offers the best match to the observed one (see Burnham & Anderson 2002 for examples). In our case, there is a draw between two patterns – the data are unable to differentiate between them.

Even if the data were to identify a relative "winner", however, the task of identifying a likely explanation may be complicated by two factors that were discussed above: For one, a pattern may be compatible with different theories, and therefore with different explanations. We encountered this situation multiple times in this study. We then end up with several plausible explanations for our empirical observations. In other words, our analytic inferences are inconclusive. The second problematic aspect is assumption uncertainty. When going from predicted patterns back to theory, we again have to cross the grey area, that is, we rely on the validity of our auxiliary assumptions. If observational evidence fails to match the predictions we derived from a particular theory, it does not follow logically that the theory provides an inadequate account of the data; the fault may in fact lie with (at least) one of the background premises. The validity of analytic inferences and explanatory statements hinges on these premises.

In the philosophy of science, these issues are linked to the *underdetermination of scientific theory*. Some scholars (e.g. Stanford 2017) distinguish between two types of underdetermination, which correspond to the two issues we have discussed:

- *Contrastive underdetermination* refers to the fact that a single set of data may be consistent with different theories, and therefore different explanations. The diverse landscape of L2 speech theory nicely illustrates this form of underdetermination. It should be noted, however, that contrastive underdetermination also refers to other potential explanations or theories that have not been thought of or formulated yet.
- *Holist underdetermination*, on the other hand, refers to the fact that hypotheses cannot be tested in isolation. Rather, to derive empirical predictions, a (possibly large) set of additional auxiliary assumptions must

be entertained. Theory is then underdetermined in the sense that, upon mismatch between data and prediction, we cannot infer that it was the theory that was in error – it could be the case, in principle, that one of the auxiliary statements was wrong. The additional premises on which we work remain relevant even if prediction and observation match, as this fit may collapse if a refinement of our auxiliary assumptions is in order.

The fact that multiple explanations may be consistent with observed data then creates interpretative problems that may hinder deeper insight into the underlying processes that govern the acquisition of the structure under study. Further, in empirical work setting out to ”test” a particular hypothesis, it may be incautious to report ”support” for the corresponding theory; we should be more circumspect and acknowledge that the observational evidence may be equally consistent with other known theories.

11.4 Concluding remarks

The purpose of research on L2 phonology is to account for patterns of variation found in learner speech. As the survey of theoretical contributions in Chapter 2 illustrated, the field has produced a multitude of theories, which are nested in different schools of thought. Empirical data are a testbed for these theoretical conjectures and therefore provide essential ground on which the field can proceed. In the changing currents of L2 phonological theorizing, data on non-native speech production remain a valuable source of information for future research. One way in which we can contribute sustainably to the field is by making our data available to the scientific community.

To conclude, insights provided by this study suggest that the following points are worthy of attention in future work on IL phonology:

- When looking to theoretical work for guidance, one way to proceed is to determine the type of variation a study investigates. Tables 11.1 and Chapter 2 could then be consulted as pointers to relevant frameworks.
- Theory relies on additional assumptions to explicitly direct empirical work and the interpretation of data. Such premises about constraints should be viewed as preliminary and stated clearly. Assumption uncertainty should warn us against (i) offering definite explanations for empirical patterns and (ii) drawing overly strong conclusions about theory, based on data.
- Judging from the insights provided by this study, contrastive underdetermination seems to be a recurrent issue in L2 phonological research. Observed IL patterns may be consistent with multiple explanations if quantitative findings map equally well onto different theories, or process models.
- Empirical data can fruitfully inform and restrain theory development and are therefore of enduring value to L2 speech research. They should be made available to the scientific community.

A
Appendix

A.1 Materials for the recordings

A.1.1 The current study

Word list people, bad, video, place, animals, Germany, back, voice, exercises, help, plate, change, weather, languages, save, jump, village, control, very good, played, friend, cheese, apples, windows, alive, safe, champions, usually, bottles, visitor, bag, sentences, called, television, always, milk, vegetables, messages, plays, water, June, happy

Phrases The following pairs of words were embedded in the same carrier sentence (I said "...", not "...".):

- *Contrasts*: see/saw, get/got, set/sat, tea/two, heart/hot, kid/could, shoot/shut, sit/sat, put/pot, dark/duck, stars/stores, start/heart
- *Rhymes*: feet/heat, got/hot, good/hood, cat/hat, card/hard, bird/heard, bad/had, sit/hit, bed/head, come/some, zoo/who, Sir/her, door/for
- *Rhymes (with nonce words)*: pet/het, bought/hought, god/hod, food/who'd, but/hut, did/hid, foot/hoot, need/heed

Short dialogues The following pairs of utterances were read out by each subject:

- Do you want to drink something? – Oh, yeah. Can I get another cup of tea, please?
- Where is your friend Peter from? – Peter? He's from the north of Germany.
- Is your brother home? – No. He said that he's going to be back at eight o'clock.
- Oh no! We haven't got any sugar. I want to bake a cake. – No problem. I can get some sugar from the market.
- Is Sally from England? – She lives in England. But she was born in America.
- Can I walk to the city centre from here? – It's too far. You must take the bus to the city centre.
- How was your trip to England? – Great! The weather was sunny and we took a lot of pictures.

- Where is the car? – I parked it at the end of the street.
- What is your sister doing at the moment? – Becky? She's writing an article for the school magazine.
- How can I help you? – Can you tell me the way to the cinema, please?
- Did you like the book? – I did. But the second part of the book was better than the first.

A.1.2 Wunder 2012

Reading passage Harry and Helen are very much in love. Both names start with the letter "H". The couple lives on a farm and enjoys life. They are about the same age. In the morning, Harry usually goes for a walk with the pup. Helen usually feeds the pig. This morning, Harry is lying in bed. He opens his eyes. He takes a breath. He gets up and has breakfast. After breakfast, he puts on his cap. Then he takes his bag. He has to go to work. He is an author, he always wanted to write. Then he leaves the house and takes a cab. He prefers a car ride to a train ride. On the way, he listens to radio plays. After work, Harry and his friends meet near a lake. They always meet at the same place. In summer, this lake is so nice that they come here to bathe. In winter, the friends like skating and so they get onto the ice. Suddenly, Harry falls on his back. He cannot breathe. His friends laugh. Luckily, Harry is not injured, so they decide to make a bet. Who wins the next match? The winner is the girl Madge. After skating, Harry goes home and takes a bath. In the evening, he goes to a pub. There are so many dishes on the menu that he does not know what to pick. After dinner, he goes home and watches a football game that is broadcast live.

A.1.3 Rank 2016

Word list welcome, run, vote, video, ball, develop, university, fast, listen, work, prevent, attempt, individual, quick, aware, conduct, value, environment, participant, previous, heavy, cute, window, television, prison, interrupt, investigation, would, woman, courage, belief, target, violence, sugar, illusion, advantage, interesting, always, unless, weak, huge, often, various, intend, united, we, tissue, responsible, available, very, attitude, anyone, involve, betray, visit, customer, measure, voice, assume, what, judge, love, activity, match, without, gain, convince, marriage, twice, beyond, interview, soul, reveal, victim, faith, treason, tiny, observation, sweat, provide, colour, eventually, though, weight, garbage, velvet, abuse, desire

Reading passage Harry was lifting heavy weights in his basement when the bell rang. Angry because he had to interrupt the workout, he opened the door. "Good evening sir, are you Mr. Harry Preston?", a uniformed man said politely while his partner examined him critically. "Yes", Harry replied confused, swiping off the sweat on his forehead. "Please excuse our late visit, we are here to investigate a case with a missing person. You do know Miss Rachel Coon?" "I do", Harry swallowed twice before continuing, "we've been dating the previous year but we broke up. What happened to her?" "So far we can only assume. However, it looks like your friend has become the victim of a violent crime." Harry started to sweat again. When he didn't react to the officer's words, his partner said: "Maybe we should sit down, would you let us in the living room? We only have a few quick questions." "Alright." Harry's voice was suddenly very weak. "So when was the last time you spoke to Miss Coon?" "Not in weeks. I called her office recently but someone told me she was unavailable." "Are you sure you're telling us the truth?", the second officer said suspiciously. "Because we happen to have the university's observation video from yesterday, showing a woman whose description matches Rachel Coon's. She was wearing a very distinctive velvet scarf. A scarf that looks extremely alike the one that's hidden behind this pillow on your sofa. So I'm just guessing you are somehow involved in this."

A.1.4 Rank 2018

Word list give, sun, until, car, friendship, comment, damage, star, condition, mature, deny, bird, language, secured, success, beware, involve, ignore, official, establish, stairs, anyone, disappeared, classroom, argue, priority, publish, sir, immediately, attend, reassured, previous, prepared, contribution, near, island, scientific, air, solution, analyse, cure, accident, landscape, fur, vegetable, narrow, weird, plenty, scared, honesty, tension, clear, walk

Reading passage "Good morning, Andy!", I greet my favourite resident at the nursing home. Poor Andy has suffered a stroke and has been staying with us for a few weeks to rehabilitate. Just like a lot of elderly people, 80-year-old Andy has this fear of being alone. As far as I can remember, his wife visited him every day since he got here, and she always stayed until a doctor told her to leave. A few minutes ago, I recognised her car on the street and came to tell Andy she's here. Of course, it's my job to take care of him, but you grow pretty attached to the patients after only a short amount of time. I love it when people tell me more about their lives; however, Andy still has a hard time speaking fluently. But he's making progress, I'm sure of it. "Oranges or apples?" He offers me some fruit-flavoured candy. "I prefer apples, thank you." – "That's good, I prefer oranges!" He turns his face to me and grins. It's hard not to stare at the scar on his forehead. From what I know, he has had quite a lot of operations over the years. Part of the reason I like him so much is that despite all the misfortune, he has never lost his positive, up-beat attitude toward life. Meeting someone like that is rare. I wonder if these thoughts ever occur in his mind, if he's aware of how much of an inspiration he is to others. "You're a brave man, Andy." I somehow feel the urge to say that now. I think he didn't even hear it. "Alright, come on! Don't forget to take your scarf with you, it's cold today. These gloves are yours as well." Before I leave, I make sure I haven't left the door unlocked and follow Andy outside.

A.2 Calculation of variety scores

Each learner received a score from –1 to +1 for each of the five features. The variety score is the simple (i.e. unweighted) average over these five scores. To determine the sub-score for a given feature, points were assigned to the individual renditions of relevant items. Table A.1 documents how points were assigned to variants, and which words in the materials were coded. An AmE rendition was coded with +1, a BrE rendition with –1, and a variant that reflects L1 transfer was coded as 0 or "NA" (not available, and excluded from calculations). It was coded as 0 if the L1 variant is distinct from either target variety. The score then signals that the learner does not lean toward one of the accents. If the transferred variant was identical to the BrE or the AmE variant, it was coded as "NA". If all cases for a given feature received "NA"-codes, this feature was dropped from the calculations. This was necessary since we cannot decide whether we should be considering the surface realization as an L1 structure or an L2 structure (in OPM terms, see §2.4). If, however, at least one of the renditions in the set for a feature did indicate one of the target accents (and therefore received a score of +1 or –1), the 'NA's were changed to 0s, to reflect the fact that the target variety is only approximated in a certain proportion of renditions.

Table A.1: Calculation of the variety score for each learner: Coding procedure.

Feature	+1	0	–1	NA	Tokens in data
Clear vs. dark /l/	[ɫ]			[l]	Dialogues: *Sally, lives, a lot, like* Word list: *village, television, languages, alive*
/ɒ/ vs. /ɑː ɔː/	/ɑː/	/ɔ/	/ɒ/		Dialogues: *want to, from, o'clock, got, want, problem, a lot of* Word list: *bottles* Phrases: *got* (2x), *hot* (2x), *pot, god, hod*
/əʊ/ vs. /oʊ/			/əʊ/	/oʊ/	Sentences: *moment, home, oh, no* (2x) Word list: *video, control, windows*
/ɜː/ vs. /ɝː/	/ɝː/	/ œ ɐ/	/ɜː/		Dialogues: *Germany, first* Word list: *Germany* Phrases: *bird, heard, Sir, her*
Postvocalic *r*	[ɹ ɻ]	[ɐ]	[ə] ∅		Dialogues: *Peter, north, part, better* Word list: *visitor, weather, water* Phrases: *heart, card, hard, stars, stores, start, hart, foor, for*

A.3 The TRAP-DRESS contrast

The data sets analyzed in Chapter 4 are openly available via *TROLLing*[1] (Sönning 2020c). The OSF project accompanying this study includes the *Stan* code for the native speakers[2] and the learners[3], and the *R* script[4] for running the analyses.

[1] https://doi.org/10.18710/ATIRRV

[2] https://osf.io/8csqh/

[3] https://osf.io/ju5ct/

[4] https://osf.io/jqhyx/

A.3.1 Native speakers

Model definition For each vowel token, we took three measurements: F_1, F_2, and the log duration of the vocalic interval. The index $[i]$ refers to vowel token i in our data (i.e. one out of the $n = 720$). We used a multivariate regression model, which means that the three quantities were modeled jointly, i.e. as three outcome variables in a single analysis. From a linguistic perspective, this means that each vocalic event was modeled "as a whole". This allows us to take into account potential correlations between these three measures.[5] Further, instead of the default normal distribution, we used a Student-*t* distribution (see Lambert 2018: 159–162) to make the model more robust against outliers, i.e. deviant formant or duration measurements. The ν parameter, which adjusts the tolerance of the distribution to aberrant data points, was set to 7.[6]

[5] A correlation between F_1 and F_2, for instance, would be reflected in a tilted joint distribution of these two measurements. The ellipse describing this distribution in the F_1-by-F_2 plane would therefore be rotated.

[6] See Figure A.3 below for illustration.

$$\begin{bmatrix} \text{F1}_i \\ \text{F2}_i \\ \text{Dn}_i \end{bmatrix} \sim \text{MVStudent-t}\left(\begin{bmatrix} \mu_{\text{F1},i} \\ \mu_{\text{F2},i} \\ \mu_{\text{Dn},i} \end{bmatrix}, \mathbf{S}, \nu \right)$$

The next part describes how the average value of the outcome measurements (F_1, F_2, log duration) varies depending on the vowel (TRAP vs. DRESS), the variety (AmE vs. BrE), and, for the log duration only, the voicing of the following consonant. We made use of index variables (see McElreath 2020: 155) to represent TRAP and DRESS in AmE and BrE, as this makes it easier to specify priors. Each $\bar{\alpha}$ parameter then represents four conditions: DRESS in AmE ($\bar{\alpha}^{\dots}_{\text{AmE, e}}$)[7], TRAP in AmE ($\bar{\alpha}^{\dots}_{\text{AmE, æ}}$), DRESS in BrE ($\bar{\alpha}^{\dots}_{\text{BrE, e}}$), and TRAP in BrE ($\bar{\alpha}^{\dots}_{\text{BrE, æ}}$).[8] The α parameters denote the varying effects for subjects ($\alpha^{\dots}_{\text{s}}$) and words ($\alpha^{\dots}_{\text{w}}$). They represent how much each speaker, or word, deviates, on average, from the conditional mean denoted by the $\bar{\alpha}$ parameters. Here, $\text{s}[i]$ is short for "the subject uttering vowel token i"; $\text{w}[i]$ refers to "the word carrying vowel token i". For the log duration, there is an additional predictor, the voicing of the following consonant. The parameter $\bar{\beta}$ denotes the average difference in log duration between vowels preceding a voiced vs. a voiceless segment. This is the average over all speakers, which means it ignores the target variety. The β_s parameters are the varying coefficients for subjects and express how much each speaker deviates, on average, from the average difference.

[7] The "..." in the notation is a placeholder for the three outcome quantities F_1, F_2, and the log duration.

[8] Recall that the $[i]$-indexes refer to the vowel tokens. The notation $\text{Var}[i]$, then, is short for "the variety of the speaker uttering token i". $\text{Vow}[i]$, in turn, is short for "the vowel category of token i".

$$\begin{aligned}
\mu_{\text{F1},i} &= \bar{\alpha}^{\text{F1}}_{\text{Var}[i],\text{Vow}[i]} + \alpha^{\text{F1}}_{\text{s}[i],\text{Vow}[i]} + \alpha^{\text{F1}}_{\text{w}[i],\text{Var}[i]} \\
\mu_{\text{F2},i} &= \bar{\alpha}^{\text{F2}}_{\text{Var}[i],\text{Vow}[i]} + \alpha^{\text{F2}}_{\text{s}[i],\text{Vow}[i]} + \alpha^{\text{F2}}_{\text{w}[i],\text{Var}[i]} \\
\mu_{\text{Dn},i} &= \bar{\alpha}^{\text{Dn}}_{\text{Var}[i],\text{Vow}[i]} + \alpha^{\text{Dn}}_{\text{s}[i],\text{Vow}[i]} + \alpha^{\text{Dn}}_{\text{w}[i],\text{Var}[i]} + (\bar{\beta}_{\text{Voiced}} + \beta_{\text{s}[i],\text{Voiced}})\text{Voiced}_i
\end{aligned}$$

This next part describes the joint distribution of the varying effects. To save space, we arranged them in multiple columns – they would normally need to be listed in a single column. We used a Student *t*-distribution to make the conditional averages described above more robust to deviant subjects and/or words. The ν parameter was set to 7. **S** refers to the covariance matrix, which describes the dispersion of each varying effect, as well as pairwise correlations between them.

$$\begin{bmatrix} \alpha^{\text{F1}}_{\text{s, e}} & \alpha^{\text{F1}}_{\text{s, æ}} & \beta_{\text{Voiced}} \\ \alpha^{\text{F2}}_{\text{s, e}} & \alpha^{\text{F2}}_{\text{s, æ}} & \\ \alpha^{\text{Dn}}_{\text{s, e}} & \alpha^{\text{Dn}}_{\text{s, æ}} & \end{bmatrix} \sim \text{MVStudent-t}\left(\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & \\ 0 & 0 & \end{bmatrix}, \mathbf{S}_{\text{s}}, \nu_{\text{s}} \right)$$

$$\begin{bmatrix} \alpha^{\text{F1}}_{\text{w, AmE}} & \alpha^{\text{F1}}_{\text{w, BrE}} \\ \alpha^{\text{F2}}_{\text{w, AmE}} & \alpha^{\text{F2}}_{\text{w, BrE}} \\ \alpha^{\text{Dn}}_{\text{w, AmE}} & \alpha^{\text{Dn}}_{\text{w, BrE}} \end{bmatrix} \sim \text{MVStudent-t}\left(\begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \mathbf{S}_{\text{w}}, \nu_{\text{w}} \right)$$

The next part lists the components of the covariance matrix, in the form of standard deviations (for the dispersion of each varying effect) and a correlation matrix **R**, which hosts the correlation coefficients describing all pairwise associations.[9]

[9] We again take the liberty of simplifying notation and omit the operation by which the set of standard deviations and the correlation matrix **R** would be translated into a covariance matrix. See McElreath (2020: 441) for a complete representation.

$$\mathbf{S} = \text{CovarianceMatrix}\left(\sigma_{\text{F1}}, \sigma_{\text{F2}}, \sigma_{\text{Dn}}, \mathbf{R}\right)$$

$$\mathbf{S}_s = \text{CovarianceMatrix}\left(\begin{matrix} \sigma_{\alpha^{\text{F1}}_{\text{s, e}}} & \sigma_{\alpha^{\text{F1}}_{\text{s, æ}}} & \sigma_{\beta_{\text{Voiced}}} \\ \sigma_{\alpha^{\text{F2}}_{\text{s, e}}} & \sigma_{\alpha^{\text{F2}}_{\text{s, æ}}} & \\ \sigma_{\alpha^{\text{Dn}}_{\text{s, e}}} & \sigma_{\alpha^{\text{Dn}}_{\text{s, æ}}} & \end{matrix}, \mathbf{R}_{\text{s}} \right)$$

$$\mathbf{S}_{\text{w}} = \text{CovarianceMatrix}\left(\begin{matrix} \sigma_{\alpha^{\text{F1}}_{\text{w, AmE}}} & \sigma_{\alpha^{\text{F1}}_{\text{w, BrE}}} \\ \sigma_{\alpha^{\text{F2}}_{\text{w, AmE}}} & \sigma_{\alpha^{\text{F2}}_{\text{w, BrE}}} \\ \sigma_{\alpha^{\text{Dn}}_{\text{w, AmE}}} & \sigma_{\alpha^{\text{Dn}}_{\text{w, BrE}}} \end{matrix}, \mathbf{R}_{\text{w}} \right)$$

This concludes the model structure, and we now turn to the specification of the prior distributions over the parameters (see §4.6). The priors for the standard deviation parameters differ for formants and the log duration, since these are measured on different scales. In general, these parameters describe the dispersion of (i) the residuals, (ii) the by-speaker varying intercepts and slopes, and (iii) the by-word varying intercepts. They can be interpreted as roughly signaling the average deviation from the expectation. All standard deviation parameters were given weakly informative priors. Figure A.1a shows the prior for F_1 and F_2, an exponential distribution with scale parameter 2. Vertical lines mark the right boundary of the intervals that contain 50% and 90% of the probability mass. For the formants, these limits are 0.35 and 1.15. These values seems reasonable since formant measurements are scaled using the method proposed by Lobanov (i.e. z-scores). Thus, a value of 1 is unrealistically high, as it would mean that a given dispersion would be equal to the dispersion of all vowel category means along the F_1 or F_2 dimension. Figure A.1a shows the prior for the log duration, an exponential distribution with scale parameter 3. For the log duration, the 50% and 90% probability bounds are 0.23 and 0.77, which corresponds to ratios of roughly 1.30 and 2.20. This prior encodes the assumption that duration ratios do not vary by more than a factor of 2 to 1,

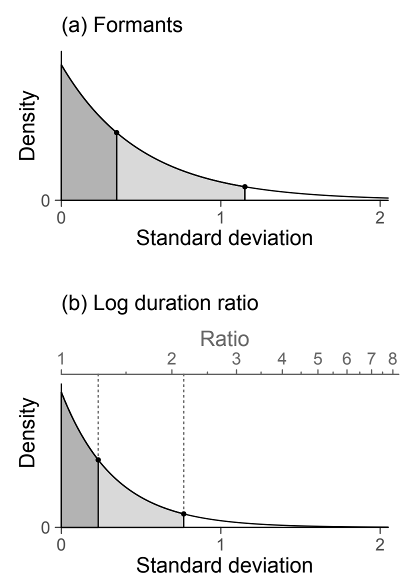


Figure A.1: Prior densities for the standard deviation parameters: (a) for the formants, an exponential distribution with scale parameter 2; and (b) for the log duration ratio, an exponential distribution with scale parameter 3.

on average.

$$\sigma^{\dots}_{\text{F1, F2}} \sim \text{Exponential}(2)$$
$$\sigma^{\dots}_{\text{Dn}} \sim \text{Exponential}(3)$$

Advance estimates for the correlation between the varying effects are difficult to formulate and we therefore default to a fairly diffuse regularizing prior, which is expressed using an LKJ distribution (see Lambert 2018: 185–189) with the shape parameter η set to 4. Figure A.2 shows how the probability mass is concentrated over the interval from –1 to +1. The shaded regions show where the central 50% and 90% of the probability mass are located. These are the intervals [–.24; +.24] and [–.55, +.55], respectively.

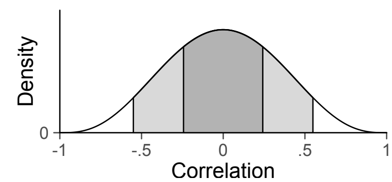


Figure A.2: Prior density for the varying effects correlation parameters: LKJ distribution with shape parameter (η) 4. The shaded areas denote where the central 50% and 90% of the probability mass are located.

$$\mathbf{R}_{\dots} \sim \text{LKJcorr}(4)$$

To make the model more resistant to (i) deviant measurements for individual vowel tokens, (ii) unusual individuals, and (iii) unusual words, we use *t*-distributed rather than normal errors (see Gelman et al. 2014: 437). We decided to set the degrees-of-freedom parameter ν to 7, to embrace the occasional outlier. Figure A.3 shows how this Student-*t* distribution compares to a normal distribution. It has thicker tails, thus accomodating aberrant varying effects estimates.

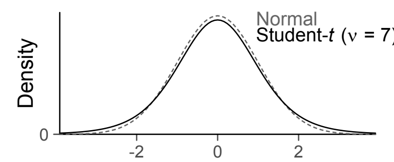


Figure A.3: Comparison of a normal distribution (grey, dotted) with a Student-*t* distribution with $\nu = 7$.

$$\nu_{\dots} = 7$$

The prior densities for the $\bar{\alpha}$ parameters are constructed based on our literature survey (see §4.6 and Figure 4.25).

$$\bar{\alpha}^{\text{F1}}_{\text{BrE, e}} \sim \text{Normal}(0.9, 0.8)$$
$$\bar{\alpha}^{\text{F2}}_{\text{BrE, e}} \sim \text{Normal}(0.4, 0.8)$$
$$\bar{\alpha}^{\text{Dn}}_{\text{BrE, e}} \sim \text{Normal}(4.9, 0.5)$$

$$\bar{\alpha}^{\text{F1}}_{\text{BrE, æ}} \sim \text{Normal}(1.5, 0.8)$$
$$\bar{\alpha}^{\text{F2}}_{\text{BrE, æ}} \sim \text{Normal}(0.45, 0.8)$$
$$\bar{\alpha}^{\text{Dn}}_{\text{BrE, æ}} \sim \text{Normal}(5.2, 0.5)$$

$$\bar{\alpha}^{\text{F1}}_{\text{AmE, e}} \sim \text{Normal}(0.75, 0.8)$$
$$\bar{\alpha}^{\text{F2}}_{\text{AmE, e}} \sim \text{Normal}(0.4, 0.8)$$
$$\bar{\alpha}^{\text{Dn}}_{\text{AmE, e}} \sim \text{Normal}(5.2, 0.5)$$

$$\bar{\alpha}^{\text{F1}}_{\text{AmE, æ}} \sim \text{Normal}(1.2, 0.8)$$
$$\bar{\alpha}^{\text{F2}}_{\text{AmE, æ}} \sim \text{Normal}(0.6, 0.8)$$
$$\bar{\alpha}^{\text{Dn}}_{\text{AmE, æ}} \sim \text{Normal}(5.5, 0.5)$$

Finally, for the $\bar{\beta}$ parameter we chose a very mildly regularizing prior centered at zero.

$$\beta_{\text{Voiced}} \sim \text{Normal}(0,1)$$

Stan code Listed next is the *Stan* code for the native speaker model.

```
data {
  int<lower=0> N;
  vector[3] Y[N];
  int<lower=0> subj[N];
  int<lower=0> word[N];
  int<lower=0> n_subj;
  int<lower=0> n_word;
  int v_id[N];
  int trap[N];
  int dress[N];
  int var_id[N];
  vector[N] voiced;
  int n_varcoef_subj;
  int n_varcoef_word;
  int n_cor_subj;
  int n_cor_word;
  int set_nu;
}
parameters {
  matrix<lower=-1, upper=2>[2,2] a_f1;
  matrix<lower=4, upper=7>[2,2] a_dn;

  real<lower=-1, upper=1> b_voiced;

  vector<lower=0>[n_varcoef_subj] sigma_subj;
  vector<lower=0>[n_varcoef_word] sigma_word;

  matrix[n_varcoef_subj, n_subj] z_subj;
  matrix[n_varcoef_word, n_word] z_word;

  cholesky_factor_corr[n_varcoef_subj] L_Rho_subj;
  cholesky_factor_corr[n_varcoef_word] L_Rho_word;
  cholesky_factor_corr[3] L_Rho_resid;

  real<lower=0> sigma_f1;
  real<lower=0> sigma_f2;
  real<lower=0> sigma_dn;

  real<lower=0> udf_word;
  real<lower=0> udf_subj;
}
transformed parameters {
  matrix[n_subj, n_varcoef_subj] r_s = sqrt(set_nu * udf_subj) * (diag_pre_multiply(sigma_subj, L_Rho_subj) * z_subj)';
  matrix[n_word, n_varcoef_word] r_w = sqrt(set_nu * udf_word) * (diag_pre_multiply(sigma_word, L_Rho_word) * z_word)';

  vector[N] mu_f1;
  vector[N] mu_f2;
  vector[N] mu_dn;

  vector[3] Mu[N];

  vector[3] sigma = [sigma_f1, sigma_f2, sigma_dn]';
  matrix[3,3] Sigma = multiply_lower_tri_self_transpose(diag_pre_multiply(sigma, L_Rho_resid));

  for(n in 1:N) {
    mu_f1[n] = a_f1[var_id[n],v_id[n]] + r_s[subj[n],1]*dress[n] + r_s[subj[n],2]*trap[n] + r_w[word[n],  var_id[n]];
    mu_f2[n] = a_f2[var_id[n],v_id[n]] + r_s[subj[n],3]*dress[n] + r_s[subj[n],4]*trap[n] + r_w[word[n],2+var_id[n]];
    mu_dn[n] = a_dn[var_id[n],v_id[n]] + r_s[subj[n],5]*dress[n] + r_s[subj[n],6]*trap[n] + r_w[word[n],3+var_id[n]] +
      (b_voiced + r_s[subj[n],7])*voiced[n];

    Mu[n] = [mu_f1[n], mu_f2[n], mu_dn[n]]';
  }
}
model {
  sigma_f1 ~ exponential(2);
  sigma_f2 ~ exponential(2);
  sigma_dn ~ exponential(3);

  L_Rho_resid ~ lkj_corr_cholesky(4);
  L_Rho_subj  ~ lkj_corr_cholesky(4);
  L_Rho_word  ~ lkj_corr_cholesky(4);

  sigma_subj[1:4] ~ exponential(2);
  sigma_subj[5:7] ~ exponential(3);
  sigma_word ~ exponential(2);

  to_vector(z_subj)  ~ normal(0,1);
  to_vector(z_word)  ~ normal(0,1);
```

```
  a_f1[2,2] ~ normal( .9 , .8);
  a_f2[2,2] ~ normal( .4 , .8);
  a_dn[2,2] ~ normal(4.9 , .5);
  a_f1[2,1] ~ normal(1.5 , .8);
  a_f2[2,1] ~ normal( .45, .8);
  a_dn[2,1] ~ normal(5.2 , .5);
  a_f1[1,2] ~ normal( .75, .8);
  a_f2[1,2] ~ normal( .4 , .8);
  a_dn[1,2] ~ normal(5.2 , .5);
  a_f1[1,1] ~ normal(1.2 , .8);
  a_f2[1,1] ~ normal( .6 , .8);
  a_dn[1,1] ~ normal(5.5 , .5);
  b_voiced  ~ normal(0,1);

  udf_word ~ inv_chi_square(set_nu);
  udf_subj ~ inv_chi_square(set_nu);

  target += multi_student_t_lpdf(Y | set_nu, Mu, Sigma);
}
generated quantities {
  matrix[n_varcoef_subj, n_varcoef_subj] Rho_subj;
  matrix[n_varcoef_word, n_varcoef_word] Rho_word;
  matrix[3,3] Rho_resid;

  int pos_subj = 1;
  int pos_word = 1;

  vector<lower=-1, upper=1>[3] cor_resid;
  vector<lower=-1, upper=1>[n_cor_subj] cor_subj;
  vector<lower=-1, upper=1>[n_cor_word] cor_word;

  Rho_subj = multiply_lower_tri_self_transpose(L_Rho_subj);
  Rho_word = multiply_lower_tri_self_transpose(L_Rho_word);
  Rho_resid = multiply_lower_tri_self_transpose(L_Rho_resid);

  cor_resid[1] = Rho_resid[1,2];
  cor_resid[2] = Rho_resid[1,3];
  cor_resid[3] = Rho_resid[2,3];

  for(i in 1:(n_varcoef_subj-1)){
    for(j in (i+1):n_varcoef_subj){
      cor_subj[pos_subj] = Rho_subj[i, j];
      pos_subj += 1;
    }
  }
  for(i in 1:(n_varcoef_word-1)){
    for(j in (i+1):n_varcoef_word){
      cor_word[pos_word] = Rho_word[i, j];
      pos_word += 1;
    }
  }
}
```

Convergence diagnostics Our visual diagnostics in Figure A.4 indicate that the MCMC chains seem to have converged. All $\hat{R}$ values are below 1.01, the profiles in the rank plot (middle) are flat and, judging from the quantile plot (right-most display), all key parameters are supported by a sufficiently large effective sample size (i.e. all above $n = 400$).

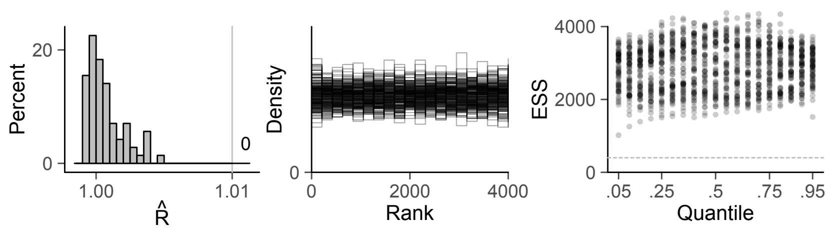


Figure A.4: TRAP-DRESS model for native speakers: Convergence diagnostics.

Posterior distribution of model parameters Table A.2 gives a summary of the posterior distribution of the parameters for the fixed part of the model and Table A.3 reports on the varying effects in the model.

Table A.2: TRAP-DRESS model for the native speakers: Posterior distribution of the fixed effects parameters.

			Posterior quantiles				
Parameter	Mean	SD	.05	.50	.95	N_{eff}	$\hat{R}$
$\bar{\alpha}^{\text{F1}}_{\text{AmE, e}}$	1.45	0.12	1.25	1.45	1.65	1784	1
$\bar{\alpha}^{\text{F1}}_{\text{AmE, e}}$	0.62	0.10	0.47	0.63	0.78	1737	1
$\bar{\alpha}^{\text{F1}}_{\text{BrE, æ}}$	1.52	0.10	1.35	1.52	1.69	1337	1
$\bar{\alpha}^{\text{F1}}_{\text{BrE, æ}}$	0.87	0.08	0.74	0.87	1.00	1734	1
$\bar{\alpha}^{\text{F2}}_{\text{AmE, e}}$	0.57	0.08	0.44	0.57	0.71	1567	1
$\bar{\alpha}^{\text{F2}}_{\text{AmE, e}}$	0.78	0.05	0.69	0.78	0.87	1665	1
$\bar{\alpha}^{\text{F2}}_{\text{BrE, æ}}$	0.09	0.06	–0.01	0.09	0.19	1538	1
$\bar{\alpha}^{\text{F2}}_{\text{BrE, æ}}$	0.66	0.04	0.60	0.66	0.73	1751	1
$\bar{\alpha}^{\text{Dn}}_{\text{AmE, e}}$	5.42	0.05	5.33	5.42	5.50	2870	1
$\bar{\alpha}^{\text{Dn}}_{\text{AmE, e}}$	4.98	0.05	4.90	4.98	5.07	2370	1
$\bar{\alpha}^{\text{Dn}}_{\text{BrE, æ}}$	5.14	0.09	4.99	5.14	5.28	1843	1
$\bar{\alpha}^{\text{Dn}}_{\text{BrE, æ}}$	4.95	0.08	4.81	4.95	5.08	2184	1
$\bar{\beta}_{\text{Voiced}}$	0.28	0.03	0.23	0.28	0.32	2560	1
Residuals							
σ_{F1}	0.23	0.01	0.20	0.23	0.25	3522	1
σ_{F2}	0.12	0.01	0.11	0.12	0.13	3400	1
σ_{Dn}	0.18	0.01	0.16	0.17	0.19	2108	1
$\rho(\text{F1, F2})$	0.07	0.07	–0.05	0.07	0.19	4892	1
$\rho(\text{F1, Dn})$	0.04	0.09	–0.10	0.04	0.18	3600	1
$\rho(\text{Dn, F2})$	–0.09	0.09	–0.22	–0.09	0.06	3796	1

A.3.2 Learners

Model definition The model for learners is very similar in structure to that for the native speakers. We will therefore only comment on differences here.

$$\begin{bmatrix} \text{F1}_i \\ \text{F2}_i \\ \text{Dn}_i \end{bmatrix} \sim \text{MVStudent-t}\left(\begin{bmatrix} \mu_{\text{F1},i} \\ \mu_{\text{F2},i} \\ \mu_{\text{Dn},i} \end{bmatrix}, \mathbf{S}, \nu\right)$$

The deterministic part, which models the expected average for the three outcomes (F_1, F_2, log duration) as a function of our covariates, differs in that there is an extra parameter for pronunciation proficiency (FAR). Let us first discuss the $\bar{\alpha}$ and α parameters, which are structured identically for each outcome. The FAR value of zero, which denotes a learner of intermediate pronunciation ability, is where (by definition) the intercepts are located. These are denoted by the $\bar{\alpha}$ parameters, which are indexed to refer to two values, one for TRAP and one for DRESS The shorthand symbol $\bar{\alpha}^{\text{F1}}$, for instance, indicates the average F_1 value we observe for TRAP among intermediate learners. The α parameters are varying effects and denote speaker-specific deviations from this average, one for TRAP and one for DRESS, and word-specific deviations from this average.

The FAR value of 0 functions as a breakpoint[10]: Learners below this

[10] See Vanhove (2018) for a helpful tutorial on how to code a breakpoint regression model in *Stan*.

Table A.3: TRAP-DRESS model for the native speakers: Posterior distribution of the varying effects parameters.

Parameters	Mean	SD	Posterior quantiles .05	.50	.95	N_{eff}	$\hat{R}$
Varying effects: Subject							
$\sigma_{\alpha^{F1}_{s,e}}$	0.25	0.06	0.16	0.24	0.35	995	1.00
$\sigma_{\alpha^{F2}_{s,e}}$	0.21	0.05	0.14	0.21	0.30	1163	1.00
$\sigma_{\alpha^{Dn}_{s,e}}$	0.18	0.04	0.12	0.18	0.26	1180	1.00
$\sigma_{\alpha^{F1}_{s,æ}}$	0.12	0.03	0.08	0.12	0.17	1192	1.00
$\sigma_{\alpha^{F2}_{s,æ}}$	0.23	0.05	0.15	0.22	0.32	1151	1.00
$\sigma_{\alpha^{Dn}_{s,æ}}$	0.22	0.05	0.15	0.22	0.31	1196	1.00
$\sigma_{\beta_{\text{Voiced}}}$	0.09	0.02	0.06	0.09	0.13	1206	1.00
$\rho(\alpha^{F1}_{e}, \alpha^{F2}_{e})$	0.48	0.12	0.26	0.49	0.67	1775	1.00
$\rho(\alpha^{F1}_{e}, \alpha^{Dn}_{e})$	–0.16	0.14	–0.37	–0.16	0.07	973	1.00
$\rho(\alpha^{F1}_{e}, \alpha^{F1}_{æ})$	0.06	0.14	–0.17	0.06	0.28	1438	1.00
$\rho(\alpha^{F1}_{e}, \alpha^{F2}_{æ})$	0.14	0.13	–0.09	0.14	0.35	1177	1.00
$\rho(\alpha^{F1}_{e}, \alpha^{Dn}_{æ})$	–0.02	0.13	–0.24	–0.02	0.20	1302	1.00
$\rho(\alpha^{F1}_{e}, \beta_{\text{Voiced}})$	0.09	0.15	–0.16	0.09	0.33	2263	1.00
$\rho(\alpha^{F2}_{e}, \alpha^{Dn}_{e})$	–0.05	0.13	–0.27	–0.05	0.17	853	1.00
$\rho(\alpha^{F2}_{e}, \alpha^{F1}_{æ})$	–0.24	0.13	–0.46	–0.25	–0.02	1457	1.00
$\rho(\alpha^{F2}_{e}, \alpha^{F2}_{æ})$	0.17	0.14	–0.07	0.17	0.39	1028	1.01
$\rho(\alpha^{F2}_{e}, \alpha^{Dn}_{æ})$	0.40	0.12	0.19	0.40	0.58	1799	1.00
$\rho(\alpha^{F2}_{e}, \beta_{\text{Voiced}})$	0.09	0.16	–0.17	0.09	0.34	2162	1.00
$\rho(\alpha^{Dn}_{e}, \alpha^{F1}_{æ})$	0.56	0.11	0.37	0.57	0.72	2549	1.00
$\rho(\alpha^{Dn}_{e}, \alpha^{F2}_{æ})$	0.06	0.13	–0.15	0.07	0.27	1680	1.00
$\rho(\alpha^{Dn}_{e}, \alpha^{Dn}_{æ})$	0.01	0.13	–0.19	0.01	0.22	2065	1.00
$\rho(\alpha^{Dn}_{e}, \beta_{\text{Voiced}})$	–0.01	0.15	–0.25	–0.01	0.23	2861	1.00
$\rho(\alpha^{F1}_{æ}, \alpha^{F2}_{æ})$	–0.04	0.13	–0.26	–0.04	0.18	1207	1.00
$\rho(\alpha^{F1}_{æ}, \alpha^{Dn}_{æ})$	–0.23	0.13	–0.43	–0.23	–0.01	1717	1.00
$\rho(\alpha^{F1}_{æ}, \beta_{\text{Voiced}})$	0.14	0.15	–0.11	0.14	0.38	2140	1.00
$\rho(\alpha^{F2}_{æ}, \alpha^{Dn}_{æ})$	0.63	0.09	0.47	0.64	0.78	2233	1.00
$\rho(\alpha^{F2}_{æ}, \beta_{\text{Voiced}})$	–0.21	0.14	–0.44	–0.21	0.03	2767	1.00
$\rho(\alpha^{Dn}_{æ}, \beta_{\text{Voiced}})$	–0.02	0.14	–0.25	–0.03	0.21	2223	1.00
Varying effects: Word							
$\sigma_{\alpha^{F1}_{w,\text{AmE}}}$	0.09	0.04	0.05	0.08	0.16	2032	1.00
$\sigma_{\alpha^{F2}_{w,\text{AmE}}}$	0.09	0.03	0.05	0.08	0.15	1869	1.00
$\sigma_{\alpha^{Dn}_{w,\text{AmE}}}$	0.12	0.05	0.07	0.11	0.21	1848	1.00
$\sigma_{\alpha^{F1}_{w,\text{BrE}}}$	0.03	0.02	0.00	0.02	0.07	1924	1.00
$\sigma_{\alpha^{F2}_{w,\text{BrE}}}$	0.01	0.01	0.00	0.01	0.02	3329	1.00
$\sigma_{\alpha^{Dn}_{w,\text{BrE}}}$	0.06	0.02	0.03	0.06	0.11	2347	1.00
$\rho(\alpha^{F1}_{\text{AmE}}, \alpha^{F2}_{\text{AmE}})$	–0.02	0.23	–0.40	–0.02	0.36	2772	1.00
$\rho(\alpha^{F1}_{\text{AmE}}, \alpha^{Dn}_{\text{AmE}})$	0.02	0.25	–0.39	0.03	0.43	3281	1.00
$\rho(\alpha^{F1}_{\text{AmE}}, \alpha^{F1}_{\text{BrE}})$	0.09	0.27	–0.36	0.10	0.53	6488	1.00
$\rho(\alpha^{F1}_{\text{AmE}}, \alpha^{F2}_{\text{BrE}})$	–0.09	0.27	–0.52	–0.10	0.36	8199	1.00
$\rho(\alpha^{F1}_{\text{AmE}}, \alpha^{Dn}_{\text{BrE}})$	–0.13	0.24	–0.52	–0.14	0.28	4088	1.00
$\rho(\alpha^{F2}_{\text{AmE}}, \alpha^{Dn}_{\text{AmE}})$	0.10	0.23	–0.28	0.11	0.46	3794	1.00
$\rho(\alpha^{F2}_{\text{AmE}}, \alpha^{F1}_{\text{BrE}})$	0.00	0.26	–0.43	0.00	0.41	4931	1.00
$\rho(\alpha^{F2}_{\text{AmE}}, \alpha^{F2}_{\text{BrE}})$	–0.07	0.27	–0.50	–0.08	0.38	6248	1.00
$\rho(\alpha^{F2}_{\text{AmE}}, \alpha^{Dn}_{\text{BrE}})$	0.06	0.23	–0.33	0.06	0.43	4346	1.00
$\rho(\alpha^{Dn}_{\text{AmE}}, \alpha^{F1}_{\text{BrE}})$	–0.10	0.27	–0.52	–0.10	0.34	4719	1.00
$\rho(\alpha^{Dn}_{\text{AmE}}, \alpha^{F2}_{\text{BrE}})$	–0.04	0.28	–0.50	–0.04	0.43	5847	1.00
$\rho(\alpha^{Dn}_{\text{AmE}}, \alpha^{Dn}_{\text{BrE}})$	0.31	0.23	–0.09	0.32	0.66	3028	1.00
$\rho(\alpha^{F1}_{\text{BrE}}, \alpha^{F2}_{\text{BrE}})$	–0.02	0.28	–0.47	–0.02	0.44	4503	1.00
$\rho(\alpha^{F1}_{\text{BrE}}, \alpha^{Dn}_{\text{BrE}})$	–0.13	0.27	–0.55	–0.14	0.31	2538	1.00
$\rho(\alpha^{F2}_{\text{BrE}}, \alpha^{Dn}_{\text{BrE}})$	–0.01	0.28	–0.47	–0.01	0.44	2700	1.00

threshold are treated differently from those above the threshold. This is what the curly brackets in the model formula express: If the FAR score for the learner who produced token i is below 0 (i.e. $z_i < 0$), the first line applies. If it is above 0, the second line applies. Learners below this threshold (i.e. beginner and lower intermediate levels) are considered as representing a single population of German learners. This is to say that target variety orientation is ignored. Accordingly, for each vowel category, there is one regression line per outcome quantity, which describes how the average F_1, F_2, and log duration values change across FAR scores up to 0. Thus, the shorthand symbol $\bar{\beta}^{F2}$ is the average FAR slope (averaged across words) for F_2 – in other words, it expresses how, on average, F_2 varies over FAR scores up to 0. The β parameters are varying slopes and denote word-specific deviations from this average slope.

To the right of the breakpoint, the model allows German learners to gradually form two populations with different target varieties. They share the same intercept (at FAR = 0), which also links them with the average trend to the left of the breakpoint. The regression lines are allowed to diverge from this point onwards. We therefore now have two regression lines, one for BrE-oriented and one for AmE-oriented learners. The $\bar{\beta}$ values are therefore indexed to refer not only to the two vowel categories, but also to these two populations. The parameter $\bar{\beta}^{F2}_{\text{æ, AmE}}$, for example, describes the regression line for TRAP to the right of 0, spoken by learners with an inclination toward AmE. The β parameters are varying slopes and denote word-specific deviations from this average slope. The $\bar{\beta}$ coefficients are prefixed by weights w_i, which express how faithful the speaker uttering vowel token i is to their target variety. These weights range from 0 to 1, where 1 indicates (near-)perfect realization of characteristic features of the target accent.[11]

[11] See Appendix A.2 for details.

$$\mu_{F1,i} = \bar{\alpha}^{F1}_{\text{Vow}[i]} + \alpha^{F1}_{s[i],\text{Vow}[i]} + \alpha^{F1}_{w[i]} + \begin{cases} (\bar{\beta}^{F1}_{\text{Vow}[i]} + \beta^{F1}_{w[i]})\text{FAR}_i & \text{if } z_i < 0 \\ (w_i\bar{\beta}^{F1}_{\text{Vow}[i],\text{Var}[i]} + \beta^{F1}_{w[i]})\text{FAR}_i & \text{if } z_i > 0 \end{cases}$$

$$\mu_{F2,i} = \bar{\alpha}^{F2}_{\text{Vow}[i]} + \alpha^{F2}_{s[i],\text{Vow}[i]} + \alpha^{F2}_{w[i]} + \begin{cases} (\bar{\beta}^{F2}_{\text{Vow}[i]} + \beta^{F2}_{w[i]})\text{FAR}_i & \text{if } z_i < 0 \\ (w_i\bar{\beta}^{F2}_{\text{Vow}[i],\text{Var}[i]} + \beta^{F2}_{w[i]})\text{FAR}_i & \text{if } z_i > 0 \end{cases}$$

$$\begin{aligned} \mu_{Dn,i} = {} & \bar{\alpha}^{Dn}_{\text{Vow}[i]} + \alpha^{Dn}_{s[i],\text{Vow}[i]} + \alpha^{Dn}_{w[i]} + \\ & (\bar{\beta}_{\text{Voiced}} + \beta_{s[i],\text{Voiced}})\text{Voiced}_i + \begin{cases} (\bar{\beta}^{Dn}_{\text{Vow}[i]} + \beta^{Dn}_{w[i]})\text{FAR}_i & \text{if} z_i < 0 \\ (w_i\bar{\beta}^{Dn}_{\text{Vow}[i],\text{Var}[i]} + \beta^{Dn}_{w[i]})\text{FAR}_i & \text{if } z_i > 0 \end{cases} \end{aligned}$$

The varying effects are organized in the same way as in the model for the native speakers.

$$
\begin{bmatrix} \alpha^{F1}_{s,e} & \alpha^{F1}_{s,æ} & \beta_{Voiced} \\ \alpha^{F2}_{s,e} & \alpha^{F2}_{s,æ} & \\ \alpha^{Dn}_{s,e} & \alpha^{Dn}_{s,æ} & \end{bmatrix} \sim \text{MVStudent-t}\left(\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & \\ 0 & 0 & \end{bmatrix}, \mathbf{S}_s, \nu_s \right)
$$

$$
\begin{bmatrix} \alpha^{F1}_{w} & \beta^{F1}_{w} \\ \alpha^{F2}_{w} & \beta^{F2}_{w} \\ \alpha^{Dn}_{w} & \beta^{Dn}_{w} \end{bmatrix} \sim \text{MVStudent-t}\left(\begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \mathbf{S}_w, \nu_w \right)
$$

$$
\mathbf{S} = \text{CovarianceMatrix}\left(\sigma_{F1}, \sigma_{F2}, \sigma_{Dn}, \mathbf{R}\right)
$$

$$
\mathbf{S}_s = \text{CovarianceMatrix}\left(\begin{matrix} \sigma_{\alpha^{F1}_{s,e}} & \sigma_{\alpha^{F1}_{s,æ}} & \sigma_{\beta_{Voiced}} \\ \sigma_{\alpha^{F2}_{s,e}} & \sigma_{\alpha^{F2}_{s,æ}} & \\ \sigma_{\alpha^{Dn}_{s,e}} & \sigma_{\alpha^{Dn}_{s,æ}} & \end{matrix}, \mathbf{R}_s \right)
$$

$$
\mathbf{S}_w = \text{CovarianceMatrix}\left(\begin{matrix} \sigma_{\alpha^{F1}_{w}} & \sigma_{\beta^{F1}_{w}} \\ \sigma_{\alpha^{F2}_{w}} & \sigma_{\beta^{F2}_{w}} \\ \sigma_{\alpha^{Dn}_{w}} & \sigma_{\beta^{Dn}_{w}} \end{matrix}, \mathbf{R}_w \right)
$$

The priors are also constructed along similar lines. See §4.6 for a justification of these a priori expectations. The $\bar{\beta}$ parameters are again weakly regularized via priors centered at zero.

$$
\begin{aligned}
\sigma_{\dots} &\sim \text{Exponential}(2) \\
\mathbf{R}_{\dots} &\sim \text{LKJcorr}(4) \\
\nu_{\dots} &= 7 \\
\bar{\alpha}^{F1}_{\dots} &\sim \text{Normal}(0.5, 1.5) \\
\bar{\alpha}^{F2}_{\dots} &\sim \text{Normal}(0.8, 1.5) \\
\bar{\alpha}^{Dn}_{\dots} &\sim \text{Normal}(5.25, 1) \\
\bar{\beta}^{F1}_{\dots} &\sim \text{Normal}(0, 2) \\
\bar{\beta}^{F2}_{\dots} &\sim \text{Normal}(0, 2) \\
\bar{\beta}^{Dn}_{\dots} &\sim \text{Normal}(0, 2) \\
\bar{\beta}_{Voiced} &\sim \text{Normal}(0,1)
\end{aligned}
$$

Stan code The *Stan* code for the learner model is listed below.

```
data {
  int<lower=0> N;
  vector[3] Y[N];
  int<lower=0> subj[N];
  int<lower=0> word[N];
  int<lower=0> n_subj;
  int<lower=0> n_word;
  int v_id[N];
  int var_id[N];
  real far_z[N];
  real var_score[N];
  vector[N] voiced;
  int n_varcoef_subj;
  int n_varcoef_word;
  int n_cor_subj;
  int n_cor_word;
```

```
  int set_nu;
}
parameters {
  row_vector<lower=-1, upper=2>[2] a_f1;
  row_vector<lower=-1, upper=2>[2] a_f2;
  row_vector<lower= 4, upper=7>[2] a_dn;
  row_vector<lower=-2, upper=2>[2] b_f1;
  row_vector<lower=-2, upper=2>[2] b_f2;
  row_vector<lower=-1, upper=1>[2] b_dn;
  real<lower=-1, upper=1> b_voiced;

  matrix<lower=-1, upper=1>[2,2] b_f1_var;
  matrix<lower=-1, upper=1>[2,2] b_f2_var;
  matrix<lower=-1, upper=1>[2,2] b_dn_var;

  vector<lower=0>[n_varcoef_subj] sigma_subj;
  vector<lower=0>[n_varcoef_word] sigma_word;
  matrix[n_varcoef_subj, n_subj] z_subj;
  matrix[n_varcoef_word, n_word] z_word;

  cholesky_factor_corr[n_varcoef_subj] L_Rho_subj;
  cholesky_factor_corr[n_varcoef_word] L_Rho_word;
  cholesky_factor_corr[3] L_Rho_resid;

  real<lower=0> sigma_f1;
  real<lower=0> sigma_f2;
  real<lower=0> sigma_dn;

  real<lower=0> udf_word;
  real<lower=0> udf_subj;
}
transformed parameters {
  matrix[n_subj, n_varcoef_subj] r_s = sqrt(set_nu * udf_subj) *  (diag_pre_multiply(sigma_subj, L_Rho_subj) * z_subj)';
  matrix[n_word, n_varcoef_word] r_w = sqrt(set_nu * udf_word) *  (diag_pre_multiply(sigma_word, L_Rho_word) * z_word)';
}
model {
  vector[3] sigma = [sigma_f1, sigma_f2, sigma_dn]';
  matrix[3,3] Sigma = multiply_lower_tri_self_transpose(diag_pre_multiply(sigma, L_Rho_resid));
  vector[N] mu_f1;
  vector[N] mu_f2;
  vector[N] mu_dn;
  vector[3] Mu[N];

  sigma_f1 ~ exponential(2);
  sigma_f2 ~ exponential(2);
  sigma_dn ~ exponential(3);

  L_Rho_resid ~ lkj_corr_cholesky(4);
  L_Rho_subj  ~ lkj_corr_cholesky(4);
  L_Rho_word  ~ lkj_corr_cholesky(4);

  sigma_subj[1:4] ~ exponential(2);
  sigma_subj[5:7] ~ exponential(3);
  sigma_word[1:2] ~ exponential(2);
  sigma_word[4:5] ~ exponential(2);
  sigma_word[6] ~ exponential(3);
  sigma_word[3] ~ exponential(3);

  to_vector(z_subj) ~ normal(0,1);
  to_vector(z_word) ~ normal(0,1);

  to_vector(a_f1) ~ normal(0.5, 1);
  to_vector(a_f2) ~ normal(0.7, 1);
  to_vector(a_dn) ~ normal(5.25,1);
  to_vector(b_f1) ~ normal(0,1);
  to_vector(b_f2) ~ normal(0,1);
  to_vector(b_dn) ~ normal(0,1);
  to_vector(b_f1_var) ~ normal(0,1);
  to_vector(b_f2_var) ~ normal(0,1);
  to_vector(b_dn_var) ~ normal(0,1);
  b_voiced ~ normal(0,1);

  udf_word ~ inv_chi_square(set_nu);
  udf_subj ~ inv_chi_square(set_nu);

  for (n in 1:N) {
    if(far_z[n] < 0){
      mu_f1[n] = a_f1[v_id[n]]  + (b_f1[v_id[n]] + r_w[word[n],4])*far_z[n] +  r_s[subj[n], v_id[n]]   + r_w[word[n], 1];
      mu_f2[n] = a_f2[v_id[n]]  + (b_f2[v_id[n]] + r_w[word[n],5])*far_z[n] +  r_s[subj[n], v_id[n]+2] + r_w[word[n], 2];
      mu_dn[n] = a_dn[v_id[n]]  + (b_dn[v_id[n]] + r_w[word[n],6])*far_z[n] +  r_s[subj[n], v_id[n]+4] + r_w[word[n], 3]  + (
           b_voiced + r_s[subj[n], 7])*voiced[n];
```

```
      Mu[n] = [mu_f1[n], mu_f2[n], mu_dn[n]]';
        } else {
      mu_f1[n] = a_f1[v_id[n]]  + (b_f1_var[var_id[n], v_id[n]]*var_score[n] + r_w[word[n],4])*far_z[n] +
        r_s[subj[n], v_id[n]]   + r_w[word[n], 1];
      mu_f2[n] = a_f2[v_id[n]]  + (b_f2_var[var_id[n], v_id[n]]*var_score[n] + r_w[word[n],5])*far_z[n] +
        r_s[subj[n], v_id[n]+2] + r_w[word[n], 2];
      mu_dn[n] = a_dn[v_id[n]]  + (b_dn_var[var_id[n], v_id[n]]*var_score[n] + r_w[word[n],6])*far_z[n] +
        r_s[subj[n], v_id[n]+4] + r_w[word[n], 3]
        + (b_voiced + r_s[subj[n], 7])*voiced[n];

      Mu[n] = [mu_f1[n], mu_f2[n], mu_dn[n]]';
    }
  }
  target += multi_student_t_lpdf(Y | set_nu, Mu, Sigma);
}
generated quantities {
  matrix[n_varcoef_subj, n_varcoef_subj] Rho_subj;
  matrix[n_varcoef_word, n_varcoef_word] Rho_word;
  matrix[3,3] Rho_resid;

  int pos_subj = 1;
  int pos_word = 1;

  vector<lower=-1, upper=1>[3] cor_resid;
  vector<lower=-1, upper=1>[n_cor_subj] cor_subj;
  vector<lower=-1, upper=1>[n_cor_word] cor_word;

  Rho_subj = multiply_lower_tri_self_transpose(L_Rho_subj);
  Rho_word = multiply_lower_tri_self_transpose(L_Rho_word);
  Rho_resid = multiply_lower_tri_self_transpose(L_Rho_resid);

  cor_resid[1] = Rho_resid[1,2];
  cor_resid[2] = Rho_resid[1,3];
  cor_resid[3] = Rho_resid[2,3];

  for(i in 1:(n_varcoef_subj-1)){
        for(j in (i+1):n_varcoef_subj){
          cor_subj[pos_subj] = Rho_subj[i, j];
          pos_subj += 1;
    }
  }
  for(i in 1:(n_varcoef_word-1)){
        for(j in (i+1):n_varcoef_word){
          cor_word[pos_word] = Rho_word[i, j];
          pos_word += 1;
    }
  }
}
```

Convergence diagnostics The diagnostics in Figure A.5 indicate that the algorithm appears to have converged. A few $\hat{R}$ values (left column) are above the 1.01 benchmark. All of these are auxiliary parameters and therefore not central to our model interpretation. The profile of the rank plot (middle) is flat and, judging from the quantile plot (right-most display), all key parameters are supported by a sufficiently large effective sample size (i.e. all above $n = 400$).

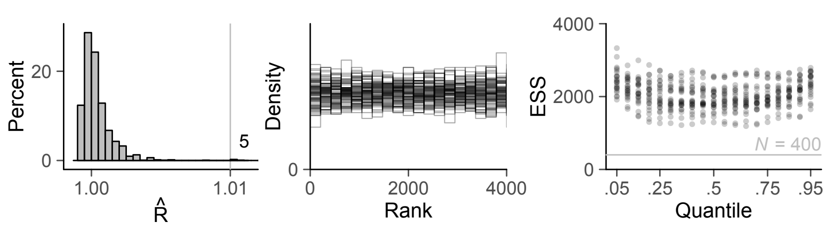


Figure A.5: TRAP-DRESS model for the learners: Convergence diagnostics.

Posterior distribution of model parameters Table A.4 gives a summary of the posterior distribution of the key parameters in the model and Table A.5 reports the posterior distribution of the varying effects parameters.

Table A.4: TRAP-DRESS model for German learner: Posterior distribution of the fixed effects parameters.

			Posterior quantiles				
Parameter	Mean	SD	.05	.50	.95	N_{eff}	$\hat{R}$
$\bar{\alpha}_{\text{e}}^{\text{F1}}$	0.78	0.06	0.69	0.78	0.88	1370	1
$\bar{\alpha}_{\text{æ}}^{\text{F1}}$	0.65	0.05	0.56	0.65	0.73	1442	1
$\bar{\alpha}_{\text{e}}^{\text{F2}}$	0.94	0.05	0.86	0.94	1.02	979	1
$\bar{\alpha}_{\text{æ}}^{\text{F2}}$	0.99	0.04	0.93	0.99	1.05	1332	1
$\bar{\alpha}_{\text{e}}^{\text{Dn}}$	5.13	0.06	5.03	5.13	5.23	1307	1
$\bar{\alpha}_{\text{æ}}^{\text{Dn}}$	5.02	0.06	4.92	5.01	5.12	1288	1
$\bar{\beta}_{\text{e}}^{\text{F1}}$	0.14	0.07	0.03	0.14	0.25	1380	1
$\bar{\beta}_{\text{æ}}^{\text{F1}}$	0.09	0.06	–0.01	0.09	0.18	1570	1
$\bar{\beta}_{\text{e}}^{\text{F2}}$	0.03	0.04	–0.04	0.03	0.10	952	1
$\bar{\beta}_{\text{æ}}^{\text{F2}}$	0.05	0.03	–0.01	0.05	0.10	1294	1
$\bar{\beta}_{\text{e}}^{\text{Dn}}$	–0.10	0.06	–0.20	–0.10	0.00	1312	1
$\bar{\beta}_{\text{æ}}^{\text{Dn}}$	–0.02	0.06	–0.12	–0.02	0.08	1385	1
$\bar{\alpha}_{\text{e, BrE}}^{\text{F1}}$	0.35	0.11	0.17	0.35	0.53	2210	1
$\bar{\alpha}_{\text{e, AmE}}^{\text{F1}}$	0.18	0.10	0.02	0.18	0.33	2139	1
$\bar{\alpha}_{\text{æ, BrE}}^{\text{F1}}$	0.45	0.07	0.33	0.45	0.56	1677	1
$\bar{\alpha}_{\text{æ, AmE}}^{\text{F1}}$	0.08	0.07	–0.03	0.08	0.19	1738	1
$\bar{\alpha}_{\text{e, BrE}}^{\text{F2}}$	–0.21	0.08	–0.33	–0.21	–0.08	1400	1
$\bar{\alpha}_{\text{e, AmE}}^{\text{F2}}$	–0.14	0.06	–0.24	–0.14	–0.05	1999	1
$\bar{\alpha}_{\text{æ, BrE}}^{\text{F2}}$	-0.27	0.05	–0.36	–0.27	–0.19	1204	1
$\bar{\alpha}_{\text{æ, AmE}}^{\text{F2}}$	–0.12	0.04	–0.18	–0.12	–0.05	1695	1
$\bar{\alpha}_{\text{e, BrE}}^{\text{Dn}}$	0.10	0.10	–0.06	0.10	0.26	1711	1
$\bar{\alpha}_{\text{e, AmE}}^{\text{Dn}}$	0.02	0.10	–0.14	0.03	0.18	1806	1
$\bar{\alpha}_{\text{æ, BrE}}^{\text{Dn}}$	–0.01	0.07	–0.12	–0.01	0.11	1464	1
$\bar{\alpha}_{\text{æ, AmE}}^{\text{Dn}}$	–0.08	0.07	–0.18	–0.08	0.03	1731	1
$\bar{\beta}_{\text{Voiced}}$	0.23	0.03	0.17	0.23	0.28	1602	1
Residuals							
σ_{F1}	0.21	0.01	0.20	0.21	0.22	5826	1
σ_{F2}	0.11	0.00	0.10	0.11	0.12	5946	1
σ_{Dn}	0.14	0.01	0.13	0.14	0.15	3645	1
$\rho(\text{F1, F2})$	–0.01	0.04	–0.08	–0.01	0.06	5846	1
$\rho(\text{F1, Dn})$	0.05	0.05	–0.03	0.05	0.12	5015	1
$\rho(\text{Dn, F2})$	0.13	0.05	0.05	0.13	0.21	5168	1

Table A.5: TRAP-DRESS model for the German learners: Posterior distribution of the varying effects parameters.

Parameters	Mean	SD	.05	.50	.95	N_{eff}	$\hat{R}$
Subject: Standard deviation parameters							
$\sigma_{\alpha^{\text{F1}}_{\text{s, e}}}$	0.25	0.06	0.16	0.24	0.35	995	1.00
$\sigma_{\alpha^{\text{F2}}_{\text{s, e}}}$	0.21	0.05	0.14	0.21	0.30	1163	1.00
$\sigma_{\alpha^{\text{Dn}}_{\text{s, e}}}$	0.18	0.04	0.12	0.18	0.26	1180	1.00
$\sigma_{\alpha^{\text{F1}}_{\text{s, æ}}}$	0.12	0.03	0.08	0.12	0.17	1192	1.00
$\sigma_{\alpha^{\text{F2}}_{\text{s, æ}}}$	0.23	0.05	0.15	0.22	0.32	1151	1.00
$\sigma_{\alpha^{\text{Dn}}_{\text{s, æ}}}$	0.22	0.05	0.15	0.22	0.31	1196	1.00
$\sigma_{\beta_{\text{Voiced}}}$	0.09	0.02	0.06	0.09	0.13	1206	1.00
Subject: Correlation parameters							
$\rho(\alpha^{\text{F1}}_{\text{e}}, \alpha^{\text{F2}}_{\text{e}})$	0.48	0.12	0.26	0.49	0.67	1775	1.00
$\rho(\alpha^{\text{F1}}_{\text{e}}, \alpha^{\text{Dn}}_{\text{e}})$	–0.16	0.14	–0.37	–0.16	0.07	973	1.00
$\rho(\alpha^{\text{F1}}_{\text{e}}, \alpha^{\text{F1}}_{\text{æ}})$	0.06	0.14	–0.17	0.06	0.28	1438	1.00
$\rho(\alpha^{\text{F1}}_{\text{e}}, \alpha^{\text{F2}}_{\text{æ}})$	0.14	0.13	–0.09	0.14	0.35	1177	1.00
$\rho(\alpha^{\text{F1}}_{\text{e}}, \alpha^{\text{Dn}}_{\text{æ}})$	–0.02	0.13	–0.24	–0.02	0.20	1302	1.00
$\rho(\alpha^{\text{F1}}_{\text{e}}, \beta_{\text{Voiced}})$	0.09	0.15	–0.16	0.09	0.33	2263	1.00
$\rho(\alpha^{\text{F2}}_{\text{e}}, \alpha^{\text{Dn}}_{\text{e}})$	–0.05	0.13	–0.27	–0.05	0.17	853	1.00
$\rho(\alpha^{\text{F2}}_{\text{e}}, \alpha^{\text{F1}}_{\text{æ}})$	–0.24	0.13	–0.46	–0.25	–0.02	1457	1.00
$\rho(\alpha^{\text{F2}}_{\text{e}}, \alpha^{\text{F2}}_{\text{æ}})$	0.17	0.14	–0.07	0.17	0.39	1028	1.01
$\rho(\alpha^{\text{F2}}_{\text{e}}, \alpha^{\text{Dn}}_{\text{æ}})$	0.40	0.12	0.19	0.40	0.58	1799	1.00
$\rho(\alpha^{\text{F2}}_{\text{e}}, \beta_{\text{Voiced}})$	0.09	0.16	–0.17	0.09	0.34	2162	1.00
$\rho(\alpha^{\text{Dn}}_{\text{e}}, \alpha^{\text{F1}}_{\text{æ}})$	0.56	0.11	0.37	0.57	0.72	2549	1.00
$\rho(\alpha^{\text{Dn}}_{\text{e}}, \alpha^{\text{F2}}_{\text{æ}})$	0.06	0.13	–0.15	0.07	0.27	1680	1.00
$\rho(\alpha^{\text{Dn}}_{\text{e}}, \alpha^{\text{Dn}}_{\text{æ}})$	0.01	0.13	–0.19	0.01	0.22	2065	1.00
$\rho(\alpha^{\text{Dn}}_{\text{e}}, \beta_{\text{Voiced}})$	–0.01	0.15	–0.25	–0.01	0.23	2861	1.00
$\rho(\alpha^{\text{F1}}_{\text{æ}}, \alpha^{\text{F2}}_{\text{æ}})$	–0.04	0.13	–0.26	–0.04	0.18	1207	1.00
$\rho(\alpha^{\text{F1}}_{\text{æ}}, \alpha^{\text{Dn}}_{\text{æ}})$	–0.23	0.13	–0.43	–0.23	–0.01	1717	1.00
$\rho(\alpha^{\text{F1}}_{\text{æ}}, \beta_{\text{Voiced}})$	0.14	0.15	–0.11	0.14	0.38	2140	1.00
$\rho(\alpha^{\text{F2}}_{\text{æ}}, \alpha^{\text{Dn}}_{\text{æ}})$	0.63	0.09	0.47	0.64	0.78	2233	1.00
$\rho(\alpha^{\text{F2}}_{\text{æ}}, \beta_{\text{Voiced}})$	–0.21	0.14	–0.44	–0.21	0.03	2767	1.00
$\rho(\alpha^{\text{Dn}}_{\text{æ}}, \beta_{\text{Voiced}})$	–0.02	0.14	–0.25	–0.03	0.21	2223	1.00
Word: Standard deviation parameters							
$\sigma_{\alpha^{\text{F1}}_{\text{w, AmE}}}$	0.09	0.04	0.05	0.08	0.16	2032	1.00
$\sigma_{\alpha^{\text{F2}}_{\text{w, AmE}}}$	0.09	0.03	0.05	0.08	0.15	1869	1.00
$\sigma_{\alpha^{\text{Dn}}_{\text{w, AmE}}}$	0.12	0.05	0.07	0.11	0.21	1848	1.00
$\sigma_{\alpha^{\text{F1}}_{\text{w, BrE}}}$	0.03	0.02	0.00	0.02	0.07	1924	1.00
$\sigma_{\alpha^{\text{F2}}_{\text{w, BrE}}}$	0.01	0.01	0.00	0.01	0.02	3329	1.00
$\sigma_{\alpha^{\text{Dn}}_{\text{w, BrE}}}$	0.06	0.02	0.03	0.06	0.11	2347	1.00
Word: Correlation parameters							
$\rho(\alpha^{\text{F1}}_{\text{AmE}}, \alpha^{\text{F2}}_{\text{AmE}})$	–0.02	0.23	–0.40	–0.02	0.36	2772	1.00
$\rho(\alpha^{\text{F1}}_{\text{AmE}}, \alpha^{\text{Dn}}_{\text{AmE}})$	0.02	0.25	–0.39	0.03	0.43	3281	1.00
$\rho(\alpha^{\text{F1}}_{\text{AmE}}, \alpha^{\text{F1}}_{\text{BrE}})$	0.09	0.27	–0.36	0.10	0.53	6488	1.00
$\rho(\alpha^{\text{F1}}_{\text{AmE}}, \alpha^{\text{F2}}_{\text{BrE}})$	–0.09	0.27	–0.52	–0.10	0.36	8199	1.00
$\rho(\alpha^{\text{F1}}_{\text{AmE}}, \alpha^{\text{Dn}}_{\text{BrE}})$	–0.13	0.24	–0.52	–0.14	0.28	4088	1.00
$\rho(\alpha^{\text{F2}}_{\text{AmE}}, \alpha^{\text{Dn}}_{\text{AmE}})$	0.10	0.23	–0.28	0.11	0.46	3794	1.00
$\rho(\alpha^{\text{F2}}_{\text{AmE}}, \alpha^{\text{F1}}_{\text{BrE}})$	0.00	0.26	–0.43	0.00	0.41	4931	1.00
$\rho(\alpha^{\text{F2}}_{\text{AmE}}, \alpha^{\text{F2}}_{\text{BrE}})$	–0.07	0.27	–0.50	–0.08	0.38	6248	1.00
$\rho(\alpha^{\text{F2}}_{\text{AmE}}, \alpha^{\text{Dn}}_{\text{BrE}})$	0.06	0.23	–0.33	0.06	0.43	4346	1.00
$\rho(\alpha^{\text{Dn}}_{\text{AmE}}, \alpha^{\text{F1}}_{\text{BrE}})$	–0.10	0.27	–0.52	–0.10	0.34	4719	1.00
$\rho(\alpha^{\text{Dn}}_{\text{AmE}}, \alpha^{\text{F2}}_{\text{BrE}})$	–0.04	0.28	–0.50	–0.04	0.43	5847	1.00
$\rho(\alpha^{\text{Dn}}_{\text{AmE}}, \alpha^{\text{Dn}}_{\text{BrE}})$	0.31	0.23	–0.09	0.32	0.66	3028	1.00
$\rho(\alpha^{\text{F1}}_{\text{BrE}}, \alpha^{\text{F2}}_{\text{BrE}})$	–0.02	0.28	–0.47	–0.02	0.44	4503	1.00
$\rho(\alpha^{\text{F1}}_{\text{BrE}}, \alpha^{\text{Dn}}_{\text{BrE}})$	–0.13	0.27	–0.55	–0.14	0.31	2538	1.00
$\rho(\alpha^{\text{F2}}_{\text{BrE}}, \alpha^{\text{Dn}}_{\text{BrE}})$	–0.01	0.28	–0.47	–0.01	0.44	2700	1.00

A.4 Laterals

The data sets analyzed in Chapter 5 are openly available via *TROLLing*[12] (Sönning 2020a). The OSF repository hosts the *Stan* code[13] and the *R* script[14] for running the analyses.

[12] https://doi.org/10.18710/G6PJ5F

[13] For the native speakers, see https://osf.io/b4f7g/, for the learners see https://osf.io/94jsm/.

[14] https://osf.io/ghbm3/

A.4.1 Native speakers

Model definition We model the distribution of the Bark difference (Δ_{Bk}) scores. The subscript i refers to the individual data points, i.e. the lateral tokens measured. Bark difference$_i$ therefore refers to the outcome quantity for token i. We use a Student-t distribution with ν set to 7.[15]

[15] See Appendix A.3.1 and Figure A.3 (p. 253) for illustration.

$$\text{Bark difference}_i \sim \text{Student-t}\left(\mu_i, \sigma, \nu\right)$$

The deterministic part of the model describes how the average Δ_{Bk} score varies with the variety of the speaker (AmE vs. BrE) and the position of the lateral (prevocalic vs. non-prevocalic). There are four conditional means, which are denoted by the $\bar{\alpha}$ parameters. The $[i]$s index, for each lateral token, the variety of the speaker and the position of the segment. "Variety$[i]$", then, is short for "the variety of the speaker uttering lateral token i".[16] The α parameters denote varying effects on subjects and words. These express by how much each speaker deviates from the conditional average in prevocalic position ($\alpha_{\text{s}}^{\text{pre}}$) and non-prevocalic position ($\alpha_{\text{s}}^{\text{non}}$), and by how much each word deviates from the conditional average for AmE speakers ($\alpha_{\text{w}}^{\text{AmE}}$) and for BrE speakers ($\alpha_{\text{w}}^{\text{AmE}}$).

[16] Thus, $\bar{\alpha}_{\text{AmE,pre}}$ refers to the average Δ_{Bk} score for a prevocalic lateral produced by an AmE speaker.

$$\mu_i = \bar{\alpha}_{\text{Variety}[i],\text{Position}[i]} + \alpha_{\text{s}[i],\text{Position}[i]} + \alpha_{\text{w}[i],\text{Variety}[i]}$$

The next block specifies the joint distribution of the varying effects, which are given a multivariate Student-t distribution with ν set to 7.[17] **S** denotes the covariance matrix, which contains information about the dispersion of the varying effects and their pairwise association.

[17] See Appendix A.3.1 and Figure A.3 (p. 253) for illustration.

$$\begin{bmatrix} \alpha_{\text{s}}^{\text{pre}} \\ \alpha_{\text{s}}^{\text{non}} \end{bmatrix} \sim \text{MVStudent-t}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mathbf{S}_{\text{s}}, \nu_{\text{s}}\right)$$
$$\begin{bmatrix} \alpha_{\text{w}}^{\text{AmE}} \\ \alpha_{\text{w}}^{\text{BrE}} \end{bmatrix} \sim \text{MVStudent-t}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mathbf{S}_{\text{w}}, \nu_{\text{w}}\right)$$

The covariance matrices are defined using the standard deviation of the varying effects and their pairwise correlations. Thus, the parameter $\sigma_{\text{s}}^{\text{pre}}$ is a standard deviation (on the Δ_{Bk} scale), which describes how much speaker-specific averages vary around the conditional means for prevocalic contexts. **R** is a correlation matrix, which includes a correlation coefficient that captures how by-subject varying effects on the one hand, and by-word varying effects on the other, are correlated.[18]

[18] A positive correlation for the speaker-specific sets of varying effects would indicate that a speaker who produces a lower/higher degree of velarization in prevocalic position is also likely to show a lower/higher degree of velarization in non-prevocalic position. In other words, it would mean that speakers can be overall "darker" or "lighter". Note that a positive correlation also arises if Δ_{Bk} scores do not completely eradicate the effect of physiological differences between speakers. A positive correlation between the word-specific sets of varying effects, on the other hand, would indicate that if a given word shows a lower/higher degree of velarization in AmE, it is likely to also show a lower/higher degree of velarization in BrE. This could be due, for instance, to coarticulatory effects induced by adjacent segments, which we expect to operate similarly across varieties.

$$\mathbf{S}_{\text{s}} = \text{CovarianceMatrix}\left(\sigma_{\text{s}}^{\text{pre}}, \sigma_{\text{s}}^{\text{non}}, \mathbf{R}_{\text{s}}\right)$$

$$\mathbf{S}_{\text{w}} = \text{CovarianceMatrix}\left(\sigma_{\text{w}}^{\text{AmE}}, \sigma_{\text{w}}^{\text{BrE}}, \mathbf{R}_{\text{w}}\right)$$

Outlined next are the priors. The standard deviation parameters for the dispersion of the varying effects were given a weakly informative start. Figure A.6 shows the density of the exponential distribution with scale parameter 2. The shaded areas show the highest density intervals containing 50% and 90% of the probability mass. These limits are 0.35 and 1.15. This means that we consider it fairly unlikely that the dispersion of (i) the residuals, (ii) the by-subject varying effects, and the by-word varying effects will be greater than 1.15 Δ_{Bk}, on average.

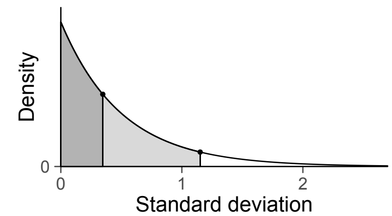


Figure A.6: Prior for the standard deviation parameters: Exponential distribution with scale parameter 2.

$$\sigma_{\ldots} \sim \text{Exponential}(2)$$

For the correlation between varying effects, we default to an LKJ distribution with η set to 4, which assigns 50% and 90% of the probability mass to the intervals [−.24; +.24] and [−.55, +.55], respectively.[19]

$$\mathbf{R}_{\ldots} \sim \text{LKJcorr}(4)$$

To guard estimates against (i) deviant Δ_{Bk} scores for individual tokens, (ii) unusual subjects, and (iii) unusual words, we use *t*-distributed rather than normal errors (Gelman et al. 2014: 437), with ν set to 7.[20]

$$\nu_{\ldots} = 7$$

For details on the following priors, see §5.6 and Figure 5.14.

$$\begin{aligned}\bar{\alpha}^{\text{pre}}_{\text{AmE}} &\sim \text{Normal}(3, 1)\\ \bar{\alpha}^{\text{pre}}_{\text{BrE}} &\sim \text{Normal}(3.5, 1)\\ \bar{\alpha}^{\text{non}}_{\text{AmE}} &\sim \text{Normal}(5, 1.5)\\ \bar{\alpha}^{\text{non}}_{\text{BrE}} &\sim \text{Normal}(7, 1.5)\end{aligned}$$

Stan code The *Stan* code for the native speaker model follows.

```
data {
  int<lower=0> N;
  vector[N] y;
  int<lower=0> subj[N];
  int<lower=0> word[N];
  int<lower=0> n_subj;
  int<lower=0> n_word;
  int pos_id[N];
  int var_id[N];
  int n_varcoef_subj;
  int n_varcoef_word;
  int n_cor_subj;
  int n_cor_word;
  int set_nu;
}
parameters {
  matrix<lower=0>[2,2] a;

  vector<lower=0>[n_varcoef_subj] sigma_subj;
  vector<lower=0>[n_varcoef_word] sigma_word;

  matrix[n_varcoef_subj, n_subj] z_subj;
  matrix[n_varcoef_word, n_word] z_word;
  cholesky_factor_corr[n_varcoef_subj] L_Rho_subj;
  cholesky_factor_corr[n_varcoef_word] L_Rho_word;

  real<lower=0> sigma;
  real<lower=0> udf_subj;
  real<lower=0> udf_word;
}
```

19 See Appendix A.3.1 and Figure A.2 (p. 253) for illustration.

20 See Appendix A.3.1 and Figure A.3 (p. 253) for illustration.

```
transformed parameters {
  matrix[n_subj, n_varcoef_subj] r_s;
  matrix[n_word, n_varcoef_word] r_w;
  r_s = sqrt(set_nu * udf_subj) * (diag_pre_multiply(sigma_subj, L_Rho_subj) * z_subj)’;
  r_w = sqrt(set_nu * udf_word) * (diag_pre_multiply(sigma_word, L_Rho_word) * z_word)’;
}
model {
  vector[N] mu;
  sigma ~ exponential(2);

  L_Rho_subj  ~ lkj_corr_cholesky(4);
  L_Rho_word  ~ lkj_corr_cholesky(4);
  sigma_subj ~ exponential(2);
  sigma_word ~ exponential(2);

  to_vector(z_subj)  ~ normal(0,1);
  to_vector(z_word)  ~ normal(0,1);

  a[1,1] ~ normal( 5   , 1.5);
  a[1,2] ~ normal( 3   , 1  );
  a[2,1] ~ normal( 7   , 1.5);
  a[2,2] ~ normal( 3.5 , 1  );

  udf_subj ~ inv_chi_square(set_nu);
  udf_word ~ inv_chi_square(set_nu);

  for(n in 1:N) {
    mu[n] = a[var_id[n],pos_id[n]] + r_s[subj[n],pos_id[n]] + r_w[word[n], var_id[n]];
  }
  target += student_t_lpdf(y | set_nu, mu, sigma);
}
generated quantities{
  matrix[n_varcoef_subj, n_varcoef_subj] Rho_subj;
  matrix[n_varcoef_word, n_varcoef_word] Rho_word;

  int pos_subj = 1;
  int pos_word = 1;

  vector<lower=-1, upper=1>[n_cor_subj] cor_subj;
  vector<lower=-1, upper=1>[n_cor_word] cor_word;

  Rho_subj = multiply_lower_tri_self_transpose(L_Rho_subj);
  Rho_word = multiply_lower_tri_self_transpose(L_Rho_word);

  for(i in 1:(n_varcoef_subj-1)){
    for(j in (i+1):n_varcoef_subj){
      cor_subj[pos_subj] = Rho_subj[i, j];
      pos_subj += 1;
    }
  }
  for(i in 1:(n_varcoef_word-1)){
    for(j in (i+1):n_varcoef_word){
      cor_word[pos_word] = Rho_word[i, j];
      pos_word += 1;
    }
  }
}
```

Convergence diagnostics Figure A.7 indicates convergence of the algorithm: All $\hat{R}$ values (left) are below 1.01, the rank plot (middle) features flat profiles, and the quantile plot (right) displays sufficiently large effective sample sizes across the board (i.e. all above $n = 400$).

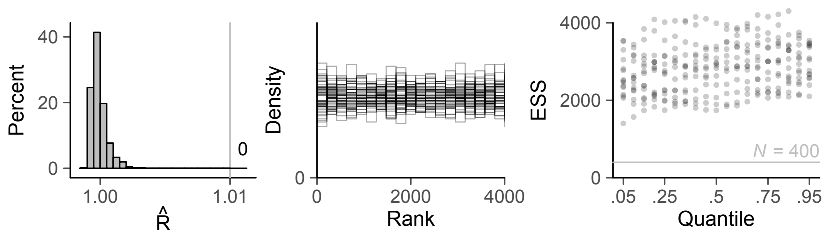


Figure A.7: Laterals model for the native speakers: Convergence diagnostics.

Posterior distribution of model parameters Table A.6 gives a summary of the posterior distribution of the parameters in the model.

Parameter	Mean	SD	Posterior quantiles .05	.50	.95	N_{eff}	$\hat{R}$
$\bar{\alpha}^{\text{pre}}_{\text{AmE}}$	5.08	0.33	4.54	5.08	5.62	1530	1
$\bar{\alpha}^{\text{non}}_{\text{AmE}}$	3.22	0.21	2.89	3.22	3.57	1557	1
$\bar{\alpha}^{\text{pre}}_{\text{BrE}}$	7.60	0.30	7.11	7.61	8.09	1117	1
$\bar{\alpha}^{\text{non}}_{\text{BrE}}$	3.88	0.21	3.54	3.88	4.21	1306	1
σ	0.63	0.03	0.58	0.63	0.68	5032	1
Varying effects: Subject							
$\sigma^{\text{pre}}_{\text{s}}$	0.76	0.23	0.44	0.74	1.17	1829	1
$\sigma^{\text{non}}_{\text{s}}$	0.50	0.15	0.29	0.49	0.78	2033	1
$\rho(\alpha^{\text{pre}}_{\text{s}}, \alpha^{\text{non}}_{\text{s}})$	0.34	0.18	0.03	0.34	0.61	1577	1
Varying effects: Word							
$\sigma^{\text{AmE}}_{\text{w}}$	0.35	0.13	0.18	0.33	0.57	2516	1
$\sigma^{\text{BrE}}_{\text{w}}$	0.45	0.15	0.24	0.43	0.72	2243	1
$\rho(\alpha^{\text{AmE}}_{\text{w}}, \alpha^{\text{BrE}}_{\text{w}})$	0.60	0.19	0.24	0.63	0.86	2109	1

Table A.6: Laterals model for the native speakers: Posterior distribution of the model parameters.

A.4.2 Learners

Model definition As the model for the learners is similar in structure, out focus will be on differences.

$$\text{Bark difference}_i \sim \text{Student-t}\left(\mu_i, \sigma, \nu\right)$$

The deterministic part includes an extra parameter for pronunciation proficiency (FAR). Let us start with the α parameters. The $\bar{\alpha}$ parameters are indexed to refer to two values, one for prevocalic and one for non-prevocalic contexts. $\bar{\alpha}^{\text{pre}}$, for instance, indicates the average Δ_{Bk} score we observe for prevocalic laterals among learners with an intermediate level of pronunciation ability (i.e. with an FAR score of 0). The α parameters are varying effects and denote speaker-specific deviations from this average, one for prevocalic and one for non-prevocalic contexts, and word-specific deviations from this average.

The FAR value of 0 again functions as a breakpoint[21]: Learners below this threshold are treated differently from those above the threshold.[22] To the right of the breakpoint, the model allows German learners to gradually form two populations with different target accents. They share the same intercept (at FAR = 0), which also links them with the average trend to the left of the breakpoint. The regression lines are allowed to diverge to the right of this point.[23] The $\bar{\beta}$ values are therefore indexed to refer not only to the two positions (pre- vs. non-prevocalic), but also to these two populations. The parameter $\bar{\beta}^{\text{pre}}_{\text{AmE}}$, for example, describes the regression line to the right of 0, for prevocalic laterals in the speech of learners with an inclination toward AmE. The β parameters are varying slopes and denote word-specific deviations from this average slope. The $\bar{\beta}$ coefficients are prefixed by weights w_i, which express how faithful the speaker uttering

[21] See Vanhove (2018) for a helpful how-to guide on breakpoint regression in *Stan*.

[22] This is what the curly brackets in the model formula express: If the standardized FAR score for the learner who produced token *i* is below 0 (i.e. $z_i < 0$), the top line applies. If it is above 0, the bottom line applies. Learners below this threshold (i.e. beginner and lower intermediate levels) are considered as representing a single population of German learners. This is to say that target variety orientation is ignored. Accordingly, for each phonetic context, there is one regression line per outcome quantity, which describes how much Δ_{Bk} scores vary, on average, across FAR scores up to 0. Thus, the shorthand symbol $\bar{\beta}^{\text{non}}$ is the average FAR slope (averaged across words) – in other words, it expresses how, on average, the outcome varies over FAR scores up to 0. The β's are varying slopes and denote word-specific deviations from this average trend.

[23] We therefore now have two regression lines for each phonetic context, one for BrE-oriented and one for AmE-oriented learners.

lateral token i is to their target variety. These weights range from 0 to 1, where one indicates (near-)perfect realization of characteristic features of the target accent.[24]

[24] See Appendix A.2 for details.

$$\mu_i = \bar{\alpha}_{\text{Pos}[i]} + \alpha_{\text{s}[i],\text{Pos}[i]} + \alpha_{\text{w}[i]} + \begin{cases} (\bar{\beta}_{\text{Pos}[i]} + \beta_{\text{w}[i]})\text{FAR}_i & \text{if } z_i < 0 \\ (\text{w}_i\bar{\beta}_{\text{Pos}[i],\text{Var}[i]} + \beta_{\text{w}[i]})\text{FAR}_i & \text{if } z_i > 0 \end{cases}$$

The joint distribution of the varying effects for subject are identical to the ones above. The varying effects for word are structured differently. Here, we have varying intercepts, α_{w}, which capture the deviations of words from the conditional means, and varying slopes, β_{w}, which express the deviation of words from the average trends across FAR scores.

$$\begin{aligned}
\begin{bmatrix} \alpha_{\text{s}}^{\text{pre}} \\ \alpha_{\text{s}}^{\text{non}} \end{bmatrix} &\sim \text{MVStudent-t}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mathbf{S}_{\text{s}}, \nu_{\text{s}}\right) \\
\begin{bmatrix} \alpha_{\text{w}} \\ \beta_{\text{w}} \end{bmatrix} &\sim \text{MVStudent-t}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mathbf{S}_{\text{w}}, \nu_{\text{w}}\right) \\
\mathbf{S}_{\text{s}} &= \text{CovarianceMatrix}\left(\sigma_{\text{s}}^{\text{pre}}, \sigma_{\text{s}}^{\text{non}}, \mathbf{R}_{\text{s}}\right) \\
\mathbf{S}_{\text{w}} &= \text{CovarianceMatrix}\left(\sigma_{\text{w}}^{\alpha}, \sigma_{\text{w}}^{\beta}, \mathbf{R}_{\text{w}}\right)
\end{aligned}$$

The priors for σ and $\mathbf{R}$ are identical to the ones above. For background on the $\bar{\alpha}$ priors, see §5.6. All $\bar{\beta}$ parameters were weakly regularized.

$$\begin{aligned}
\sigma_{\dots} &\sim \text{Exponential}(2) \\
\mathbf{R}_{\dots} &\sim \text{LKJcorr}(4) \\
(\bar{\alpha}^{\text{pre}}, \bar{\alpha}^{\text{non}}) &\sim \text{Normal}(6, 5) \\
(\bar{\beta}_{\text{AmE}}^{\text{pre}}, \bar{\beta}^{\text{non}}) &\sim \text{Normal}(0, 2) \\
(\bar{\beta}_{\text{BrE}}^{\text{pre}}, \bar{\beta}^{\text{non}}) &\sim \text{Normal}(0, 2) \\
\nu_{\dots} &= 7
\end{aligned}$$

Stan code This is a translation of the model in the *Stan* language:

```
data {
  int<lower=0> N;
  vector[N] y;
  int<lower=0> subj[N];
  int<lower=0> word[N];
  int<lower=0> n_subj;
  int<lower=0> n_word;
  int pos_id[N];
  int var_id[N];
  real var_score[N];
  real far_z[N];
  int n_varcoef_subj;
  int n_varcoef_word;
  int n_cor_subj;
  int n_cor_word;
  int set_nu;
}
parameters {
  row_vector<lower=0>[2] a;
  row_vector<lower=0>[2] b;
  matrix[2,2] b_var;

  vector<lower=0>[n_varcoef_subj] sigma_subj;
```

```
  vector<lower=0>[n_varcoef_word] sigma_word;

  matrix[n_varcoef_subj, n_subj] z_subj;
  matrix[n_varcoef_word, n_word] z_word;
  cholesky_factor_corr[n_varcoef_subj] L_Rho_subj;
  cholesky_factor_corr[n_varcoef_word] L_Rho_word;

  real<lower=0> sigma;

  real<lower=0> udf_word;
  real<lower=0> udf_subj;
}
transformed parameters {
  matrix[n_subj, n_varcoef_subj] r_s;
  matrix[n_word, n_varcoef_word] r_w;
  vector[N] mu;

  r_s = sqrt(set_nu * udf_subj) * (diag_pre_multiply(sigma_subj, L_Rho_subj) * z_subj)';
  r_w = sqrt(set_nu * udf_word) * (diag_pre_multiply(sigma_word, L_Rho_word) * z_word)';

  for (n in 1:N) {
    if(far_z[n] < 0){
      mu[n] = a[pos_id[n]] + (b[pos_id[n]] + r_w[word[n],2]) * far_z[n] +
              r_s[subj[n],pos_id[n]] + r_w[word[n], 1];
    } else {
      mu[n] = a[pos_id[n]] + (b_var[var_id[n],pos_id[n]]*var_score[n] + r_w[word[n],2]) * far_z[n] +
              r_s[subj[n],pos_id[n]] + r_w[word[n], 1];
    }
  }
}
model {
  sigma ~ exponential(2);
  L_Rho_subj  ~ lkj_corr_cholesky(4);
  L_Rho_word  ~ lkj_corr_cholesky(4);
  sigma_subj ~ exponential(2);
  sigma_word ~ exponential(2);

  to_vector(z_subj)  ~ normal(0,1);
  to_vector(z_word)  ~ normal(0,1);

  to_vector(a) ~ normal( 6 , 5 );
  to_vector(b) ~ normal( 0 , 1 );
  to_vector(b_var) ~ normal( 0 , 1 );

  udf_word ~ inv_chi_square(set_nu);
  udf_subj ~ inv_chi_square(set_nu);

  target += student_t_lpdf(y | set_nu, mu, sigma);
}
generated quantities{
  matrix[n_varcoef_subj, n_varcoef_subj] Rho_subj;
  matrix[n_varcoef_word, n_varcoef_word] Rho_word;

  int pos_subj = 1;
  int pos_word = 1;

  vector<lower=-1, upper=1>[n_cor_subj] cor_subj;
  vector<lower=-1, upper=1>[n_cor_word] cor_word;

  Rho_subj = multiply_lower_tri_self_transpose(L_Rho_subj);
  Rho_word = multiply_lower_tri_self_transpose(L_Rho_word);

  for(i in 1:(n_varcoef_subj-1)){
    for(j in (i+1):n_varcoef_subj){
      cor_subj[pos_subj] = Rho_subj[i, j];
      pos_subj += 1;
    }
  }
  for(i in 1:(n_varcoef_word-1)){
    for(j in (i+1):n_varcoef_word){
      cor_word[pos_word] = Rho_word[i, j];
      pos_word += 1;
    }
  }
}
```

Convergence diagnostics Convergence diagnostics are shown in Figure A.8, which reveals no signs of failure: All $\hat{R}$ values (left) are below 1.01, the profile of the rank plot (middle) is flat, and the quantile plot (right) places all effective sample sizes comfortably above $n = 400$.

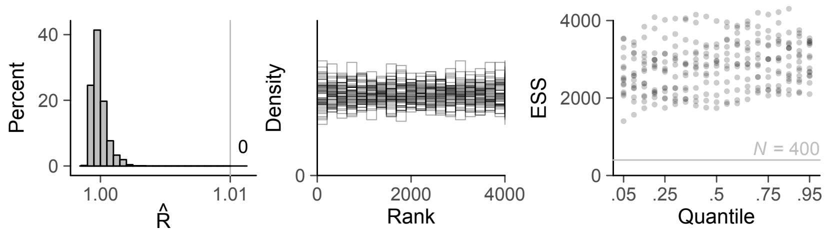


Figure A.8: Laterals model for the learners: Convergence diagnostics.

Posterior distribution of model parameters Table A.7 gives a summary of the posterior distribution of the parameters in the model.

Table A.7: Laterals model for the learners: Posterior distribution of the parameters.

			Posterior quantiles				
Parameter	Mean	SD	.05	.50	.95	N_{eff}	$\hat{R}$
$\bar{\alpha}^{\text{pre}}$	7.96	0.26	7.54	7.96	8.39	1924	1
$\bar{\alpha}^{\text{non}}$	5.69	0.20	5.36	5.68	6.02	1674	1
$\bar{\beta}^{\text{pre}}$	0.16	0.12	0.01	0.13	0.39	3007	1
$\bar{\beta}^{\text{non}}$	0.08	0.07	0.00	0.06	0.22	4786	1
$\bar{\beta}^{\text{pre}}_{\text{AmE}}$	–1.58	0.35	–2.17	–1.56	–1.01	2767	1
$\bar{\beta}^{\text{non}}_{\text{AmE}}$	–1.07	0.36	–1.66	–1.07	–0.47	2981	1
$\bar{\beta}^{\text{pre}}_{\text{BrE}}$	–0.54	0.24	–0.93	–0.54	–0.13	2252	1
$\bar{\beta}^{\text{non}}_{\text{BrE}}$	–1.19	0.25	–1.61	–1.20	–0.77	2280	1
σ	0.78	0.02	0.74	0.78	0.82	5339	1
Varying effects: Subject							
$\sigma^{\text{pre}}_{\text{s}}$	0.63	0.18	0.36	0.62	0.94	3260	1
$\sigma^{\text{non}}_{\text{s}}$	0.74	0.20	0.44	0.73	1.09	3298	1
$\rho(\alpha^{\text{pre}}_{\text{s}}, \alpha^{\text{non}}_{\text{s}})$	0.69	0.10	0.51	0.69	0.82	1368	1
Varying effects: Word							
$\sigma^{\alpha}_{\text{w}}$	0.50	0.16	0.27	0.47	0.79	2757	1
$\sigma^{\beta}_{\text{w}}$	0.12	0.06	0.03	0.11	0.23	1430	1
$\rho(\alpha_{\text{w}}, \beta_{\text{w}})$	0.28	0.27	–0.18	0.31	0.68	4380	1

A.5 *Prevocalic /r/*

The data sets analyzed in Chapter 6 are openly available via *TROLLing*[25] (Sönning et al. 2020a). This appendix provides information about the analysis of the data from the current study. Corresponding details about the analyses of the data from Pascoe (1987) and Wunder (2012) are given in the web appendix.[26] The OSF repository includes the *Stan* code for the regression models[27] and the *R* script[28] for running the analyses.

[25] https://doi.org/10.18710/YDKDFG

[26] https://osf.io/aszq6/

[27] https://osf.io/deahk/

[28] https://osf.io/j7u3d/

A.5.1 *The present study*

Model definition The realization of prevocalic /r/ was coded as a binary outcome (1 for target-like realization, 0 for other variants) and therefore modeled using a Bernoulli distribution. The parameter p denotes the probability of observing a target-like approximant rhotic, and the subscript i indexes the *r*-tokens in the data set.

$$\mathrm{r}_i \sim \text{Bernoulli}\,(1, p_i)$$

To model the probability of observing a target-like *r*-variant, we use the logit link function, which maps constrained probabilities (i.e. the interval $[0;1]$) to an unconstrained scale from $-\infty$ to $+\infty$. On this unconstrained logit scale, we model the expected logits for conditions of interest. The $\bar{\alpha}_{\text{Cmplx}}$ parameter is indexed to denote two intercepts: the logit-transformed probability of a target-like rendition among learners with an FAR score of 0 in (i) singleton contexts ($\bar{\alpha}_{\text{sing}}$) and (ii) cluster contexts ($\bar{\alpha}_{\text{clust}}$). The α coefficients represent varying intercepts for speakers and words. For speakers (α_{s}), there are two sets, one that indicates how much each learner deviates from the conditional expectation in singleton contexts ($\alpha_{\text{s}}^{\text{sing}}$), and one for cluster contexts ($\alpha_{\text{s}}^{\text{clust}}$). By-word varying intercepts (α_{w}) reflect by how much each word is deflected from the expected logit. The $\bar{\beta}^{\text{FAR}}$ parameter traces how target accuracy changes across proficiency levels, with indexes referring to the two contexts (singletons: $\bar{\beta}_{\text{sing}}^{\text{FAR}}$; clusters: $\bar{\beta}_{\text{sing}}^{\text{FAR}}$).

$$\text{logit}(p_i) = \bar{\alpha}_{\text{Cmplx}[i]} + \alpha_{\text{s}[i],\text{Cmplx}[i]} + \alpha_{\text{w}[i]} + (\bar{\beta}_{\text{Cmplx}[i]}^{\text{FAR}} + \beta_{\text{w}[i]}^{\text{FAR}})\text{FAR}_i$$

The next block defines the joint distribution of the varying effects. Both by-word and by-subject sets are given a joint multivariate Student-*t* distribution. As this probability distribution allows for the occasional outlier, the estimates of interest (i.e. the $\bar{\alpha}$ and $\bar{\beta}$ parameters) are less sensitive to deviant speakers and words. The degrees-of-freedom parameter ν was set to 7.[29] **S** denotes the covariance matrices, which detail how the varying effects in each set are distributed.

[29] See Appendix A.3.1 and Figure A.3 (p. 253) for illustration.

$$\begin{bmatrix} \alpha_{\text{s}}^{\text{sing}} \\ \alpha_{\text{s}}^{\text{clust}} \end{bmatrix} \sim \text{MVStudent-t}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mathbf{S}_{\text{s}}, \nu_{\text{s}} \right)$$

$$\begin{bmatrix} \alpha_{\text{w}} \\ \beta_{\text{w}}^{\text{FAR}} \end{bmatrix} \sim \text{MVStudent-t}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mathbf{S}_{\text{w}}, \nu_{\text{w}} \right)$$

We descibe the covariance matrices with two quantities, the standard deviation parameters, which specify how much each varying effect is spread out along the outcome (logit) scale, and a correlation matrix **R**, which contains correlation coefficients that reflect the association between the batches of varying effects in each set.[30]

[30] If the standard deviation for the by-subject intercepts were higher in singleton contexts ($\sigma_{\alpha_s^{\text{sing}}}$) compared to cluster contexts ($\sigma_{\alpha_s^{\text{clust}}}$), for instance, this would indicate that, on the logit scale, the variation among learners was greater in this context. A positive correlation between the by-subject varying effects α_s^{sing} and α_s^{clust} would tell us that a learner who performed "below-average" in singleton contexts would be expected to also perform below-average in cluster contexts.

$$\mathbf{S}_\text{s} = \text{CovarianceMatrix}\left(\sigma_{\alpha_\text{s}^{\text{sing}}}, \sigma_{\alpha_\text{s}^{\text{clust}}}, \mathbf{R}_\text{s}\right)$$

$$\mathbf{S}_\text{w} = \text{CovarianceMatrix}\left(\sigma_{\alpha_\text{w}}, \sigma_{\beta_\text{w}^{\text{FAR}}}, \mathbf{R}_\text{w}\right)$$

As for the priors, we default to a weakly regularizing statement for the σ parameters. Figure A.9 shows the density of the exponential distribution with scale parameter 2. The shaded areas show the highest density intervals containing 50% and 90% of the probability mass. These limits are 0.35 and 1.15.

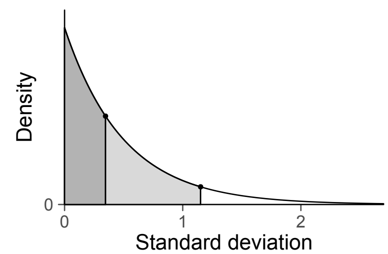


Figure A.9: Prevocalic *r*: Prior for the standard deviation parameters: Exponential distribution with scale parameter 2.

$$\sigma_{\dots} \sim \text{Exponential}(2)$$

We have no a priori hints at the direction and magnitude of the correlation between varying coefficients and therefore specify an LKJ prior with the shape parameter η set to 4. It assigns the central 50% and 90% of the probability mass to the intervals [–.24; +.24] and [–.55, +.55], respectively.[31]

[31] See Appendix A.3.1 and Figure A.2 (p. 253) for illustration.

$$\mathbf{R}_{\dots} \sim \text{LKJcorr}(4)$$

For the average accuracy rate in singleton and cluster contexts, we specified a prior for the marginal (population-averaged) distribution of the $\bar{\alpha}$ parameters. As such, multilevel models by design produce subject-specific estimates. We, however, are interested in marginal estimates.[32] Thus, our model returns parameter estimates that describe how the average subject (in terms of skill) in our sample performs when producing the average word (in terms of difficulty) in our sample. We, however, are interested in averages across learners and across words on the proportion scale. Our priors describe these marginal estimates, and we therefore need to translate these to the subject-specific quantities required by the model. We use the formula[33] provided by Molenberghs & Verbeke (2005: 300). The prior distributions given on the right-hand side were motivated in §6.6 (see Figure 6.6).

[32] See §3.4.1 for a brief discussion.

[33] In this formula, $c = 16\sqrt{3}/(15\pi)$. There are two unknown quantities in this prior, the standard deviation of the varying intercepts for subjects and words. These also need to be estimated by the model. Fortunately, *Stan* enables us to build unknown parameters into a prior.

$$\frac{\bar{\alpha}_{\text{sing}}}{\sqrt{c^2(\sigma_{\alpha_s^{\text{sing}}}{}^2 + \sigma_{\alpha_w}{}^2) + 1}} \sim \text{Normal}(2, 2)$$

$$\frac{\bar{\alpha}_{\text{clust}}}{\sqrt{c^2(\sigma_{\alpha_s^{\text{clust}}}{}^2 + \sigma_{\alpha_w}{}^2) + 1}} \sim \text{Normal}(2, 2)$$

The $\bar{\beta}$ parameters were given weakly regularizing priors centered at 0.

$$\bar{\beta}_{\text{sing}}^{\text{FAR}} \sim \text{Normal}(0, 2)$$

$$\bar{\beta}_{\text{clust}}^{\text{FAR}} \sim \text{Normal}(0, 2)$$

To make the model more robust against unusual subjects and unusual words, we use *t*-distributed rather than normal errors. The degrees-of-freedom parameter ν was set to 7.[34]

[34] See Appendix A.3.1 and Figure A.3 (p. 253) for illustration.

$$\nu_{\dots} = 7$$

Stan code The *Stan* code for the model is listed next.

```
data {
  int<lower=0> N;
  int y[N];
  int<lower=0> subj[N];
  int<lower=0> word[N];
  int<lower=0> n_subj;
  int<lower=0> n_word;
  int cmplx_id[N];
  real far_z[N];
  int n_varcoef_subj;
  int n_varcoef_word;
  int n_cor_subj;
  int n_cor_word;
  int set_nu;
}
parameters {
  row_vector[2] a;
  row_vector[2] b;

  vector<lower=0>[n_varcoef_subj] sigma_subj;
  vector<lower=0>[n_varcoef_word] sigma_word;

  matrix[n_varcoef_subj, n_subj] z_subj;
  matrix[n_varcoef_word, n_word] z_word;

  cholesky_factor_corr[n_varcoef_subj] L_Rho_subj;
  cholesky_factor_corr[n_varcoef_word] L_Rho_word;

  real<lower=0> udf_word;
  real<lower=0> udf_subj;
}
transformed parameters {
  matrix[n_subj, n_varcoef_subj] r_s;
  matrix[n_word, n_varcoef_word] r_w;

  vector[N] mu;

  r_s = sqrt(set_nu * udf_subj) * (diag_pre_multiply(sigma_subj, L_Rho_subj) * z_subj)';
  r_w = sqrt(set_nu * udf_word) * (diag_pre_multiply(sigma_word, L_Rho_word) * z_word)';

  for (n in 1:N) {
    mu[n] = a[cmplx_id[n]] + (b[cmplx_id[n]] + r_w[word[n],2]) * far_z[n] + r_s[subj[n],cmplx_id[n]] + r_w[word[n], 1];
  }
}
model {
  L_Rho_subj  ~ lkj_corr_cholesky(4);
  L_Rho_word  ~ lkj_corr_cholesky(4);
  sigma_subj ~ exponential(2);
  sigma_word ~ exponential(2);

  to_vector(z_subj)  ~ normal(0,1);
  to_vector(z_word)  ~ normal(0,1);

  to_vector(b) ~ normal( 0 , 2 );

  a[1]/sqrt((16*sqrt(3)/(15*pi())) * (sigma_subj[1]^2 + sigma_word[1]^2) + 1) ~ normal(2,2);
  a[2]/sqrt((16*sqrt(3)/(15*pi())) * (sigma_subj[1]^2 + sigma_word[1]^2) + 1) ~ normal(2,2);

  udf_word ~ inv_chi_square(set_nu);
  udf_subj ~ inv_chi_square(set_nu);

  target += bernoulli_logit_lpmf(y | mu);
}
generated quantities{
  matrix[n_varcoef_subj, n_varcoef_subj] Rho_subj;
  matrix[n_varcoef_word, n_varcoef_word] Rho_word;

  int pos_subj = 1;
```

```
  int pos_word = 1;

  vector<lower=-1, upper=1>[n_cor_subj] cor_subj;
  vector<lower=-1, upper=1>[n_cor_word] cor_word;

  Rho_subj = multiply_lower_tri_self_transpose(L_Rho_subj);
  Rho_word = multiply_lower_tri_self_transpose(L_Rho_word);

  for(i in 1:(n_varcoef_subj-1)){
    for(j in (i+1):n_varcoef_subj){
      cor_subj[pos_subj] = Rho_subj[i, j];
      pos_subj += 1;
    }
  }
  for(i in 1:(n_varcoef_word-1)){
    for(j in (i+1):n_varcoef_word){
      cor_word[pos_word] = Rho_word[i, j];
      pos_word += 1;
    }
  }
}
```

Convergence diagnostics Our audit of the MCMC algorithm is displayed in Figure A.10. We see no indications of non-convergence: All $\hat{R}$ values (left) are below 1.01, the profile of the rank plot (middle) is flat, and the quantile plot (right) suggests that all key parameters are supported by a sufficiently large effective sample size (i.e. all above $n = 400$).

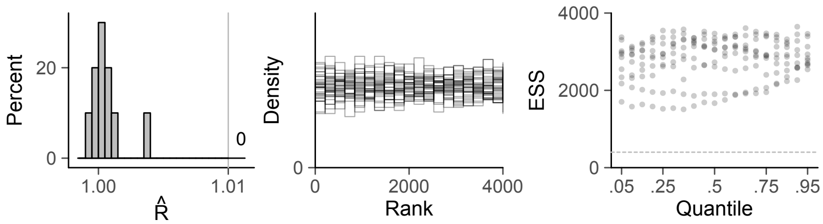


Figure A.10: Prevocalic /r/ model for the data from the present study: Convergence diagnostics.

Posterior distribution of model parameters Table A.8 reports the posterior distribution of the model parameters.

Table A.8: Prevocalic /r/ model for the data from the current study: Posterior distribution of parameters.

			Posterior quantiles				
Parameter	Mean	SD	.05	.50	.95	N_{eff}	$\hat{R}$
α^{sing}	3.51	0.80	2.32	3.45	4.94	3018	1
α^{clust}	4.29	0.61	3.40	4.24	5.37	2649	1
β^{sing}	2.35	0.79	1.09	2.32	3.73	2874	1
β^{clust}	1.79	0.57	0.89	1.78	2.75	3310	1
Varying effects: Subject							
$\sigma_{\text{s}}^{\text{sing}}$	1.39	0.50	0.69	1.32	2.32	2488	1
$\sigma_{\text{s}}^{\text{clust}}$	1.62	0.51	0.88	1.56	2.55	3131	1
$\rho(\alpha_{\text{s}}^{\text{sing}}, \alpha_{\text{s}}^{\text{clust}})$	0.76	0.11	0.55	0.77	0.91	1511	1
Varying effects: Word							
$\sigma_{\text{w}}^{\alpha}$	0.55	0.35	0.06	0.51	1.17	1400	1
$\sigma_{\text{w}}^{\beta_{\text{FAR}}}$	0.87	0.35	0.41	0.82	1.53	2944	1
$\rho(\alpha_{\text{w}}, \beta_{\text{w}}^{\text{FAR}})$	0.06	0.29	–0.44	0.06	0.52	2128	1

A.6 Postvocalic /r/

The data sets analyzed in Chapter 6 are openly available via *TROLLing*[35] (Sönning et al. 2020a). In what follows, we provide additional information about the analysis of the data from the current study. Details about the analyses of the data by Pascoe (1987), Wunder (2012), and Rank (2018) are deferred to the web appendix.[36] Note that the structure of the models fit to these data sets is different, as documented in the supplementary files. The OSF project accompanying this study includes the *Stan* code for the regression models[37] and the *R* script[38] for running the analyses.

[35] https://doi.org/10.18710/YDKDFG

[36] https://osf.io/aszq6/

[37] This study: https://osf.io/7q8sz/; Pascoe 1987 and Wunder 2012: https://osf.io/y8vr6/; Rank 2018: https://osf.io/y4tzn/

[38] https://osf.io/h6cjg/

A.6.1 The present study

Model definition The realization of postvocalic /r/ was coded with three categories, and we therefore used a multinomial regression model. The observed counts for the three variants are modeled with a multinomial distribution, where the observed outcome value depends on the probabilities of the individual events (in our case, the probability of each variant, i.e. $p^{\text{ɹ}}$, $p^{\text{ɐ}}$, and $p^{\varnothing}$).

$$\mathrm{r}_i \sim \text{Categorical}\left(p_i^{\text{ɹ}}, p_i^{\text{ɐ}}, p_i^{\varnothing}\right)$$

The purpose of our model is to describe how these probabilities change across conditions (i.e. contextual and speaker-specific attributes). Similar to logistic regression models, these probabilities are not modeled directly, but through a link function. Here, this is the softmax function, which translates the constrained probabilities (which can range from 0 to 1) to an unconstrained scale. For technical reasons, one of the outcome events serves as a reference category. We chose $\varnothing$ as the benchmark. In practical terms, this means that, instead of three probabilities, we are modeling just two quantities: (i) the (log-transformed) ratio of the probability of [ɹ] divided by that of the reference category (i.e. $\log(p^{\text{ɹ}}/p^{\varnothing})$) and (ii) the (log-transformed) ratio of the probability of [ɐ] divided by that of the reference category (i.e. $\log(p^{\text{ɐ}}/p^{\varnothing})$). We will refer to these quantities as *log probability-ratios*.[39] Our regression model therefore deals with two log ratios, which is why we are modeling two conditional means, one for $\log(p^{\text{ɹ}}/p^{\varnothing})$ and one for $\log(p^{\text{ɐ}}/p^{\varnothing})$.

[39] The hyphen clarifies the structure of the compound (i.e. log [probability ratio]).

The next model block details the patterns we wish to extract from the data. Note that it consists of two identically structured components. This part of the model is rather long and we will take it step by step. The $\bar{\alpha}_{\dots}$ parameters express the expected log probability-ratio when all predictors are set to 0. The α coefficients represent the varying intercepts for subjects ($\alpha_{\mathrm{S}}^{\dots}$) and words ($\alpha_{\mathrm{W}}^{\dots}$). Thereby, each speaker and word is allowed to assume a higher or lower ratio for each pair of probabilities. The next four lines describe how the probability ratios denoted by the $\bar{\alpha}$ parameters shift as we compare different conditions. The predictors are coded as follows:

- Link (linking context): Whether postvocalic /r/ occurred in a linking context (Link = 1) or elsewhere (Link = 0).
- Str (structural strength): Whether *r* occurred in a strong context following a full vowel (Str = +1) or a weak context, following [ə] or

occuring in a function word (Str = -1).

- Clu (cluster context): Whether *r* occurred in a cluster (Clu = $+1$) or as a singleton (Clu = -1).
- Con (connected speech): Whether *r* occurred in the reading passage (Con = $+1$) or the word list (Con = -1).

$$
\begin{aligned}
\log\left(\frac{p_i^{\text{ɹ}}}{p_i^{\varnothing}}\right) = {} & \bar{\alpha}^{\text{ɹ}} + \alpha^{\text{ɹ}}_{\mathrm{s}[i]} + \alpha^{\text{ɹ}}_{\mathrm{w}[i]} + \\
& (\bar{\gamma}^{\text{ɹ}}_{\text{Link}} + \gamma^{\text{ɹ}}_{\text{Link, s}[i]})\text{Link}_i + \\
& (\bar{\gamma}^{\text{ɹ}}_{\text{Str}} + \gamma^{\text{ɹ}}_{\text{Str, s}[i]})\text{Str}_i + \\
& (\bar{\gamma}^{\text{ɹ}}_{\text{Clu}} + \gamma^{\text{ɹ}}_{\text{Clu, s}[i]})\text{Clu}_i + \\
& (\bar{\gamma}^{\text{ɹ}}_{\text{Con}} + \gamma^{\text{ɹ}}_{\text{Con, s}[i]})\text{Con}_i + \\
& \begin{cases}
(\bar{\beta}^{\text{ɹ}}_{\text{FAR}} + \bar{\beta}^{\text{ɹ}}_{\text{FAR:Link}}\text{Link}_i + \bar{\beta}^{\text{ɹ}}_{\text{FAR:Str}}\text{Str}_i + \\
\bar{\beta}^{\text{ɹ}}_{\text{FAR:Clu}}\text{Clu}_i + \bar{\beta}^{\text{ɹ}}_{\text{FAR:Con}}\text{Con}_i + \beta^{\text{ɹ}}_{\text{FAR, w}[i]})\text{FAR}_i & \text{if } z_i < 0 \\
(\mathrm{w}_i\bar{\beta}^{\text{ɹ}}_{\text{FAR, Var}[i]} + \mathrm{w}_i\bar{\beta}^{\text{ɹ}}_{\text{FAR:Link, Var}[i]}\text{Link}_i + \mathrm{w}_i\bar{\beta}^{\text{ɹ}}_{\text{FAR:Str, Var}[i]}\text{Str}_i + \\
\mathrm{w}_i\bar{\beta}^{\text{ɹ}}_{\text{FAR:Clu, Var}[i]}\text{Clu}_i + \mathrm{w}_i\bar{\beta}^{\text{ɹ}}_{\text{FAR:Con, Var}[i]}\text{Con}_i + \beta^{\text{ɹ}}_{\text{FAR, w}[i]})\text{FAR}_i & \text{if } z_i > 0
\end{cases}
\end{aligned}
$$

$$
\begin{aligned}
\log\left(\frac{p_i^{\text{ɐ}}}{p_i^{\varnothing}}\right) = {} & \bar{\alpha}^{\text{ɐ}} + \alpha^{\text{ɐ}}_{\mathrm{s}[i]} + \alpha^{\text{ɐ}}_{\mathrm{w}[i]} + \\
& (\bar{\gamma}^{\text{ɐ}}_{\text{Link}} + \gamma^{\text{ɐ}}_{\text{Link, s}[i]})\text{Link}_i + \\
& (\bar{\gamma}^{\text{ɐ}}_{\text{Str}} + \gamma^{\text{ɐ}}_{\text{Str, s}[i]})\text{Str}_i + \\
& (\bar{\gamma}^{\text{ɐ}}_{\text{Clu}} + \gamma^{\text{ɐ}}_{\text{Clu, s}[i]})\text{Clu}_i + \\
& (\bar{\gamma}^{\text{ɐ}}_{\text{Con}} + \gamma^{\text{ɐ}}_{\text{Con, s}[i]})\text{Con}_i + \\
& \begin{cases}
(\bar{\beta}^{\text{ɐ}}_{\text{FAR}} + \bar{\beta}^{\text{ɐ}}_{\text{FAR:Link}}\text{Link}_i + \bar{\beta}^{\text{ɐ}}_{\text{FAR:Str}}\text{Str}_i + \\
\bar{\beta}^{\text{ɐ}}_{\text{FAR:Clu}}\text{Clu}_i + \bar{\beta}^{\text{ɐ}}_{\text{FAR:Con}}\text{Con}_i + \beta^{\text{ɐ}}_{\text{FAR, w}[i]})\text{FAR}_i & \text{if } z_i < 0 \\
(\mathrm{w}_i\bar{\beta}^{\text{ɐ}}_{\text{FAR, Var}[i]} + \mathrm{w}_i\bar{\beta}^{\text{ɐ}}_{\text{FAR:Link, Var}[i]}\text{Link}_i + \mathrm{w}_i\bar{\beta}^{\text{ɐ}}_{\text{FAR:Str, Var}[i]}\text{Str}_i + \\
\mathrm{w}_i\bar{\beta}^{\text{ɐ}}_{\text{FAR:Clu, Var}[i]}\text{Clu}_i + \mathrm{w}_i\bar{\beta}^{\text{ɐ}}_{\text{FAR:Con, Var}[i]}\text{Con}_i + \beta^{\text{ɐ}}_{\text{FAR, w}[i]})\text{FAR}_i & \text{if } z_i > 0
\end{cases}
\end{aligned}
$$

The $\bar{\gamma}$ parameters express by how much the log probability-ratios change, on average, as we compare the conditions described by these binary predictors. Thus, the parameter $\bar{\gamma}^{\text{ɹ}}_{\text{Link}}$ describes the change in the log probability-ratio when going from a non-linking to a linking context. We expect this parameter to be positive, because the probability of observing [ɹ] relative to ∅ should be greater in linking contexts. The γ coefficients are varying slopes and capture by how much a speaker deviates from the average change in each log probability-ratio.

The final block describes how the outcome probabilities vary over proficiency levels. The FAR value of 0 functions as a breakpoint.[40] Learners below this threshold (i.e. beginner and lower intermediate levels) are considered as representing a single population of German learners. This

[40] Learners below this threshold are treated differently from those above the threshold. This is what the curly brackets in the model formula express: If the standardized FAR score for the learner who produced token i is below 0 (i.e. $z_i < 0$), the upper part applies. If it is above 0, the lower part applies.

is to say that target variety orientation is ignored. Accordingly, all learners are represented by a single regression line. The $\bar{\beta}^{\cdots}_{\text{FAR}}$ parameter specifies by how much, on average, a log probability-ratio changes as we move up the FAR scale by 1. The $\bar{\beta}_{\text{FAR:...}}$ coefficients capture how this trendline changes across our binary predictors[41], i.e. how much of a difference we observe when comparing strong and weak contexts, etc. The $\beta^{\cdots}_{\text{FAR, w}}$ parameters are varying slopes and represent how much each word differs from the average FAR slope for a particular condition.

To the right of the breakpoint, the model allows German learners to gradually form two populations with different target varieties. They share the same intercept (at FAR $= 0$), which also links them with the average trend to the left of the breakpoint. They regression lines are allowed to diverge from this point onwards. We therefore now have two values for each of the coefficients described in the preceding paragraph, one for BrE-oriented and one for AmE-oriented learners. The $\bar{\beta}$ and β values are therefore indexed to also refer to these subgroups.[42] The βs are again varying slopes and denote word-specific deviations from the average slopes. The $\bar{\beta}$ coefficients are prefixed by weights w_i, which express how faithful the speaker uttering vowel token i is to their target variety. These weights range from 0 to 1, where one indicates (near-)perfect realization of characteristic features of the target accent.[43]

The joint distributions of the varying effects are assumed to follow a multivariate Student-t distribution, with ν set to 7.[44] **S** denotes the covariance matrix, which contains information about the dispersion of the varying effects and their pairwise association.[45]

$$\begin{bmatrix} \alpha^{\text{ɹ}}_{\text{s}} & \gamma^{\text{ɹ}}_{\text{Link,s}} & \gamma^{\text{ɹ}}_{\text{Str, s}} & \gamma^{\text{ɹ}}_{\text{Clu, s}} & \gamma^{\text{ɹ}}_{\text{Con, s}} \\ \alpha^{\text{ɐ}}_{\text{s}} & \gamma^{\text{ɐ}}_{\text{Link,s}} & \gamma^{\text{ɐ}}_{\text{Str, s}} & \gamma^{\text{ɐ}}_{\text{Clu, s}} & \gamma^{\text{ɐ}}_{\text{Con, s}} \end{bmatrix} \sim \text{MVStudent-t}\left(\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \mathbf{S}_{\text{s}}, \nu_{\text{s}} \right)$$

$$\begin{bmatrix} \alpha^{\text{ɹ}}_{\text{w}} & \beta^{\text{ɹ}}_{\text{FAR, w}} \\ \alpha^{\text{ɐ}}_{\text{w}} & \beta^{\text{ɐ}}_{\text{FAR, w}} \end{bmatrix} \sim \text{MVStudent-t}\left(\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \mathbf{S}_{\text{w}}, \nu_{\text{w}} \right)$$

The covariance matrices are defined using (instead of variances and covariances) the standard deviation of the varying effects sets, and their pairwise correlations. Thus, the parameter $\sigma_{\alpha^{\text{ɹ}}_{\text{w}}}$ is a standard deviation (on the log ratio scale), which describes how much speakers varied around the average log probability-ratio of [ɹ] over ∅. **R** is a correlation matrix, which includes correlation effects that capture how by-subject varying effects on the one hand, and by-word varying coefficients on the other, are correlated.[46]

$$\mathbf{S}_{\text{s}} = \text{CovarianceMatrix}\begin{pmatrix} \sigma_{\alpha^{\text{ɹ}}_{\text{s}}} & \sigma_{\gamma^{\text{ɹ}}_{\text{Link,s}}} & \sigma_{\gamma^{\text{ɹ}}_{\text{Str, s}}} & \sigma_{\gamma^{\text{ɹ}}_{\text{Clu, s}}} & \sigma_{\gamma^{\text{ɹ}}_{\text{Con, s}}} & \\ \sigma_{\alpha^{\text{ɐ}}_{\text{s}}} & \sigma_{\gamma^{\text{ɐ}}_{\text{Link,s}}} & \sigma_{\gamma^{\text{ɐ}}_{\text{Str, s}}} & \sigma_{\gamma^{\text{ɐ}}_{\text{Clu, s}}} & \sigma_{\gamma^{\text{ɐ}}_{\text{Con, s}}} & , \mathbf{R}_{\text{s}} \end{pmatrix}$$

$$\mathbf{S}_{\text{w}} = \text{CovarianceMatrix}\begin{pmatrix} \sigma_{\alpha^{\text{ɹ}}_{\text{w}}} & \sigma_{\beta^{\text{ɹ}}_{\text{FAR, w}}} & \\ \sigma_{\alpha^{\text{ɐ}}_{\text{w}}} & \sigma_{\beta^{\text{ɐ}}_{\text{FAR, w}}} & , \mathbf{R}_{\text{w}} \end{pmatrix}$$

We now turn to the priors. The standard deviation parameters for the dispersion of the varying effects were given a weakly informative specification, in the form of an exponential distribution with scale parameter

41 In statistical terms, the model includes an interaction between the FAR score and the predictors linking context, structural strength, cluster context, and connected speech.

42 The parameter $\bar{\beta}^{\text{ɐ}}_{\text{FAR, AmE}}$, for example, describes how, for AmE-oriented learners, the log probability-ratio of vocalized vs. null realizations varies as the FAR score increases by 1.

43 See Appendix A.2 for details.

44 See Appendix A.3.1 and Figure A.3 (p. 253) for illustration.

45 To save space, we arranged them in multiple columns – they would normally need to be listed in a single column.

46 We again simplify notation and omit the statement translating the set of standard deviations and the correlation matrix **R** into a covariance matrix. See McElreath (2020: 441) for a complete representation.

2. In Figure A.11, the shaded areas show the highest density intervals containing 50% and 90% of the probability mass. These limits are 0.35 and 1.15, which translates into ratios of roughly 1.4 and 3.2. The prior thus states that the probability ratios, which the model allows to vary across subjects and words are within a factor of 3.2 of the predicted values with 90% probability.

$$\sigma_{...} \sim \text{Exponential}(2)$$

We have no direct information about candidate values for the correlation between varying coefficients, and therefore settle for an LKJ prior with the shape parameter η set to 4.[47]

$$\mathbf{R}_{...} \sim \text{LKJcorr}(4)$$

To specify reasonable priors for the two log probability-ratios, two steps are necessary: First, we need to convert the prior distribution over probabilities shown in Figure 6.7 to log probability-ratios. Then, we need to translate these estimates, which have a marginal ('population-averaged') interpretation to their conditional ('subject-specific') analogues.

For step 1, we relied on simulation. First, we expressed our prior information about the probability of observing each of the three variants on the probability scale.[48] To this end, we used a Dirichlet distribution with parameters 3 ([ɹ]), 4 (∅), and 1 ([ɐ]) to generate these values.[49] Figure A.12a shows how, for each variant, the a priori probability mass is concentrated over the interval [0; 1]. We then divided the distributions for [ɹ] and [ɐ] by that of ∅, which gave us two distributions of ratios.[50] This intermediate step is shown in panel (b). Since both variants are assumed to be less likely than ∅, most of the probability mass is found below 1. Panel (c) shows these ratios on the log scale and brings us to the quantity that is modeled (i.e. log probability-ratios). These distributions are slightly negatively skewed. To approximate this shape, we specified a skew-normal distribution as a prior. The solid lines in panel (c) are the simulated log ratios, and the dashed curves are the prior distributions we used. These log probability-ratios are constructed based the probability distributions shown in panel (a). We attach to these probabilities a marginal (or population-averaged) interpretation. This means that we consider them as representing

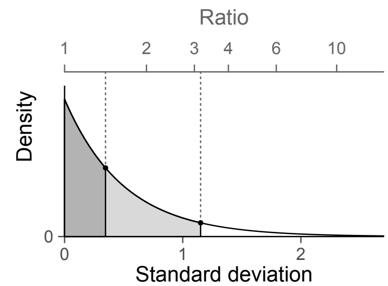


Figure A.11: Prior for the standard deviation parameters: Exponential distribution with scale parameter 2.

[47] See Appendix A.3.1 and Figure A.2 (p. 253) for illustration.

[48] See §6.6 for motivation.

[49] The Dirichlet distribution has a convenient interpretation: The sum of these parameter values, 8, directly expresses the amount of inferential information encoded by the prior distribution. Thus, the pre-data information we are supplying is "worth" $n = 8$ observations. It contains the same amount of statistical information as an imaginary sample of $n = 8$ postvocalic *r*-tokens, with 3 [ɹ]'s, 4 ∅'s, and 1 [ɐ].

[50] Since we used a Dirichlet distribution to simulate the probabilities, the correlation between the individual values is taken into account. Thus, if the simulated value for, say, [ɹ] is high, the values for the other variants must be lower in this draw.

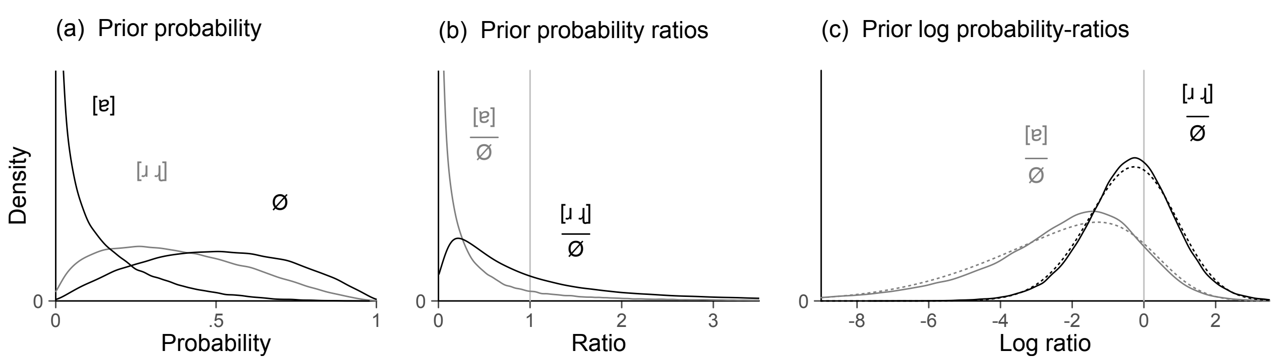


Figure A.12: Postvocalic *r*: Construction of the priors for the log probability-ratios based on the prior probabilities of the three variants. Solid lines show simulated values for (a) probabilities, (b) probability ratios, and (c) log probability-ratios. The dashed curves in panel (c) are the parametric prior distributions specified to approximate the simulated values.

averages computed over a distribution of subject-specific and word-specific probabilities. Since the regression model estimates conditional (i.e. subject-/word-specific) estimates and parameters, we need to translate the conditional $\bar{\alpha}_{\ldots}$ parameters into a marginal form, to which we can then attach our prior distributions (see Liu 2016: 349, 355).[51]

$$\bar{\alpha}^{\text{ɹ}} + \frac{{\sigma_{\alpha_{\mathrm{s}}^{\text{ɹ}}}}^2 + {\sigma_{\alpha_{\mathrm{w}}^{\text{ɹ}}}}^2}{2} \sim \text{SkewNormal}(0.55, 1.57, –1.13)$$
$$\bar{\alpha}^{\text{ɐ}} + \frac{{\sigma_{\alpha_{\mathrm{s}}^{\text{ɐ}}}}^2 + {\sigma_{\alpha_{\mathrm{w}}^{\text{ɐ}}}}^2}{2} \sim \text{SkewNormal}(0.32, 3.61, –3.48)$$

[51] The tricky part here is that the mapping between marginal and conditional parameter values depends on the variance (or standard deviation) of the varying intercepts, which also need to be estimated by the model. Fortunately, *Stan* allows us to incorporate unknown parameters into a prior.

Finally, for the $\bar{\beta}$ and $\bar{\gamma}$ parameters we chose mildly regularizing priors centered at zero.

$$\bar{\gamma}_{\ldots} \sim \text{Normal}(0, 1)$$
$$\bar{\beta}_{\ldots} \sim \text{Normal}(0, 1)$$

To guard estimates against deviant subjects and words, we use *t*-distributed rather than normal errors (Gelman et al. 2014: 437). The degrees-of-freedom parameter ν was set to 7.[52]

$$\nu_{\ldots} = 7$$

[52] See Appendix A.3.1 and Figure A.3 (p. 253) for illustration.

Stan code Translated into the *Stan* language, the model statement reads as follows.

```
data {
  int<lower=0> N;
  int y[N];
  int K;
  int<lower=0> subj[N];
  int<lower=0> word[N];
  int<lower=0> n_subj;
  int<lower=0> n_word;
  real link[N];
  real strong_c[N];
  real clust_c[N];
  real con_c[N];
  int var_id[N];
  real var_score[N];
  real far_z[N];
  int n_varcoef_subj;
  int n_varcoef_word;
  int n_cor_subj;
  int n_cor_word;
  int set_nu;
}
parameters {
  row_vector[K-1] a;
  row_vector[K-1] b;

  vector[K-1] g_link;
  vector[K-1] g_strong;
  vector[K-1] g_clust;
  vector[K-1] g_con;

  vector[K-1] b_link;
  vector[K-1] b_strong;
  vector[K-1] b_clust;
  vector[K-1] b_con;

  matrix[2,K-1] b_var;
  matrix[2,K-1] b_var_link;
  matrix[2,K-1] b_var_strong;
  matrix[2,K-1] b_var_clust;
  matrix[2,K-1] b_var_con;
```

```
  vector<lower=0>[n_varcoef_subj] sigma_subj;
  vector<lower=0>[n_varcoef_word] sigma_word;

  matrix[n_varcoef_subj, n_subj] z_subj;
  matrix[n_varcoef_word, n_word] z_word;

  cholesky_factor_corr[n_varcoef_subj] L_Rho_subj;
  cholesky_factor_corr[n_varcoef_word] L_Rho_word;

  real<lower=0> udf_word;
  real<lower=0> udf_subj;
}
transformed parameters {
  matrix[n_subj, n_varcoef_subj] r_s;
  matrix[n_word, n_varcoef_word] r_w;

  matrix[N,K] mu;

  r_s = sqrt(set_nu * udf_subj) * (diag_pre_multiply(sigma_subj, L_Rho_subj) * z_subj)';
  r_w = sqrt(set_nu * udf_word) * (diag_pre_multiply(sigma_word, L_Rho_word) * z_word)';

  for (n in 1:N) {
    if(far_z[n] < 0){
      mu[n,1] = a[1]         + r_s[subj[n],1]  + r_w[word[n],1] +
                (g_link[1]   + r_s[subj[n],2]) * link[n] +
                (g_strong[1] + r_s[subj[n],3]) * strong_c[n] +
                (g_clust[1]  + r_s[subj[n],4]) * clust_c[n] +
                (g_con[1]    + r_s[subj[n],5]) * con_c[n] +
                (b[1] + b_link[1]   * link[n]
                      + b_strong[1] * strong_c[n]
                      + b_clust[1]  * clust_c[n]
                      + b_con[1]    * con_c[n]
                      + r_w[word[n], 2]) * far_z[n];
      mu[n,2] = a[2]         + r_s[subj[n],6]  + r_w[word[n],3] +
                (g_link[2]   + r_s[subj[n],7]) * link[n] +
                (g_strong[2] + r_s[subj[n],8]) * strong_c[n] +
                (g_clust[2]  + r_s[subj[n],9]) * clust_c[n] +
                (g_con[2]    + r_s[subj[n],10]) * con_c[n] +
                (b[2] + b_link[2]   * link[n]
                      + b_strong[2] * strong_c[n]
                      + b_clust[2]  * clust_c[n]
                      + b_con[2]    * con_c[n]
                      + r_w[word[n], 4]) * far_z[n];
      mu[n,3] = 0;
    }
    else{
      mu[n,1] = a[1]         + r_s[subj[n],1]  + r_w[word[n],1] +
                (g_link[1]   + r_s[subj[n],2]) * link[n] +
                (g_strong[1] + r_s[subj[n],3]) * strong_c[n] +
                (g_clust[1]  + r_s[subj[n],4]) * clust_c[n] +
                (g_con[1]    + r_s[subj[n],5]) * con_c[n] +
                (b_var[var_id[n],1]         * var_score[n]
                + b_var_link[var_id[n],1]   * var_score[n] * link[n]
                + b_var_strong[var_id[n],1] * var_score[n] * strong_c[n]
                + b_var_clust[var_id[n],1]  * var_score[n] * clust_c[n]
                + b_var_con[var_id[n],1]    * var_score[n] * con_c[n]
                + r_w[word[n],2]) * far_z[n];

      mu[n,2] = a[2]        + r_s[subj[n],6]  + r_w[word[n],3] +
                (g_link[2]   + r_s[subj[n],7]) * link[n] +
                (g_strong[2] + r_s[subj[n],8]) * strong_c[n] +
                (g_clust[2]  + r_s[subj[n],9]) * clust_c[n] +
                (g_con[2]    + r_s[subj[n],10]) * con_c[n] +
                (b_var[var_id[n],2]         * var_score[n]
                + b_var_link[var_id[n],2]   * var_score[n] * link[n]
                + b_var_strong[var_id[n],2] * var_score[n] * strong_c[n]
                + b_var_clust[var_id[n],2]  * var_score[n] * clust_c[n]
                + b_var_con[var_id[n],2]    * var_score[n] * con_c[n]
                + r_w[word[n],4]) * far_z[n];

      mu[n,3] = 0;
    }
  }
}
model {
  L_Rho_subj  ~ lkj_corr_cholesky(4);
  L_Rho_word  ~ lkj_corr_cholesky(4);

  sigma_subj ~ exponential(2);
  sigma_word ~ exponential(2);

  to_vector(z_subj)  ~ normal(0,1);
```

```
  to_vector(z_word)  ~ normal(0,1);

  (a[1] + ((sigma_subj[1]^2 + sigma_word[1]^2)/2)) ~ skew_normal( 0.55 , 1.57, -1.13 );
  (a[2] + ((sigma_subj[6]^2 + sigma_word[3]^2)/2)) ~ skew_normal( 0.32 , 3.61, -3.48 );

  to_vector(b) ~ normal( 0 , 1 );

  to_vector(g_link) ~ normal( 0 , 1 );
  to_vector(g_strong) ~ normal( 0 , 1 );
  to_vector(g_clust) ~ normal( 0 , 1 );
  to_vector(g_con) ~ normal( 0 , 1 );

  to_vector(b_link) ~ normal( 0 , 1 );
  to_vector(b_strong) ~ normal( 0 , 1 );
  to_vector(b_clust) ~ normal( 0 , 1 );
  to_vector(b_con) ~ normal( 0 , 1 );

  to_vector(b_var) ~ normal( 0 , 1 );

  to_vector(b_var_link) ~ normal( 0 , 1 );
  to_vector(b_var_strong) ~ normal( 0 , 1 );
  to_vector(b_var_clust) ~ normal( 0 , 1 );
  to_vector(b_var_con) ~ normal( 0 , 1 );

  udf_word ~ inv_chi_square(set_nu);
  udf_subj ~ inv_chi_square(set_nu);

  for(n in 1:N) y[n] ~ categorical_logit(mu[n]’);
}
generated quantities{
  matrix[n_varcoef_subj, n_varcoef_subj] Rho_subj;
  matrix[n_varcoef_word, n_varcoef_word] Rho_word;

  int pos_subj = 1;
  int pos_word = 1;

  vector<lower=-1, upper=1>[n_cor_subj] cor_subj;
  vector<lower=-1, upper=1>[n_cor_word] cor_word;

  Rho_subj = multiply_lower_tri_self_transpose(L_Rho_subj);
  Rho_word = multiply_lower_tri_self_transpose(L_Rho_word);

  for(i in 1:(n_varcoef_subj-1)){
    for(j in (i+1):n_varcoef_subj){
      cor_subj[pos_subj] = Rho_subj[i, j];
      pos_subj += 1;
    }
  }
  for(i in 1:(n_varcoef_word-1)){
    for(j in (i+1):n_varcoef_word){
      cor_word[pos_word] = Rho_word[i, j];
      pos_word += 1;
    }
  }
}
```

Convergence diagnostics Our visual convergence checks, which are displayed in Figure A.13, show no signs of malperformance: All $\hat{R}$ values (left) are below 1.01, the profile of the rank plot (middle) is flat, and the quantile plot (right) suggests that all key parameters are supported by a sufficiently large effective sample size (i.e. all above $n = 400$).

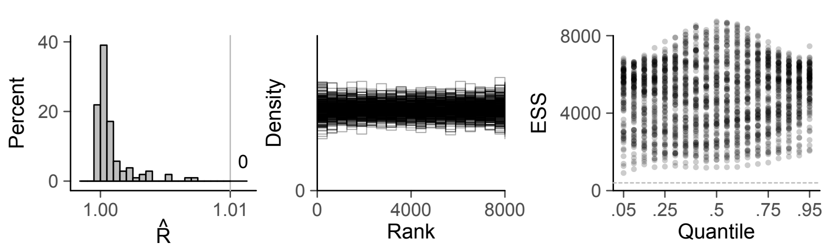


Figure A.13: Postvocalic /r/ model for the data from the present study: Convergence diagnostics.

Posterior distribution of model parameters The posterior distribution of the model parameters is reported in Table A.9 (fixed effects), Table A.10 (varying effects for subject: standard deviations), Table A.11 (varying effects for subject: correlations), and Table A.12 (varying effects for word).

Table A.9: Postvocalic /r/ model for the data from the current study: Posterior distribution of the fixed effects parameters.

			Posterior quantiles				
Parameter	Mean	SD	.05	.50	.95	N_{eff}	$\hat{R}$
$\bar{\alpha}^{\text{ɹ}}$	–2.44	0.59	–3.46	–2.41	–1.49	880	1.01
$\bar{\alpha}^{\text{ɐ}}$	–6.66	0.92	–8.28	–6.60	–5.31	2284	1.00
$\bar{\beta}^{\text{ɹ}}$	1.19	0.67	0.08	1.20	2.27	1046	1.01
$\bar{\beta}^{\text{ɐ}}$	–0.81	0.49	–1.60	–0.82	0.00	3514	1.00
$\bar{\gamma}^{\text{ɹ}}_{\text{Link}}$	3.37	0.66	2.28	3.36	4.46	5625	1.00
$\bar{\gamma}^{\text{ɐ}}_{\text{Link}}$	0.32	0.92	–1.22	0.33	1.82	8001	1.00
$\bar{\gamma}^{\text{ɹ}}_{\text{Str}}$	1.41	0.32	0.88	1.41	1.94	1997	1.00
$\bar{\gamma}^{\text{ɐ}}_{\text{Str}}$	1.36	0.46	0.60	1.36	2.11	4054	1.00
$\bar{\gamma}^{\text{ɹ}}_{\text{Clu}}$	0.34	0.29	–0.15	0.34	0.80	1599	1.00
$\bar{\gamma}^{\text{ɐ}}_{\text{Clu}}$	–0.30	0.59	–1.25	–0.31	0.68	3060	1.00
$\bar{\gamma}^{\text{ɹ}}_{\text{Con}}$	–0.86	0.29	–1.34	–0.87	–0.38	2018	1.00
$\bar{\gamma}^{\text{ɐ}}_{\text{Con}}$	–0.47	0.31	–1.00	–0.47	0.03	4263	1.00
$\bar{\beta}^{\text{ɹ}}_{\text{Link}}$	–0.45	0.78	–1.74	–0.44	0.82	6450	1.00
$\bar{\beta}^{\text{ɐ}}_{\text{Link}}$	–0.23	0.93	–1.73	–0.21	1.33	8560	1.00
$\bar{\beta}^{\text{ɹ}}_{\text{Str}}$	0.05	0.35	–0.52	0.05	0.62	3018	1.00
$\bar{\beta}^{\text{ɐ}}_{\text{Str}}$	–0.45	0.46	–1.23	–0.44	0.30	3917	1.00
$\bar{\beta}^{\text{ɹ}}_{\text{Clu}}$	0.25	0.27	–0.20	0.25	0.69	3765	1.00
$\bar{\beta}^{\text{ɐ}}_{\text{Clu}}$	–0.07	0.32	–0.59	–0.07	0.46	4831	1.00
$\bar{\beta}^{\text{ɹ}}_{\text{Con}}$	–0.41	0.34	–0.97	–0.41	0.15	2798	1.00
$\bar{\beta}^{\text{ɐ}}_{\text{Con}}$	0.13	0.28	–0.32	0.13	0.57	4616	1.00
$\bar{\beta}^{\text{ɹ}}_{\text{AmE}}$	2.22	0.90	0.72	2.25	3.69	2946	1.00
$\bar{\beta}^{\text{ɐ}}_{\text{AmE}}$	–0.46	0.88	–1.91	–0.45	0.99	7061	1.00
$\bar{\beta}^{\text{ɹ}}_{\text{BrE}}$	–1.63	0.76	–2.84	–1.65	–0.36	2116	1.00
$\bar{\beta}^{\text{ɐ}}_{\text{BrE}}$	–1.94	0.71	–3.12	–1.92	–0.81	5035	1.00
$\bar{\beta}^{\text{ɹ}}_{\text{Link, AmE}}$	0.01	0.98	–1.59	0.00	1.63	9255	1.00
$\bar{\beta}^{\text{ɐ}}_{\text{Link, AmE}}$	0.02	0.99	–1.61	0.03	1.64	8556	1.00
$\bar{\beta}^{\text{ɹ}}_{\text{Link, BrE}}$	0.70	0.94	–0.84	0.70	2.25	8856	1.00
$\bar{\beta}^{\text{ɐ}}_{\text{Link, BrE}}$	–0.02	0.99	–1.63	–0.03	1.63	9619	1.00
$\bar{\beta}^{\text{ɹ}}_{\text{Str, AmE}}$	0.45	0.56	–0.46	0.45	1.37	3934	1.00
$\bar{\beta}^{\text{ɐ}}_{\text{Str, AmE}}$	0.13	0.86	–1.28	0.14	1.57	7022	1.00
$\bar{\beta}^{\text{ɹ}}_{\text{Str, BrE}}$	–0.13	0.50	–0.94	–0.14	0.71	3738	1.00
$\bar{\beta}^{\text{ɐ}}_{\text{Str, BrE}}$	–1.06	0.70	–2.20	–1.08	0.11	6173	1.00
$\bar{\beta}^{\text{ɹ}}_{\text{Clu, AmE}}$	0.08	0.55	–0.82	0.07	1.01	5314	1.00
$\bar{\beta}^{\text{ɐ}}_{\text{Clu, AmE}}$	0.30	0.85	–1.09	0.30	1.70	8096	1.00
$\bar{\beta}^{\text{ɹ}}_{\text{Clu, BrE}}$	0.20	0.36	–0.38	0.20	0.80	4732	1.00
$\bar{\beta}^{\text{ɐ}}_{\text{Clu, BrE}}$	0.96	0.60	–0.02	0.96	1.93	6492	1.00
$\bar{\beta}^{\text{ɹ}}_{\text{Con, AmE}}$	–0.47	0.60	–1.47	–0.47	0.51	4469	1.00
$\bar{\beta}^{\text{ɐ}}_{\text{Con, AmE}}$	–0.03	0.85	–1.42	–0.02	1.35	7779	1.00
$\bar{\beta}^{\text{ɹ}}_{\text{Con, BrE}}$	–0.35	0.45	–1.11	–0.34	0.37	3563	1.00
$\bar{\beta}^{\text{ɐ}}_{\text{Con, BrE}}$	0.12	0.56	–0.81	0.13	1.04	6736	1.00

Table A.10: Postvocalic /r/ model for the data from the current study: Posterior distribution of varying effects parameters for subject (standard deviation parameters).

Parameter	Mean	SD	Quantiles .05	.95	N_{eff}	$\hat{R}$
$\sigma_{\alpha_s^{ɹ}}$	1.95	0.39	1.31	2.59	1841	1
$\sigma_{\gamma^{ɹ}_{\text{Link, s}}}$	0.64	0.51	0.04	1.59	1592	1
$\sigma_{\gamma^{ɹ}_{\text{Str, s}}}$	0.51	0.14	0.30	0.77	2041	1
$\sigma_{\gamma^{ɹ}_{\text{Clu, s}}}$	0.27	0.11	0.09	0.46	1538	1
$\sigma_{\gamma^{ɹ}_{\text{Con, s}}}$	0.57	0.15	0.34	0.84	1851	1
$\sigma_{\alpha_s^{ɐ}}$	0.48	0.21	0.13	0.84	1391	1
$\sigma_{\gamma^{ɐ}_{\text{Link,s}}}$	0.67	0.59	0.03	1.85	4618	1
$\sigma_{\gamma^{ɐ}_{\text{Str, s}}}$	0.70	0.22	0.37	1.08	1715	1
$\sigma_{\gamma^{ɐ}_{\text{Clu, s}}}$	0.36	0.15	0.12	0.62	1618	1
$\sigma_{\gamma^{ɐ}_{\text{Con, s}}}$	0.16	0.11	0.01	0.38	2391	1

Table A.11: Postvocalic /r/ model for the data from the current study: Correlation matrix for the by-subject varying effects parameters. The lower triangle reports posterior medians with .05 and .95 quantiles (i.e. 90% uncertainty intervals). Superscripts reflect the upper limit, subscripts the lower limit. The upper triangle lists $\hat{R}$ statistics and the effective sample size for each correlation parameter.

Parameter	$\alpha_s^{ɹ}$	$\gamma^{ɹ}_{\text{Link, s}}$	$\gamma^{ɹ}_{\text{Str, s}}$	$\gamma^{ɹ}_{\text{Clu, s}}$	$\gamma^{ɹ}_{\text{Con, s}}$	$\alpha_s^{ɐ}$	$\gamma^{ɐ}_{\text{Link,s}}$	$\gamma^{ɐ}_{\text{Str, s}}$	$\gamma^{ɐ}_{\text{Clu, s}}$	$\gamma^{ɐ}_{\text{Con, s}}$
$\alpha_s^{ɹ}$		$^{1}_{7590}$	$^{1}_{4942}$	$^{1}_{6299}$	$^{1}_{4572}$	$^{1}_{5346}$	$^{1}_{8938}$	$^{1}_{4443}$	$^{1}_{6052}$	$^{1}_{7350}$
$\gamma^{ɹ}_{\text{Link, s}}$	$-.04^{+.36}_{-.42}$		$^{1}_{1282}$	$^{1}_{2378}$	$^{1}_{1255}$	$^{1}_{2668}$	$^{1}_{8707}$	$^{1.01}_{1246}$	$^{1}_{2967}$	$^{1}_{5784}$
$\gamma^{ɹ}_{\text{Str, s}}$	$+.08^{+.40}_{-.24}$	$+.03^{+.40}_{-.36}$		$^{1}_{4494}$	$^{1}_{2339}$	$^{1}_{4991}$	$^{1}_{7547}$	$^{1}_{2926}$	$^{1}_{4528}$	$^{1}_{6935}$
$\gamma^{ɹ}_{\text{Clu, s}}$	$+.10^{+.46}_{-.27}$	$-.04^{+.34}_{-.43}$	$+.13^{+.47}_{-.22}$		$^{1.01}_{1496}$	$^{1}_{3482}$	$^{1}_{6239}$	$^{1}_{2700}$	$^{1}_{4037}$	$^{1}_{6003}$
$\gamma^{ɹ}_{\text{Con, s}}$	$-.29^{+.03}_{-.56}$	$+.02^{+.38}_{-.35}$	$+.01^{+.30}_{-.28}$	$-.28^{+.08}_{-.57}$		$^{1}_{5242}$	$^{1}_{7845}$	$^{1}_{4143}$	$^{1}_{5989}$	$^{1}_{7177}$
$\alpha_s^{ɐ}$	$-.20^{+.15}_{-.49}$	$+.08^{+.46}_{-.32}$	$+.09^{+.44}_{-.26}$	$-.08^{+.32}_{-.43}$	$+.21^{+.52}_{-.14}$		$^{1}_{7443}$	$^{1}_{2638}$	$^{1}_{4087}$	$^{1}_{4655}$
$\gamma^{ɐ}_{\text{Link,s}}$	$+.05^{+.44}_{-.35}$	$.00^{+.39}_{-.40}$	$+.04^{+.43}_{-.36}$	$+.06^{+.45}_{-.34}$	$-.09^{+.32}_{-.48}$	$-.03^{+.38}_{-.42}$		$^{1}_{2361}$	$^{1}_{3163}$	$^{1}_{5581}$
$\gamma^{ɐ}_{\text{Str, s}}$	$-.23^{+.05}_{-.49}$	$+.06^{+.43}_{-.32}$	$+.25^{+.54}_{-.08}$	$+.14^{+.47}_{-.22}$	$+.12^{+.41}_{-.20}$	$+.08^{+.44}_{-.30}$	$.00^{+.39}_{-.39}$		$^{1}_{4131}$	$^{1}_{7067}$
$\gamma^{ɐ}_{\text{Clu, s}}$	$+.15^{+.45}_{-.19}$	$-.03^{+.37}_{-.41}$	$-.13^{+.23}_{-.47}$	$-.06^{+.31}_{-.42}$	$-.02^{+.32}_{-.34}$	$-.07^{+.31}_{-.42}$	$-.02^{+.38}_{-.41}$	$-.13^{+.23}_{-.47}$		$^{1}_{5137}$
$\gamma^{ɐ}_{\text{Con, s}}$	$+.11^{+.46}_{-.29}$	$-.04^{+.36}_{-.43}$	$-.05^{+.34}_{-.43}$	$+.04^{+.42}_{-.35}$	$-.05^{+.34}_{-.43}$	$-.11^{+.30}_{-.49}$	$.00^{+.40}_{-.40}$	$-.05^{+.34}_{-.43}$	$+.10^{+.47}_{-.30}$	

Table A.12: Postvocalic /r/ model for the data from the current study: Posterior distribution of the varying effects parameters for word.

Parameter	Mean	SD	Posterior quantiles .05	.50	.95	N_{eff}	$\hat{R}$
Standard deviation parameters							
$\sigma_{\alpha_w^{ɹ}}$	0.98	0.27	0.58	0.96	1.45	2501	1
$\sigma_{\beta^{ɹ}_{\text{FAR, w}}}$	0.34	0.12	0.17	0.33	0.56	2719	1
$\sigma_{\alpha_w^{ɐ}}$	2.37	0.64	1.35	2.35	3.47	2622	1
$\sigma_{\beta^{ɐ}_{\text{FAR, w}}}$	0.24	0.18	0.02	0.20	0.59	2878	1
Correlation parameters							
$\rho(\alpha_w^{ɹ}, \beta^{ɹ}_{\text{FAR, w}})$	–0.19	0.21	–0.52	–0.19	0.18	4997	1
$\rho(\alpha_w^{ɹ}, \alpha_w^{ɐ})$	–0.05	0.17	–0.33	–0.05	0.23	2848	1
$\rho(\alpha_w^{ɹ}, \beta^{ɐ}_{\text{FAR, w}})$	–0.02	0.28	–0.48	–0.01	0.43	6894	1
$\rho(\beta_w^{ɹ}, \alpha_w^{ɐ})$	–0.42	0.20	–0.71	–0.44	–0.07	1406	1
$\rho(\beta_w^{ɹ}, \beta^{ɐ}_{\text{FAR, w}})$	0.13	0.29	–0.37	0.15	0.58	5975	1
$\rho(\alpha_w^{ɐ}, \beta^{ɐ}_{\text{FAR, w}})$	–0.18	0.30	–0.64	–0.20	0.35	5146	1

A.7 The labio-velar glide /w/

Here we provide additional information about the analysis of the data from Pascoe (1987). Details about the analyses of the data by Wunder (2012), Rank (2016), and the current study can be found in the web appendix.[53] The structure of the models fit to these data sets is different, as documented in these supplements. The data sets analyzed in Chapter 7 are openly available via *TROLLing*[54] (Sönning et al. 2020b). The OSF project accompanying this study includes the *Stan* code for the regression models[55] and the *R* script[56] for running the analyses.

[53] https://osf.io/aszq6/

[54] https://doi.org/10.18710/F1A34O

[55] This study: https://osf.io/shm59/; Pascoe 1987: https://osf.io/a37js/; Wunder 2012: https://osf.io/vnbzs/; Rank 2016: https://osf.io/8nms3/

[56] https://osf.io/f3msb/

A.7.1 Pascoe 1987

Model definition The realization of /w/ was coded as a binary response (1 for target-like realization, 0 for other variants) and therefore modeled using a Bernoulli distribution. The parameter p denotes the probability of observing target-like [w], and the subscript i indexes the *w*-tokens in the data set.

$$\mathrm{w}_i \sim \mathrm{Bernoulli}\,(1, p_i)$$

To model the probability of observing target-like [w], we use the logit link function.[57] The $\bar{\alpha}$ parameter denotes the logit-transformed probability of a target-like rendition when all predictors are held at 0. The α coefficients represent varying intercepts for speakers (α_{s}), and for word (α_{w}). These reflect by how much each speaker/word deviates from the average logit. The $\bar{\beta}^{\mathrm{FAR}}$ parameter is the slope of the regression line that captures how target accuracy varies across proficiency levels. The $\beta_{\mathrm{w}}^{\mathrm{FAR}}$ parameters convey by how much each word departs from this slope.

[57] This link function is commonly used in regression models for binary outcomes. It maps constrained probabilities (i.e. the interval $[0; 1]$) to an unconstrained scale from $-\infty$ to $+\infty$. On this unconstrained logit scale, we model the expected logits for conditions of interest.

$$\begin{aligned}\mathrm{logit}(p_i) = \;& \bar{\alpha} + \alpha_{\mathrm{s}[i]} + \alpha_{\mathrm{w}[i]} + (\bar{\beta}^{\mathrm{FAR}} + \beta^{\mathrm{FAR}}_{\mathrm{w}[i]})\mathrm{FAR}_i + \\ & (\bar{\beta}^{\mathrm{a}} + \beta^{\mathrm{a}}_{\mathrm{s}[i]})\mathrm{a}_i + (\bar{\beta}^{\mathrm{e}} + \beta^{\mathrm{e}}_{\mathrm{s}[i]})\mathrm{e}_i + (\bar{\beta}^{\mathrm{o}} + \beta^{\mathrm{o}}_{\mathrm{s}[i]})\mathrm{o}_i + \\ & (\bar{\beta}^{\mathrm{u}} + \beta^{\mathrm{u}}_{\mathrm{s}[i]})\mathrm{u}_i + (\bar{\beta}^{\mathrm{c}} + \beta^{\mathrm{c}}_{\mathrm{s}[i]})\mathrm{c}_i + \\ & (\bar{\beta}^{\mathrm{Vs}} + \beta^{\mathrm{Vs}}_{\mathrm{s}[i]})\mathrm{Vs}_i + (\bar{\beta}^{\mathrm{Vr}} + \beta^{\mathrm{Vr}}_{\mathrm{s}[i]})\mathrm{Vr}_i + (\bar{\beta}_{\mathrm{Cv}} + \beta^{\mathrm{Cv}}_{\mathrm{s}[i]})\mathrm{Cv}_i + \\ & (\bar{\beta}^{\mathrm{Cvl}} + \beta^{\mathrm{Cvl}}_{\mathrm{s}[i]})\mathrm{Cvl}_i + (\bar{\beta}^{\mathrm{P}} + \beta^{\mathrm{P}}_{\mathrm{s}[i]})\mathrm{P}_i\end{aligned}$$

The next two lines specify the variation depending on the following vowel. We distinguish 6 vowel categories, which means that we may use 5 contrasts to represent them in this model. The aim of the representation we chose is to make estimates comparable across the four data sets. In other words, we would like the intercept, which reflects the probability of a target-like rendition when all contrasts are set to 0, to have the same interpretation in all analyses. The four data sets share three vowel categories: *e*, *i*, and *o*. Our contrast coding maps to the intercept the average accuracy rate across these three categories. In order to apply these custom contrasts, we followed the guidelines given by Schad et al. (2020). Table A.13 reports the coding we applied, along with the corresponding hypothesis matrix[58], which makes transparent the difference signaled by each coefficient.

[58] The hypothesis matrix gives a more transparent expression of the meaning of the intercept and the contrasts. The weights for the intercept reflect the meaning of this parameter. In the present case, it is the average over 3 vowel categories (*e, i, o*). For the contrasts, the hypothesis matrix shows the comparison expressed by each coefficient. Each contrast sums to zero and the positive and negative weights show which categories are being compared.

For the preceding segment, we applied the same rationale. Here, the

	Contrasts					Hypothesis matrix					
Vowel	a	e	o	u	c	Intercept	a	e	o	u	c
a [a ʌ ɑː]	−1	0	0	0	0	0	1	0	0	0	0
e [e eɪ æ]	0	−1	0	0	0	1/3	1/3	−2/3	1/3	1/3	1/3
i [iː ɪ]	0	1	1	0	0	1/3	1/3	1/3	1/3	1/3	1/3
o [ɒ ɔ ɔː]	0	0	−1	0	0	1/3	1/3	1/3	−2/3	1/3	1/3
u [ʊ uː]	0	0	0	−1	0	0	0	0	0	−1	0
c [ə ɜː]	0	0	0	0	−1	0	0	0	0	0	−1

Table A.13: /w/ model for the data from the current study: Custom contrasts for the following vowel. With this coding, the intercept represents the average over the three vowel categories that are represented in all three data sets (*e i o*).

four data sets share two preceding segment categories: neutral vowel and voiced consonant. Accordingly, the intercept was defined as the average accuracy rate over these two contexts. Table A.14 lists the contrast coding and the corresponding hypothesis matrix.

	Contrasts					Hypothesis matrix					
Context	Vs	Vr	Cv	Cvl	P	Intercept	Vs	Vr	Cv	Cvl	P
Spread V	−1	0	0	0	0	0	−1	0	0	0	0
Neutral V	0	−1	0	0	0	1/2	1/2	−1/2	1/2	1/2	1/2
Round V	0	0	−1	0	0	0	0	0	−1	0	0
Voiced C	0	1	0	0	0	1/2	1/2	1/2	1/2	1/2	1/2
Voiceless C	0	0	0	−1	0	0	0	0	0	−1	0
Pause	0	0	0	0	−1	0	0	0	0	0	−1

Table A.14: /w/ model for the data from the current study: Custom contrasts for the preceding segment. With this coding, the intercept represents the average over the two segment classes that are represented in all three data sets (Neutral V, voiceless C).

The $\bar{\beta}$ coefficients therefore represent these contrasts, and the β_s parameters are varying slopes, which reflect by-speaker deviations from the average difference instantiated by a given contrast.

The next block defines the joint distribution of the varying effects. Components for word and subject are each given a joint multivariate Student-*t* distribution with the degrees-of-freedom parameter ν set to 7.[59] **S** denotes the two covariance matrices, which detail how the varying effects in each set are distributed.[60]

$$\begin{bmatrix} \alpha_s & & & & \\ \beta_s^a & \beta_s^e & \beta_s^o & \beta_s^u & \beta_s^c \\ \beta_s^{Vs} & \beta_s^{Vr} & \beta_s^{Cv} & \beta_s^{Cvl} & \beta_s^{P} \end{bmatrix} \sim \text{MVStudent-t}\left(\begin{bmatrix} 0 & & & & \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \mathbf{S}_s, \nu_s \right)$$

$$\begin{bmatrix} \alpha_w \\ \beta_w^{FAR} \end{bmatrix} \sim \text{MVStudent-t}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mathbf{S}_w, \nu_w \right)$$

The covariance matrices[61] are described with two quantities, the standard deviation parameters, which specify how much each varying effect is spread out along the outcome (logit) scale, and a correlation matrix **R**, which contains correlation coefficients that reflect the association between the batches of varying effects in each set.[62]

[59] See Appendix A.3.1 and Figure A.3 (p. 253) for illustration.

[60] To save space, we arranged them in multiple columns – they would normally be listed in a single column.

[61] We simplify notation and omit the translation of standard deviations and the correlation matrix **R** into a covariance matrix. See McElreath (2020: 441) for a complete representation.

[62] If, for instance, the standard deviation for the varying intercepts for subjects (σ_{α_s}) were higher than that for words (σ_{α_w}), this would indicate that, when holding constant all predictors in the model, the conditional variation among subjects is higher than that among words. In other words, heterogeneity among speakers would be more pronounced than among words, indicating that there is more residual variation among speakers that we haven't managed to account for with our model. A positive correlation between, say, the by-word varying effects α_w and β_w^{FAR} would tell us that the accuracy rate of difficult words (i.e. items with a downward deflection in terms of overall accuracy rate) would increase more slowly across proficiency levels (i.e. the slopes are also downwardly deflected).

$$\mathbf{S}_{\mathrm{s}} = \mathrm{CovarianceMatrix}\left(\begin{array}{lllll} \sigma_{\alpha_{\mathrm{s}}} & & & & \\ \sigma_{\beta_{\mathrm{s}}^{\mathrm{Vs}}} & \sigma_{\beta_{\mathrm{s}}^{\mathrm{Vr}}} & \sigma_{\beta_{\mathrm{s}}^{\mathrm{Cv}}} & \sigma_{\beta_{\mathrm{s}}^{\mathrm{Cvl}}} & \sigma_{\beta_{\mathrm{s}}^{\mathrm{P}}} \\ \sigma_{\beta_{\mathrm{s}}^{\mathrm{a}}} & \sigma_{\beta_{\mathrm{s}}^{\mathrm{e}}} & \sigma_{\beta_{\mathrm{s}}^{\mathrm{o}}} & \sigma_{\beta_{\mathrm{s}}^{\mathrm{u}}} & \sigma_{\beta_{\mathrm{s}}^{\mathrm{c}}} \end{array}, \mathbf{R}_{\mathrm{s}}\right)$$

$$\mathbf{S}_{\mathrm{w}} = \mathrm{CovarianceMatrix}\left(\sigma_{\alpha_{\mathrm{w}}}, \sigma_{\beta_{\mathrm{w}}^{\mathrm{FAR}}}, \mathbf{R}_{\mathrm{w}}\right)$$

As for the priors, we chose a weakly regularizing a priori statement for the σ parameters in the form of an exponential distribution with scale parameter 2.[63]

$$\sigma_{\ldots} \sim \mathrm{Exponential}(2)$$

For the correlation between varying effects, we are relatively ignorant and therefore default to an LKJ prior the shape parameter η set to 4.[64]

$$\mathbf{R}_{\ldots} \sim \mathrm{LKJcorr}(4)$$

For the average accuracy rate when all predictors are held at 0, the priors derived from our survey of the research literature (see §7.6) apply to the marginal distribution of the $\bar{\alpha}$ parameters. We need to link these to the subject-specific quantities required by the model.[65]

$$\frac{\bar{\alpha}}{\sqrt{c^2({\sigma_{\alpha_s}}^2 + {\sigma_{\alpha_w}}^2) + 1}} \sim \mathrm{Normal}(2, 1.5)$$

The $\bar{\beta}$ parameters were given weakly regularizing priors centered at 0.

$$\bar{\beta}^{\cdots} \sim \mathrm{Normal}(0, 2)$$

To make the model more robust against unusual subjects and unusual words, we use t-distributed rather than normal errors (Gelman et al. 2014: 437). The degrees-of-freedom parameter ν was set to 7.[66].

$$\nu_{\ldots} = 7$$

[63] See Appendix A.5.1 and Figure A.9.

[64] See Appendix A.3.1 and Figure A.2 (p. 253) for illustration.

[65] See §3.4.1 (p. 43) for a brief discussion and Appendix A.5.1 (p. 271) for details about the procedure used here.

[66] see Appendix A.3.1 and Figure A.3 (p. 253) for illustration.

Stan code This is the *Stan* code for the model:

```
data {
  int<lower=0> N;
  int y[N];
  int<lower=0> subj[N];
  int<lower=0> word[N];
  int<lower=0> n_subj;
  int<lower=0> n_word;
  real far_z[N];
  real C_a[N];
  real C_e[N];
  real C_o[N];
  real C_u[N];
  real C_c[N];
  real C_Vs[N];
  real C_Vr[N];
  real C_Cv[N];
  real C_Cvl[N];
  real C_P[N];
  int n_varcoef_subj;
  int n_varcoef_word;
  int n_cor_subj;
  int n_cor_word;
  int set_nu;
}
parameters {
  real a;
```

```
  real b_far;
  real b_a;
  real b_e;
  real b_o;
  real b_u;
  real b_c;
  real b_Vs;
  real b_Vr;
  real b_Cv;
  real b_Cvl;
  real b_P;

  vector<lower=0>[n_varcoef_subj] sigma_subj;
  vector<lower=0>[n_varcoef_word] sigma_word;

  matrix[n_varcoef_subj, n_subj] z_subj;
  matrix[n_varcoef_word, n_word] z_word;
  cholesky_factor_corr[n_varcoef_subj] L_Rho_subj;
  cholesky_factor_corr[n_varcoef_word] L_Rho_word;

  real<lower=0> udf_word;
  real<lower=0> udf_subj;
}
transformed parameters {
  matrix[n_subj, n_varcoef_subj] r_s;
  matrix[n_word, n_varcoef_word] r_w;
  vector[N] mu;

  r_s = sqrt(set_nu * udf_subj) * (diag_pre_multiply(sigma_subj, L_Rho_subj) * z_subj)';
  r_w = sqrt(set_nu * udf_word) * (diag_pre_multiply(sigma_word, L_Rho_word) * z_word)';

  for (n in 1:N) {
    mu[n] = a     + r_s[subj[n],1] + r_w[word[n], 1]
         + (b_far +                 r_w[word[n],2]) * far_z[n]
         + (b_a   + r_s[subj[n],2])                 * C_a[n]
         + (b_e   + r_s[subj[n],3])                 * C_e[n]
         + (b_o   + r_s[subj[n],4])                 * C_o[n]
         + (b_u   + r_s[subj[n],5])                 * C_u[n]
         + (b_c   + r_s[subj[n],6])                 * C_c[n]

         + (b_Vs  + r_s[subj[n],7])                 * C_Vs[n]
         + (b_Vr  + r_s[subj[n],8])                 * C_Vr[n]
         + (b_Cv  + r_s[subj[n],9])                 * C_Cv[n]
         + (b_Cvl + r_s[subj[n],10])                * C_Cvl[n]
         + (b_P   + r_s[subj[n],11])                * C_P[n];
        }
}
model {
  L_Rho_subj  ~ lkj_corr_cholesky(4);
  L_Rho_word  ~ lkj_corr_cholesky(4);
  sigma_subj ~ exponential(2);
  sigma_word ~ exponential(2);

  to_vector(z_subj)  ~ normal(0,1);
  to_vector(z_word)  ~ normal(0,1);

  a/sqrt((16*sqrt(3)/(15*pi())) * (sigma_subj[1]^2 + sigma_word[1]^2) + 1) ~ normal(2, 1.5);

  b_far ~ normal( 0 , 2 );
  b_a ~ normal( 0 , 2 );
  b_e ~ normal( 0 , 2 );
  b_o ~ normal( 0 , 2 );
  b_u ~ normal( 0 , 2 );
  b_c ~ normal( 0 , 2 );

  b_Vs ~ normal( 0 , 2 );
  b_Vr ~ normal( 0 , 2 );
  b_Cv ~ normal( 0 , 2 );
  b_Cvl ~ normal( 0 , 2 );
  b_P ~ normal( 0 , 2 );

  udf_word ~ inv_chi_square(set_nu);
  udf_subj ~ inv_chi_square(set_nu);

  target += bernoulli_logit_lpmf(y | mu);
}
generated quantities{
  matrix[n_varcoef_subj, n_varcoef_subj] Rho_subj;
  matrix[n_varcoef_word, n_varcoef_word] Rho_word;

  int pos_subj = 1;
  int pos_word = 1;
```

```
  vector<lower=-1, upper=1>[n_cor_subj] cor_subj;
  vector<lower=-1, upper=1>[n_cor_word] cor_word;

  Rho_subj = multiply_lower_tri_self_transpose(L_Rho_subj);
  Rho_word = multiply_lower_tri_self_transpose(L_Rho_word);

  for(i in 1:(n_varcoef_subj-1)){
    for(j in (i+1):n_varcoef_subj){
      cor_subj[pos_subj] = Rho_subj[i, j];
      pos_subj += 1;
    }
  }
  for(i in 1:(n_varcoef_word-1)){
    for(j in (i+1):n_varcoef_word){
      cor_word[pos_word] = Rho_word[i, j];
      pos_word += 1;
    }
  }
}
```

Convergence diagnostics Figure A.14 indicates convergence of the algorithm: All $\hat{R}$ values (left) are below 1.01, the profile of the rank plot (middle) is flat, and the quantile plot (right) suggests that all key parameters are supported by a sufficiently large effective sample size.

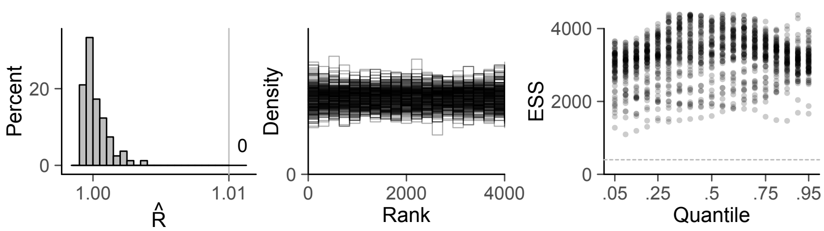


Figure A.14: /w/-model for the data from Pascoe (1987): Convergence diagnostics.

Posterior distribution of model parameters A summary of the posterior distribution of the parameters can be found in Table A.15 (fixed parameters), Table A.16 (varying effects; for subject: standard deviation of varying effects), and Table A.17 (correlation of varying effects for subject).

			Posterior quantiles				
Parameter	Mean	SD	.05	.50	.95	N_{eff}	$\hat{R}$
$\bar{\alpha}$	0.75	0.71	–0.32	0.70	2.00	1316	1
$\bar{\beta}^{\text{FAR}}$	3.58	0.73	2.48	3.55	4.83	1771	1
$\bar{\beta}^{\text{a}}$	–2.32	0.90	–3.80	–2.30	–0.86	2862	1
$\bar{\beta}^{\text{e}}$	1.26	0.54	0.42	1.25	2.17	2626	1
$\bar{\beta}^{\text{o}}$	–2.11	0.60	–3.13	–2.09	–1.18	2020	1
$\bar{\beta}^{\text{u}}$	0.14	1.42	–2.24	0.15	2.43	4219	1
$\bar{\beta}^{\text{c}}$	–0.10	1.22	–2.17	–0.06	1.85	4418	1
$\bar{\beta}^{\text{Vs}}$	0.31	0.72	–0.89	0.29	1.50	2407	1
$\bar{\beta}^{\text{Vr}}$	0.01	0.52	–0.84	0.01	0.85	3028	1
$\bar{\beta}^{\text{Cv}}$	0.23	1.19	–1.77	0.25	2.17	4049	1
$\bar{\beta}^{\text{Cvl}}$	–1.13	0.77	–2.43	–1.13	0.12	3259	1
$\bar{\beta}^{\text{P}}$	–1.50	0.66	–2.62	–1.47	–0.44	2652	1

Table A.15: /w/-model for the data from Pascoe (1987): Posterior distribution of the fixed effects parameters.

Parameter	Mean	SD	Posterior quantiles .05	.50	.95	N_{eff}	$\hat{R}$
σ_{α_s}	1.60	0.56	0.79	1.53	2.59	1745	1
$\sigma_{\beta_s^a}$	0.38	0.35	0.02	0.27	1.07	3584	1
$\sigma_{\beta_s^e}$	0.46	0.37	0.04	0.38	1.20	1616	1
$\sigma_{\beta_s^o}$	0.76	0.41	0.15	0.71	1.48	1450	1
$\sigma_{\beta_s^u}$	0.55	0.54	0.03	0.38	1.64	4369	1
$\sigma_{\beta_s^c}$	0.45	0.44	0.02	0.31	1.33	5018	1
$\sigma_{\beta_s^{Vs}}$	0.45	0.40	0.03	0.34	1.24	1982	1
$\sigma_{\beta_s^{Vr}}$	0.47	0.42	0.03	0.36	1.28	1893	1
$\sigma_{\beta_s^{Cv}}$	0.46	0.42	0.02	0.33	1.29	4728	1
$\sigma_{\beta_s^{Cvl}}$	0.48	0.43	0.02	0.37	1.30	3017	1
$\sigma_{\beta_s^P}$	0.41	0.37	0.02	0.30	1.13	1916	1
σ_{α_w}	0.41	0.33	0.03	0.33	1.07	1706	1
$\sigma_{\beta_w^{FAR}}$	0.62	0.46	0.04	0.55	1.49	1582	1
$\rho(\alpha_w, \beta_w^{FAR})$	-0.04	0.33	-0.56	-0.05	0.51	3259	1

Table A.16: /w/-model for the data from Pascoe (1987): Posterior distribution of varying effects parameters for subject (standard deviation parameters only) and word.

Table A.17: /w/-model for the data from Pascoe (1987): Correlation matrix for the by-subject varying effects parameters. The lower triangle reports posterior medians with .05 and .95 quantiles (i.e. 90% uncertainty intervals). Superscripts reflect the upper limit, subscripts the lower limit. The upper triangle lists $\hat{R}$ statistics and the effective sample size for each correlation parameter.

Param.	α_s	β_s^a	β_s^e	β_s^o	β_s^u	β_s^c	β_s^{Vs}	β_s^{Vr}	β_s^{Cv}	β_s^{Cvl}	β_s^P
α_s		5699^{1}	5132^{1}	3827^{1}	5876^{1}	5334^{1}	4785^{1}	3955^{1}	6589^{1}	5531^{1}	4779^{1}
β_s^a	$-.01^{+.37}_{-.39}$		3924^{1}	2723^{1}	4601^{1}	4551^{1}	4191^{1}	3868^{1}	5278^{1}	5034^{1}	4641^{1}
β_s^e	$-.03^{+.36}_{-.41}$	$.00^{+.37}_{-.39}$		2688^{1}	4775^{1}	5013^{1}	3774^{1}	4097^{1}	4339^{1}	4366^{1}	4276^{1}
β_s^o	$+.06^{+.41}_{-.32}$	$-.02^{+.37}_{-.40}$	$+.03^{+.39}_{-.35}$		4824^{1}	4046^{1}	3959^{1}	4777^{1}	4702^{1}	4035^{1}	4277^{1}
β_s^u	$+.01^{+.41}_{-.37}$	$.00^{+.38}_{-.39}$	$.00^{+.38}_{-.38}$	$.00^{+.39}_{-.39}$		3251^{1}	3351^{1}	3410^{1}	3516^{1}	3411^{1}	3442^{1}
β_s^c	$.00^{+.38}_{-.37}$	$.00^{+.39}_{-.38}$	$+.01^{+.41}_{-.36}$	$+.01^{+.40}_{-.37}$	$+.01^{+.40}_{-.39}$		3644^{1}	3618^{1}	3961^{1}	3635^{1}	3461^{1}
β_s^{Vs}	$-.02^{+.38}_{-.39}$	$-.01^{+.39}_{-.39}$	$+.01^{+.39}_{-.39}$	$+.01^{+.39}_{-.36}$	$.00^{+.40}_{-.38}$	$-.01^{+.38}_{-.39}$		2852^{1}	3509^{1}	3118^{1}	3522^{1}
β_s^{Vr}	$+.05^{+.41}_{-.35}$	$.00^{+.39}_{-.38}$	$+.03^{+.41}_{-.38}$	$+.02^{+.40}_{-.37}$	$.00^{+.40}_{-.40}$	$+.02^{+.40}_{-.38}$	$-.02^{+.37}_{-.40}$		2890^{1}	3300^{1}	2991^{1}
β_s^{Cv}	$+.01^{+.39}_{-.39}$	$-.01^{+.39}_{-.38}$	$.00^{+.38}_{-.37}$	$.00^{+.37}_{-.40}$	$+.01^{+.41}_{-.39}$	$.00^{+.39}_{-.39}$	$-.01^{+.39}_{-.39}$	$+.01^{+.39}_{-.37}$		2950^{1}	2843^{1}
β_s^{Cvl}	$-.01^{+.38}_{-.38}$	$.00^{+.39}_{-.37}$	$+.01^{+.40}_{-.36}$	$.00^{+.37}_{-.37}$	$-.01^{+.38}_{-.39}$	$.00^{+.39}_{-.38}$	$+.01^{+.40}_{-.37}$	$-.01^{+.37}_{-.40}$	$.00^{+.40}_{-.39}$		2719^{1}
β_s^P	$+.01^{+.38}_{-.37}$	$+.01^{+.40}_{-.38}$	$+.02^{+.40}_{-.37}$	$+.03^{+.41}_{-.36}$	$.00^{+.38}_{-.39}$	$.00^{+.39}_{-.38}$	$+.01^{+.39}_{-.41}$	$+.01^{+.39}_{-.38}$	$+.01^{+.39}_{-.38}$	$-.01^{+.39}_{-.39}$	

A.8 The labiodental fricative /v/

The data sets analyzed in Chapter 8 are openly available via *TROLLing*[67] (Sönning & Rank 2020). The OSF project includes the *Stan* code for the regression models[68] and the *R* script[69] for running the analyses.

[67] https://doi.org/10.18710/B276ZX

[68] This study: https://osf.io/bkjuh/; Rank 2016: https://osf.io/rsd2g/

[69] https://osf.io/zj36f/

A.8.1 This study

Model definition As the realization of /v/ was coded with three categories, we applied a multinomial regression model for analysis. We model the probability of each variant (p^{v}, p^{w}, p^{υ}) as a function of proficiency level.

$$\text{v}_i \sim \text{Categorical}\,(p_i^{\upsilon}, p_i^{\text{v}}, p_i^{\text{w}})$$

Our model describes how these probabilities (p^{υ}, p^{v}, and p^{w}) change across proficiency levels. Similar to logistic regression models, however, these probabilities are not modeled directly, but through a link function. Here, this is the softmax function, which translates constrained probabilities to an unconstrained scale. For technical reasons, one of the outcome events serves as the reference category. We chose [ʋ] as the benchmark. In practical terms, this means that we are modeling just two quantities: (i) the log ratio of the probability of [v] divided by that of the reference category (i.e. $\log(p^{\text{v}}/p^{\upsilon})$) and (ii) the log ratio of the probability of [w] divided by that of the reference category (i.e. $\log(p^{\text{w}}/p^{\upsilon})$). We will refer to these quantities as *log probability-ratios*[70]. Our regression model therefore simultaneously models two log ratios, which is why we are modeling two conditional means, one for $\log(p^{\text{v}}/p^{\upsilon})$ and one for $\log(p^{\text{w}}/p^{\upsilon})$.

[70] The hyphen clarifies the structure of the compound (i.e. log [probability ratio]).

$$\begin{aligned}\log\left(\frac{p_i^{\text{v}}}{p_i^{\upsilon}}\right) = {} & \bar{\alpha}^{\text{v}} + \alpha^{\text{v}}_{\text{s}[i]} + \alpha^{\text{v}}_{\text{w}[i]} + \\ & (\bar{\beta}^{\text{v}}_{\text{FAR}} + \beta^{\text{v}}_{\text{FAR, w}[i]})\text{FAR}_i + \\ & (\bar{\beta}^{\text{v}}_{\text{FAR}^2} + \beta^{\text{v}}_{\text{FAR}^2\text{,w}[i]})\text{FAR}_i^2\end{aligned}$$

$$\begin{aligned}\log\left(\frac{p_i^{\text{w}}}{p_i^{\upsilon}}\right) = {} & \bar{\alpha}^{\text{w}} + \alpha^{\text{w}}_{\text{s}[i]} + \alpha^{\text{w}}_{\text{w}[i]} + \\ & (\bar{\beta}^{\text{w}}_{\text{FAR}} + \beta^{\text{w}}_{\text{FAR, w}[i]})\text{FAR}_i + \\ & (\bar{\beta}^{\text{w}}_{\text{FAR}^2} + \beta^{\text{w}}_{\text{FAR}^2\text{,w}[i]})\text{FAR}_i^2\end{aligned}$$

The $\bar{\alpha}_{\dots}$ parameters express the expected probability ratio when FAR is set to 0. The α coefficients represent the varying intercepts for subjects ($\alpha_s^{\dots}$) and words ($\alpha_w^{\dots}$). The next two lines describe how probability ratios change as we compare different proficiency levels. The $\bar{\beta}$ parameters express by how much the ratios change, on average, as we shift the FAR scale upwards by 1.[71] The $\bar{\beta}_{\text{FAR}^2}$ coefficients reflect the quadratic component of the link between FAR score and the probability of the outcome variants. The $\beta^{\dots}_{\text{FAR, w}}$ and $\beta^{\dots}_{\text{FAR}^2\text{,w}}$ parameters are varying slopes and represent how much each word differs from the average FAR trend for a particular log probability-ratio.

[71] Thus, the parameter $\bar{\beta}^{\text{v}}_{\text{FAR}}$ describes the change in the log probability-ratio $\log(p^{\text{v}}/p^{\upsilon})$ when comparing speakers with an average FAR score of 1 vs. 0 (or −2 vs. −1, etc.). We expect this parameter to be positive, because the probability of observing target-like [v] vs. L1-transferred [ʋ] should increase with higher proficiency levels.

The joint distributions of the varying effects are assumed to follow a multivariate Student-*t* distribution, which guards our key estimates against unusual subjects and words. The ν parameter was set to 7.[72] **S** denotes the covariance matrix, which contains information about the dispersion of the varying effects and their pairwise association.[73]

$$\begin{bmatrix} \alpha_s{}^{v} \\ \alpha_s{}^{w} \end{bmatrix} \sim \text{MVStudent-t}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mathbf{S}_s, \nu_s \right)$$

$$\begin{bmatrix} \alpha_w^{v} & \beta_{\text{FAR, w}}^{v} & \beta_{\text{FAR}^2\text{,w}}^{v} \\ \alpha_w^{w} & \beta_{\text{FAR, w}}^{w} & \beta_{\text{FAR}^2\text{,w}}^{w} \end{bmatrix} \sim \text{MVStudent-t}\left(\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \mathbf{S}_w, \nu_w \right)$$

The covariance matrices are defined using the standard deviation of the varying effects sets, and their pairwise correlations. Thus, the parameter $\sigma_{\alpha_w^{w}}$ is a standard deviation (on the log ratio scale), which describes how much speakers varied around the average log probability-ratio of [w] over [ʋ]. **R** is a correlation matrix, which includes correlation coefficients that capture how by-subject varying effects on the one hand, and by-word varying effects on the other, are correlated.[74]

$$\mathbf{S}_s = \text{CovarianceMatrix}\left(\begin{matrix} \sigma_{\alpha_s{}^{v}} \\ \sigma_{\alpha_s{}^{w}} \end{matrix}, \mathbf{R}_s \right)$$

$$\mathbf{S}_w = \text{CovarianceMatrix}\left(\begin{matrix} \sigma_{\alpha_w^{v}} & \sigma_{\beta_{\text{FAR, w}}^{v}} & \sigma_{\beta_{\text{FAR}^2\text{,w}}^{v}} \\ \sigma_{\alpha_w^{w}} & \sigma_{\beta_{\text{FAR, w}}^{w}} & \sigma_{\beta_{\text{FAR}^2\text{,w}}^{w}} \end{matrix}, \mathbf{R}_w \right)$$

Next we list the priors. The standard deviation parameters for the dispersion of the varying effects were given a weakly informative prior – an exponential distribution with scale parameter 2.[75]

$$\sigma_{\dots} \sim \text{Exponential}(2)$$

For the correlation between the varying effects, we have no information to build on and therefore default to an LKJ prior with η set to 4.[76]

$$\mathbf{R}_{\dots} \sim \text{LKJcorr}(4)$$

To specify reasonable priors for the two log probability-ratios, we followed the procedure outlined in Appendix A.6.1 (p. 275). We first expressed our prior state of information about the probability of observing each of the three variants on the probability scale (see §8.6). To this end, we used a Dirichlet distribution with parameters 0.8 ([v]), 1 ([w]), and 1.2 ([ʋ]) to generate these values.[77] Figure A.15a shows how, for each variant, the a priori probability mass is concentrated over the interval [0; 1]. We then divided the distributions for [v] and [w] by that of [ʋ], which gave us two distributions of ratios. This intermediate step is shown in panel (b). Since both variants are assumed to be less likely than [ʋ], most of the probability mass is below 1. Panel (c) shows these ratios on the log scale and brings us to the quantity that is modeled (i.e. log probability-ratios). These distributions are slightly negatively skewed. To approximate this shape, we specified a skew-normal distribution as a prior. The solid lines

[72] see Appendix A.3.1 and Figure A.3 (p. 253) for illustration.

[73] To save space, we arranged the parameters for word in multiple columns - they would normally need to be listed in a single column.

[74] We simplify notation and omit the translation of standard deviations and the correlation matrix **R** into a covariance matrix. See McElreath (2020: 441) for a complete representation.

[75] See Appendix A.6.1 and Figure A.11 for details.

[76] See Appendix A.3.1 and Figure A.2 (p. 253) for illustration.

[77] The prior distribution therefore contains the amount of inferential information in a sample of n = 3 /v/-tokens. In other words, the pre-data information we are supplying is "worth" n = 3 observations.

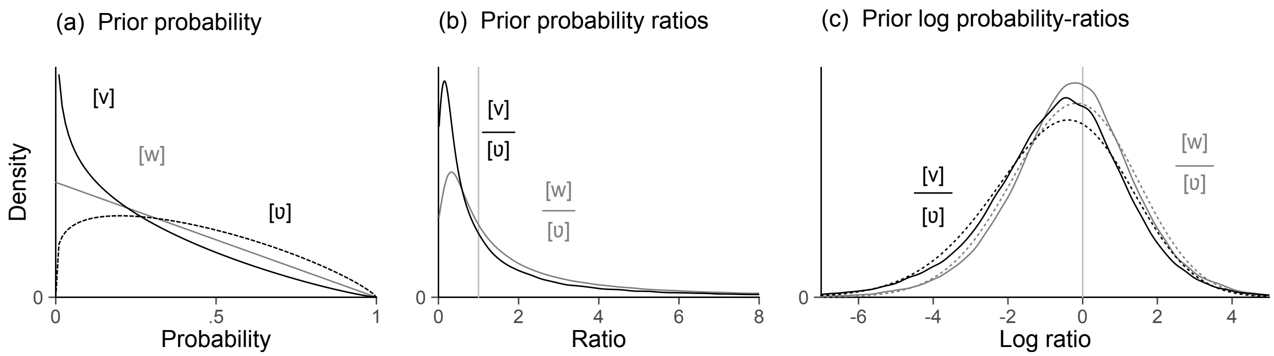


Figure A.15: Construction of the priors for the log probability-ratios based on the prior probabilities of the three variants. Solid lines show simulated values for (a) probabilities, (b) probability ratios, and (c) log probability-ratios. The dashed curves in panel (c) are the parametric prior distributions specified to approximate the simulated values.

in panel (c) are the simulated log ratios, and the dashed curves are the prior distributions we used. These ratios were then translated into the conditional ('subject-specific') parameter values required by the model. The procedure was outlined above in Appendix A.6.1.

$$\bar{\alpha}^{\,\mathrm{v}} + \frac{\sigma_{\alpha_{\mathrm{s}}^{\mathrm{v}}}{}^{2} + \sigma_{\alpha_{\mathrm{w}}^{\mathrm{v}}}{}^{2}}{2} \sim \text{SkewNormal}(0.97, 2.51, -1.47)$$

$$\bar{\alpha}^{\,\mathrm{w}} + \frac{\sigma_{\alpha_{\mathrm{s}}^{\mathrm{w}}}{}^{2} + \sigma_{\alpha_{\mathrm{w}}^{\mathrm{w}}}{}^{2}}{2} \sim \text{SkewNormal}(0.89, 2.07, -1.02)$$

Finally, for the $\bar{\beta}$ parameters we chose mildly regularizing priors.

$$\bar{\beta}^{\cdots} \sim \text{Normal}(0, 2)$$

To guard estimates against deviant subjects and words, we use *t*-distributed rather than normal errors (Gelman et al. 2014: 437) and set ν to 7.[78]

$$\nu_{\dots} = 7$$

[78] See Appendix A.3.1 and Figure A.3 (p. 253) for illustration.

Stan code Listed below is the *Stan* code for the model.

```
data {
  int<lower=0> N;
  int y[N];
  int K;
  int<lower=0> subj[N];
  int<lower=0> word[N];
  int<lower=0> n_subj;
  int<lower=0> n_word;
  real far_z[N];
  real far_z2[N];
  int n_varcoef_subj;
  int n_varcoef_word;
  int n_cor_subj;
  int n_cor_word;
  int set_nu;
}
parameters {
  row_vector[K-1] a;
  row_vector[K-1] b_far;
  row_vector[K-1] b_far2;

  vector<lower=0>[n_varcoef_subj] sigma_subj;
  vector<lower=0>[n_varcoef_word] sigma_word;

  matrix[n_varcoef_subj, n_subj] z_subj;
  matrix[n_varcoef_word, n_word] z_word;
  cholesky_factor_corr[n_varcoef_subj] L_Rho_subj;
  cholesky_factor_corr[n_varcoef_word] L_Rho_word;
```

```
  real<lower=0> udf_word;
  real<lower=0> udf_subj;
}
transformed parameters {
  matrix[n_subj, n_varcoef_subj] r_s;
  matrix[n_word, n_varcoef_word] r_w;
  matrix[N,K] mu;

  r_s = sqrt(set_nu * udf_subj) * (diag_pre_multiply(sigma_subj, L_Rho_subj) * z_subj)';
  r_w = sqrt(set_nu * udf_word) * (diag_pre_multiply(sigma_word, L_Rho_word) * z_word)';

  for (n in 1:N) {
    mu[n,1] = 0;
    mu[n,2] = a[1] + r_s[subj[n],1] + r_w[word[n],1] +
             (b_far[1]              + r_w[word[n],2]) * far_z[n] +
             (b_far2[1]             + r_w[word[n],3]) * far_z2[n];
    mu[n,3] = a[2] + r_s[subj[n],2] + r_w[word[n],4] +
             (b_far[2]              + r_w[word[n],5]) * far_z[n] +
             (b_far2[2]             + r_w[word[n],6]) * far_z2[n];
  }
}
model {
  L_Rho_subj  ~ lkj_corr_cholesky(4);
  L_Rho_word  ~ lkj_corr_cholesky(4);
  sigma_subj ~ exponential(2);
  sigma_word ~ exponential(2);

  to_vector(z_subj)  ~ normal(0,1);
  to_vector(z_word)  ~ normal(0,1);

  (a[1] + ((sigma_subj[1]^2 + sigma_word[1]^2)/2)) ~ skew_normal( 0.97 , 2.51, -1.47 );
  (a[2] + ((sigma_subj[2]^2 + sigma_word[4]^2)/2)) ~ skew_normal( 0.89 , 2.07, -1.02 );

  b_far ~ normal( 0 , 2 );
  b_far2 ~ normal( 0 , 2 );

  udf_word ~ inv_chi_square(set_nu);
  udf_subj ~ inv_chi_square(set_nu);

  for(n in 1:N) y[n] ~ categorical_logit(mu[n]');
}
generated quantities{
  matrix[n_varcoef_subj, n_varcoef_subj] Rho_subj;
  matrix[n_varcoef_word, n_varcoef_word] Rho_word;

  int pos_subj = 1;
  int pos_word = 1;

  vector<lower=-1, upper=1>[n_cor_subj] cor_subj;
  vector<lower=-1, upper=1>[n_cor_word] cor_word;

  Rho_subj = multiply_lower_tri_self_transpose(L_Rho_subj);
  Rho_word = multiply_lower_tri_self_transpose(L_Rho_word);

  for(i in 1:(n_varcoef_subj-1)){
    for(j in (i+1):n_varcoef_subj){
      cor_subj[pos_subj] = Rho_subj[i, j];
      pos_subj += 1;
    }
  }
  for(i in 1:(n_varcoef_word-1)){
    for(j in (i+1):n_varcoef_word){
      cor_word[pos_word] = Rho_word[i, j];
      pos_word += 1;
    }
  }
}
```

Convergence diagnostics The diagnostics shown Figure A.16 indicate convergence of the algorithm: All $\hat{R}$ values (left) are below 1.01, the profile of the rank plot (middle) is flat, and the quantile plot (right) suggests that all key parameters are supported by a sufficiently large effective sample size (i.e. all above $n = 400$).

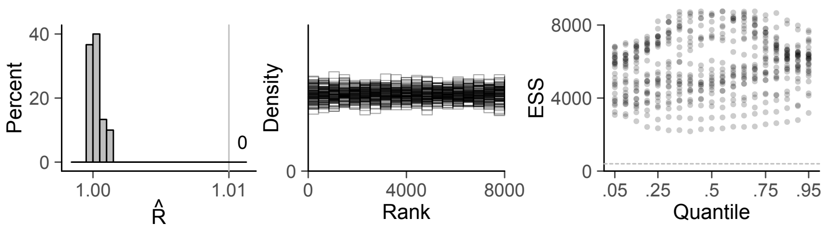


Figure A.16: /v/-model for the data from the present study: Convergence diagnostics.

Posterior distribution of model parameters The posterior distribution of the model parameters for the data from the current study is reported in Tables A.18 (fixed effects) and A.19 (varying effects).

Table A.18: /v/-model for the data from the present study: Posterior distribution of the fixed effects parameters.

			Posterior quantiles				
Parameter	Mean	SD	.05	.50	.95	N_{eff}	$\hat{R}$
$\bar{\alpha}^{\text{v}}$	–0.86	0.58	–1.85	–0.83	0.06	2338	1
$\bar{\alpha}^{\text{w}}$	–1.58	0.75	–2.87	–1.52	–0.44	2722	1
$\bar{\beta}^{\text{v}}_{\text{FAR}}$	1.29	0.37	0.70	1.28	1.91	3459	1
$\bar{\beta}^{\text{w}}_{\text{FAR}}$	–1.33	0.59	–2.34	–1.31	–0.39	4019	1
$\bar{\beta}^{\text{v}}_{\text{FAR}^2}$	–0.24	0.65	–1.34	–0.22	0.83	2667	1
$\bar{\beta}^{\text{w}}_{\text{FAR}^2}$	–3.09	1.11	–4.99	–3.07	–1.32	4534	1

A.8.2 Rank 2016

Model definition The model we fit to the data from Rank (2016) includes two predictor variables, which are coded as follows:

- Structural strength (Str): +1 for strong contexts, –1 for weak contexts;
- Connected speech (Con): +1 for connected speech, –1 for the word list.

The $\bar{\alpha}$ parameters again denote the intercepts for the two log probability-ratios, and define the probability of observing the three variants when all predictors are held at zero, i.e. averaging over the two predictor variables. The α parameters are the varying intercepts for subject (α_{s}) and word (α_{w}). The $\bar{\beta}$ coefficients represent the observed changes in the probability of observing the three variants for a 1-unit change in the predictor variable. These are accompanied by varying slopes on words, which capture the extent to which the items in the data deviate from the average trend described by the $\bar{\beta}$ parameters. Thus, the set of $\beta_{\ddot{\text{Str}},\text{ s}}$ coefficients, one for each word, describes variability among the lexical items with regard to the difference we observe between strong and weak contexts. The $\beta_{\ddot{\text{Con}},\text{ s}}$ coefficients capture differences among words in the extent to which speaking style is associated with changes in the outcome probabilities. The remainder of the model is structured analogously to the model for the data from the current study.

Parameter	Mean	SD	Posterior quantiles .05	.50	.95	N_{eff}	$\hat{R}$
Subject							
$\sigma_{\alpha_{\text{s}}^{\text{v}}}$	1.17	0.35	0.64	1.14	1.79	4126	1
$\sigma_{\beta_{\text{FAR}}^{\text{v}}}$	1.47	0.44	0.80	1.44	2.25	4043	1
$\sigma_{\beta_{\text{FAR}}^{\text{v}}}$	–0.30	0.19	–0.60	–0.31	0.03	1647	1
$\sigma_{\alpha_{\text{s}}^{\text{w}}}$	0.97	0.34	0.50	0.93	1.59	4567	1
$\sigma_{\beta_{\text{FAR}^2}^{\text{w}}}$	0.25	0.22	0.01	0.20	0.68	4033	1
$\sigma_{\beta_{\text{FAR}^2}^{\text{w}}}$	0.29	0.26	0.02	0.21	0.80	5314	1
$\rho(\alpha_{\text{s}}^{\text{v}}, \beta_{\text{FAR}}^{\text{v}})$	0.91	0.35	0.41	0.86	1.57	3857	1
$\rho(\alpha_{\text{s}}^{\text{v}}, \beta_{\text{FAR}^2}^{\text{v}})$	0.21	0.19	0.01	0.16	0.57	7948	1
$\rho(\alpha_{\text{s}}^{\text{v}}, \alpha_{\text{s}}^{\text{w}})$	0.36	0.34	0.02	0.26	1.05	8300	1
$\rho(\alpha_{\text{s}}^{\text{v}}, \beta_{\text{FAR}}^{\text{w}})$	0.07	0.27	–0.38	0.08	0.51	8466	1
$\rho(\alpha_{\text{s}}^{\text{v}}, \beta_{\text{FAR}^2}^{\text{w}})$	–0.03	0.28	–0.48	–0.03	0.44	10805	1
$\rho(\beta_{\text{FAR}}^{\text{v}}, \beta_{\text{FAR}^2}^{\text{v}})$	0.10	0.23	–0.29	0.10	0.47	7009	1
$\rho(\beta_{\text{FAR}}^{\text{v}}, \alpha_{\text{s}}^{\text{w}})$	–0.03	0.27	–0.48	–0.03	0.43	11560	1
$\rho(\beta_{\text{FAR}}^{\text{v}}, \beta_{\text{FAR}}^{\text{w}})$	0.03	0.28	–0.43	0.04	0.48	11629	1
$\rho(\beta_{\text{FAR}}^{\text{v}}, \beta_{\text{FAR}^2}^{\text{w}})$	–0.01	0.27	–0.45	–0.01	0.45	9251	1
$\rho(\beta_{\text{FAR}^2}^{\text{v}}, \alpha_{\text{s}}^{\text{w}})$	–0.02	0.26	–0.46	–0.02	0.41	4870	1
$\rho(\beta_{\text{FAR}^2}^{\text{v}}, \beta_{\text{FAR}}^{\text{w}})$	0.01	0.28	–0.45	0.01	0.46	12214	1
$\rho(\beta_{\text{FAR}^2}^{\text{v}}, \beta_{\text{FAR}^2}^{\text{w}})$	0.00	0.28	–0.45	0.01	0.46	10115	1
$\rho(\alpha_{\text{s}}^{\text{w}}, \beta_{\text{FAR}}^{\text{w}})$	–0.05	0.28	–0.50	–0.05	0.43	4044	1
$\rho(\alpha_{\text{s}}^{\text{w}}, \beta_{\text{FAR}^2}^{\text{w}})$	–0.01	0.28	–0.47	–0.01	0.47	8235	1
$\rho(\beta_{\text{FAR}}^{\text{w}}, \beta_{\text{FAR}^2}^{\text{w}})$	0.00	0.28	–0.47	0.00	0.47	8953	1
Word							
$\sigma_{\alpha_{\text{w}}^{\text{v}}}$	0.00	0.28	–0.45	0.00	0.46	8638	1
$\sigma_{\alpha_{\text{w}}^{\text{w}}}$	0.01	0.27	–0.45	0.01	0.45	8888	1
$\rho(\alpha_{\text{w}}^{\text{v}}, \alpha_{\text{w}}^{\text{w}})$	0.01	0.28	–0.44	0.01	0.47	6629	1

Table A.19: /v/-model for the data from the present study: Posterior distribution of the varying effects parameters.

$$
\begin{aligned}
\text{v}_i &\sim \text{Categorical}\left(p_i^{\upsilon}, p_i^{\text{v}}, p_i^{\text{w}}\right) \\
\log\left(\frac{p_i^{\text{v}}}{p_i^{\upsilon}}\right) &= \bar{\alpha}^{\text{v}} + \alpha_{\text{s}[i]}^{\text{v}} + \alpha_{\text{w}[i]}^{\text{v}} + \\
&\quad (\bar{\beta}_{\text{Str}}^{\text{v}} + \beta_{\text{Str, w}[i]}^{\text{v}})\text{Str}_i + \\
&\quad (\bar{\beta}_{\text{Con}}^{\text{v}} + \beta_{\text{Con, w}[i]}^{\text{v}})\text{Con}_i \\
\log\left(\frac{p_i^{\text{w}}}{p_i^{\upsilon}}\right) &= \bar{\alpha}^{\text{w}} + \alpha_{\text{s}[i]}^{\text{w}} + \alpha_{\text{w}[i]}^{\text{w}} + \\
&\quad (\bar{\beta}_{\text{Str}}^{\text{w}} + \beta_{\text{Str, w}[i]}^{\text{w}})\text{Str}_i + \\
&\quad (\bar{\beta}_{\text{Con}}^{\text{w}} + \beta_{\text{Con, w}[i]}^{\text{w}})\text{Con}_i \\
\begin{bmatrix} \alpha_{\text{s}}^{\text{v}} & \beta_{\text{Str, s}}^{\text{v}} & \beta_{\text{Con, s}}^{\text{v}} \\ \alpha_{\text{s}}^{\text{w}} & \beta_{\text{Str, s}}^{\text{w}} & \beta_{\text{Con, s}}^{\text{w}} \end{bmatrix} &\sim \text{MVStudent-t}\left(\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \mathbf{S}_{\text{s}}, \nu_{\text{s}}\right) \\
\begin{bmatrix} \alpha_{\text{w}}^{\text{v}} \\ \alpha_{\text{w}}^{\text{w}} \end{bmatrix} &\sim \text{MVStudent-t}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mathbf{S}_{\text{w}}, \nu_{\text{w}}\right)
\end{aligned}
$$

$$\mathbf{S}_{\mathrm{S}} = \text{CovarianceMatrix}\left(\begin{matrix} \sigma_{\alpha_{\mathrm{s}}^{\mathrm{v}}} & \sigma_{\beta^{\mathrm{v}}_{\mathrm{Str,\,s}}} & \sigma_{\beta^{\mathrm{v}}_{\mathrm{Con,\,s}}} \\ \sigma_{\alpha_{\mathrm{s}}^{\mathrm{w}}} & \sigma_{\beta^{\mathrm{w}}_{\mathrm{Str,\,s}}} & \sigma_{\beta^{\mathrm{w}}_{\mathrm{Con,\,s}}} \end{matrix}, \mathbf{R}_{\mathrm{S}}\right)$$

$$\mathbf{S}_{\mathrm{W}} = \text{CovarianceMatrix}\left(\begin{matrix} \sigma_{\alpha_{\mathrm{w}}^{\mathrm{v}}} \\ \sigma_{\alpha_{\mathrm{w}}^{\mathrm{w}}} \end{matrix}, \mathbf{R}_{\mathrm{W}}\right)$$

$$\sigma_{\dots} \sim \text{Exponential}(2)$$

$$\mathbf{R}_{\dots} \sim \text{LKJcorr}(4)$$

$$\bar{\alpha}^{\mathrm{v}} + \frac{{\sigma_{\alpha_{\mathrm{s}}^{\mathrm{v}}}}^2 + {\sigma_{\alpha_{\mathrm{w}}^{\mathrm{v}}}}^2}{2} \sim \text{SkewNormal}(0.97, 2.51, -1.47)$$

$$\bar{\alpha}^{\mathrm{w}} + \frac{{\sigma_{\alpha_{\mathrm{s}}^{\mathrm{w}}}}^2 + {\sigma_{\alpha_{\mathrm{w}}^{\mathrm{w}}}}^2}{2} \sim \text{SkewNormal}(0.89, 2.07, -1.02)$$

$$\bar{\beta}^{\dots} \sim \text{Normal}(0, 2)$$

$$\nu_{\dots} = 7$$

Stan code The *Stan* code for the model follows.

```
data {
  int<lower=0> N;
  int y[N];
  int K;
  int<lower=0> subj[N];
  int<lower=0> word[N];
  int<lower=0> n_subj;
  int<lower=0> n_word;
  real strong_c[N];
  real wordlist_c[N];
  int n_varcoef_subj;
  int n_varcoef_word;
  int n_cor_subj;
  int n_cor_word;
  int set_nu;
}
parameters {
  row_vector[K-1] a;

  row_vector[K-1] b_strong;
  row_vector[K-1] b_wordlist;

  vector<lower=0>[n_varcoef_subj] sigma_subj;
  vector<lower=0>[n_varcoef_word] sigma_word;

  matrix[n_varcoef_subj, n_subj] z_subj;
  matrix[n_varcoef_word, n_word] z_word;

  cholesky_factor_corr[n_varcoef_subj] L_Rho_subj;
  cholesky_factor_corr[n_varcoef_word] L_Rho_word;

  real<lower=0> udf_word;
  real<lower=0> udf_subj;
}
transformed parameters {
  matrix[n_subj, n_varcoef_subj] r_s;
  matrix[n_word, n_varcoef_word] r_w;

  matrix[N,K] mu;

  r_s = sqrt(set_nu * udf_subj) * (diag_pre_multiply(sigma_subj, L_Rho_subj) * z_subj)';
  r_w = sqrt(set_nu * udf_word) * (diag_pre_multiply(sigma_word, L_Rho_word) * z_word)';

  for (n in 1:N) {
    mu[n,1] = 0;

    mu[n,2] = a[1]           + r_s[subj[n],1] + r_w[word[n],1] +
             (b_strong[1]    + r_s[subj[n],2])  *  strong_c[n] +
             (b_wordlist[1]  + r_s[subj[n],3])  *  wordlist_c[n];
```

```
    mu[n,3] = a[2]           + r_s[subj[n],4] + r_w[word[n],2] +
             (b_strong[2]    + r_s[subj[n],5])  *  strong_c[n] +
             (b_wordlist[2]  + r_s[subj[n],6])  *  wordlist_c[n];
  }
}
model {
  L_Rho_subj  ~ lkj_corr_cholesky(4);
  L_Rho_word  ~ lkj_corr_cholesky(4);

  sigma_subj ~ exponential(2);
  sigma_word ~ exponential(2);

  to_vector(z_subj)  ~ normal(0,1);
  to_vector(z_word)  ~ normal(0,1);

  (a[1] + ((sigma_subj[1]^2 + sigma_word[1]^2)/2)) ~ skew_normal( 0.97 , 2.51, -1.47 );
  (a[2] + ((sigma_subj[4]^2 + sigma_word[2]^2)/2)) ~ skew_normal( 0.89 , 2.07, -1.02 );

  b_strong ~ normal( 0 , 2 );
  b_wordlist ~ normal( 0 , 2 );

  udf_word ~ inv_chi_square(set_nu);
  udf_subj ~ inv_chi_square(set_nu);

  for(n in 1:N) y[n] ~ categorical_logit(mu[n]');
}
generated quantities{
  matrix[n_varcoef_subj, n_varcoef_subj] Rho_subj;
  matrix[n_varcoef_word, n_varcoef_word] Rho_word;

  int pos_subj = 1;
  int pos_word = 1;

  vector<lower=-1, upper=1>[n_cor_subj] cor_subj;
  vector<lower=-1, upper=1>[n_cor_word] cor_word;

  Rho_subj = multiply_lower_tri_self_transpose(L_Rho_subj);
  Rho_word = multiply_lower_tri_self_transpose(L_Rho_word);

  for(i in 1:(n_varcoef_subj-1)){
    for(j in (i+1):n_varcoef_subj){
      cor_subj[pos_subj] = Rho_subj[i, j];
      pos_subj += 1;
    }
  }
  for(i in 1:(n_varcoef_word-1)){
    for(j in (i+1):n_varcoef_word){
      cor_word[pos_word] = Rho_word[i, j];
      pos_word += 1;
    }
  }
}
```

Convergence diagnostics Our graphical check of the MCMC chains is shown in Figure A.17. There is no indication of convergence problems: All $\hat{R}$ values (left) are below 1.01, the the rank plot profiles (middle) are flat, and the quantile plot (right) suggests that all key parameters are backed by a sufficiently large effective sample size (i.e. all above $n = 400$).

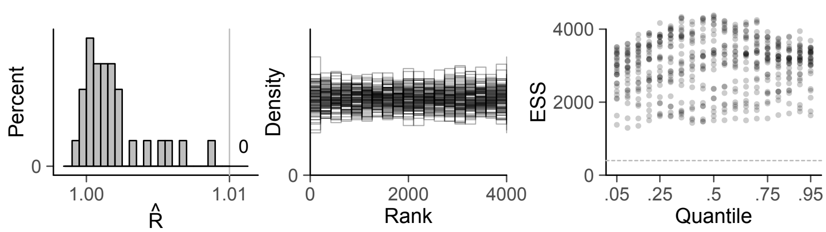


Figure A.17: /v/-model for the data from Rank 2016: Convergence diagnostics.

Posterior distribution of model parameters The output of the model is reported in Tables A.20 (fixed effects) and A.21 (varying effects).

			Posterior quantiles				
Parameter	Mean	SD	.05	.50	.95	N_{eff}	$\hat{R}$
$\bar{\alpha}^{\text{v}}$	2.00	0.31	1.49	2.00	2.50	2250	1
$\bar{\alpha}^{\text{w}}$	–1.26	0.45	–2.01	–1.24	–0.57	2282	1
$\bar{\beta}^{\text{v}}_{\text{Str}}$	0.11	0.24	–0.29	0.11	0.51	2533	1
$\bar{\beta}^{\text{w}}_{\text{Str}}$	1.31	0.35	0.76	1.30	1.92	2770	1
$\bar{\beta}^{\text{v}}_{\text{Con}}$	0.32	0.13	0.10	0.32	0.54	4343	1
$\bar{\beta}^{\text{w}}_{\text{Con}}$	0.08	0.17	–0.20	0.08	0.35	3412	1

Table A.20: /v/-model for the data from Rank (2016): Posterior distribution of the fixed parameters.

			Posterior quantiles				
Parameter	Mean	SD	.05	.50	.95	N_{eff}	$\hat{R}$
Subject							
$\sigma_{\alpha_s^{\text{v}}}$	0.93	0.25	0.57	0.91	1.37	3147	1.00
$\sigma_{\beta^{\text{v}}_{\text{Str}}}$	0.27	0.16	0.03	0.25	0.56	1617	1.00
$\sigma_{\beta^{\text{v}}_{\text{Con}}}$	0.17	0.13	0.01	0.14	0.40	1596	1.00
$\sigma_{\alpha_s^{\text{w}}}$	1.05	0.31	0.58	1.02	1.62	2876	1.00
$\sigma_{\beta^{\text{w}}_{\text{Str}}}$	0.23	0.19	0.02	0.18	0.59	1904	1.00
$\sigma_{\beta^{\text{w}}_{\text{Con}}}$	0.24	0.18	0.02	0.21	0.57	1405	1.00
$\rho(\alpha_s^{\text{v}}, \beta^{\text{v}}_{\text{Str}})$	0.14	0.25	–0.29	0.15	0.55	4472	1.00
$\rho(\alpha_s^{\text{v}}, \beta^{\text{v}}_{\text{Con}})$	0.06	0.26	–0.38	0.07	0.49	5395	1.00
$\rho(\alpha_s^{\text{v}}, \alpha_s^{\text{w}})$	–0.09	0.20	–0.42	–0.09	0.25	2939	1.00
$\rho(\alpha_s^{\text{v}}, \beta^{\text{w}}_{\text{Str}})$	0.05	0.27	–0.41	0.05	0.48	4128	1.00
$\rho(\alpha_s^{\text{v}}, \beta^{\text{w}}_{\text{Con}})$	–0.09	0.25	–0.49	–0.09	0.34	4919	1.00
$\rho(\beta^{\text{v}}_{\text{Str}}, \beta^{\text{v}}_{\text{Con}})$	0.06	0.27	–0.40	0.06	0.50	4014	1.00
$\rho(\beta^{\text{v}}_{\text{Str}}, \alpha_s^{\text{w}})$	–0.12	0.25	–0.52	–0.13	0.31	1144	1.01
$\rho(\beta^{\text{v}}_{\text{Str}}, \beta^{\text{w}}_{\text{Str}})$	0.04	0.28	–0.44	0.04	0.49	4385	1.00
$\rho(\beta^{\text{v}}_{\text{Str}}, \beta^{\text{w}}_{\text{Con}})$	–0.07	0.27	–0.51	–0.08	0.39	4056	1.00
$\rho(\beta^{\text{v}}_{\text{Con}}, \alpha_s^{\text{w}})$	–0.01	0.26	–0.45	–0.01	0.43	1271	1.00
$\rho(\beta^{\text{v}}_{\text{Con}}, \beta^{\text{w}}_{\text{Str}})$	0.01	0.28	–0.46	0.02	0.48	4159	1.00
$\rho(\beta^{\text{v}}_{\text{Con}}, \beta^{\text{w}}_{\text{Con}})$	–0.05	0.27	–0.49	–0.05	0.41	4182	1.00
$\rho(\alpha_s^{\text{w}}, \beta^{\text{w}}_{\text{Str}})$	0.00	0.28	–0.46	0.00	0.45	4757	1.00
$\rho(\alpha_s^{\text{w}}, \beta^{\text{w}}_{\text{Con}})$	–0.05	0.25	–0.47	–0.05	0.37	4093	1.00
$\rho(\beta^{\text{w}}_{\text{Str}}, \beta^{\text{w}}_{\text{Con}})$	–0.05	0.28	–0.49	–0.04	0.42	3105	1.00
Word							
$\sigma_{\alpha_w^{\text{v}}}$	0.68	0.21	0.37	0.66	1.06	3201	1.00
$\sigma_{\alpha_w^{\text{w}}}$	0.48	0.22	0.14	0.46	0.88	1620	1.00
$\rho(\alpha_w^{\text{v}}, \alpha_w^{\text{w}})$	–0.04	0.27	–0.48	–0.03	0.41	2982	1.00

Table A.21: /v/-model for the data from Rank (2016): Posterior distribution of the varying effects parameters.

A.9 *Dental fricatives*

This appendix gives information about the analysis of the data from Wunder (2012). Details about the analyses of the data by Pascoe (1987) and the present study are given in the web appendix.[79] The data sets analyzed in Chapter 9 are openly available via *TROLLing*[80] (Sönning et al. 2020c). The OSF project accompanying this study includes the *Stan* code for the regression models[81] and the *R* script[82] for running the analyses.

[79] https://osf.io/aszq6/

[80] https://doi.org/10.18710/DYAGZG

[81] This study: https://osf.io/ks764/; Pascoe 1987: https://osf.io/7qdjx/; Wunder 2012: https://osf.io/tbw2z/

[82] https://osf.io/zuheq/

A.9.1 *Wunder 2012*

Model definition As the realization of /ð/ was coded with 4 categories, we fit a multinomial regression model. We model the probability of each variant, i.e. p^{d} (for *d*-types), p^{z} (for *z*-types), p^{other}, and $p^{\text{ð}}$, for a given /ð/-token *i*.

$$\text{th}_i \sim \text{Categorical}\left(p_i^{\text{d}}, p_i^{\text{z}}, p_i^{\text{other}}, p_i^{\text{ð}}\right)$$

Our model describes how these probabilities change depending on (i) whether the preceding sound is a sibilant and (ii) proficiency level. These probabilities are not modeled directly, but through a link function (the softmax function), which translates constrained probabilities to an unconstrained scale. For technical reasons, one of the outcome events serves as the reference category ([ð] in our case), and we therefore model only three quantities: (i) the log ratio of the probability of *d*-types divided by that of the reference category [ð] (i.e. $\log(p^{\text{d}}/p^{\text{ð}})$); (ii) the log ratio of the probability of *z*-types divided by that of the reference category [ð] (i.e. $\log(p^{\text{z}}/p^{\text{ð}})$); and (iii) the log ratio of the probability of other variants divided by that of the reference category (i.e. $\log(p^{\text{other}}/p^{\text{ð}})$). We will refer to the quantities as *log probability-ratios*.[83] Our regression model therefore simultaneously models three log ratios, which is why we are modeling three conditional means, one for each log probability-ratio.

[83] The hyphen clarifies the structure of the compound (i.e. log [probability ratio]).

The $\bar{\alpha}_{\dots}$ parameters express the expected probability ratio when FAR is set to 0. The α coefficients represent the varying intercepts for subjects ($\alpha_s^{\dots}$) and words ($\alpha_w^{\dots}$). Thereby, each speaker and word is allowed to assume a higher or lower ratio for each pair of probabilities. The next two lines describe how probability ratios change with the preceding segment type (sibilant vs. other sound) and with the proficiency level of the speaker. The $\bar{\beta}_{\text{FAR}}^{\dots}$ parameters express by how much the probability ratios change, on average, as we shift the FAR scale upwards by 1 unit. Thus, the parameter $\bar{\beta}_{\text{FAR}}^{\text{z}}$ describes the change in the log probability-ratio $\log(p^{\text{z}}/p^{\text{ð}})$ when comparing speakers who differ by 1 point on the FAR scale. We expect this parameter to be negative, because the probability of observing target-like [ð] should increase toward higher proficiency levels. The $\beta_{\text{FAR, w}}^{\dots}$ parameters are varying slopes and represent how much each word differs from the average FAR trend for a log probability-ratio. The $\bar{\beta}_{\text{Sibilant}}^{\dots}$ parameters express by how much the log probability ratios change, on average, as we change the preceding segment from a non-sibilant to a sibilant. Thus, the parameter $\bar{\beta}_{\text{Sibilant}}^{\text{d}}$ describes the change in the log probability-ratio $\log(p^{\text{d}}/p^{\text{ð}})$ when comparing sibilant vs. other contexts.

Assuming that sibilants cause difficulty and reduce the overall accuracy rate, we expect this parameter to be positive, because the probability of observing L1-transferred *z*-type vs. target-like [ð] should increase as we change to a sibilant context.

$$\log\left(\frac{p_i^{\text{d}}}{p_i^{\text{ð}}}\right) = \bar{\alpha}^{\text{d}} + \alpha^{\text{d}}_{\text{s}[i]} + \alpha^{\text{d}}_{\text{w}[i]} + (\bar{\beta}^{\text{d}}_{\text{Sibilant}} + \beta^{\text{d}}_{\text{Sibilant, s}[i]})\text{Sibilant}_i + (\bar{\beta}^{\text{d}}_{\text{FAR}} + \bar{\beta}^{\text{d}}_{\text{FAR:Sibilant}} + \beta^{\text{d}}_{\text{FAR, w}[i]})\text{FAR}_i$$

$$\log\left(\frac{p_i^{\text{z}}}{p_i^{\text{ð}}}\right) = \bar{\alpha}^{\text{z}} + \alpha^{\text{z}}_{\text{s}[i]} + \alpha^{\text{z}}_{\text{w}[i]} + (\bar{\beta}^{\text{z}}_{\text{Sibilant}} + \beta^{\text{z}}_{\text{Sibilant, s}[i]})\text{Sibilant}_i + (\bar{\beta}^{\text{z}}_{\text{FAR}} + \bar{\beta}^{\text{z}}_{\text{FAR:Sibilant}} + \beta^{\text{z}}_{\text{FAR, w}[i]})\text{FAR}_i$$

$$\log\left(\frac{p_i^{\text{other}}}{p_i^{\text{ð}}}\right) = \bar{\alpha}^{\text{other}} + \alpha^{\text{other}}_{\text{s}[i]} + \alpha^{\text{other}}_{\text{w}[i]} + (\bar{\beta}^{\text{other}}_{\text{Sibilant}} + \beta^{\text{other}}_{\text{Sibilant, s}[i]})\text{Sibilant}_i + (\bar{\beta}^{\text{other}}_{\text{FAR}} + \bar{\beta}^{\text{other}}_{\text{FAR:Sibilant}} + \beta^{\text{other}}_{\text{FAR, w}[i]})\text{FAR}_i$$

The varying effects are assumed to follow a multivariate Student-*t* distribution, which guards our key estimates against unusual subjects and words. The ν parameter was set to 7.[84] **S** denotes the covariance matrix, which contains information about the dispersion of the varying coefficients and their pairwise associations.[85]

$$\begin{bmatrix} \alpha_{\text{s}}^{\text{d}} & \beta^{\text{d}}_{\text{Sibilant, s}} \\ \alpha_{\text{s}}^{\text{z}} & \beta^{\text{z}}_{\text{Sibilant, s}} \\ \alpha_{\text{s}}^{\text{other}} & \beta^{\text{other}}_{\text{Sibilant, s}} \end{bmatrix} \sim \text{MVStudent-t}\left(\begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \mathbf{S}_{\text{s}}, \nu_{\text{s}}\right)$$

$$\begin{bmatrix} \alpha_{\text{w}}^{\text{d}} & \beta^{\text{d}}_{\text{w}} \\ \alpha_{\text{w}}^{\text{z}} & \beta^{\text{z}}_{\text{w}} \\ \alpha_{\text{w}}^{\text{other}} & \beta^{\text{other}}_{\text{w}} \end{bmatrix} \sim \text{MVStudent-t}\left(\begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \mathbf{S}_{\text{w}}, \nu_{\text{w}}\right)$$

The covariance matrices are defined using the standard deviation of the varying effects and their pairwise correlations. Thus, the parameter $\sigma_{\alpha_{\text{w}}^{\text{d}}}$ is a standard deviation (on the log ratio scale), which describes how much speakers varied around the average log probability-ratio of *d*-types over [ð]. **R** is a correlation matrix, which includes correlation coefficients that capture how by-subject varying effects on the one hand, and by-word varying effects on the other, are correlated.[86]

[84] See Appendix A.3.1 and Figure A.3 (p. 253) for illustration.

[85] To save space, we arranged the parameters for word in multiple columns - they would normally need to be listed in a single column.

[86] We simplify notation and omit the translation of standard deviations and the correlation matrix **R** into a covariance matrix. See McElreath (2020: 441) for a complete representation.

$$\mathbf{S}_{\mathrm{s}} = \text{CovarianceMatrix}\begin{pmatrix} \sigma_{\alpha^{\mathrm{d}}_{\mathrm{s}}} & \sigma_{\beta^{\mathrm{d}}_{\text{Sibilant, s}}} \\ \sigma_{\alpha^{\mathrm{z}}_{\mathrm{s}}} & \sigma_{\beta^{\mathrm{z}}_{\text{Sibilant, s}}} \\ \sigma_{\alpha^{\text{other}}_{\mathrm{s}}} & \sigma_{\beta^{\text{other}}_{\text{Sibilant, s}}} \end{pmatrix}, \mathbf{R}_{\mathrm{s}}$$

$$\mathbf{S}_{\mathrm{w}} = \text{CovarianceMatrix}\begin{pmatrix} \sigma_{\alpha^{\mathrm{d}}_{\mathrm{w}}} & \sigma_{\beta^{\mathrm{d}}_{\text{FAR, w}}} \\ \sigma_{\alpha^{\mathrm{z}}_{\mathrm{w}}} & \sigma_{\beta^{\mathrm{z}}_{\text{FAR, w}}} \\ \sigma_{\alpha^{\text{other}}_{\mathrm{w}}} & \sigma_{\beta^{\text{other}}_{\text{FAR, w}}} \end{pmatrix}, \mathbf{R}_{\mathrm{w}}$$

Next, the priors. The standard deviation parameters for the dispersion of the varying effects were given weakly informative starts, in the form of an exponential distribution with scale parameter 2.[87]

[87] See Appendix A.6.1 and Figure A.11 for details.

$$\sigma_{\dots} \sim \text{Exponential}(2)$$

For the correlation between the varying effects, we have no pre-data information and resort to an LKJ prior with η set to 4.[88]

[88] See Appendix A.3.1 and Figure A.2 (p. 253) for illustration.

$$\mathbf{R}_{\dots} \sim \text{LKJcorr}(4)$$

To specify reasonable priors for the three log probability-ratios, we first formulated a conservative state of information about the probability of observing each variant on the probability scale (see §9.6 for a derivation of these priors.). To this end, we used a Dirichlet distribution with parameters 0.2 (other variants), 0.5 (*z*-types), 1.2 (*d*-types) and 2.2 ([ð]) to generate these values.[89] Figure A.18a shows how, for each variant, the prior probability mass is concentrated over the interval [0; 1]. We then divided the distributions for *d*-types, *z*-types, and other variants by that of [ð], which gave us three distributions of ratios. This intermediate step is shown in panel (b). Since all three variants are assumed to be less likely than [ð], most of the probability mass is found below 1. Panel (c) shows these ratios on the log scale and brings us to the quantity that is being modeled (i.e. log probability-ratios). These distributions are slightly negatively skewed. To approximate this shape, we used skew-normal distributions to define priors. The solid lines in panel (c) are the simulated log ratios, and the dashed curves are the priors we used.

[89] The prior distribution therefore contains the amount of inferential information in a sample of $n = 4$ /ð/-tokens. In other words, the pre-data information we are supplying is "worth" $n = 4$ observations.

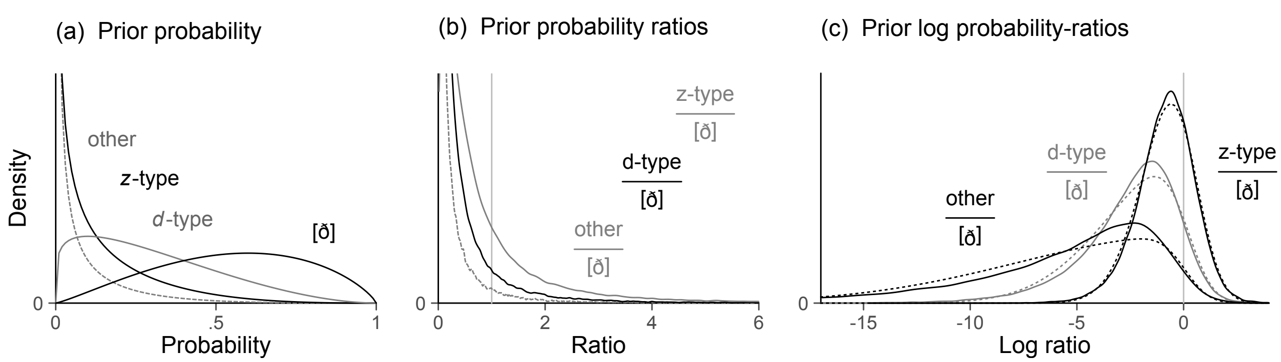


Figure A.18: Construction of the priors for the log probability-ratios based on the prior probabilities of the four variants. Solid lines show simulated values for (a) probabilities, (b) probability ratios, and (c) log probability-ratios. The dashed curves in panel (c) are the parametric prior distributions specified to approximate the simulated values.

$$\bar{\alpha}^{\,\mathrm{d}} + \frac{\sigma_{\alpha_s^{\mathrm{d}}}{}^2 + \sigma_{\alpha_w^{\mathrm{d}}}{}^2}{2} \sim \mathrm{SkewNormal}(0.40, 1.83, –1.60)$$
$$\bar{\alpha}^{\,\mathrm{z}} + \frac{\sigma_{\alpha_s^{\mathrm{z}}}{}^2 + \sigma_{\alpha_w^{\mathrm{z}}}{}^2}{2} \sim \mathrm{SkewNormal}(0.18, 3.56, –3.69)$$
$$\bar{\alpha}^{\mathrm{other}} + \frac{\sigma_{\alpha_s^{\mathrm{other}}}{}^2 + \sigma_{\alpha_w^{\mathrm{other}}}{}^2}{2} \sim \mathrm{SkewNormal}(0.06, 7.79, –8.66)$$

Finally, for the $\bar{\beta}$ parameters we chose mildly regularizing priors centered at zero.

$$\bar{\beta}^{\cdots} \sim \mathrm{Normal}(0, 2)$$

To make the model more robust against unusual subjects and unusual words, we use *t*-distributed rather than normal errors (Gelman et al. 2014: 437). The degrees-of-freedom parameter ν was set to 7.[90]

[90] See Appendix A.3.1 and Figure A.3 (p. 253) for illustration.

$$\nu_{\ldots} = 7$$

Stan code Listed next is the *Stan* code for the native speaker model.

```
data {
  int<lower=0> N;
  int y[N];
  int K;
  int<lower=0> subj[N];
  int<lower=0> word[N];
  int<lower=0> n_subj;
  int<lower=0> n_word;
  real sz[N];
  real far_z[N];
  int n_varcoef_subj;
  int n_varcoef_word;
  int n_cor_subj;
  int n_cor_word;
  int set_nu;
}
parameters {
  vector[K-1] a;

  vector[K-1] b_sz;
  vector[K-1] b_far;
  vector[K-1] b_far_sz;

  vector<lower=0>[n_varcoef_subj] sigma_subj;
  vector<lower=0>[n_varcoef_word] sigma_word;

  matrix[n_varcoef_subj, n_subj] z_subj;
  matrix[n_varcoef_word, n_word] z_word;

  cholesky_factor_corr[n_varcoef_subj] L_Rho_subj;
  cholesky_factor_corr[n_varcoef_word] L_Rho_word;

  real<lower=0> udf_word;
  real<lower=0> udf_subj;
}
transformed parameters {
  matrix[n_subj, n_varcoef_subj] r_s;
  matrix[n_word, n_varcoef_word] r_w;

  matrix[N, K] mu;

  r_s = sqrt(set_nu * udf_subj) * (diag_pre_multiply(sigma_subj, L_Rho_subj) * z_subj)';
  r_w = sqrt(set_nu * udf_word) * (diag_pre_multiply(sigma_word, L_Rho_word) * z_word)';

  for (n in 1:N) {
    mu[n,1] = a[1]                         + r_s[subj[n],1]      + r_w[word[n],1] +
             (b_sz[1]                      + r_s[subj[n],2]                      ) * sz[n] +
             (b_far[1]  + (b_far_sz[1]*sz[n])                    + r_w[word[n],2]) * far_z[n];

    mu[n,2] = a[2]                         + r_s[subj[n],3]      + r_w[word[n],3] +
             (b_sz[2]                      + r_s[subj[n],4]                      ) * sz[n] +
```

```
            (b_far[2]  + (b_far_sz[2]*sz[n])                       + r_w[word[n],4]) * far_z[n];

    mu[n,3] = a[3]                            + r_s[subj[n],5]      + r_w[word[n],5] +
              (b_sz[3]                        + r_s[subj[n],6]                       ) * sz[n] +
              (b_far[3]  + (b_far_sz[3]*sz[n])                       + r_w[word[n],6]) * far_z[n];

    mu[n,4] = 0;
  }
}
model {
  L_Rho_subj  ~ lkj_corr_cholesky(4);
  L_Rho_word  ~ lkj_corr_cholesky(4);

  sigma_subj ~ exponential(2);
  sigma_word ~ exponential(2);

  to_vector(z_subj)  ~ normal(0,1);
  to_vector(z_word)  ~ normal(0,1);

  (a[1] + ((sigma_subj[1]^2 + sigma_word[1]^2)/2)) ~ skew_normal( 0.40 , 1.83, -1.60 );
  (a[2] + ((sigma_subj[3]^2 + sigma_word[3]^2)/2)) ~ skew_normal( 0.18 , 3.56, -3.69 );
  (a[3] + ((sigma_subj[5]^2 + sigma_word[5]^2)/2)) ~ skew_normal( 0.06 , 7.79, -8.66 );

  to_vector(b_sz) ~ normal( 0 , 2 );
  to_vector(b_far) ~ normal( 0 , 2 );
  to_vector(b_far_sz) ~ normal( 0 , 2 );

  udf_word ~ inv_chi_square(set_nu);
  udf_subj ~ inv_chi_square(set_nu);

  for(n in 1:N) y[n] ~ categorical_logit(mu[n]');
}
generated quantities{
  matrix[n_varcoef_subj, n_varcoef_subj] Rho_subj;
  matrix[n_varcoef_word, n_varcoef_word] Rho_word;

  int pos_subj = 1;
  int pos_word = 1;

  vector<lower=-1, upper=1>[n_cor_subj] cor_subj;
  vector<lower=-1, upper=1>[n_cor_word] cor_word;

  Rho_subj = multiply_lower_tri_self_transpose(L_Rho_subj);
  Rho_word = multiply_lower_tri_self_transpose(L_Rho_word);

  for(i in 1:(n_varcoef_subj-1)){
    for(j in (i+1):n_varcoef_subj){
      cor_subj[pos_subj] = Rho_subj[i, j];
      pos_subj += 1;
    }
  }
  for(i in 1:(n_varcoef_word-1)){
    for(j in (i+1):n_varcoef_word){
      cor_word[pos_word] = Rho_word[i, j];
      pos_word += 1;
    }
  }
}
```

Convergence diagnostics The MCMC diagnostics, which are displayed in Figure A.19, indicate convergence of the algorithm: All $\hat{R}$ values (left) are below 1.01, the profile of the rank plot (middle) is flat, and the quantile plot (right) shows sufficiently large effective sample size for all key parameters (i.e. all above $n = 400$).

Posterior distribution of model parameters The posterior distribution of the model parameters for the data from the current study is reported in Tables A.22 (fixed effects) and A.23 (varying effects).

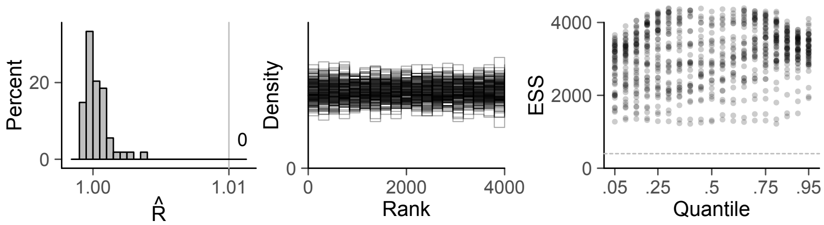


Figure A.19: /ð/ model for the data from Wunder 2012: Convergence diagnostics.

			Posterior quantiles				
Parameter	Mean	SD	.05	.50	.95	N_{eff}	$\hat{R}$
$\bar{\alpha}^{\text{d}}$	–3.16	0.75	–4.44	–3.11	–2.02	1480	1
$\bar{\alpha}^{\text{z}}$	–5.62	1.16	–7.71	–5.48	–4.03	1921	1
$\bar{\alpha}^{\text{other}}$	–4.41	0.78	–5.80	–4.34	–3.26	2815	1
$\bar{\beta}^{\text{d}}_{\text{Sibilant}}$	1.60	0.45	0.85	1.61	2.33	5019	1
$\bar{\beta}^{\text{z}}_{\text{Sibilant}}$	1.24	1.04	–0.56	1.28	2.83	3715	1
$\bar{\beta}^{\text{other}}_{\text{Sibilant}}$	1.59	0.62	0.55	1.60	2.57	4157	1
$\bar{\beta}^{\text{d}}_{\text{FAR}}$	–1.54	0.70	–2.69	–1.54	–0.41	1738	1
$\bar{\beta}^{\text{z}}_{\text{FAR}}$	–2.21	0.88	–3.68	–2.21	–0.76	2721	1
$\bar{\beta}^{\text{other}}_{\text{FAR}}$	–1.40	0.66	–2.46	–1.38	–0.36	3185	1
$\bar{\beta}^{\text{d}}_{\text{FAR:Sibilant}}$	1.22	0.58	0.30	1.20	2.18	3765	1
$\bar{\beta}^{\text{z}}_{\text{FAR:Sibilant}}$	–1.50	1.06	–3.33	–1.46	0.16	4365	1
$\bar{\beta}^{\text{other}}_{\text{FAR:Sibilant}}$	0.24	0.61	–0.79	0.25	1.21	5223	1

Table A.22: /ð/ model for the data from Wunder 2012: Posterior distribution of the fixed effects parameters.

Table A.23: /ð/ model for the data from Wunder 2012: Posterior distribution of the varying effects parameters.

Parameter	Mean	SD	Posterior quantiles .05	.50	.95	N_{eff}	$\hat{R}$
Subject: Standard deviation parameters							
$\sigma_{\alpha_s^d}$	1.66	0.46	0.96	1.63	2.46	1917	1
$\sigma_{\beta^d_{\text{Sibilant, s}}}$	0.30	0.25	0.02	0.24	0.78	2290	1
$\sigma_{\alpha_s^z}$	1.66	0.56	0.89	1.58	2.67	2111	1
$\sigma_{\beta^z_{\text{Sibilant, s}}}$	0.80	0.55	0.07	0.72	1.82	1755	1
$\sigma_{\alpha_s^{\text{other}}}$	1.23	0.40	0.68	1.18	1.96	2410	1
$\sigma_{\beta^{\text{other}}_{\text{Sibilant, s}}}$	0.31	0.27	0.02	0.23	0.84	2493	1
Subject: Correlation parameters							
$\rho(\alpha_s^d, \beta^d_{\text{Sibilant, s}})$	0.00	0.28	–0.47	0.00	0.47	7722	1
$\rho(\alpha_s^d, \alpha_s^z)$	–0.03	0.20	–0.34	–0.04	0.31	2633	1
$\rho(\alpha_s^d, \beta^z_{\text{Sibilant, s}})$	–0.13	0.25	–0.53	–0.14	0.30	6388	1
$\rho(\alpha_s^d, \alpha_s^{\text{other}})$	0.02	0.20	–0.31	0.02	0.35	2714	1
$\rho(\alpha_s^d, \beta^{\text{other}}_{\text{Sibilant, s}})$	–0.04	0.27	–0.49	–0.04	0.42	7607	1
$\rho(\beta^d_{\text{Sibilant, s}}, \alpha_s^z)$	–0.05	0.28	–0.50	–0.06	0.42	881	1
$\rho(\beta^d_{\text{Sibilant, s}}, \beta^z_{\text{Sibilant, s}})$	0.00	0.28	–0.47	0.00	0.46	3794	1
$\rho(\beta^d_{\text{Sibilant, s}}, \alpha_s^{\text{other}})$	–0.04	0.28	–0.49	–0.05	0.43	931	1
$\rho(\beta^d_{\text{Sibilant, s}}, \beta^{\text{other}}_{\text{Sibilant, s}})$	0.03	0.27	–0.43	0.03	0.47	4758	1
$\rho(\alpha_s^z), \beta^z_{\text{Sibilant, s}})$	–0.02	0.26	–0.45	–0.02	0.40	6914	1
$\rho(\alpha_s^z), \alpha_s^{\text{other}})$	0.23	0.20	–0.11	0.24	0.54	2979	1
$\rho(\alpha_s^z), \beta^{\text{other}}_{\text{Sibilant, s}})$	0.00	0.28	–0.46	0.00	0.45	7113	1
$\rho(\beta^z_{\text{Sibilant, s}}, \alpha_s^{\text{other}})$	0.10	0.25	–0.32	0.11	0.49	1418	1
$\rho(\beta^z_{\text{Sibilant, s}}, \beta^{\text{other}}_{\text{Sibilant, s}})$	–0.02	0.27	–0.47	–0.02	0.43	4807	1
$\rho(\alpha_s^{\text{other}}, \beta^{\text{other}}_{\text{Sibilant, s}})$	–0.03	0.27	–0.48	–0.04	0.43	6194	1
Word: Standard deviation parameters							
$\sigma_{\alpha_w^d}$	0.91	0.41	0.34	0.86	1.65	2381	1
$\sigma_{\beta^d_{\text{FAR, w}}}$	0.45	0.30	0.05	0.41	1.00	2308	1
$\sigma_{\alpha_w^z}$	0.65	0.40	0.10	0.60	1.38	2421	1
$\sigma_{\beta^z_{\text{FAR, w}}}$	0.39	0.34	0.02	0.30	1.04	2679	1
$\sigma_{\alpha_w^{\text{other}}}$	0.66	0.40	0.09	0.61	1.39	2229	1
$\sigma_{\beta^{\text{other}}_{\text{FAR, w}}}$	0.31	0.26	0.02	0.25	0.82	3944	1
Word: Correlation parameters							
$\rho(\alpha_w^d, \beta^d_{\text{FAR, w}})$	–0.01	0.27	–0.45	–0.01	0.44	8411	1
$\rho(\alpha_w^d, \alpha_w^z)$	–0.03	0.26	–0.47	–0.03	0.41	5892	1
$\rho(\alpha_w^d, \beta^z_{\text{FAR, w}})$	–0.07	0.27	–0.51	–0.07	0.39	8296	1
$\rho(\alpha_w^d, \alpha_w^{\text{other}})$	0.02	0.26	–0.41	0.02	0.44	6989	1
$\rho(\alpha_w^d, \beta^{\text{other}}_{\text{FAR, w}})$	0.02	0.27	–0.42	0.03	0.46	9771	1
$\rho(\beta^d_{\text{FAR, w}}, \alpha_w^z)$	0.09	0.27	–0.36	0.10	0.53	5193	1
$\rho(\beta^d_{\text{FAR, w}}, \beta^z_{\text{FAR, w}})$	0.05	0.27	–0.40	0.06	0.48	5918	1
$\rho(\beta^d_{\text{FAR, w}}, \alpha_w^{\text{other}})$	0.03	0.27	–0.41	0.03	0.46	3802	1
$\rho(\beta^d_{\text{FAR, w}}, \beta^{\text{other}}_{\text{FAR, w}})$	–0.07	0.27	–0.50	–0.07	0.38	5761	1
$\rho(\alpha_w^z), \beta^z_{\text{FAR, w}})$	0.06	0.28	–0.40	0.07	0.50	5442	1
$\rho(\alpha_w^z), \alpha_w^{\text{other}})$	0.05	0.27	–0.40	0.05	0.49	4898	1
$\rho(\alpha_w^z), \beta^{\text{other}}_{\text{FAR, w}})$	–0.04	0.27	–0.49	–0.05	0.40	5256	1
$\rho(\beta^z_{\text{FAR, w}}, \alpha_w^{\text{other}})$	0.03	0.28	–0.42	0.03	0.49	3959	1
$\rho(\beta^z_{\text{FAR, w}}, \beta^{\text{other}}_{\text{FAR, w}})$	–0.03	0.28	–0.50	–0.03	0.44	4630	1
$\rho(\alpha_w^{\text{other}}, \beta^{\text{other}}_{\text{FAR, w}})$	0.02	0.27	–0.43	0.02	0.47	3797	1

A.10 Final voiced obstruents: Acoustic analysis

The data sets analyzed in Chapter 10 are openly available via *TROLLing*[91] (Sönning & Pascoe 2020). The OSF repository hosts the *Stan* code for the native speakers[92] and the learners[93], and the *R* script[94] for running the analyses.

[91] https://doi.org/10.18710/DKIGE5

[92] https://osf.io/us4a5/

[93] https://osf.io/uy45k/

[94] https://osf.io/4rtn9/

A.10.1 Native speakers

Model definition The PVD ratio for each word pair was first transformed to the log scale. These log ratios were then modeled using a Student-*t* distribution with the degrees-of-freedom parameter ν set to 7.[95]

[95] See Appendix A.3.1 and Figure A.3 (p. 253) for illustration.

$$\text{Log PVD ratio}_i \sim \text{Student-t}\left(\mu_i, \sigma, \nu\right)$$

The log PVD ratio was modeled conditional on the variety of the native speaker, which is signaled by the index variable $\bar{\alpha}$.[96] Thus, $\bar{\alpha}_{\text{AmE}}$ is the average log PVD ratio observed among the US native speakers recorded for the current study. The α parameters are varying effects for subject and word. The α_{s} set denotes by how much each speaker deviates from the average for their variety. The α_{w} component consists of two sets, which indicate by how much each word (pair) is deflected from the expected log PVD ratio, with $\alpha_{\text{w}}^{\text{AmE}}$ signaling the deflection in AmE speech, and $\alpha_{\text{w}}^{\text{BrE}}$ that in BrE speech.

[96] On the use of index variables in regression modeling, see McElreath (2020: 155, 447).

$$\mu_i = \bar{\alpha}_{\text{Var[i]}} + \alpha_{\text{w}[i],\text{Var}[i]} + \alpha_{\text{s}[i]}$$

The varying effects are modeled using a (multivariate) Student-*t* distribution, with degrees-of-freedom parameter ν set to 7.

$$\alpha_{\text{s}} \sim \text{Student-t}\left(0, \sigma_{\text{s}}, \nu_{\text{s}}\right)$$

$$\begin{bmatrix} \alpha_{\text{w}}^{\text{AmE}} \\ \alpha_{\text{w}}^{\text{BrE}} \end{bmatrix} \sim \text{MVStudent-t}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mathbf{S}_{\text{w}}, \nu_{\text{w}}\right)$$

The covariance matrix for the by-word varying effects distribution is constructed from the standard deviation and correlation parameters:

$$\mathbf{S}_{\text{w}} = \begin{pmatrix} \sigma_{\alpha_{\text{w}}^{\text{AmE}}} & 0 \\ 0 & \sigma_{\alpha_{\text{w}}^{\text{BrE}}} \end{pmatrix} \mathbf{R}_{\text{w}} \begin{pmatrix} \sigma_{\alpha_{\text{w}}^{\text{AmE}}} & 0 \\ 0 & \sigma_{\alpha_{\text{w}}^{\text{BrE}}} \end{pmatrix}$$

Thus, the parameter $\sigma_{\alpha_{\text{w}}^{\text{AmE}}}$ is a standard deviation (on the log ratio scale), which describes how much the word-specific averages produced by the AmE informants vary around the conditional means. $\mathbf{R}_{\text{w}}$ is a correlation matrix, which includes a correlation coefficient that captures how the by-word varying effects are correlated. The standard deviation parameters for the dispersion of the varying effects were given a weakly informative outset, that is, an exponential distribution with scale parameter 2.[97] This prior states that, on average, (average) PVD ratios are within a factor of about 3.0 of the prediction – for the observation-level residuals (σ), across subjects (σ_{s}), and across words ($\sigma_{\alpha_{\text{w}}^{\text{AmE}}}$, $\sigma_{\alpha_{\text{w}}^{\text{BrE}}}$) – with 90% probability.

[97] See Appendix A.6.1 and Figure A.11 for details.

$$\sigma_{\dots} \sim \text{Exponential}(2)$$

There is no ancillary information on the correlation between the varying coefficients, and we therefore specify an LKJ prior with η set to 4.[98]

98 See Appendix A.3.1 and Figure A.2 (p. 253) for illustration.

$$\mathbf{R}_{\mathrm{w}} \sim \mathrm{LKJcorr}(4)$$

The priors for the expected average log PVD ratio in BrE and AmE were developed in §10.6 (see Figure 10.9).

$$\bar{\alpha}_{\mathrm{BrE}} \sim \mathrm{Normal}(0.5, 0.2)$$
$$\bar{\alpha}_{\mathrm{AmE}} \sim \mathrm{Normal}(0.4, 0.17)$$

The ν parameter was set to 7.[99]

99 See Appendix A.3.1 and Figure A.3 (p. 253) for illustration.

$$\nu_{\dots} = 7$$

Stan code A translation of the model statement in the the *Stan*follows.

```
data {
  int<lower=0> N;
  vector[N] y;
  int<lower=0> subj[N];
  int<lower=0> word[N];
  int<lower=0> n_subj;
  int<lower=0> n_word;
  int var_id[N];
  int n_varcoef_subj;
  int n_varcoef_word;
  int n_cor_word;
  int set_nu;
}
parameters {
  row_vector[2] a;

  real<lower=0> sigma_subj;
  vector<lower=0>[n_varcoef_word] sigma_word;

  matrix[n_varcoef_word, n_word] z_word;
  cholesky_factor_corr[n_varcoef_word] L_Rho_word;

  real<lower=0> sigma;

  real<lower=0> udf_word;
  vector[n_subj] r_s;
}
transformed parameters {
  matrix[n_word, n_varcoef_word] r_w;
  vector[N] mu;

  r_w = sqrt(set_nu * udf_word) * (diag_pre_multiply(sigma_word, L_Rho_word) * z_word)';

  for (n in 1:N) {
    mu[n] = a[var_id[n]] + r_s[subj[n]] + r_w[word[n], var_id[n]];
  }
}
model {
  sigma ~ exponential(2);
  L_Rho_word  ~ lkj_corr_cholesky(4);
  sigma_subj ~ exponential(2);
  sigma_word ~ exponential(2);

  to_vector(z_word)  ~ normal(0,1);

  a[1] ~ normal( 0.4 , 0.17 );
  a[2] ~ normal( 0.5 , 0.3 );

  r_s ~ student_t(set_nu, 0, sigma_subj);
  udf_word ~ inv_chi_square(set_nu);

  target += student_t_lpdf(y | set_nu, mu, sigma);
}
generated quantities{
  matrix[n_varcoef_word, n_varcoef_word] Rho_word;
```

```
  int pos_word = 1;
  vector<lower=-1, upper=1>[n_cor_word] cor_word;
  Rho_word = multiply_lower_tri_self_transpose(L_Rho_word);
  for(i in 1:(n_varcoef_word-1)){
    for(j in (i+1):n_varcoef_word){
      cor_word[pos_word] = Rho_word[i, j];
      pos_word += 1;
    }
  }
}
```

Convergence diagnostics Figure A.20 indicates convergence of the Markov chains: All $\hat{R}$ values (left) are below 1.01, the profile of the rank plot (middle) is flat, and the quantile plot (right) suggests that all key parameters are supported by an effective sample size greater than 400.

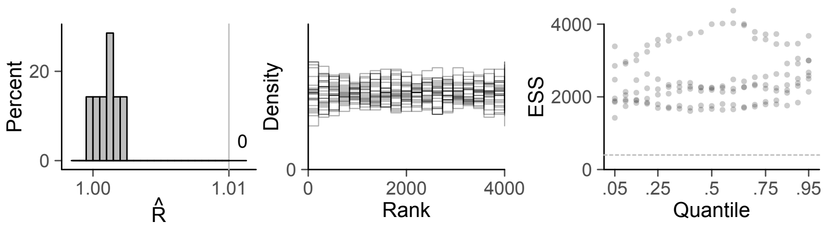


Figure A.20: PVD ratio model for the native speakers: Convergence diagnostics.

Posterior distribution of model parameters Table A.24 reports the posterior distribution of the parameters in the model.

			Posterior quantiles				
Parameter	Mean	SD	.05	.50	.95	N_{eff}	$\hat{R}$
Fixed effects							
$\bar{\alpha}_{\text{AmE}}$	0.57	0.05	0.48	0.57	0.66	1640	1
$\bar{\alpha}_{\text{BrE}}$	0.40	0.07	0.29	0.40	0.52	1327	1
σ	0.19	0.01	0.17	0.19	0.21	5012	1
Varying effects							
σ_{s}	0.12	0.03	0.08	0.11	0.16	3531	1
$\sigma_{\text{w}}^{\text{AmE}}$	0.10	0.05	0.03	0.09	0.19	1844	1
$\sigma_{\text{w}}^{\text{BrE}}$	0.21	0.08	0.11	0.19	0.35	1816	1
$\rho(\alpha_{\text{w}}^{\text{AmE}}, \alpha_{\text{w}}^{\text{BrE}})$	0.26	0.27	–0.21	0.28	0.66	1305	1

Table A.24: PVD ratio model for the native speakers: Posterior distribution of the model parameters.

A.10.2 Learners

Model definition For the learner data, the structure of the model is similar. The average log PVD ratio was modeled as a function of proficiency. The intercept parameter $\bar{\alpha}$ denotes the log PVD ratio for intermediate-level learners with a FAR score of 0. The α parameters are varying intercepts for subject (α_{s}) and word (α_{w}). $\bar{\beta}$ reflects the increment in the outcome for a 1-unit change on the FAR scale. The β_{w} parameters are by-word varying slopes for proficiency: They reflect by how much each word deviates from the average cline across proficiency levels. The remainder of the model closely aligns with that specified for the native speakers. For the prior on the intercept parameter $\bar{\alpha}$, see Figure 10.9.

$$
\begin{aligned}
\text{Log PVD ratio}_i &\sim \text{Student-t}\,(\mu_i, \sigma, \nu) \\
\mu_i &= \bar{\alpha} + \alpha_{\text{s}[i]} + \alpha_{\text{w}[i]} + (\bar{\beta} + \beta_{\text{w}[i]})\text{FAR}_i \\
\alpha_{\text{s}} &\sim \text{Student-t}\,(0, \sigma_{\text{s}}, \nu_{\text{s}}) \\
\begin{bmatrix} \alpha_{\text{w}} \\ \beta_{\text{w}} \end{bmatrix} &\sim \text{MVStudent-t}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mathbf{S}_{\text{w}}, \nu_{\text{w}} \right) \\
\mathbf{S}_{\text{w}} &= \begin{pmatrix} \sigma_{\alpha_{\text{w}}} & 0 \\ 0 & \sigma_{\beta_{\text{w}}} \end{pmatrix} \mathbf{R}_{\text{w}} \begin{pmatrix} \sigma_{\alpha_{\text{w}}} & 0 \\ 0 & \sigma_{\beta_{\text{w}}} \end{pmatrix} \\
\sigma_{\ldots} &\sim \text{Exponential}(2) \\
\mathbf{R}_{\text{w}} &\sim \text{LKJcorr}(4) \\
\bar{\alpha} &\sim \text{Normal}(0.34, 0.3) \\
\bar{\beta} &\sim \text{Normal}(0, 1) \\
\nu_{\ldots} &= 7
\end{aligned}
$$

Stan code This is the *Stan* code for the learner model.

```
data {
  int<lower=0> N;
  vector[N] y;
  int<lower=0> subj[N];
  int<lower=0> word[N];
  int<lower=0> n_subj;
  int<lower=0> n_word;
  real far_z[N];
  int n_varcoef_subj;
  int n_varcoef_word;
  int n_cor_word;
  int set_nu;
}
parameters {
  real a;
  real b;

  real<lower=0> sigma_subj;
  vector<lower=0>[2] sigma_word;

  matrix[2, n_word] z_word;
  cholesky_factor_corr[2] L_Rho_word;

  real<lower=0> sigma;

  real<lower=0> udf_word;
  vector[n_subj] r_s;
}
transformed parameters {
  matrix[n_word, 2] r_w;
  vector[N] mu;

  r_w = sqrt(set_nu * udf_word) * (diag_pre_multiply(sigma_word, L_Rho_word) * z_word)’;

  for (n in 1:N) {
    mu[n] = a + (b + r_w[word[n],2]) * far_z[n] +
      r_s[subj[n]] + r_w[word[n], 1];
  }
}
model {
  sigma ~ exponential(2);
  L_Rho_word  ~ lkj_corr_cholesky(4);
  sigma_subj ~ exponential(2);
```

```
  sigma_word ~ exponential(2);

  to_vector(z_word)  ~ normal(0,1);

  a ~ normal( 0.34 , 0.3 );
  b ~ normal( 0 , 0.25 );

  udf_word ~ inv_chi_square(set_nu);

  r_s ~ student_t(set_nu, 0, sigma_subj);

  target += student_t_lpdf(y | set_nu, mu, sigma);
}
generated quantities{
  matrix[n_varcoef_word, n_varcoef_word] Rho_word;
  int pos_word = 1;
  vector<lower=-1, upper=1>[n_cor_word] cor_word;
  Rho_word = multiply_lower_tri_self_transpose(L_Rho_word);
  for(i in 1:(n_varcoef_word-1)){
    for(j in (i+1):n_varcoef_word){
      cor_word[pos_word] = Rho_word[i, j];
      pos_word += 1;
    }
  }
}
```

Convergence diagnostics Convergence diagnostics, shown Figure A.21, indicate convergence of the algorithm: All $\hat{R}$ values (left) are below 1.01, the profile of the rank plot (middle) is flat, and the quantile plot (right) suggests that all key parameters are supported by a sufficiently large effective sample size (i.e. all above $N = 400$).

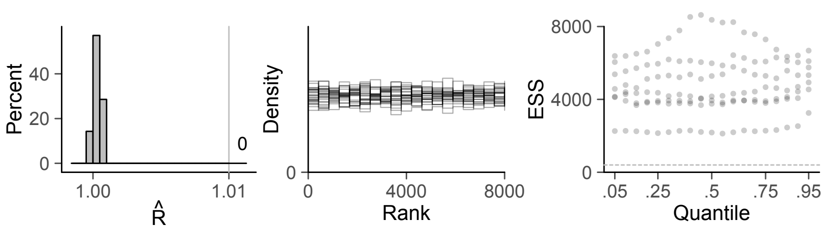


Figure A.21: PVD ratio model for the learners: Convergence diagnostics.

Posterior distribution of model parameters Table A.25 gives a summary of the posterior distribution of the parameters in the model.

			Posterior quantiles				
Parameter	Mean	SD	.05	.50	.95	N_{eff}	$\hat{R}$
Fixed effects							
$\bar{\alpha}$	0.22	0.06	0.12	0.22	0.32	1694	1
$\bar{\beta}$	0.09	0.02	0.05	0.09	0.13	3653	1
σ	0.19	0.01	0.18	0.19	0.20	8581	1
Varying effects							
σ_{s}	0.09	0.01	0.07	0.09	0.12	4745	1
$\sigma_{\alpha_{\text{w}}}$	0.21	0.08	0.11	0.20	0.35	3168	1
$\sigma_{\alpha_{\text{w}}}$	0.06	0.03	0.03	0.06	0.11	3352	1
$\rho(\alpha_{\text{w}}, \beta_{\text{w}})$	0.20	0.25	–0.23	0.21	0.59	5390	1

Table A.25: PVD ratio model for the learners: Posterior distribution of the model parameters.

A.11 *Final voiced obstruents: Auditory analysis*

The data from Chapter 10 are openly available via *TROLLing*[100] (Sönning & Pascoe 2020). The *Stan*[101] and *R* code[102] can be found on the OSF.

[100] https://doi.org/10.18710/DKIGE5

[101] https://osf.io/d74qm/

[102] https://osf.io/4rtn9/

A.11.1 *Pascoe 1987*

Model definition The realization of final voiced obstruents was coded with three categories, and we therefore applied a multinomial regression model for analysis. We model the probability of each variant, i.e. p^{voiced}, p^{devoiced}, and $p^{\text{voiceless}}$, as a function of our predictor variables.

$$\text{FVO}_i \sim \text{Categorical}(p_i^{\text{voiced}}, p_i^{\text{devoiced}}, p_i^{\text{voiceless}})$$

Our aim is to describe how these probabilities change with (i) proficiency level, (ii) characteristics of the obstruent, and (iii) the following phonetic context. The model does not address these probabilities directly, but through a link function (the softmax function), which translates constrained probabilities to an unconstrained scale. For technical reasons, one of the outcome events serves as the reference category (in our case the voiceless realization), and we therefore model only two quantities: (i) the log ratio of the probability of voiced variants divided by that of the reference category (i.e. $\log(p^{\text{voiced}} / p^{\text{voiceless}})$) and (ii) the log ratio of the probability of devoiced variant divided by that of the reference category (i.e. $\log(p^{\text{devoiced}} / p^{\text{voiceless}})$). Our regression model therefore simultaneously models two log ratios, which is why we are modeling two conditional means, one for each log probability-ratio.

The $\bar{\alpha}_{\dots}$ parameters express the expected probability ratio when FAR is held at 0 and both obstruent characteristics and following contexts are held at their means.[103] The α coefficients represent the varying intercepts for subjects ($\alpha_{\text{s}}^{\dots}$) and words ($\alpha_{\text{w}}^{\dots}$). Thereby, each speaker and word is allowed to assume a higher or lower ratio for each pair of probabilities. The next lines describe how probability ratios change with proficiency level. The $\bar{\beta}_{\text{FAR}}^{\dots}$ parameters express by how much a given probability ratio changes, on average, as we shift the FAR scale upwards by 1 unit.[104] The $\beta_{\text{FAR, w}}^{\dots}$ parameters are varying slopes and represent how much each word differs from the average FAR trend for a particular log probability-ratio.

The next two lines address obstruent characteristics. Obstruents were classified according to the underlying phoneme (/d g v z/). These four categories were coded with nested contrasts (see Schad et al. 2020), where place of articulation is nested in manner of articulation. The coding and the corresponding hypothesis matrix[105] are shown in Table A.26.

The final two lines are concerned with the sonority of the following context, which was classified with 5 ordered categories: Vowel, voiced consonant, voiceless consonant, glottal stop, and pause. These categories were coded with repeated contrasts (see Schad et al. 2020), which are shown in Table A.27, along with the corresponding hypothesis matrix.

[103] In this analysis, this is the simple (i.e. unweighted) average over the levels of the predictors.

[104] Thus, the parameter $\bar{\beta}_{\text{FAR}}^{\text{voiced}}$ describes the change in the log probability-ratio $\log(p^{\text{voiced}} / p^{\text{voiceless}})$ when comparing speakers who differ by 1 point on the FAR scale. We expect this parameter to be positive, because the probability of observing target-like voiced variants vs. L1-transferred voiceless renditions should increase with proficiency.

[105] The hypothesis matrix gives a more transparent expression of the meaning of the intercept and the different contrasts. The weights for the intercept reflect the meaning of this parameter. In the present case, it is the average over all 5 categories. For the contrasts, the hypothesis matrix shows the comparison expressed by each coefficient. Each contrast sums to zero and the positive and negative weights show which categories are being compared.

	Contrasts			Hypothesis matrix			
Obstruent	dg/vz	d/g	v/z	Intercept	dg/vz	d/g	v/z
/d/	−1/2	−1/2	0	1/4	−1/2	−1	0
/g/	−1/2	1/2	0	1/4	−1/2	1	0
/v/	1/2	0	−1/2	1/4	1/2	0	−1
/z/	1/2	0	1/2	1/4	1/2	0	1

Table A.26: FVO model for the data from Pascoe (1987): Custom nested contrasts for obstruent type.

	Contrasts				Hypothesis matrix				
Context	V/vC	vC/vlC	vlC/G	G/P	Intercept	V/vC	vC/vlC	vlC/G	G/P
Vowel	−4/5	−3/5	−2/5	−1/5	1/5	−1	0	0	0
Voiced C	1/5	−3/5	−2/5	−1/5	1/5	1	−1	0	0
Voiceless C	1/5	2/5	−2/5	−1/5	1/5	0	1	−1	0
Glottal stop	1/5	2/5	3/5	−1/5	1/5	0	0	1	−1
Pause	1/5	2/5	3/5	4/5	1/5	0	0	0	1

Table A.27: FVO model for the data from Pascoe (1987): Custom contrasts for the following context.

$$
\begin{aligned}
\log\left(\frac{p_i^{\text{voiced}}}{p_i^{\text{voiceless}}}\right) = {} & \bar{\alpha}^{\text{voiced}} + \alpha^{\text{voiced}}_{\text{s}[i]} + \alpha^{\text{voiced}}_{\text{w}[i]} + \\
& (\bar{\beta}^{\text{voiced}}_{\text{FAR}} + \beta^{\text{voiced}}_{\text{FAR, w}[i]})\text{FAR}_i + \\
& (\bar{\beta}^{\text{voiced}}_{\text{dg/vz}} + \beta^{\text{voiced}}_{\text{dg/vz, s}[i]})\text{dg/vz}_i + \\
& (\bar{\beta}^{\text{voiced}}_{\text{d/g}} + \beta^{\text{voiced}}_{\text{d/g, s}[i]})\text{d/g}_i + (\bar{\beta}^{\text{voiced}}_{\text{v/z}} + \beta^{\text{voiced}}_{\text{v/z, s}[i]})\text{v/z}_i + \\
& (\bar{\beta}^{\text{voiced}}_{\text{V/vC}} + \beta^{\text{voiced}}_{\text{V/vC, s}[i]})\text{V/vC}_i + (\bar{\beta}^{\text{voiced}}_{\text{vC/vlC}} + \beta^{\text{voiced}}_{\text{vC/vlC, s}[i]})\text{vC/vlC}_i + \\
& (\bar{\beta}^{\text{voiced}}_{\text{vlC/G}} + \beta^{\text{voiced}}_{\text{vlC/G, s}[i]})\text{vlC/G}_i + (\bar{\beta}^{\text{voiced}}_{\text{G/P}} + \beta^{\text{voiced}}_{\text{G/P, s}[i]})\text{G/P}_i \\
\log\left(\frac{p_i^{\text{devoiced}}}{p_i^{\text{voiceless}}}\right) = {} & \bar{\alpha}^{\text{devoiced}} + \alpha^{\text{devoiced}}_{\text{s}[i]} + \alpha^{\text{devoiced}}_{\text{w}[i]} + \\
& (\bar{\beta}^{\text{devoiced}}_{\text{FAR}} + \beta^{\text{devoiced}}_{\text{FAR, s}[i]})\text{FAR}_i + \\
& (\bar{\beta}^{\text{devoiced}}_{\text{dg/vz}} + \beta^{\text{devoiced}}_{\text{dg/vz, s}[i]})\text{dg/vz}_i + \\
& (\bar{\beta}^{\text{devoiced}}_{\text{d/g}} + \beta^{\text{devoiced}}_{\text{d/g, s}[i]})\text{d/g}_i + (\bar{\beta}^{\text{devoiced}}_{\text{v/z}} + \beta^{\text{devoiced}}_{\text{v/z, s}[i]})\text{v/z}_i + \\
& (\bar{\beta}^{\text{devoiced}}_{\text{V/vC}} + \beta^{\text{devoiced}}_{\text{V/vC, s}[i]})\text{V/vC}_i + (\bar{\beta}^{\text{devoiced}}_{\text{vC/vlC}} + \beta^{\text{devoiced}}_{\text{vC/vlC, s}[i]})\text{vC/vlC}_i + \\
& (\bar{\beta}^{\text{devoiced}}_{\text{vlC/G}} + \beta^{\text{devoiced}}_{\text{vlC/G, s}[i]})\text{vlC/G}_i + (\bar{\beta}^{\text{devoiced}}_{\text{G/P}} + \beta^{\text{devoiced}}_{\text{G/P, s}[i]})\text{G/P}_i
\end{aligned}
$$

The varying effects are assumed to follow a multivariate Student-t distribution, which allows for slightly more robust inference. The ν parameter was set to 7.[106] **S** denotes the covariance matrix, which contains information about the dispersion of the varying effects and their pairwise associations.

[106] See Appendix A.3.1 and Figure A.3 (p. 253) for illustration.

$$
\begin{bmatrix} \alpha^{\text{voiced}}_{\text{w}} & \alpha^{\text{devoiced}}_{\text{w}} \\ \beta^{\text{voiced}}_{\text{FAR, w}} & \beta^{\text{devoiced}}_{\text{FAR, w}} \end{bmatrix} \sim \text{MVStudent-t}\left(\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \mathbf{S}_{\text{w}}, \nu_{\text{w}} \right)
$$

$$\begin{bmatrix} \alpha_{s}^{\text{voiced}} & \alpha_{s}^{\text{devoiced}} \\ \beta_{\text{dg/vz, s}}^{\text{voiced}} & \beta_{\text{dg/vz, s}}^{\text{devoiced}} \\ \beta_{\text{v/z, s}}^{\text{voiced}} & \beta_{\text{v/z, s}}^{\text{devoiced}} \\ \beta_{\text{d/g, s}}^{\text{voiced}} & \beta_{\text{d/g, s}}^{\text{devoiced}} \\ \beta_{\text{V/vC, s}}^{\text{voiced}} & \beta_{\text{V/vC, s}}^{\text{devoiced}} \\ \beta_{\text{vC/vlC, s}}^{\text{voiced}} & \beta_{\text{vC/vlC, s}}^{\text{devoiced}} \\ \beta_{\text{vlC/G, s}}^{\text{voiced}} & \beta_{\text{vlC/G, s}}^{\text{devoiced}} \\ \beta_{\text{G/P, s}}^{\text{voiced}} & \beta_{\text{G/P, s}}^{\text{devoiced}} \end{bmatrix} \sim \text{MVStudent-t}\left(\begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \mathbf{S}_{\text{s}}, \nu_{\text{s}}\right)$$

The covariance matrices are defined using (instead of variances and covariances) the standard deviation of the varying effects and their pairwise correlations. Thus, the parameter $\sigma_{\alpha_{w}^{\text{voiced}}}$ is a standard deviation (on the log ratio scale), which describes how much speakers varied around the average log probability-ratio of voiced over voiceless realizations. **R** is a correlation matrix, which includes correlation coefficients that capture how by-subject varying effects on the one hand, and by-word varying effects on the other, are correlated.

$$\mathbf{S}_{\text{s}} = \text{CovarianceMatrix}\begin{pmatrix} \sigma_{\alpha_{w}^{\text{voiced}}} & \sigma_{\alpha_{w}^{\text{devoiced}}} \\ \sigma_{\beta_{\text{FAR, w}}^{\text{voiced}}} & \sigma_{\beta_{\text{FAR, w}}^{\text{devoiced}}} \end{pmatrix}, \mathbf{R}_{\text{s}})$$

$$\mathbf{S}_{\text{w}} = \text{CovarianceMatrix}\begin{pmatrix} \sigma_{\alpha_{s}^{\text{voiced}}} & \sigma_{\alpha_{s}^{\text{devoiced}}} \\ \sigma_{\beta_{\text{dg/vz, s}}^{\text{voiced}}} & \sigma_{\beta_{\text{dg/vz, s}}^{\text{devoiced}}} \\ \sigma_{\beta_{\text{d/g, s}}^{\text{voiced}}} & \sigma_{\beta_{\text{d/g, s}}^{\text{devoiced}}} \\ \sigma_{\beta_{\text{v/z, s}}^{\text{voiced}}} & \sigma_{\beta_{\text{v/z, s}}^{\text{devoiced}}} \\ \sigma_{\beta_{\text{V/vC, s}}^{\text{voiced}}} & \sigma_{\beta_{\text{V/vC, s}}^{\text{devoiced}}} \\ \sigma_{\beta_{\text{vC/vlC, s}}^{\text{voiced}}} & \sigma_{\beta_{\text{vC/vlC, s}}^{\text{devoiced}}} \\ \sigma_{\beta_{\text{vlC/G, s}}^{\text{voiced}}} & \sigma_{\beta_{\text{vlC/G, s}}^{\text{devoiced}}} \\ \sigma_{\beta_{\text{G/P, s}}^{\text{voiced}}} & \sigma_{\beta_{\text{G/P, s}}^{\text{devoiced}}} \end{pmatrix}, \mathbf{R}_{\text{w}}$$

Finally, the priors. The standard deviation parameters for the dispersion of the varying effects were given a weakly informative specification in the form of an exponential distribution with scale parameter 2.[107]

$$\sigma_{\dots} \sim \text{Exponential}(2)$$

In the absence of information on the correlation between the varying coefficients, we rely on an LKJ prior with shape parameter η set to 4.[108]

$$\mathbf{R}_{\dots} \sim \text{LKJcorr}(4)$$

To specify priors for the log probability-ratio intercept parameters, we first determined the state of information on the probability of observing each variant on the probability scale (see §10.6 for motivation). To this end, we used a Dirichlet distribution with parameters 0.8 (voiced variants), 1.2 (devoiced variants), and 2 (voiceless variants) to generate these values.[109] Figure A.22a shows how, for each variant, the a priori probability mass is

[107] See Appendix A.3.1 and Figure A.2 (p. 253) for illustration.

[108] See Appendix A.3.1 and Figure A.2 (p. 253) for illustration.

[109] The prior distribution therefore contains the amount of inferential information offered by a sample of $n = 4$ /ð/-tokens. In other words, the pre-data information we are supplying is "worth" $n = 4$ observations.

concentrated over the interval [0; 1]. We then divided the distributions for voiced and devoiced variants by that of voiceless realizations, which gave us two distributions of ratios. This intermediate step is shown in panel (b). Since voiced and devoiced variants are assumed to be less likely than voiceless renditions, most of the probability mass is found below 1. Panel (c) shows these ratios on the log scale and brings us to the quantity that is being modeled (i.e. log probability-ratios). These distributions are slightly negatively skewed. To approximate this shape, we specified a skew-normal distribution as a prior. The solid lines in panel (c) are the simulated log ratios, and the dotted curves are the prior distributions we used.

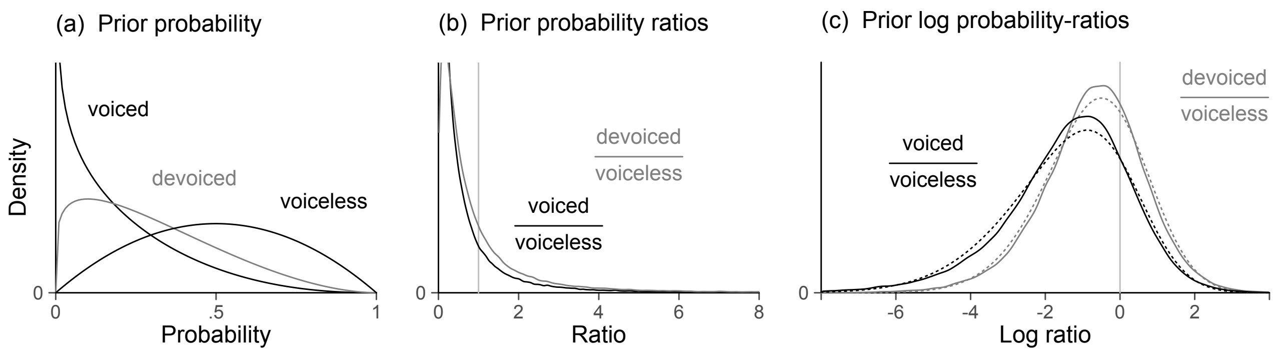


Figure A.22: Construction of the priors for the log probability-ratios based on the prior probabilities of the three variants. Solid lines show simulated values for (a) probabilities, (b) probability ratios, and (c) log probability-ratios. The dashed curves in panel (c) are the parametric prior distributions specified to approximate the simulated values.

$$\bar{\alpha}^{\text{voiced}} + \frac{{\sigma_{\alpha_{\text{s}}^{\text{voiced}}}}^2 + {\sigma_{\alpha_{\text{w}}^{\text{voiced}}}}^2}{2} \sim \text{SkewNormal}(0.40, 2.48, \text{-}2.25)$$
$$\bar{\alpha}^{\text{devoiced}} + \frac{{\sigma_{\alpha_{\text{s}}^{\text{devoiced}}}}^2 + {\sigma_{\alpha_{\text{w}}^{\text{devoiced}}}}^2}{2} \sim \text{SkewNormal}(0.50, 1.84, \text{-}1.49)$$

Finally, for the $\bar{\beta}$ parameters we chose mildly regularizing priors.

$$\bar{\beta}^{\cdots} \sim \text{Normal}(0, 1)$$

The ν parameter was set to 7.[110]

$$\nu_{\dots} = 7$$

[110] See Appendix A.3.1 and Figure A.3 (p. 253) for illustration.

Stan code Listed next is the *Stan* code for the model.

```
data {
  int<lower=0> N;
  int K;
  int y[N];
  int<lower=0> subj[N];
  int<lower=0> word[N];
  int<lower=0> n_subj;
  int<lower=0> n_word;
  real far_z[N];
  real c_V_vC[N];
  real c_vC_vlC[N];
  real c_vlC_G[N];
  real c_G_P[N];
  real c_dg_vz[N];
  real c_d_g[N];
  real c_v_z[N];
  int n_varcoef_subj;
  int n_varcoef_word;
```

```
  int n_cor_subj;
  int n_cor_word;
  int set_nu;
}
parameters {
  row_vector[K-1] a;
  row_vector[K-1] b_far;
  row_vector[K-1] b_V_vC;
  row_vector[K-1] b_vC_vlC;
  row_vector[K-1] b_vlC_G;
  row_vector[K-1] b_G_P;
  row_vector[K-1] b_dg_vz;
  row_vector[K-1] b_d_g;
  row_vector[K-1] b_v_z;
  vector<lower=0>[n_varcoef_subj] sigma_subj;
  vector<lower=0>[n_varcoef_word] sigma_word;
  matrix[n_varcoef_subj, n_subj] z_subj;
  matrix[n_varcoef_word, n_word] z_word;
  cholesky_factor_corr[n_varcoef_subj] L_Rho_subj;
  cholesky_factor_corr[n_varcoef_word] L_Rho_word;
  real<lower=0> udf_word;
  real<lower=0> udf_subj;
}
transformed parameters {
  matrix[n_subj, n_varcoef_subj] r_s;
  matrix[n_word, n_varcoef_word] r_w;
  matrix[N, K] mu;
  r_s = sqrt(set_nu * udf_subj) * (diag_pre_multiply(sigma_subj, L_Rho_subj) * z_subj)';
  r_w = sqrt(set_nu * udf_word) * (diag_pre_multiply(sigma_word, L_Rho_word) * z_word)';

  for (n in 1:N) {
    mu[n,1] = a[1] +        r_s[subj[n], 1] + r_w[word[n], 1]
           + (b_far[1]                      + r_w[word[n], 2]) * far_z[n]
           + (b_dg_vz[1]  + r_s[subj[n], 2]) * c_dg_vz[n]
           + (b_d_g[1]    + r_s[subj[n], 3]) * c_d_g[n]
           + (b_v_z[1]    + r_s[subj[n], 4]) * c_v_z[n]
           + (b_V_vC[1] +   r_s[subj[n], 5]) * c_V_vC[n]
           + (b_vC_vlC[1] + r_s[subj[n], 6]) * c_vC_vlC[n]
           + (b_vlC_G[1] +  r_s[subj[n], 7]) * c_vlC_G[n]
           + (b_G_P[1] +    r_s[subj[n], 8]) * c_G_P[n];

    mu[n,2] = a[2] +        r_s[subj[n], 9] + r_w[word[n], 3]
           + (b_far[2]                      + r_w[word[n], 4]) * far_z[n]
           + (b_dg_vz[2]  + r_s[subj[n], 10]) * c_dg_vz[n]
           + (b_d_g[2]    + r_s[subj[n], 11]) * c_d_g[n]
           + (b_v_z[2]    + r_s[subj[n], 12]) * c_v_z[n]
           + (b_V_vC[2] +   r_s[subj[n], 13]) * c_V_vC[n]
           + (b_vC_vlC[2] + r_s[subj[n], 14]) * c_vC_vlC[n]
           + (b_vlC_G[2] +  r_s[subj[n], 15]) * c_vlC_G[n]
           + (b_G_P[2] +    r_s[subj[n], 16]) * c_G_P[n];

    mu[n,3] = 0;
  }
}
model {
  L_Rho_subj  ~ lkj_corr_cholesky(4);
  L_Rho_word  ~ lkj_corr_cholesky;
  sigma_subj ~ exponential(2);
  sigma_word ~ exponential(2);
  to_vector(z_subj)  ~ normal(0,1);
  to_vector(z_word)  ~ normal(0,1);
  (a[1] + ((sigma_subj[1]^2 + sigma_word[1]^2)/2)) ~ skew_normal( 0.40 , 2.48, -2.25 );
  (a[2] + ((sigma_subj[9]^2 + sigma_word[3]^2)/2)) ~ skew_normal( 0.50 , 1.84, -1.49 );

  to_vector(b_far)  ~ normal( 0 , 2 );
  to_vector(b_V_vC) ~ normal( 0 , 2 );
  to_vector(b_vC_vlC) ~ normal( 0 , 2 );
  to_vector(b_vlC_G) ~ normal( 0 , 2 );
  to_vector(b_G_P) ~ normal( 0 , 2 );
  to_vector(b_dg_vz) ~ normal( 0 , 2 );
  to_vector(b_d_g) ~ normal( 0 , 2 );
  to_vector(b_v_z) ~ normal( 0 , 2 );

  udf_word ~ inv_chi_square(set_nu);
  udf_subj ~ inv_chi_square(set_nu);

  for(n in 1:N) y[n] ~ categorical_logit(mu[n]');
}
generated quantities{
  matrix[n_varcoef_subj, n_varcoef_subj] Rho_subj;
  matrix[n_varcoef_word, n_varcoef_word] Rho_word;
  int pos_subj = 1;
```

```
  int pos_word = 1;
  vector<lower=-1, upper=1>[n_cor_subj] cor_subj;
  vector<lower=-1, upper=1>[n_cor_word] cor_word;
  Rho_subj = multiply_lower_tri_self_transpose(L_Rho_subj);
  Rho_word = multiply_lower_tri_self_transpose(L_Rho_word);
  for(i in 1:(n_varcoef_subj-1)){
    for(j in (i+1):n_varcoef_subj){
      cor_subj[pos_subj] = Rho_subj[i, j];
      pos_subj += 1;
    }
  }
  for(i in 1:(n_varcoef_word-1)){
    for(j in (i+1):n_varcoef_word){
      cor_word[pos_word] = Rho_word[i, j];
      pos_word += 1;
    }
  }
}
```

Convergence diagnostics The diagnostics shown in Figure A.23 indicate convergence of the algorithm: All $\hat{R}$ values (left) are below 1.01, the profile of the rank plot (middle) is flat, and the quantile plot (right) suggests sufficiently large effective sample sizes for all key parameters.

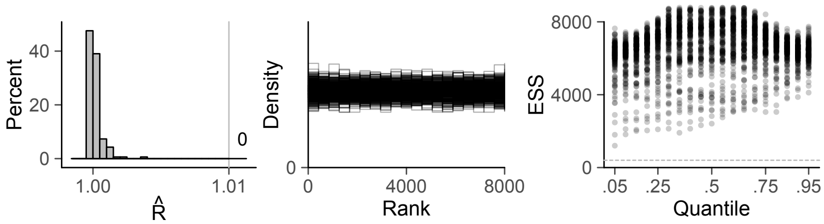


Figure A.23: FVO model for the data from Pascoe (1987): Convergence diagnostics.

Posterior distribution of model parameters The model output is summarized in Tables A.28 (varying effects parameters for word), A.29 (fixed effects parameters), A.30 (varying effects for subject: standard deviation parameters), and A.31 (varying effects for subject: correlation parameters).

Parameter	Mean	SD	.05	.50	.95	N_{eff}	$\hat{R}$
$\sigma_{\alpha_{\text{w}}^{\text{voiced}}}$	0.68	0.25	0.32	0.66	1.13	3790	1
$\sigma_{\beta_{\text{FAR, w}}^{\text{voiced}}}$	0.54	0.29	0.13	0.50	1.07	1950	1
$\sigma_{\alpha_{\text{w}}^{\text{devoiced}}}$	1.26	0.38	0.69	1.23	1.95	4862	1
$\sigma_{\beta_{\text{FAR, s}}^{\text{devoiced}}}$	0.59	0.32	0.10	0.56	1.15	2145	1
$\rho(\alpha_{\text{w}}^{\text{voiced}}, \beta_{\text{FAR, w}}^{\text{voiced}})$	–0.02	0.26	–0.45	–0.02	0.42	4883	1
$\rho(\alpha_{\text{w}}^{\text{voiced}}, \alpha_{\text{w}}^{\text{devoiced}})$	0.25	0.24	–0.17	0.26	0.62	2509	1
$\rho(\alpha_{\text{w}}^{\text{voiced}}, \beta_{\text{FAR, w}}^{\text{devoiced}})$	–0.10	0.28	–0.56	–0.11	0.37	4696	1
$\rho(\beta_{\text{FAR, w}}^{\text{voiced}}, \alpha_{\text{w}}^{\text{devoiced}})$	0.31	0.25	–0.14	0.33	0.69	1901	1
$\rho(\beta_{\text{FAR, w}}^{\text{voiced}}, \beta_{\text{FAR, w}}^{\text{devoiced}})$	–0.02	0.28	–0.47	–0.02	0.45	4711	1
$\rho(\alpha_{\text{w}}^{\text{devoiced}}, \beta_{\text{FAR, w}}^{\text{devoiced}})$	–0.08	0.26	–0.50	–0.08	0.36	7452	1

Table A.28: FVO model for the data from Pascoe (1987): Posterior distribution of varying effects parameters for word.

Table A.31: (Opposite page) FVO model for the data from Pascoe (1987): Correlation matrix for the by-subject varying effects parameters. The lower triangle reports posterior medians with .05 and .95 quantiles (i.e. 90 percent uncertainty intervals). Superscripts reflect the upper limit, subscripts the lower limit. The upper triangle lists Rhat statistics and the effective sample size for each correlation parameter.

Param.	α_{s}^{voiced}	$\beta_{dg/vz,\,s}^{voiced}$	$\beta_{d/g,\,s}^{voiced}$	$\beta_{v/z,\,s}^{voiced}$	$\beta_{V/vC,\,s}^{voiced}$	$\beta_{vC/vlC,\,s}^{voiced}$	$\beta_{vlC/G,\,s}^{voiced}$	$\beta_{G/P,\,s}^{voiced}$	$\alpha_{s}^{devoiced}$	$\beta_{dg/vz,\,s}^{devoiced}$	$\beta_{v/z,\,s}^{devoiced}$	$\beta_{d/g,\,s}^{devoiced}$	$\beta_{V/vC,\,s}^{devoiced}$	$\beta_{vC/vlC,\,s}^{devoiced}$	$\beta_{vlC/G,\,s}^{devoiced}$	$\beta_{G/P,\,s}^{devoiced}$
α_{s}^{voiced}		10722^{1}	10911^{1}	10108^{1}	11737^{1}	11944^{1}	11178^{1}	12219^{1}	10096^{1}	10673^{1}	11951^{1}	12032^{1}	11424^{1}	10806^{1}	11357^{1}	9956^{1}
$\beta_{dg/vz,\,s}^{voiced}$	$+.02^{+.36}_{-.31}$		10178^{1}	11033^{1}	10679^{1}	10505^{1}	9397^{1}	8215^{1}	7759^{1}	9663^{1}	10957^{1}	11055^{1}	11211^{1}	11364^{1}	9754^{1}	9406^{1}
$\beta_{d/g,\,s}^{voiced}$	$.00^{+.34}_{-.35}$	$-.01^{+.33}_{-.36}$		9508^{1}	9597^{1}	8800^{1}	9100^{1}	8855^{1}	7867^{1}	8839^{1}	9869^{1}	10424^{1}	9907^{1}	10034^{1}	9233^{1}	9426^{1}
$\beta_{v/z,\,s}^{voiced}$	$-.01^{+.34}_{-.36}$	$.00^{+.35}_{-.34}$	$.00^{+.34}_{-.34}$		9571^{1}	9814^{1}	8174^{1}	8873^{1}	7809^{1}	8391^{1}	8806^{1}	8990^{1}	8781^{1}	8293^{1}	9380^{1}	7626^{1}
$\beta_{V/vC,\,s}^{voiced}$	$+.03^{+.37}_{-.32}$	$.00^{+.35}_{-.34}$	$-.02^{+.33}_{-.35}$	$+.01^{+.35}_{-.33}$		8005^{1}	7815^{1}	7795^{1}	7032^{1}	7637^{1}	8097^{1}	9032^{1}	8940^{1}	8492^{1}	8519^{1}	8517^{1}
$\beta_{vC/vlC,\,s}^{voiced}$	$+.01^{+.35}_{-.34}$	$+.01^{+.36}_{-.33}$	$.00^{+.34}_{-.34}$	$.00^{+.34}_{-.34}$	$.00^{+.35}_{-.34}$		8334^{1}	8072^{1}	7742^{1}	7694^{1}	8636^{1}	7744^{1}	7622^{1}	8660^{1}	8012^{1}	7884^{1}
$\beta_{vlC/G,\,s}^{voiced}$	$+.02^{+.37}_{-.32}$	$+.01^{+.35}_{-.33}$	$-.02^{+.32}_{-.36}$	$+.01^{+.37}_{-.33}$	$+.01^{+.35}_{-.33}$	$.00^{+.34}_{-.35}$		7809^{1}	7391^{1}	6679^{1}	7688^{1}	7794^{1}	7601^{1}	7319^{1}	8192^{1}	7456^{1}
$\beta_{G/P,\,s}^{voiced}$	$+.01^{+.35}_{-.33}$	$+.05^{+.39}_{-.30}$	$+.02^{+.36}_{-.35}$	$+.02^{+.36}_{-.33}$	$-.01^{+.34}_{-.35}$	$-.01^{+.34}_{-.34}$	$-.01^{+.33}_{-.35}$		7055^{1}	7096^{1}	7361^{1}	7201^{1}	7105^{1}	7030^{1}	6712^{1}	6694^{1}
$\alpha_{s}^{devoiced}$	$-.02^{+.32}_{-.36}$	$+.04^{+.36}_{-.31}$	$+.04^{+.37}_{-.31}$	$+.01^{+.36}_{-.33}$	$-.02^{+.33}_{-.36}$	$-.01^{+.33}_{-.36}$	$.00^{+.33}_{-.34}$	$+.03^{+.36}_{-.31}$		7107^{1}	7979^{1}	7654^{1}	7029^{1}	7765^{1}	7417^{1}	6811^{1}
$\beta_{dg/vz,\,s}^{devoiced}$	$-.02^{+.32}_{-.37}$	$+.01^{+.35}_{-.34}$	$+.01^{+.35}_{-.33}$	$+.02^{+.36}_{-.33}$	$-.02^{+.33}_{-.36}$	$.00^{+.34}_{-.34}$	$-.01^{+.32}_{-.35}$	$.00^{+.34}_{-.34}$	$-.03^{+.31}_{-.36}$		6975^{1}	6666^{1}	6787^{1}	6897^{1}	6456^{1}	6332^{1}
$\beta_{v/z,\,s}^{devoiced}$	$+.03^{+.38}_{-.32}$	$+.01^{+.35}_{-.33}$	$-.01^{+.34}_{-.35}$	$+.01^{+.35}_{-.33}$	$.00^{+.35}_{-.34}$	$-.01^{+.34}_{-.35}$	$-.01^{+.34}_{-.35}$	$.00^{+.34}_{-.34}$	$.00^{+.34}_{-.34}$	$-.01^{+.33}_{-.36}$		5863^{1}	6094^{1}	5012^{1}	6469^{1}	6275^{1}
$\beta_{d/g,\,s}^{devoiced}$	$-.03^{+.31}_{-.38}$	$-.01^{+.34}_{-.35}$	$+.01^{+.36}_{-.33}$	$+.01^{+.34}_{-.34}$	$-.01^{+.34}_{-.35}$	$-.01^{+.35}_{-.35}$	$-.01^{+.32}_{-.35}$	$.00^{+.35}_{-.34}$	$.00^{+.33}_{-.34}$	$.00^{+.34}_{-.35}$	$-.01^{+.34}_{-.35}$		5596^{1}	5512^{1}	6033^{1}	5892^{1}
$\beta_{V/vC,\,s}^{devoiced}$	$+.02^{+.36}_{-.32}$	$-.01^{+.34}_{-.34}$	$-.01^{+.33}_{-.36}$	$-.01^{+.33}_{-.35}$	$.00^{+.35}_{-.34}$	$.00^{+.35}_{-.33}$	$-.01^{+.33}_{-.33}$	$.00^{+.34}_{-.34}$	$-.04^{+.30}_{-.38}$	$+.02^{+.36}_{-.33}$	$+.01^{+.35}_{-.34}$	$.00^{+.34}_{-.34}$		5500^{1}	5198^{1}	5889^{1}
$\beta_{vC/vlC,\,s}^{devoiced}$	$+.02^{+.36}_{-.32}$	$+.02^{+.36}_{-.33}$	$.00^{+.34}_{-.34}$	$-.01^{+.34}_{-.35}$	$-.01^{+.33}_{-.36}$	$.00^{+.34}_{-.35}$	$.00^{+.34}_{-.34}$	$+.01^{+.35}_{-.33}$	$+.01^{+.35}_{-.33}$	$.00^{+.34}_{-.35}$	$+.01^{+.35}_{-.33}$	$.00^{+.34}_{-.35}$	$.00^{+.34}_{-.34}$		5363^{1}	5694^{1}
$\beta_{vlC/G,\,s}^{devoiced}$	$+.03^{+.36}_{-.31}$	$+.02^{+.36}_{-.31}$	$.00^{+.33}_{-.34}$	$.00^{+.34}_{-.35}$	$.00^{+.34}_{-.34}$	$-.01^{+.34}_{-.35}$	$.00^{+.35}_{-.35}$	$+.02^{+.36}_{-.32}$	$+.01^{+.34}_{-.33}$	$-.02^{+.34}_{-.36}$	$+.01^{+.35}_{-.34}$	$-.02^{+.33}_{-.35}$	$.00^{+.35}_{-.34}$	$.00^{+.34}_{-.35}$		5380^{1}
$\beta_{G/P,\,s}^{devoiced}$	$+.03^{+.36}_{-.31}$	$+.01^{+.35}_{-.33}$	$.00^{+.34}_{-.33}$	$.00^{+.34}_{-.34}$	$-.01^{+.34}_{-.35}$	$-.03^{+.32}_{-.37}$	$-.02^{+.33}_{-.35}$	$+.02^{+.37}_{-.33}$	$+.02^{+.35}_{-.32}$	$-.01^{+.32}_{-.35}$	$+.01^{+.35}_{-.34}$	$-.01^{+.33}_{-.34}$	$+.01^{+.34}_{-.34}$	$+.01^{+.35}_{-.33}$	$+.01^{+.35}_{-.35}$	

Parameter	Mean	SD	Posterior quantiles .05	.50	.95	N_{eff}	$\hat{R}$
$\bar{\alpha}^{\text{voiced}}$	–0.43	0.33	–0.99	–0.44	0.12	5503	1
$\bar{\alpha}^{\text{devoiced}}$	–2.02	0.52	–2.91	–2.00	–1.21	5047	1
$\bar{\beta}_{\text{FAR}}^{\text{voiced}}$	1.48	0.26	1.09	1.46	1.93	3702	1
$\bar{\beta}_{\text{FAR}}^{\text{devoiced}}$	0.18	0.31	–0.35	0.19	0.67	5910	1
$\bar{\beta}_{\text{dg/vz}}^{\text{voiced}}$	–0.46	0.65	–1.49	–0.47	0.62	4978	1
$\bar{\beta}_{\text{dg/vz}}^{\text{devoiced}}$	–1.47	0.86	–2.88	–1.47	–0.05	5625	1
$\bar{\beta}_{\text{d/g}}^{\text{voiced}}$	–1.15	1.04	–2.86	–1.15	0.55	5775	1
$\bar{\beta}_{\text{d/g}}^{\text{devoiced}}$	0.31	1.22	–1.78	0.33	2.27	5677	1
$\bar{\beta}_{\text{v/z}}^{\text{voiced}}$	–1.59	0.65	–2.69	–1.58	–0.56	6724	1
$\bar{\beta}_{\text{v/z}}^{\text{devoiced}}$	1.63	1.06	–0.10	1.62	3.42	6978	1
$\bar{\beta}_{\text{V/vC}}^{\text{voiced}}$	–0.12	0.47	–0.89	–0.12	0.65	9393	1
$\bar{\beta}_{\text{V/vC}}^{\text{devoiced}}$	0.88	0.71	–0.27	0.86	2.08	7968	1
$\bar{\beta}_{\text{VC/vlC}}^{\text{voiced}}$	–1.30	0.35	–1.90	–1.30	–0.74	7407	1
$\bar{\beta}_{\text{VC/vlC}}^{\text{devoiced}}$	–0.94	0.51	–1.78	–0.95	–0.09	7127	1
$\bar{\beta}_{\text{vlC/G}}^{\text{voiced}}$	–0.60	0.41	–1.28	–0.61	0.05	8333	1
$\bar{\beta}_{\text{vlC/G}}^{\text{devoiced}}$	0.72	0.53	–0.15	0.72	1.60	7465	1
$\bar{\beta}_{\text{G/P}}^{\text{voiced}}$	–0.69	0.50	–1.51	–0.68	0.11	7932	1
$\bar{\beta}_{\text{G/P}}^{\text{devoiced}}$	–0.07	0.52	–0.93	–0.06	0.78	8386	1

Table A.29: FVO model for the data from Pascoe (1987): Posterior distribution of fixed effects parameters.

Parameter	Mean	SD	Posterior quantiles .05	.50	.95	N_{eff}	$\hat{R}$
$\sigma_{\alpha_s^{\text{voiced}}}$	0.39	0.24	0.05	0.37	0.82	2628	1
$\sigma_{\beta_{\text{dg/vz, s}}^{\text{voiced}}}$	0.42	0.33	0.03	0.35	1.05	4495	1
$\sigma_{\beta_{\text{d/g, s}}^{\text{voiced}}}$	0.50	0.43	0.03	0.40	1.35	3118	1
$\sigma_{\beta_{\text{v/z, s}}^{\text{voiced}}}$	0.45	0.41	0.03	0.35	1.28	4723	1
$\sigma_{\beta_{\text{V/vC, s}}^{\text{voiced}}}$	0.42	0.37	0.02	0.33	1.14	5191	1
$\sigma_{\beta_{\text{vC/vlC, s}}^{\text{voiced}}}$	0.32	0.27	0.02	0.25	0.86	5280	1
$\sigma_{\beta_{\text{vlC/G, s}}^{\text{voiced}}}$	0.47	0.37	0.03	0.38	1.18	3484	1
$\sigma_{\beta_{\text{G/P, s}}^{\text{voiced}}}$	0.56	0.45	0.03	0.45	1.44	3640	1
$\sigma_{\alpha_s^{\text{devoiced}}}$	0.53	0.30	0.08	0.51	1.07	3096	1
$\sigma_{\beta_{\text{dg/vz, s}}^{\text{devoiced}}}$	0.66	0.49	0.04	0.58	1.56	3012	1
$\sigma_{\beta_{\text{v/z, s}}^{\text{devoiced}}}$	0.39	0.35	0.02	0.29	1.09	6182	1
$\sigma_{\beta_{\text{d/g, s}}^{\text{devoiced}}}$	0.51	0.45	0.03	0.38	1.43	5633	1
$\sigma_{\beta_{\text{V/vC, s}}^{\text{devoiced}}}$	0.49	0.45	0.03	0.35	1.38	5361	1
$\sigma_{\beta_{\text{vC/vlC, s}}^{\text{devoiced}}}$	0.34	0.30	0.02	0.25	0.94	6141	1
$\sigma_{\beta_{\text{vlC/G, s}}^{\text{devoiced}}}$	0.37	0.33	0.02	0.28	1.02	5510	1
$\sigma_{\beta_{\text{G/P, s}}^{\text{devoiced}}}$	0.57	0.48	0.03	0.45	1.51	3417	1

Table A.30: FVO model for the data from Pascoe (1987): Posterior distribution of varying effects standard deviation parameters for subject.

B
Bibliography

Adank, Patti, Roel Smith & Roeland van Hout. 2004. A comparison of vowel normalization procedures for language variation research. *Journal of the Acoustical Society of America* 116(5). 3099–3107.

Agresti, Alan. 2013. *Categorical data analysis*. Hoboken, NJ: John Wiley & Sons.

Amayreh, Mousa M. & Alice T. Dyson. 1998. The acquisition of Arabic consonants. *Journal of Speech, Language and Hearing Research* 41. 642–653.

Anderson, Janet I. 1987. The markedness differential hypothesis and syllable structure difficulty. In Georgette Ioup & Steven Weinberger (eds.), *Interlanguage phonology: The acquisition of a second language sound system*, 279–291. New York: Harper and Row.

Ankerstein, Carrie A. & Raphael Morschett. 2013. Do you hear what I hear? A comparison of phoneme perception in native and Saarlandian German nonnative speakers of English. *Saarland Working Papers in Linguistics* 4. 1–8.

Archibald, John. 1994. A formal model of learning L2 prosodic phonology. *Second Language Research* 10. 215–240.

Archibald, John. 1998. *Second language phonology*. Amsterdam: John Benjamins.

Arnold, Roland & Klaus Hansen. 1974. *Englische Phonetik*. Leipzig: VEB Verlag Enzyklopädie Leipzig.

Assmann, Peter F. & William F. Katz. 2000. Time-varying spectral change in the vowels of children and adults. *Journal of the Acoustical Society of America* 108. 1856–1866.

Baayen, R. Harald, Douglas J. Davidson & Douglas M. Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59(4). 390–412.

Bailey, Charles-James N. 1973. The patterning of language variation. In Richard W. Bailey & Jay L. Robinson (eds.), *Varieties of present-day English*, 156–189. New York: Macmillan.

Barlow, Jessica A. 2014. Age of acquisition and allophony in Spanish-English bilinguals. *Frontiers in Psychology* 5(288).

Barr, Dale J., Roger Levy, Christoph Scheepers & Harry Tily. 2013. Random effects structure in confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68(3). 225–278.

Barry, William J. 1977. *Ausspracheabweichungen: Eine experimentalphonetische Untersuchung an deutschen Englischlernenden*, vol. 7. Kiel: University of Kiel Arbeitsberichte des Instituts für Phonetik der Universität Kiel edn.

Barry, William J. & Christoph Gutknecht. 1973. *Practical English phonetics*. Hamburg: Buske.

Bauer, Laurie. 1984. Linking /r/ in RP: Some facts. *Journal of the International Phonetic Association* 14(2). 74–79.

Bauer, Laurie. 1985. Tracing phonetic change in the Received Pronunciation of British English. *Journal of Phonetics* 13. 61–81.

Baumotte, Henrike. 2009. Segmental factors in language proficiency: Coarticulatory resistance as a signature of talent. In Grzegorz Dogil & Susanne M. Reiterer (eds.), *Language talent (and brain activity)*, 279–304. Berlin: De Gruyter.

Baumotte, Henrike. 2011. *Segmental factors in language proficiency: Degree of velarization, coarticulatory resistance and vowel formant frequency distribution as a signature of talent*: University of Stuttgart dissertation.

Bayley, Robert. 1996. Competing constraints on variation in the speech of Chinese learners of English. In Robert Bayley & Dennis R. Preston (eds.), *Second language acquisition and linguistic variation*, 97–120. Amsterdam: John Benjamins.

Becker, Thomas. 1998. *Das Vokalsystem der deutschen Standardsprache*. Frankfurt: Peter Lang.

Benware, Wilbur A. 1986. *Phonetics and phonology of Modern German*. Washington: Georgetown University Press.

Best, Catherine T. 1995. A direct realist view of cross-language speech perception. In Winifred Strange (ed.), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research*, 171–206.

Baltimore: York Press.

Best, Catherine T. & Winifred Strange. 1992. Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics* 20. 305–330.

Best, Catherine T. & Michael D. Tyler. 2007. Nonnative and second-language speech perception: Commonalities and complementarities. In Murray J. Munro & Ocke-Schwen Bohn (eds.), *Language experience in second language speech learning*, 13–34. Amsterdam: Benjamins.

Bickerton, Derek. 1971. Inherent variability and variable rules. *Foundations of Language* 7. 457–492.

Bickerton, Derek. 1975. *Dynamics of a creole system*. New York: Cambridge University Press.

Black, John W. 1949. Natural frequency, duration, and intensity of vowels in reading. *Journal of Speech and Hearing Disorders* 14. 216–221.

Bladon, R. Anthony W. & Ameen Al-Bamerni. 1976. Coarticulation resistance in English /l/. *Journal of Phonetics* 4. 137–150.

Blevins, Juliette. 2004. *Evolutionary phonology: The emergence of sound patterns*. Cambridge: Cambridge University Press.

Blevins, Juliette. 2006. New perspectives on English sound patterns: "Natural" and "unnatural" in evolutionary phonology. *Journal of English Linguistics* 34(1). 6–25.

Boersma, Paul. 1993. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Proceedings of the Institute of Phonetic Sciences, University of Amsterdam* 17. 97–110.

Boersma, Paul. 1998. *Functional phonology*. The Hague: Holland Academic Graphics.

Boersma, Paul & David Weenink. 2019. Praat: Doing phonetics by computer. URL http://www.praat.org/.

Bohn, Ocke-Schwen. 1995. Cross-language speech perception in adults: First language transfer doesn't tell it all. In Winifred Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research*, 273–304. Baltimore: York Press.

Bohn, Ocke-Schwen. 2002. On phonetic similarity. In Petra Burmeister, Thorsten Piske & Andreas Rohde (eds.), *An integrated view of language development: Papers in honour of Henning Wode*, 191–216. Trier: Wissenschaftlicher Verlag Trier.

Bohn, Ocke-Schwen & Catherine T. Best. 2012. Native-language phonetic and phonological influences on perception of American English approximants by Danish and German listeners. *Journal of Phonetics* 40(1). 109–128.

Bohn, Ocke-Schwen & James E. Flege. 1990. Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics* 11. 303–328.

Bohn, Ocke-Schwen & James E. Flege. 1992. The production of new and similar vowels by adult German learners of English. *Studies in Second Language Acquisition* 14. 131–158.

Boyce, Suzanne & Carol Y. Espy-Wilson. 1997. Coarticulatory stability in American English /r/. *Journal of the Acoustical Society of America* 101(6). 3741–3753.

Bradlow, Ann R. 1995. A comparative acoustic study of English and Spanish vowels. *Journal of the Acoustical Society of America* 97(3). 1916–1924.

Brannen, Kathleen J. 2011. *The perception and production of interdental fircatives in second language acquisition*: McGill University, Montréal dissertation.

Briere, Eugène J. 1966. An investigation of phonological interference. *Language* 42. 768–96.

Broersma, Mirjam. 2005. Perception of final fricative voicing: Native and non-native listeners' use of vowel duration. *Journal of the Acoustical Society of America* 127(3). 1636–1644.

Broselow, Ellen & Yoonjung Kang. 2013. Phonology and speech. In Julia Herschensohn & Martha Young-Scholten (eds.), *The Cmbridge handbook of second language acquisition*, 529–553. Cambridge: Cambridge University Press.

Browman, Catherine P. & Louis Goldstein. 1992. 'Targetless' schwa: An articulatory analysis. In Gerald J. Docherty & D. Robert Ladd (eds.), *Papers in laboratory phonology II*, 26–56. Cambridge: Cambridge University Press.

Brown, Cynthia A. 1998. The role of the L1 grammar in the acquisition of segmental structure. *Second Language Research* 14. 139–193.

Buder, Eugene H. & Carol Stoel-Gammon. 2002. American and Swedish children's acquisition of vowel duration: Effects of vowel identity and final stop voicing. *Journal of the Acoustical Society of America* 111(4). 1854–1864.

Bürkner, Paul-Christian. 2017. brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software* 80(1). 1–28.

Burnham, Kenneth P. & David R. Anderson. 2002. *Model selection and multimodel inference: A practical information-theoretic approach*. New York: Springer.

Butcher, Andrew. 1976. *The influence of the native language on the perception of vowel quality* Arbeitsberichte des Instituts für Phonetik der Universität Kiel. Kiel: University of Kiel.

Carlisle, Robert S. 1991. The influence of environment on vowel epenthesis in Spanish/English interphonology. *Applied Linguistics* 12. 76–95.

Carpenter, Bob, Andrew Gelman, Matthew D Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li & Allen Riddell. 2017. Stan: A probabilistic programming language. *Journal of statistical software* 76(1).

Carter, Paul. 1999. Abstractness in phonology and extrinsic phonetic interpretation: The case of liquids in English. In John J. Ohala, Yoko Hasegawa, Manjari Ohala, D. Granville & A. C. Bailey (eds.), *Proceedings of the 14th international congress of the phonetic sciences*, 105–108. New York: American Institute of Physics Press.

Carter, Paul & John Local. 2007. F2 variation in Newcastle and Leeds English liquid systems. *Journal of the International Phonetic Association* 37. 183–199.

Cebrian, Juli. 2006. Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics* 34. 372–387.

Chen, Matthew. 1970. Vowel length variation as a function of the voicing of the consonant environment. *Phonetica* 22. 129–159.

Chen, Matthew Y. & William S.-Y. Wang. 1975. Sound change: Actuation and implementation. *Language* 51. 255–281.

Childers, Donald G. & Ke Wu. 1991. Gender recognition from speech. Part II: Fine analysis. *Journal of the Acoustical Society of America* 90. 1841–1856.

Chomsky, Noam. 1959. *Syntactic structures*. The Hague: Mouton.

Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge: MIT Press.

Chomsky, Noam. 1981. *Lectures on government and binding*. Dordrecht: Foris.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.

Clark, John, Colin Yallop & Janet Fletcher. 2007. *An introduction to phonetics and phonology*. Oxford: Blackwell.

Claßen, Kathrin, Grzegorz Dogil, Michael Jessen, Krzysztof Marasek & Wolfgang Wokurek. 1998. Stimmqualität und Wortbetonung im Deutschen. *Linguistische Berichte* 174. 202–245.

Clements, G. Nick. 1999. Affricates as noncontoured stops. In Osamu Fujimura, Brian D. Joseph & Bohumil Palek (eds.), *Item order in language and speech*, 271–299. Prag: The Karolinum Press.

Cleveland, William S. & Robert McGill. 1984. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association* 79. 531–554.

Colantoni, Laura & Jeffrey Steele. 2008. Integrating articulatory constraints into models of second language phonological acquisition. *Applied Psycholinguistics* 29(3). 489–534.

Colantoni, Laura, Jeffrey Steele & Paola Escudero. 2015. *Second language speech: Theory and practice*. Cambridge: Cambridge University Press.

Collins, Beverley & Inger M. Mees. 2013. *Practical phonetics and phonology: A resource book for students*. New York: Routledge.

Comrie, Bernard. 1989. *Language universals and linguistic typology*. Chicago: The University of Chicago Press.

Corder, Stephen P. 1967. The significance of learners' errors. *International Review of Applied Linguistics in Language Teaching* 5. 161–169.

Corder, Stephen P. 1981. *Error analysis and interlanguage*. Oxford: Oxford University Press.

Cox, David R. & Christl A. Donnelly. 2011. *Principles of applied statistics*. Cambridge: Cambridge University Press.

Crowther, Court S. & Virginia Mann. 1992. Native language factors affecting use of vocalic cues to final consonant voicing in English. *Journal of the Acoustical Society of America* 92(2). 711–722.

Cruttenden, Alan. 2014. *Gimson's pronunciation of English*. New York: Routledge.

Cutler, Anne, Andrea Weber, Roel Smits & Nicole Cooper. 2004. Patterns of English phoneme confusions by native and non-native listeners. *Journal of the Acoustical Society of America* 116(6). 3668–3678.

Dalston, Rodger M. 1975. Acoustic characteristics of English /w,r,l/ spoken correctly by young children and adults. *Journal of the Acoustical Society of America* 57. 462–469.

Davenport, Mike & S. J. Hannahs. 2005. *Introducing phonetics and phonology*. London: Hodder Arnold.

Davies, Mark. 2008-. The Corpus of Contemporary American English (COCA): 600 million words, 1990-present. URL https://www.english-corpora.org/coca/.

De Bot, Kees & Diane Larsen-Freeman. 2011. Research in second language development from a dynamic systems theory perspective. In Marjolijn Verspoor, Kees De Bot & Wander Lowie (eds.), *A dynamic approach to second language development*, 5–23. Amsterdam: John Benjamins.

De Bot, Kees, Wander Lowie & Marjolijn Verspoor. 2007. A dynamic systems theory approach to second language acquisition. *Bilingualism: Language and Cognition* 10(1). 7–21.

Dehaene, Stanislas. 2003. The neural basis of the Weber-Fechner law: A logarithmic mental number line. *Trends in cognitive sciences* .

Delattre, Pierre C. 1965. *Comparing the phonetics features of English, German, Spanish and French*. Heidelberg: Groos.

Delattre, Pierre C. & Donald C. Freeman. 1968. A dialect study of American r's by x-ray motion picture. *Linguistics* 44. 29–68.

Denes, Peter B. 1955. Effect of duration on the perception of voicing. *Journal of the Acoustical Society of America* 27(4). 761–764.

Deterding, David H. 1990. *Speaker normalization for automatic speech recognition*: Cambridge University dissertation.

Deterding, David H. 1997. The formants of monophthong vowels in standard southern British English pronunciation. *Journal of the International Phonetic Association* 27. 47–55.

Di Paolo, Marianna, Malcah Yaeger-Dror & Alicia Beckford Wassink. 2011. Analyzing vowels. In Marianna Di Paolo & Malcah Yaeger-Dror (eds.), *Sociophonetics: A student's guide*, 87–106. New York: Routledge.

Dimitrieva, Olga, Allard Jongman & Joan Sereno. 2010. Phonetic neutralization by native and non-native speakers: The case of Russian final devoicing. *Journal of Phonetics* 38. 483–492.

Dinnsen, Daniel A. & Fred R. Eckman. 1975. A functional explanation of some phonological typologies. In Robin E. Grossman, L. James San & Timothy J. Vance (eds.), *Functionalism*, 126–139. Chicago: Chicago Linguistics Society.

Dodd, Barbara, Alison Holm, Zhu Hua & Sharon Crosbie. 2003. Phonological development: A normative study of British English-speaking children. *Clinical Linguistics and Phonetics* 17(8). 617–643.

van Dommelen, Wim A. 1982. A contrastive investigation of vowel duration in German and Dutch. *Phonetica* 39. 29–35.

Donegan, Patricia. 1978. *On the natural phonology of vowels*: Ohio State University dissertation.

Dresher, B. Elan & Jonathan Kaye. 1990. A computational learning model for metrical phonology. *Cognition* 34. 137–195.

Dretzke, Burkhard. 1998. *Modern British and American English pronunciation*. Stuttgart: UTB.

Dziubalska-Kołaczyk, Katarzyna. 1990a. A natural model of acquisition of second language phonology and the notion of relative markedness. In Julián M. Dosuna & Carmen Pensado (eds.), *Naturalists at Krems*, 43–51. Salamanca: Ediciones Universidad Salamanca.

Dziubalska-Kołaczyk, Katarzyna. 1990b. *A theory of second language acquisition within the framework of natural phonology*. Poznan: AMU Press.

Eckert, Hartwig & William Barry. 2005. *The phonetics and phonology of English pronunciation*. Trier: Wissenschaftlicher Verlag Trier.

Eckman, Fred R. 1977. Markedness and the contrastive analysis hypothesis. *Language Learning* 27. 315–330.

Eckman, Fred R. 1981. On predicting phonological difficulty in second language acquisition. *Studies in Second Language Acquisition* 4. 18–30.

Eckman, Fred R. 1991. The structural conformity hypothesis and the acquisition of consonant clusters in the interlanguage of ESL learners. *Studies in Second Language Acquisition* 13. 23–41.

Eckman, Fred R. 2008. Typological markedness and second language acquisition. In Jette G. Hansen Edwards & Mary L. Zampini (eds.), *Phonology and second language acquisition*, 95–115. Amsterdam: John Benjamins.

Eckman, Fred R. 2012. Second language phonology. In Susan M. Gass & Alison Mackey (eds.), *The Routledge handbook of second language acquisition*, 91–105. London: Routledge.

Eckman, Fred R., Abdullah Elreyes & Gregory K. Iverson. 2003. Some principles on second language phonology. *Second Language Research* 19. 169–208.

Edge, Beverly A. 1991. The production of word-final voiced obstruents in English by L1 speakers of Japanese and Cantonese. *Studies in Second Language Acquisition* 13. 377–393.

Elert, Claes C. 1964. *Phonological studies of quantity in Swedish*. Uppsala: Almqvist & Wiksells.

Elliott, A. Raymond. 1995. Foreign language phonology: Field dependence, attitude, and success of formal instruction in Spanish pronunciation. *Modern Language Journal* 79. 530–542.

Elliott, A. Raymond. 1997. On the teaching and acquisition of pronunciation with a communicative approach. *Hispania* 80. 96–108.

Elsen, Hilke. 1991. *Erstspracherwerb: Der Erwerb des deutschen Lautsystems*. Wiesbaden: DUV.

Elsendoorn, Ben A. G. 1984. *Tolerances of durational properties in British English vowels*: Utrecht University dissertation.

Engstrand, Olle, Johan Frid & Björn Lindblom. 2007. A perceptual bridge between coronal and dorsal /r/. In Maria-Josep Sole, Patrice Speeter Beddor & Manjari Ohala (eds.), *Experimental approaches to phonology*, 175–191.

Oxford: Oxford University Press.

Escudero, Paola & Paul Boersma. 2004. Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition* 26. 551–585.

Esposito, Anna. 2002. On vowel height and consonantal voicing effects: Data from Italian. *Phonetica* 59. 197–231.

Espy-Wilson, Carol Y. 1992. Acoustic measures for linguistic features distinguishing the semivowels /w j r l/ in American English. *Journal of the Acoustical Society of America* 92(2). 736–757.

Fasold, Ralph W. & Dennis R. Preston. 2007. The psycholinguistic unity of inherent variability: Old Occam whips out his razor. In Robert Bayley & Ceil Lucas (eds.), *Sociolinguistic variation: Theories, methods, and applications*, 45–69. Cambridge: Cambridge University Press.

Faure, Georges. 1972. Analyse acoustique de deux allophones du l final anglais. In Albert Valdman (ed.), *Papers in linguistics and phonetics to the memory of Pierre Delattre*, 117–127. The Hague: Mouton.

Ferragne, Emmanuel & François Pellegrino. 2010. Formant frequencies of vowels in 13 accents of the British Isles. *Journal of the International Phonetic Association* 40(1). 1–34.

Fintoft, Knut. 1962. The duration of some Norwegian speech sounds. *Phonetica* 7. 19–39.

Flege, James E. 1987. The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics* 15. 47–65.

Flege, James E. 1995. Second language speech learning: Theory, findings, and problems. In W Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research*, 233–277. Baltimore: York Press.

Flege, James E., Ocke-Schwen Bohn & Sunyoung Jang. 1997. Effects of experience on nonnative speakers' production and perception of English vowels. *Journal of Phonetics* 25. 437–470.

Flege, James E. & Robert Port. 1981. Cross-language phonetic interference. *Language and Speech* 24(2). 125–146.

Flege, James E., Naoyuki Takagi & Virginia Mann. 1995. Japanese adults can learn to produce English /ɹ/ and /l/ accurately. *Language and Speech* 38(1). 25–55.

Flynn, Nicholas. 2011. Comparing vowel formant normalization procedures. *York Papers in Linguistics* (*Series 2*) 11. 1–28.

Folkes, Paul & Gerard J. Docherty. 2000. Another chapter in the story of /r/: 'Labiodental' variants in in British English. *Journal of Sociolinguistics* 4. 30–59.

Fromkin, Victoria & Robert Rodman. 1978. *An introduction to language*. New York: Holt McDougal.

Garnica, Olga K. & Robert K. Herbert. 1979. Some phonological errors in second language learning: Interference doesn't tell it all. *International Journal of Psycholinguistics* 6. 5–19.

Gatbonton, Elizabeth. 1978. Patterned phonetic variability in second-language speech: A gradual diffusion model. *Canadian Modern Language Review* 34. 335–347.

Gelman, Andrew, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari & Donald B. Rubin. 2014. *Bayesian data analysis*. Boca Raton: CRC Press.

Gick, Bryan, Penelope Bacsfalvi, Barbara M. Bernhardt, Sunyoung Oh, Slade Stolar & Ian Wilson. 2008. A motor differentiation model for liquid substitutions in children's speech. *Proceedings of Meetings on Acoustics* 1. 1–9.

Giegerich, Heinz J. 1992. *English phonology*. Cambridge: Cambridge University Press.

Gonet, Wiktor & Lidia Stadnicka. 2006. Vowel clipping in English. *Speech and Language Technology* 8. 77–86.

Green, Rebekah J. 2013. *Vowel lengthening in L2: A strategy for some coda devoicers?* California State University, Long Beach MA thesis.

Greenberg, Joseph H. 1970. Some generalizations concerning glottalic consonants, especially implosives. *International Journal of American Linguistics* 36. 123–145.

Greenberg, Joseph H. 1976. *Language universals*. The Hague: Mouton.

Gusshoven, Carlos. 1992. Dutch. *Journal of the International Phonetic Association* 22. 45–47.

Gut, Ulrike. 2009. *Non-native speech: A corpus-based analysis of phonological and phonetic properties of L2 English and German*. Frankfurt: Peter Lang.

Hall, T. Alan. 1993. The phonology of German /ʀ/. *Phonology* 10. 83–105.

Hall, T. Alan. 2011. *Phonologie: Eine Einführung*. Berlin: De Gruyter.

Hamann, Silke, Paul Boersma & Małgorzata Cavar. 2010. Language-specific differences in the weighting of perceptual cues for labiodentals. In Katarzyna Dziubalska-Kołaczyk, Magdalena Wrembel & Małgorzata Kul (eds.), *Proceedings of the 6th International Symposium on the Acquisition of Second Language Speech* (*New Sounds 2010*), 173–178. Poznan: Adam Mickiewicz University.

Hamann, Silke & Anke Sennema. 2005a. Acoustic differences between German and Dutch labiodentals. *ZAS Papers* *in Linguistics* 42. 33–41.

Hamann, Silke & Anke Sennema. 2005b. Voiced labiodental fricatives or glides – all the same to Germans? In Valerie Hazan & Paul Iverson (eds.), *Proceedings of the Conference on Plasticity in Speech Processing*, 164–167. London: University College London.

Hammerly, Hector. 1982. Contrastive phonology and error analysis. *International Review of Applied Linguistics* 20. 17–32.

Hancin-Bhatt, Barbara & Anita Govindjee. 1999. A computational model of feature competition in L2 phonology. In Peter Broeder & Jaap Murre (eds.), *Language and thought in development: Crosslinguistic studies*, 145–162. Tübingen: Narr.

Hancin-Bhatt, Barbara J. 1994. Segment transfer: A consequence of a dynamic system. *Second Language Research* 10. 241–269.

Hannisdal, Bente R. 2006. *Variability and change in Received Pronunciation: A study of six phonological variables in the speech of television newsreaders*: University of Bergen dissertation.

Hansen Edwards, Jette G. 2001. Linguistic constraints on the acquisition of English syllable codas by native speakers of Mandarin Chinese. *Applied Linguistics* 22. 338–365.

Hanulíková, Adriana & Andrea Weber. 2010. Production of English interdental fricatives by Dutch, German and English speakers. In Katarzyna Dziubalska-Kołaczyk, Magdalena Wrembel & Małgorzata Kul (eds.), *New sounds: Proceedings of the 6th International Symposium on the Acquisition of Second Language Speech*, 173–178. Poznan: Adam Mickiewicz University.

Hanulíková, Adriana & Andrea Weber. 2012. Sink positive: Linguistic experience with th substitutions influences nonnative word recognition. *Attention, Perception and Psychophysics* 74. 613–629.

Hardy, Jane E. 1993. Phonological learning and retention in second language acquisition. In Fred R. Eckman (ed.), *Confluence: Linguistics, L2 acquisition, speech pathology*, 235–248. Amsterdam: John Benjamins.

Harrington, Johnathan & Steve Cassidy. 1999. *Techniques in speech acoustics*. Dordrecht: Kluwer.

Harrington, Johnathan, Sallyanne Palethorpe & Catherine Watson. 2000. Monophthongal changes in the vowels of Received Pronunciation: An acoustic analysis of the Queen's Christmas broadcasts. *Journal of the International Phonetic Association* 30. 63–78.

Haspelmath, Martin. 2006. Against markedness (and what to replace it with). *Journal of Linguistics* 42(1). 25–70.

Hawkins, Sarah & Johnathan Midgley. 2005. Formant frequencies of RP monophthongs in four age groups of speakers. *Journal of the International Phonetic Association* 35(2). 183–199.

Hawkins, Sarah & Noël Nguyen. 2004. Influence of syllable-coda voicing on the acoustic properties of syllable-onset /l/ in English. *Journal of Phonetics* 32(2). 199–231.

Hay, Jennifer, Paul Warren & Katie Drager. 2006. Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics* 34(4). 458–484.

Hellyer-Jones, Rosemary, Marion Horner & Robert Parr. 2003. *Green Line New 1: Ausgabe für Bayern*. Stuttgart: Ernst Klett.

Henton, Caroline G. 1983. Changes in the vowels of Received Pronunciation. *Journal of Phonetics* 11. 353–371.

Hillenbrand, James M., Michael J. Clark & Terrance M. Nearey. 2001. Effects of consonant environment on vowel formant patterns. *Journal of the Acoustical Society of America* 109. 748–763.

Hillenbrand, James M., Laura A. Getty, Michael J. Clark & Kimberlee Wheeler. 1995. Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America* 97. 3099–3111.

Holm, John. 2000. *An introduction to pidgins and creoles*. Cambridge: Cambridge University Press.

Honikman, Beatrice. 1964. Articulatory settings. In David Abercrombie, Dennis B. Fry, P. A. D. MacCarthy, N. C. Scott & J. L. M. Trim (eds.), *In honour of Daniel Jones: Papers contributed on the occasion of his eightieth birthday, 12 September 1961*, 73–84. London: Longmans, Green & Co.

Hoole, Philip. 1999. Laryngeal coarticulation. Section A: Coarticulatory investigations of the devoicing gesture. In W. H. Hardcastle & N. Newlett (eds.), *Coarticulation: Theory, data and techniques*, 105–121. Cambridge: Cambridge University Press.

House, Arthur S. 1961. On vowel duration in English. *Journal of the Acoustical Society of America* 33. 1174–1178.

Huffman, Marie K. 1997. Phonetic variation in intervocalic onset /l/'s in English. *Journal of Phonetics* 25. 115–141.

Hughes, Arthur, Peter Trudgill & Dominic Watt. 2005. *English accents and dialects*. New York: Routledge.

Hume, Elizabeth. 2004. Deconstructing markedness: A predictability-based approach. In *Proceedings of the Annual Meeting of the Berkeley Linguistics Society 13*, 182–198.

Hussein, Lutfi. 1994. *Voicing-dependent vowel duration in standard Arabic and its acquisition by adult American students*:

Ohio State University dissertation.

Ingram, David. 1989. *First language acquisition: Method, description and explanation*. Cambridge: Cambridge University Press.

Ingram, David, Lynda Christensen, Sharon Veach & Brendan Webster. 1980. The acquisition of word-initial fricatives and affricates in English by children between 2 and 6 years. In Grace H. Yeni-Komshian, James F. Kavanagh & Charles A. Ferguson (eds.), *Child phonology. Volume 1: Production*, 169–192. New York: Academic Press.

Iverson, Paul, Dulika Ekanayake, Silke Hamann, Anke Sennema & Bronwen G. Evans. 2008. Category and perceptual interference in second-language phoneme learning: An examination of English /w/-/v/ learning by Sinhala, German, and Dutch speakers. *Journal of Experimental Psychology* 35. 1305–1316.

Jakobson, Roman. 1968. *Child language, aphasia, and phonological universals*. The Hague: Mouton.

Jakobson, Roman & Morris Halle. 1956. *Fundamentals of language*. The Hague: Mouton.

James, Allan R. 1986. *Suprasegmental phonology and segmental form*. Tübingen: Niemeyer.

James, Allan R. 1988. *The acquisition of a second language phonology*. Tübingen: Narr.

James, Allan R. 1990. Phonological structure and the contextual variability of interlanguage segments. In Jacek Fisiak (ed.), *Further insights into contrastive linguistics*, 231–240. Amsterdam: John Benjamins.

Jassem, Wiktor. 2003. Polish. *Journal of the International Phonetic Association* 33. 103–108.

Jekiel, Mateusz. 2012. *The evolution of English dental fricatives: Variation and change*. Adam Mickiewicz University, Poland MA thesis.

Jesney, Karen. 2004. *The use of global foreign accent rating in studies of L2 acquisition*. University of Calgary MA thesis.

Jessen, Michael. 1998. *Phonetics and phonology of tense and lax obstruents in German*. Amsterdam: John Benjamins.

Johns, Donnell F. & Frederic L. Darley. 1970. Phonemic variability in apraxia of speech. *Journal of Speech and Hearing Research* 13. 556–583.

Johnson, Daniel E. 2014. Progress in regression: Why natural language data calls for mixed-effects models. Unpublished manuscript. URL www.danielezrajohnson.com/johnson_2014b.pdf.

Johnson, Wyn & David Britain. 2007. L-vocalisation as a natural phenomenon: Explorations in sociophonology. *Language Sciences* 29. 294–315.

Jokisch, Oliver, Hongwei Ding & Rüdiger Hoffmann. 2011. Acoustic analysis of postvocalic /l/ in Chinese learners of German in the context of an overall perception experiment. In *Proceedings of Workshop on Speech and Language Technology for Education (SLaTE)*, 129–132. Florence, Italy.

Jones, Daniel. 1918. *An outline of English phonetics*. Cambridge: Heffer.

Jones, Daniel, Peter Roach, Jane Setter & John Esling. 2011. *Cambridge English pronouncing dictionary*. Cambridge: Cambridge University Press.

Joos, Martin. 1948. *Acoustic phonetics*. Baltimore: Linguistic Society of America.

Kager, René. 1999. *Optimality theory*. Cambridge: Cambridge University Press.

Kautzsch, Alexander. 2010. Exploring L1 transfer in German learners of English: High front vowels, high back vowels and the bed/bad distinction. *Research in Language* 8. 63–84.

Kautzsch, Alexander. 2017. *The attainment of an English accent: British and American features in advanced German learners*. Frankfurt: Peter Lang.

Keating, Patricia A. 1980. *A phonetic study of a voicing contrast in Polish*: Brown University dissertation.

Keutsch, Wilfried. 1974. *Praxis der englischen Aussprache*. Tübingen: Niemeyer.

Khattab, Ghada. 2011. Acquisition of Lebanese Arabic and Yorkshire English /l/ by bilingual and monolingual children: A comparative spectrographic study. In Zeki M. Hassan & Barry Heselwood (eds.), *Instrumental studies in Arabic phonetics*, 325–353. Amsterdam: John Benjamins.

Kinnaird, Susan K. & Jennifer Zapf. 2004. An acoustical analysis of a Japanese speaker's production of English /r/ and /l/. Tech. rep. Indiana University Linguistic Club Working Papers 4. URL https://www.indiana.edu/~iulcwp/pdfs/04-kinnaird.pdf.

Klatt, Dennis H. 1973. Interaction between two factors that influence vowel duration. *Journal of the Acoustical Society of America* 54. 1102–1104.

Kluender, Keith R., Randy L. Diehl & Beverly A. Wright. 1988. Vowel-length differences before voiced and voiceless consonants: An auditory explanation. *Journal of Phonetics* 16. 153–169.

Ko, Eon-Suk. 2007. Acquisition of vowel duration in children speaking American English. In *Proceedings of Interspeech 2007*, 1881–1884. Antwerp, Belgium.

Kohler, Klaus J. 1979. Parameters in the production and perception of plosives in German and French. *Arbeitsberichte* *des Instituts für Phonetik an der Universität Kiel* 12. 261–291.

Kohler, Klaus J. 1984. Phonetic explanations in phonology: The feature foris/lenis. *Phonetica* 41. 150–174.

Kohler, Klaus J. 1995. *Einführung in die Phonetik des Deutschen*. Berlin: Erich Schmidt Verlag.

König, Ekkehard & Volker Gast. 2009. *Understanding English-German contrasts*. Berlin: Erich Schmidt.

Kornfeld, Judy R. & Henry Goehl. 1974. A new twist to an old observation: Kids know more than they say. In Anthony Bruck, Robert A. Fox & Michale W. La Galy (eds.), *Papers from parasession on Natural Phonology*, 210–219. Chicago: Linguistic Society.

Kortmann, Bernd. 1999. *Linguistics: Essentials*. Berlin: Cornelsen.

Koziol, Herbert. 1959. *Die Aussprache des Englischen*. Stuttgart: Braumüller.

Krug, Manfred & Katrin Sell. 2013. Designing and conducting interviews and questionnaires. In Manfred Krug & Julia Schlüter (eds.), *Research methods in language variation and change*, 69–98. Cambridge: Cambridge University Press.

Kruschke, John K. 2015. *Doing Bayesian data analysis*. London: Academic Press.

Kucharek, Richard. 1988. *Untersuchungen zur englischen Schüleraussprache an Gymnasien in Oberbayern* Augsburger I-&-I Schriften. Augsburg: University of Augsburg.

Kufner, Herbert L. 1971. *Kontrastive Phonologie: Deutsch-Englisch*. Stuttgart: Ernst Klett Verlag.

Kuhl, Patricia K., Karen A. Williams, Francisco Lacerda, Kenneth N. Stevens & Björn Lindblom. 1992. Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* 255. 606–608.

Labov, William. 1969. Contraction, deletion, and inherent variability of the English copula. *Language* 45(4). 715–762.

Labov, William, Sharon Ash & Charles Boberg. 2006. *The atlas of North American English*. Berlin: Mouton de Gruyter.

Ladefoged, Peter. 1999. American English. In *Handbook of the IPA*, 41–44. Cambridge: Cambridge University Press.

Ladefoged, Peter & Keith Johnson. 2011. *A course in phonetics*. Boston, MA: Wadsworth.

Ladefoged, Peter & Ian Maddieson. 1996. *The sounds of the world's languages*. Oxford: Blackwell.

Lado, Robert. 1957. *Linguistics across cultures*. Ann Arbor, MI: University of Michigan Press.

Laeufer, Christiane. 1992. Patterns of voicing-conditioned vowel duration in French and English. *Journal of Phonetics* 20. 411–440.

Lambert, Ben. 2018. *A student's guide to Bayesian statistics*. London: Sage.

Langguth, Annkatrin. 2010. *Phonological data from a Bavarian grammar school and their relation to Received Pronunciation and General American*. University of Bamberg, Germany MA thesis.

Langguth, Benedikt. 2009. *The pronunciation of English in Bamberg*. University of Bamberg, Germany MA thesis.

Larsen-Freeman, Diane. 1997. Chaos/complexity science and second language acquisition. *Applied Linguistics* 18. 141–165.

Larsen-Freeman, Diane. 2005. Second language acquisition and the issue of fossilization: There is no end and there is no end state. In Zhaohong Han & Terence Odlin (eds.), *Studies of fossilization in second language acquisition*, 189–200. Clevedon: Multilingual matters.

Lass, Roger. 1984. *Phonology: An introduction to basic concepts*. Cambridge: Cambridge University Press.

Leather, James. 1999. Second language speech: An introduction. *Language Learning* 39(Supplement 1). 1–56.

Lee, Sungbok, Alexandros Potamianos & Shrikanth Narayanan. 1999. Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *Journal of the Acoustical Society of America* 105(3). 1455–1468.

Lee-Kim, Sang-Im, Lisa Davidson & Sangjin Hwang. 2013. Morphological effects on the darkness of English intervocalic /l/. *Laboratory Phonology* 4(2). 475–511.

de Leeuw, Esther, Ineke Mennen & James M. Scobbie. 2012. Dynamic systems, maturational constraints and L1 phonetic attrition. *International Journal of Bilingualism* 17(6). 683–700.

Lehiste, Ilse. 1964. *Acoustical characteristics of selected English consonants*. Indiana University.

Lesnoff, Matthieu & Renaud Lancelot. 2019. *aod: Analysis of overdispersed data*. URL https://CRAN.R-project.org/package=aod. R package version 1.3.1.

Levelt, Clara C., Niels O. Schiller & Willem J. Levelt. 1999/2000. The acquisition of syllable types. *Language Acquisition* 8. 237–264.

Lindau, Mona. 1985. The story of /r/. In Victoria Fromkin (ed.), *Phonetic linguistics: Essays in honour of Peter Ladefoged*, 157–168. Orlando: Academic Press.

Liu, Chang, Su-Hyun Jin & Chia-Tsen Chen. 2014. Durations of American English vowels by native and non-native speakers: Acoustic analyses and perceptual effects. *Language and Speech* 57(2). 238–253.

Liu, Xian. 2016. *Methods and applications of longitudinal data analysis*. Amsterdam: Elsevier.

Lleó, Conxita, Imme Kuchenbrandt, Margaret Kehoe & Cristina Trujillo. 2003. Syllable final consonants in Spanish and German monolingual and bilingual acquisition. In Natascha Müller (ed.), *(In)vulnerable domains in multilingualism*, 191–220. Amsterdam: John Benjamins.

Lobanov, Boris M. 1971. Classification of Russian vowels spoken by different speakers. *Journal of the Acoustical Society of America* 49. 606–608.

Locke, John L. 1983. *Phonological acquisition and change*. New York: Academic Press.

Long, J. Scott. 1997. *Regression models for categorical and limited dependent variables*. Thousand Oaks: Sage.

Long, J. Scott & Jeremy Freese. 2014. *Regression models for categorical dependent variables using Stata*. College Station: Stata Press.

Long, J. Scott & Sarah A. Mustillo. 2018. Using predictions and marginal effects to compare groups in regression models for binary outcomes. *Sociological Methods and Research* DOI: 10.1177/0049124118799374.

Luce, Paul A. & Jan Charles-Luce. 1985. Contextual effects on vowel duration, closure duration, and the consonant/vowel ratio in speech production. *Journal of the Acoustical Society of America* 78(6). 1949–1957.

Machač, Pavel & Radek Skarnitzl. 2009. *Principles of phonetic segmentation*. Praha: Epocha Publishing House.

Maddieson, Ian. 1984. *Patterns of sounds*. Cambridge: Cambridge University Press.

Maddieson, Ian. 2013. Syllable structure. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*, Leipzig: Max Planck Institute for Evolutionary Anthropology. URL http://wals.info/chapter/12.

Mair, Christian. 1995. *Englisch für Anglisten*. Tübingen: Stauffenburg.

Mair, Christian. 2006. *Twentieth century English: History, variation and standardization*. Cambridge: Cambridge University Press.

Major, Roy C. 1987. Phonological similarity, markedness, and rate of L2 acquisition. *Studies in Second Language Acquisition* 9. 63–82.

Major, Roy C. 1997. Further evidence for the Similarity Differential Rate Hypothesis. In James Leather & Allan James (eds.), *New sounds 97: Proceedings of the Third International Symposium on the Acquisition of Second-language Speech*, 215–222. Klagenfurt: University of Klagenfurt.

Major, Roy C. 2001. *Foreign accent: The ontogeny and phylogeny of second language phonology*. Mahwah: Erlbaum.

Major, Roy C. 2008. Transfer in second language phonology: A review. In Jette G. Hansen Edwards & Mary L. Zampini (eds.), *Phonology and second language acquisition*, 63–94. Amsterdam: John Benjamins.

Major, Roy C. & Michael C. Faudree. 1996. Markedness universals and the acquisition of voicing contrasts by Korean speakers of English. *Studies in Second Language Acquisition* 18. 69–90.

Major, Roy C. & Eunyi Kim. 1996. The similarity differential rate hypothesis. *Language Learning* 46. 465–496.

Mangold, Max. 2015. *Duden Aussprachewörterbuch*. Mannheim: Dudenverlag.

Mathôt, Sebastiaan, Daniel Schreij & Jan Theeuwes. 2012. Opensesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods* 44(2). 314–324.

Maurer, Anne-Julie. 2014. *Acquisition of subphonetic variation by German L2 learners of English*: University of Freiburg dissertation.

McAllister, Robert, James E. Flege & Thorsten Piske. 2002. The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics* 30. 229–258.

McElreath, Richard. 2020. *Statistical rethinking: A Bayesian course with examples in R and Stan*. Boca Raton: CRC Press.

McGurk, Harry & John MacDonald. 1976. Hearing lips and seeing voices. *Nature* 264. 746–748.

Miller, George A. & Patricia E. Nicely. 1955. An analysis of perceptual confusion among some English consonants. *Journal of the Acoustical Society of America* 27. 338–346.

Mines, M. Ardussi, Barbara F. Hanson & June E. Shoup. 1978. Frequency of occurrence of phonemes in conversational English. *Language and Speech* 21. 221–241.

Mitchell, Rosamond, Florence Myles & Emma Marsden. 2012. *Second language learning theories*. New York: Routledge.

Molenberghs, Geert & Geert Verbeke. 2005. *Models for discrete longitudinal data*. New York: Springer.

Mompeán-Gonzalez, Jose & Pilar Mompeán-Guillamón. 2009. /r/-liaison in English: An empirical study. *Cognitive Linguistics* 20(4). 733–776.

Moulton, William G. 1962. *The sounds of English and German*. Chicago: University of Chicago Press.

Müller, Daniela. 2015. Cue weighting in the perception of phonemic and allophonic laterals along the darkness continuum: Evidence from Greek and Albanian. *Albanohellenica* 6. URL http://www.romanistik.uni-muenchen.de/personen/wiss_ma/mueller/publikationen/mueller2015_albanohellenica.pdf).

Naeser, Margaret A. 1970. *The American child's acquisition of differential vowel duration*: University of Wisconsin, Madison dissertation.

Nemser, William. 1971a. Approximate systems of foreign language learners. *International Review of Applied Linguistics* 9. 115–123.

Nemser, William. 1971b. *An experimental study of phonological interference in the English of Hungarians*. Bloomington: Indiana University Press.

Nicenboim, Bruno, Timo B. Roettger & Shravan Vasishth. 2018. Using meta-analysis for evidence synthesis: The case of incomplete neutralization in German. *Journal of Phonetics* 70. 39–55.

Nolan, Francis. 1983. *The phonetic bases of speaker recognition*. Cambridge: Cambridge University Press.

Ohala, John J. 1971. The role of physiological and acoutic models in explaining the direction of sound change. *Project on Linguistic Analysis Reports* 15. 25–40.

Ohala, John J. 1993. The origin of sound patterns in vocal tract constraints. In Peter F. MacNeilage (ed.), *The production of speech*, 189–216. New York: Springer.

Ohala, John J. & Haruko Kawasaki. 1984. Prosodic phonology and phonetics. *Phonology Yearbook* 1. 113–127.

Olmsted, David L. 1971. *Out of the mouth of babes*. The Hague: Mouton.

Olson, Chester L. 1976. On choosing a test statistic in multivariate analysis of variance. *Psychological Bulletin* 83. 579–586.

Oxley, Judith, Nancye Roussel & Hugh Buckingham. 2007. Contextual variability in American English dark-l. *Clinical Linguistics and Phonetics* 21(7). 523–542.

Pascoe, Graham. 1987. *Die Aussprache des Englischen an bayerischen Schulen*. München: Profil.

Peterson, Gordon E. & Harold L. Barney. 1952. Control methods used in a study of the vowels. *Journal of the Acoustical Society of America* 24(2). 175–184.

Peterson, Gordon E. & Ilse Lehiste. 1960. Duration of syllable nuclei in English. *Journal of the Acoustical Society of America* 32. 693–703.

Pickering, Lucy. 2012. Second language speech production. In Susan M. Gass & Alison Mackey (eds.), *The Routledge handbook of second language acquisition*, 335–348. London: Routledge.

Piske, Thorsten, Ian R. A. MacKay & James E. Flege. 2001. Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics* 29. 191–215.

Polka, Linda & Janet F. Werker. 1994. Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology Human Perception and Performance* 20. 421–435.

Port, Robert F. 1981. Linguistic timing factors in combination. *Journal of the Acoustical Society of America* 62. 1051–1054.

Prather, Elizabeth M., Dona Lee Hedrick & Carolyn A. Kern. 1975. Articulation development in children aged two to four years. *Journal of Speech and Hearing Disorders* 40(2). 179–191.

Preston, Dennis R. 1996. Variationist perspectives on second language acquisition. In Robert Bayley & Dennis R. Preston (eds.), *Second language acquisition and linguistic variation*, 1–45. Amsterdam: John Benjamins.

Prince, Alan & Paul Smolensky. 1993. Optimality Theory: Constraint interaction in generative grammar. Unpublished manuscript, Rutgers University, New Brunswick, NJ and University of Colorado, Boulder.

R Core Team. 2020. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing Vienna, Austria. URL https://www.R-project.org/.

Rabe-Hesketh, Sophia & Anders Skrondal. 2012. *Multilevel and longitudinal modeling using Stata*. College Station: Stata Press.

Ramers, Karl Heinz. 1988. *Vokalquantität und -qualität im Deutschen*. Tübingen: Niemeyer.

Rank, Isabell. 2016. Hypercorrection of /v/ amongst advanced German L2 learners of English: Phonetic and speaker-dependent factors. Unpublished term paper, University of Bamberg, Germany.

Rank, Isabell. 2018. Realisation of post-vocalic /r/ by advanced German learners of English. Unpublished BA thesis, University of Bamberg, Germany.

Raphael, Lawrence J. 1972. Preceding vowel duration as a cue to the perception of the voicing charateristics of word-final consonants in American English. *Journal of the Acoustical Society of America* 51. 1296–1303.

Raphael, Lawrence J., Michael F. Dorman & Donna Geffner. 1980. Voicing-conditioned durational differences in vowel and consonants in the speech of three- and four-year old children. *Journal of Phonetics* 8. 335–341.

Rausch, Arsen. 1972. Untersuchungen zur Vokalartikulation im Deutschen. In Heinrich Kelz & Arsen Rausch (eds.), *Betiträge zur phonetik, ipk forschungsberichte 30*, 35–82. Hamburg: Buske.

Recasens, Daniel. 2012. A cross-language acoustic study of initial and final allophones of /l/. *Speech Communication* 54. 368–383.

Recasens, Daniel, Jordi Fontdevila & Maria Dolors Pallarès. 1995. Velarization degree and coarticulatory resistance for /l/ in Catalan and German. *Journal of Phonetics* 23. 37–52.

Rice, Keren. 2007. Markedness in phonology. In Paul de Lacy (ed.), *The Cambridge handbook of phonology*, 79–97. Cambridge: Cambridge University Press.

Rice, Keren & Peter Avery. 1995. Variability in a deterministic model of language acquisition: A theory of segmental acquisition. In John Archibald (ed.), *Phonological acquisition and phonological theory*, 23–42. Hillsdale: Lawrence Erlbaum.

Richards, Jack C. 1971. Error analysis and second language strategies. *Language Sciences* 17. 12–22.

Richtsmeier, Peter. 2010. Child phoneme errors are not substitutions. *Toronto Working Papers in Linguistics* 33. 1–15.

Roach, Peter. 2004. British English: Received Pronunciation. *Journal of the International Phonetic Association* 34. 239–245.

Roach, Peter. 2009. *English phonetics and phonology*. Cambridge: Cambridge University Press.

Roberts, Joanne E., Margaret Burchinal & Marianna M. Footo. 1990. Phonological process decline from 2½ to 8 years. *Journal of Communication Disorders* 23. 205–217.

Rojczyk, Arkadiusz. 2010. Preceding vowel duration as a cue to the consonant voicing contrast: Perception experiments with Polish-English bilinguals. In Ewa Waniek-Klimczak (ed.), *Issues in accents of English: Variability and norm*, 341–360. Newcastle upon Tyne: Cambridge Scholars Publishing.

Rositzke, Harry A. 1939. Vowel length in General American speech. *Language* 15(2). 99–109.

RStudio Team. 2020. *Rstudio: Integrated development environment for r*. RStudio, PBC. Boston, MA. URL http://www.rstudio.com/.

Saleemi, Anjun. 1992. *Universal grammar and language learnability*. Cambridge: University Press.

Sarkar, Deepayan. 2008. *Lattice: Multivariate data visualization with R*. New York: Springer.

Schad, Daniel J., Shravan Vasishth, Sven Hohenstein & Reinhold Kliegl. 2020. How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language* 110. 104038.

Scherer, Günther & Alfred Wollmann. 1986. *Englische Phonetik und Phonologie*. Berlin: Erich Schmidt Verlag.

Schmid, Stephan. 1997. The naturalness differential hypothesis: Cross-linguistic influence and universal preferences in interlanguage phonology and morphology. *Folia Linguistica* 31(3-4). 331–348.

Schmitt, Holger. 2007. The case for the epsilon symbol in RP dress. *Journal of the International Phonetic Association* 37(3). 1–56.

Schneider, Edgar W., Kate Burridge, Bernd Kortmann, Rajend Methrie & Clive Upton (eds.). 2004. *A handbook of varieties of English I: Phonology*. Berlin: Mouton de Gruyter.

Selinker, Larry. 1971. The psychologically relevant data of second language learning. In Paul Pimsleur & Terence Quinn (eds.), *The psychology of second language learning*, Cambridge: Cambridge University Press.

Selinker, Larry. 1972. Interlanguage. *International Review of Applied Linguistics in Language Teaching* 10. 209–231.

Sendlmeier, Walter F. & Julia Seebode. 2007. *Formantkarten des deutschen Vokalsystems*. URL https://www.kw.tu-berlin.de/fileadmin/a01311100/Formantkarten_des_deutschen_Vokalsystems_01.pdf.

Sharf, Donald J. 1962. Duration of post-stress intervocalic stops and preceding vowels. *Language and Speech* 5(1). 26–30.

Sherman, Donald. 1975. Stop and fricative systems: A discussion of paradigmatic gaps and the question of language sampling. *Stanford Working Papers in Language Universals* 17. 1–31.

Sieg, A. 2004. *Die Entwicklung der Phonologie von der dritten zur vierten Klasse des ersten immersiv auf Englisch unterrichteten Jahrgangs der Claus-Rixen-Schule*. University of Kiel MA thesis.

Simons, Daniel J, Yuichi Shoda & D. Stephen Lindsay. 2016. Constraints on generality (cog): A proposed addition to all empirical papers. *Perspectives on Psychological Science* 12(6). 1123–1128.

Simpson, Adrian P. 1998. *Phonetische Datenbanken des Deutschen in der empirischen Sprachforschung und der phonologischen Theoriebildung*, vol. 33 Arbeitsberichte des Instituts für Phonetik der Universität Kiel. Kiel: University of Kiel.

Slawinski, Elzbieta B. 1999. Acquisition of /r-l/ phonemic contrast by Japanese children and adults. In Maria de Graça Pinto, João Veloso & Belinda Maia (eds.), *Psycholinguistics on the threshold of the year 2000. Proceedings of the 5th International Congress of the International Society of Applied Psycholinguistics*, 583–590. Porto: Faculdade de Letras da Universidade do Porto.

Smit, Ann B. 1993. Phonologic error distributions in the Iowa-Nebraska Articulation Norms Project: Consonant singletons. *Journal of Speech and Hearing Research* 36. 533–547.

Smit, Ann B., Linda Hand, J. Joseph Freilinger, John E. Bernthal & Ann Bird. 1990. The Iowa Articulation Norms

Project and its Nebraska replication. *Journal of Speech and Hearing Disorders* 55. 779–798.

Smith, Bruce L. 1979. A phonetic analysis of consonantal devoicing in children's speech. *Journal of Child Language* 6(1). 19–28.

Smith, Bruce L. 2002. Effects of speaking rate on temporal patterns of English. *Phonetica* 59. 232–244.

Smith, Bruce L., Rachel Hayes-Harb, Michael Bruss & Amy Harker. 2009. Production and perception of voicing and devoicing in similar German and English word pairs by native speakers of German. *Journal of Phonetics* 37(3). 257–275.

Smith, Bruce L. & Carol Stoel-Gammon. 1983. A longitudinal study of the development of stop consonant production in normal and down's syndrome children. *Journal of Speech and Hearing Disorders* 48. 114–118.

Smith, James G. 2010. *Acoustic properties of English /l/ and /r/ produced by Mandarin Chinese speakers*. University of Toronto MA thesis.

Smith, Nelson V. 1973. *The acquisition of phonology: A case study*. Cambridge: Cambridge University Press.

Smith, R. Edward. 1980. *Natural phonology of Japanese*: University of Hawaii dissertation.

Sönning, Lukas. 2014. Developmental patterns in the reduction of unstressed vowels by German learners of English. In Ludmila Veselovská & Markéta Janebová (eds.), *Complex visibles out there: Proceedings of the Olomouc Linguistics Colloquium 2014: Language use and linguistic structure*, vol. 4 Olomouc modern language series, 765–778. Olomouc: Palacký University.

Sönning, Lukas. 2016. The dot plot: A graphical tool for data analysis and presentation. In Hanna Christ, Daniel Klenovšak, Lukas Sönning & Valentin Werner (eds.), *A blend of MaLT: Selected contributions from the Methods and Linguistic Theories Symposium*, 101–129. Bamberg: University of Bamberg Press.

Sönning, Lukas. 2019. Statistical inference using estimation: Application to language data. Unpublished manuscript. URL https://osf.io/9khwx/.

Sönning, Lukas. 2020a. Clear vs. dark /l/ in German Learner English: Data set for chapter 5 in "Phonological variation in German Learner English". DataverseNO, V1, UNF:6:qGMGOYpGPdJCZcXD2QAyeg== [fileUNF]. URL https://doi.org/10.18710/G6PJ5F.

Sönning, Lukas. 2020b. Drawing on principles of perception: The line plot. Manuscript submitted for publication. URL https://osf.io/uac9q/.

Sönning, Lukas. 2020c. The TRAP-DRESS contrast in German Learner English: Data set for chapter 4 in "Phonological variation in German Learner English". DataverseNO, V1, UNF:6:InisFzv6Q0s2HMUf96W1KQ== [fileUNF]. URL https://doi.org/10.18710/ATIRRV.

Sönning, Lukas. to appear. (Re-)viewing the acquisition of rhythm in the light of L2 phonological theories. In Robert Fuchs (ed.), *Speech rhythm in L2 and L3 varieties of English*, Singapore: Springer.

Sönning, Lukas & Graham Pascoe. 2020. Final voiced obstruents in German Learner English: Data set for chapter 10 in "Phonological variation in German Learner English". DataverseNO, V1, UNF:6:/8HCS5unH033K8lMayPadg== [fileUNF]. URL https://doi.org/10.18710/DKIGE5.

Sönning, Lukas, Graham Pascoe, Isabell Rank & Christina Wunder. 2020a. English /r/ in German Learner English: Data set for chapter 6 in "Phonological variation in German Learner English". DataverseNO, V1, UNF:6:7uAmFi0J1Pc/mGbzfuRwBg== [fileUNF]. URL https://doi.org/10.18710/YDKDFG.

Sönning, Lukas, Graham Pascoe, Isabell Rank & Christina Wunder. 2020b. The labio-velar glide /w/ in German Learner English: Data set for chapter 7 in "Phonological variation in German Learner English". DataverseNO, V1, UNF:6:mCb5VbBOhTzrESjozWXIIw== [fileUNF]. URL https://doi.org/10.18710/F1A34O.

Sönning, Lukas, Graham Pascoe & Christina Wunder. 2020c. The voiced dental fricative in German Learner English: Data set for chapter 9 in "Phonological variation in German Learner English". DataverseNO, V1, UNF:6:ASRuHTebbP1zQZzlFrtx3A== [fileUNF]. URL https://doi.org/10.18710/DYAGZG.

Sönning, Lukas & Isabell Rank. 2020. The labiodental fricative /v/ in German Learner English: Data set for chapter 8 in "Phonological variation in German Learner English". DataverseNO, V1, UNF:6:IPTiCn2tvSr49eurkB8w7A== [fileUNF]. URL https://doi.org/10.18710/B276ZX.

Sproat, Richard & Osamu Fujimura. 1993. Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics* 21. 291–311.

Stampe, David. 1979. *A dissertation on natural phonology*. New York: Garland Publishing.

Stan Development Team. 2018. The Stan Core Library. Version 2.18.0. URL http://mc-stan.org/.

Stan Development Team. 2020. RStan: The R interface to Stan. R package version 2.19.3. URL http://mc-stan.org/.

Stanford, Kyle. 2017. Underdetermination of scientific theory. In Edward N. Zalta (ed.), *The stanford encyclopedia of philosophy*, Metaphysics Research Lab, Stanford University winter 2017 edn.

Steinlen, Anja K. 2005. *The influence of consonants on native and non-native vowel production: A cross-linguistic study*. Tübingen: Narr.

Stevens, Kenneth N. 1998. *Acoustic phonetics*. Cambridge, MA: MIT Press.

Stockwell, Robert P. & J. Donald Bowen. 1965. *The sounds of English and Spanish*. Chicago: University of Chicago Press.

Strange, Winifred. 1992. Learning non-native phoneme contrasts: Interactions among subject, stimulus, and task variables. In Yohichi Tohkura, Eric Vatikiotis-Bateson & Yoshinori Sagisaka (eds.), *Speech perception, production and linguistic structure*, 197–219. Tokyo: Ohmsha.

Strange, Winifred & Valerie L. Shafer. 2008. Speech perception in second language learners: The re-education of selective perception. In Jette G. Hansen Edwards & Mary L. Zampini (eds.), *Phonology and second language acquisition*, 153–191. Amsterdam: John Benjamins.

Strange, Winifred, Andrea Weber, Erika S. Levy, Valeriy Shafiro, Miwako Hisagi & Kanae Nishi. 2007. Acoustic variability within and across German, French, and American English vowels: Phonetic context effects. *Journal of the Acoustical Society of America* 122(2). 1111–1129.

Stroup, Walter W. 2013. *Generalized linear mixed models: Modern concepts, methods and applications*. Boca Raton: CRC Press.

Syrdal, Ann K. & H. S. Gopal. 1986. A perceptual model of vowel recognition based on the auditory representation of American English vowels. *Journal of the Acoustical Society of America* 79. 1086–1100.

Tabain, Marija. 1998. Non-sibilant fricatives in English: Spectral information above 10 khz. *Phonetica* 55. 107–130.

Tarone, Elaine E. 1978. The phonology of interlanguage. In Jack C. Richards (ed.), *Understanding second and foreign language learning: Issues and approaches*, 15–33. Rowley, MA: Newbury.

Tauberer, Joshua I. 2010. *Learning [voice]*: University of Pennsylvania dissertation.

Templin, Mildred C. 1957. *Certain language skills in children: Their development and interrelationships*, vol. 26 Institute of Child Welfare Monograph Series. Minneapolis: University of Minnesota Press.

Tennekes, Martijn. 2020. tmap: Thematic maps. R package version 3.2. URL https://CRAN.R-project.org/package=tmap.

Thomas, Erik R. 2011. *Sociophonetics: An introduction*. New York: Palgrave Macmillan.

Tizón-Couto, David & David Lorenz. 2020. Variables are valuable: Making a case for deductive modeling. Manuscript submitted for publication.

Tomasello, Michael. 2003. *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Traunmüller, Hartmut. 1990. Analytical expressions for the tonotopic sensory scale. *Journal of the Acoustical Society of America* 88. 97–100.

Trovimovich, Pavel, Elizabeth Gatbonton & Norman Segalowitz. 2007. A dynamic look at L2 phonological learning: Seeking processing explanations for implicational phenomena. *Studies in Second Language Acquisition* 29. 407–448.

Trubetzkoy, Nikolai S. 1939. *Principles of phonology*. Paris: Klincksieck.

Umeda, Noriko. 1975. Vowel duration in American English. *Journal of the Acoustical Society of America* 58. 434–445.

Upton, Clive, William A. Kretzschmar & Rafal Konopka. 2001. *The Oxford English dictionary of pronunciation for current English*. Oxford: Oxford University Press.

Van Hofwegen, Janneke. 2009. Cross-generational change in /l/ in Chicano English. *English World-Wide* 30(3). 302–325.

Vanhove, Jan. 2018. Baby steps in Bayes: Piecewise regression [web log post]. URL https://janhove.github.io/analysis/2018/07/04/bayesian-breakpoint-model.

Vanhove, Jan. 2019. Visualizing statistical uncertainty using model-based graphs [web log post]. URL https://janhove.github.io/visualise_uncertainty/.

Vanhove, Jan. 2020. Towards simpler and more transparent research reports. *International Journal of Applied Linguistics* DOI: 10.1075/itl.20010.van.

Varshney, Lav R. & John Z. Sun. 2013. Why do we perceive logarithmically? *Significance* 10(1). 28–31.

Vehtari, Aki, Andrew Gelman, Daniel Simpson, Bob Carpenter & Paul-Christian Bürkner. 2020. Rank-normalization, folding, and localization: An improved $\widehat{R}$ for assessing convergence of MCMC. *Bayesian Analysis* URL https://doi.org/10.1214/20-BA1221.

Vennemann, Theo. 1988. *Preference laws for syllable structure and the explanation of sound change*. New York: Mouton De Gruyter.

Wang, Hongyan. 2007. *English as a lingua franca: Mutual intelligibility of Chinese, Dutch and American speakers of English*:

University of Leiden, Netherlands dissertation.

Wängler, Hans-Heinrich. 1981. *Atlas deutscher Sprachlaute*. Berlin: Akademie-Verlag.

Weiher, Eckart. 1975. *Lautwahrnehmung und Lautproduktion im Englischunterricht für Deutsche* Arbeitsberichte des Instituts für Phonetik der Universität Kiel. Kiel: University of Kiel.

Wellman, Beth L., Ida M. Case, Ida Gaarder Mengert & Dorothy E. Bradbury. 1931. Speech sounds of young children. *University of Iowa Studies in Child Welfare* 5(2). 1–82.

Wells, John C. 1962. *A study of the formants of the pure vowels of British English*. University College London MA thesis.

Wells, John C. 1982. *Accents of English*. Cambridge: Cambridge University Press.

Wells, John C. 1990. Syllabification and allophony. In Susan Ramsaran (ed.), *Studies in the pronunciation of English, a commemorative volume in honour of A.C. Gimson*, 76–86. New York: Routledge.

Wells, John C. 2008. *Longman pronunciation dictionary*. Harlow: Pearson.

Wenk, Brian J. 1979. Articulatory setting and de-fossilization. *Interlanguage Studies Bulletin* 4. 202–220.

Werker, Janet F. & Richard C. Tees. 1999. Influences on infant speech processing: Toward a new synthesis. *Annual Review of Psychology* 50. 509–535.

Wickham, Hadley. 2016. *ggplot2: Elegant graphics for data analysis*. New York: Springer.

Wieden, Wilfried & William Nemser. 1991. *The pronunciation of English in Austria*. Tübingen: Narr.

Wiener, Seth & Ya-ting Shih. 2013. Evaluating the emergence of [v] in modern spoken Mandarin. In Zhuo Jing-Schmidt (ed.), *Increased empiricism: Recent advances in Chinese linguistics*, 171–187. Amsterdam: John Benjamins.

Wiese, Richard. 1996. *The phonology of German*. Oxford: Oxford University Press.

Wiik, Kalevi. 1965. *Finnish and English vowels: A comparison with special reference to the learning problems met by native speakers of Finnish learning English*, vol. 94 Annales Universitatis Turkuensis, Series B. Turku: Turun Yliopisto.

Williams, Daniel & Paola Escudero. 2014. A cross-dialectal acoustic comparison of vowels in Northern and Southern British English. *Journal of the Acoustical Society of America* 136(5). 2751–2761.

Wilson, Ian & Bryan Gick. 2014. Bilinguals use language-specific articulatory settings. *Journal of Speech, Language and Hearing Research* 57(2). 361–373.

Winter, Bodo. 2019. *Statistics for linguistics*. New York: Routledge.

Winter, Bodo & Martine Grice. 2020. The independence assumption in linguistics: Items, repetitions, pooling, and the goal of statistical inference. Manuscript submitted for publication.

Wode, Henning. 1977. The L2 acquisition of /r/. *Phonetica* 34. 200–217.

Wode, Henning. 1978. The beginnings of non-school room L2 phonological acquisition. *International Review of Applied Linguistics* 16. 109–25.

Wode, Henning. 2009. Developing non-native pronunciation in immersion settings. In Thorsten Piske & Martha Young-Scholten (eds.), *Input matters in SLA*, 238–256. Bristol: Multilingual matters.

Wolfram, Walt. 1985. Variability in tense marking: A case for the obvious. *Language Learning* 35. 229–253.

Wolfram, Walt & Natalie Schilling. 2016. *American English: Dialects and variation*. Malden, MA: John Wiley & Sons.

Wunder, Christina. 2012. *Final devoicing in German Learner English: A study at a Franconian grammar school*. University of Bamberg, Germany MA thesis.

Xie, Yihui. 2016. *bookdown: Authoring books and technical documents with R markdown*. Boca Raton, Florida: Chapman and Hall/CRC. URL https://github.com/rstudio/bookdown.

Yang, Byunggon. 1996. A comparative study of American English and Korean vowels produced by male and female speakers. *Journal of Phonetics* 24. 245–261.

Yarkoni, Tal. 2019. The generalizability crisis. DOI: 10.31234/osf.io/jqw35. URL https://psyarxiv.com/jqw35/.

Yavaş, Mehmet. 2005. *Applied English phonology*. Malden: Blackwell.

Yumoto, Eiji, Wilbur J. Gould & Thomas Baer. 1982. Harmonics-to-noise ratio as an index of the degree of hoarseness. *Journal of the Acoustic Society of America* 71. 1544–1550.

Zampini, Mary L. 2008. L2 speech production research: Findings, issues, and advances. In Jette G. Hansen Edwards & Mary L. Zampini (eds.), *Phonology and second language acquisition*, 219–249. Amsterdam: John Benjamins.

Zhu, Hao. 2020. kableExtra: Construct complex table with 'kable' and pipe syntax. R package version 1.3.1. URL https://CRAN.R-project.org/package=kableExtra.

Zimmerman, Samuel A. & Stanley M. Sapon. 1958. Note on vowel duration seen cross-linguistically. *Journal of the Acoustical Society of America* 30. 152–153.