24 Bamberger Beiträge zur Linguistik

Nikolai Beland

The Superlative Alternation in Present-Day English

Triangulating Elicitation and Usage Data

University of Bamberg Press

Bamberger Beiträge zur Linguistik

hg. von Martin Haase, Thomas Becker (†),
Sebastian Kempgen, Manfred Krug
und Patrizia Noel Aziz Hanna

Band 24

University
of Bamberg
Press
2025

The Superlative Alternation in Present-Day English

Triangulating Elicitation and Usage Data

Nikolai Beland

University
of Bamberg
Press
2025

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.dnb.de/ abrufbar.

Diese Arbeit hat der Fakultät Geistes- und Kulturwissenschaften der Otto-Friedrich-Universität Bamberg als Dissertation vorgelegen.
Gutachterin: Prof. Dr. Julia Schlüter
Gutachterin: Prof. Dr. Britta Mondorf
Tag der mündlichen Prüfung: 21.11.2021

Herstellung und Druck: docupoint, Magdeburg
Umschlaggestaltung: University of Bamberg Press
Umschlagbild: Watzmann in den Berchtesgadener Alpen, Bayern, Deutschland, © Günter Seggebäing, CC BY-SA 3.0

University of Bamberg Press, Bamberg 2025
ROR https://ror.org/004fa0x02

ISSN: 2190-3298 (Print) eISSN: 2750-7726 (Online)
ISBN: 978-3-98989-078-7 (Print) eISBN: 978-3-98989-079-4 (Online)

URN: urn:nbn:de:bvb:473-irb-110919x
DOI: https://doi.org/10.20378/irb-110919

Acknowledgements

This book has grown out of a dissertation project at the Chair of English Linguistics at the University of Bamberg. From its earliest stages to its final form, it has benefited from the guidance, support, and encouragement of many people.

I am sincerely grateful to my head supervisor, Julia Schlüter, whose guidance, patience, and generous investment of time have shaped every stage of this project. From detailed feedback on drafts to stimulating discussions that went beyond the immediate scope of the work, her support has been unwavering. I could not have wished for a more dedicated and insightful mentor. I also wish to thank Britta Mondorf, my co-supervisor, for her constructive suggestions and steady support, which helped refine the project and bring it to completion.

I am especially thankful to the Hanns Seidel Foundation for admitting me to its doctoral fellowship program. Beyond financial support, the foundation provided a vibrant academic community that greatly enriched my research experience. Andreas Burtscheidt deserves particular recognition for his dedication in organizing meetings and events, especially the idea exchanges with fellow doctoral students at Kloster Banz, which offered both intellectual stimulation and a sense of community.

At the Chair of English Linguistics at the University of Bamberg, I received invaluable support throughout the project. Manfred Krug generously provided the academic infrastructure and encouraged my work at every stage, including the decision to include my book in the *Bamberger Beiträge zur Linguistik* series. I am particularly grateful to Fabian Vetter for his direct assistance with questionnaire design and corpus infrastructure. I also wish to thank Gabriele Knappe, Heinrich Ramisch, Ole Schützler, and Valentin Werner for their constructive advice, which greatly helped to shape the project.

I owe a special debt of gratitude to Lukas Sönning, whose contribution to this work was indispensable. He conducted the statistical analyses, provided detailed feedback on early drafts, offered critical yet supportive guidance throughout, and inspired my interest in *R* for data analysis. His expertise, insight, and careful attention were central to the development and completion of this project.

The discussions at the English Linguistics Research Seminar and the Linguistische Werkstatt have also been invaluable. In particular, Stefan Hartmann offered constructive feedback and advice that greatly enhanced my research.

I am grateful to Markus Schauer for serving on the examination committee for my dissertation defense, and to Sabine Vogt for her role as chair of the committee.

I would also like to thank the team at *University of Bamberg Press* for their guidance and support throughout the publication process.

I am thankful to Stefan Ehrlich, Jeremias Grau, Maximilian Krug, and Stephan Renker for their friendship and the many stimulating conversations we shared, both academic and personal, which provided perspective and encouragement throughout this journey.

Above all, my family has been my unwavering foundation, offering love and encouragement even while my head was spinning among the clouds of the ivory tower. To my parents, Christiane and Heinz, for their constant support, guidance, and encouragement throughout my life. To my sister, Luisa, whose insight, warmth, and presence have brightened countless moments along the way. To my wife, Stephanie, whose steadfast support, patience, and companionship have carried me through this long journey, making even the most challenging moments meaningful. And to our daughter Clara, who joined us during the revision and publication process, bringing joy and a renewed sense of wonder into our lives. It is to them that this work is dedicated.

Table of Contents

1 Introduction

When it comes to expressing the highest degree of English adjectives, language users have the choice between two strategies: One is to attach the suffix *-est* to the adjectival base (synthetic superlative), and the other is to precede the adjective with *most* (analytic superlative). Szmrecsanyi (2006: 63) remarks that while the comparative alternation has received considerable attention in empirical work, scholarly interest in the superlative has to date been rather meagre. This discrepancy may be attributable to two reasons: First, the comparative comes with greater conceptual transparency on various levels of linguistic analysis; it will be shown in Chapter 2 of this study that the superlative is frequently ambiguous between three interpretations and that the categorization of superlative constructions hinges not only on syntactic criteria but also on idiomatic and pragmatic ones. Second, it is tempting to assume that the comparative and superlative alternations are governed by the same set of factors. This claim, however, does not stand up to scrutiny in the hitherto only contrastive multifactorial account of the degree alternations (Cheung & Zhang 2016). Thus, a comprehensive analysis of the superlative alternation has long been overdue.

Besides this change of empirical focus, the present account expands upon earlier work on gradation strategy choice in three ways: First, two different methodological approaches will be taken to investigate the superlative alternation: a questionnaire-based elicitation study and a corpus analysis. Second, in addition to covering the three predictor groups that have been examined in previous accounts, i.e. adjective factors, contextual constraints and frequency variables, my analysis will shed light on a fourth category of determinants that has hitherto received relatively little attention, namely social factors such as age and gender. This empirical diversity also requires an eclectic approach on the level of linguistic theory: In assessing Mondorf's (2009b) *more*-support hypothesis in the superlative alternation, I will rely on processing theories grounded in the cognitive-functional paradigm; further, my findings will be interpreted against the backdrop of variationist sociolinguistics and usage-based Construction Grammar. Third, while previous analyses have largely been confined to alternating adjectives, the current one also considers categorical

types. In this context, it will be interesting to examine against the backdrop of usage-based theory how frequency of use relates to (non-)categoricity.

1.1 The state of the art

The current section is divided into two parts. Section 1.1.1 defines the place of English alternation phenomena in the broader context of linguistic variation. This survey sketches the conceptual history of scholarly engagement with linguistic alternations as well as the methodological pathways that have consequently been chosen by researchers in the respective traditions. Section 1.1.2 provides a review of previous work on the comparative and superlative alternations: a synopsis of the diachronic perspective and a more fine-grained description of the synchronic situation.

1.1.1 Alternation phenomena in English

Alternations have recently received substantial attention in linguistic research. The concept of *alternation*, however, while intuitively clear, has been defined along very different lines in the literature. More often than not, the notion of alternation that serves as a conceptual starting point in a particular study can only be inferred from the overall methodological framework and, more specifically, from the variables that have been included in the quantitative analysis. There is strong disagreement as to the nature of the "envelope of variation" (Bayley 2013: 91), i.e. the degree to which linguistic forms and structures must be equivalent in order to qualify as alternants (see Bresnan & Ford 2010: 170 for a discussion of this problem). Labov's (1972: 271) definition of equivalence as "two ways of saying the same thing" seems most valid in the domain of phonology.[1] Different realizations of /t/ in *water*, for instance, do not affect the semantics of the lexeme. Extending the concept of alternation to other levels of linguistic analysis, however, is a problematic undertaking (cf. Rosenbach 2002: 22). A broader definition of "sameness" is required in such contexts.

[1] Gries (2017: 8) defines the term *syntactic alternation* as "structurally and/or lexically different ways to say functionally very similar things" and points out that as such, this concept is similar to the phonological notion of minimal pair.

The notions of grammatical alternation that have been proposed in the literature can be located on a scale between two extreme definitions, i.e. a broad and a narrow one: Weiner & Labov's (1983: 31-32) postulate that alternants be "truth-conditionally equivalent and used on the whole to refer to the same state of affairs" leaves considerable room for interpretation and can thus be considered to mark the broad end of the scale. The opposite extreme is represented by Bolinger's (1977: 4) *Principle of No Synonymy*, which posits that any difference in word order also implies a difference in meaning. This narrow conception of sameness appears particularly problematic when we investigate showcases of positional variation, e.g. the dative alternation, for which early studies assumed the alternants in focus to be semantically equivalent (Larson 1988, *inter alia*). Bolinger's principle can further be considered a conceptual bedrock of Construction Grammar:[2] According to Goldberg (1995: 67), "if two constructions are syntactically distinct, they must be semantically and pragmatically distinct." Along these lines, Hilpert (2010: 25) points out that "comparatives provide a particularly good illustration of the notion 'construction', as they pair different formal characteristics (i.e. syntactic, morphological, and phonological traits) with different aspects of meaning (i.e. semantic and pragmatic characteristics)." Evidently, the same holds for superlatives. Construction grammarians would therefore contend that the competing superlative constructions (X-*est* and *most* X) that will be examined in this study should not be considered equivalents.

More recent quantitative studies of alternation phenomena have steered clear of a radical notion of equivalence to embrace a both theoretically and methodologically more malleable concept of *interchangeability*, which comes with conceptual problems of its own (cf. Ehret, Wolk & Szmrecsanyi 2014: 267). The guiding question in these research settings is: What makes a particular variant more likely or 'felicitous' (Hinrichs & Szmrecsanyi 2007: 446) than the competing one in a given context? Studies in this tradition are typically located at the crossroads of probabilistic

[2] See Uhrig (2015) for a critical discussion of this principle. While he acknowledges that a departure from the generativist notion of alternations is a conceptual sine qua non of Construction Grammar, he warns against the "undesirable effect that at least some people refuse to regard groups of semantically highly similar constructions as alternatives" (Uhrig 2015: 335).

grammar (Bresnan & Hay 2008; Bresnan & Ford 2010), comparative (Tagliamonte 2013) and variationist sociolinguistics (Tagliamonte 2012), and they endeavor to map (morpho-)syntactic variability from cross-varietal and apparent-time perspectives using state-of-the-art statistical tools. It has been reported that probabilistic grammars are remarkably similar and stable across varieties; however, there appear to be noticeable differences between Inner Circle and Outer Circle varieties (cf. Szmrecsanyi 2019). Works in this paradigm have revisited textbook examples of (morpho-)syntactic variation such as the dative (Bresnan & Hay 2008; Bresnan & Ford 2010; Jäschke & Plag 2016; Röthlisberger, Grafmiller & Szmrecsanyi 2017) and genitive alternations (Hinrichs & Szmrecsanyi 2007; Grafmiller 2014; Szmrecsanyi et al. 2017 on both alternations in a comparative perspective) or the particle verb alternation (Grafmiller & Szmrecsanyi 2018). Rhythmic aspects and their potential interactions with other, more extensively researched predictors (e.g. possessor animacy in the genitive alternation) have attracted increasing attention in recent years (Ehret, Wolk & Szmrecsanyi 2014 and Shih et al. 2015 on the genitive). Probabilistic studies have further made fruitful contributions to the discussion of the extent to which the linguistic subsystems of morphosyntax and phonology interact, employing the relative grammaticality of competing forms as a key criterion of variationist analysis.

A crucial problem raised in the context of the probabilistic approach is yet to be commented on: In what way can one alternant be considered more likely or 'felicitous' than the other? This question calls for some insight into the explanatory factors that have been adduced to account for linguistic variation in general:

> Some variation is the result of **articulatory constraints on grammatical processes**; some reflects a **variable recognition of grammatical boundaries**; some appears to be the **residue of historical processes** which persists long after the social conditions that gave rise to it have disappeared. (Weiner & Labov 1983: 31, emphasis added)

The three causes of variation mentioned here clearly are interdependent; every phenomenon of variation is somehow constrained by each of them. The former two have been channeled into functional-cognitive theories that attempt to uncover how language interacts with cognition. Studies in this tradition have focused on the idea that language production and

processing is efficiency-driven (Hawkins 1999, 2004), thereby invoking effects of economy (e.g. Rosenbach 2002) and complexity (Rohdenburg 1996; Mondorf 2009b) as explanatory factors. Rohdenburg's (1996: 151) *Complexity Principle* ties cognitive complexity to linguistic transparency:

> In the case of more or less explicit grammatical options the more explicit one(s) will tend to be favored in cognitively more complex environments.

The explanatory power of this notion in contexts of (morpho)syntactic variation is contested, primarily due to the vagueness of the concept (cf. Kortmann & Szmrecsanyi 2012: 10-13 for a summary of the complexity debate; Hawkins 1999, 2004 for the efficiency-driven notion of processing complexity in functional-cognitive accounts of grammatical variation). Szmrecsanyi (2009: 322) issues the caveat that the link between analyticity, syntheticity and language complexity "is not always backed up by hard empirical evidence, especially when it comes to processing complexity". From a diachronic angle, exclusively user-centered explanations of this kind further fail to address the question why language users should feel compelled to react to cognitive pressures with varying intensity at different points in time (Rosenbach 2002: 273-275). These limitations emphasize that empirical investigations of morphosyntactic alternations must give equal consideration to alternating and potentially categorical contexts. Despite these reservations about the complexity construct, the research conducted in this context has paved the way for studies in the probabilistic tradition, both methodologically and conceptually.

In a diachronic perspective, Szmrecsanyi (2009, 2012) draws on corpus data to chart the trajectory of English on the continuum of grammatical *syntheticity* and *analyticity*, thus zooming in on the established opposition of crosslinguistic morphological typology coined by Schlegel (1818).[3] His findings cast doubt on the commonly held view that synthetic and analytic

[3] There is a great deal of terminological variation in studies of alternation phenomena. In publications on gradation strategy choice, the *-er/-est*-variant is also referred to as 'inflectional', 'compositional', 'suffixed' or 'terminational' (Poutsma 1914: 496), the *more-/most*-variant as 'periphrastic' or 'syntactic' (Bauer, Lieber & Plag 2013: 103). Spencer & Popova (2015) provide a comprehensive theoretical treatment of the relationship between periphrasis and inflection, notably also addressing terminological aspects. The thesis at hand adopts Schlegel's terminology for the most part.

marking have been inversely correlated in English diachrony; instead, there is evidence to suggest that analyticity and syntheticity are positively correlated, and variety is assumed to play a role here, too. A binary classification of a considerable number of variation phenomena along the lines of the aforementioned opposition (i.e. synthetic vs. analytic), however, is vulnerable to methodological criticism: These phenomena range widely on the spectrum between morphological and syntactic variation (Cheshire 1987: 264); the terms *analytic* and *synthetic* are thus "more often than not seen as referring to notional concepts not amenable to rigorous quantification", as remarked by Szmrecsanyi (2012: 654-655) himself. This caveat appears to be less relevant to gradation strategy choice than to other alternation phenomena that are located closer to the syntactic extreme of the morphosyntactic spectrum, where the formal dissimilarity of alternants is often accompanied by more sizeable semantic divergence.

1.1.2 The comparative and superlative alternations

The present section provides a survey of the studies that have been conducted on the comparative and superlative alternations from a diachronic and synchronic perspective. This literature review does not cover the special case of double synthetic (*worser*) and periphrastic comparatives (*more friendlier*), which has been investigated by Schlüter (2001, 2005: 67-79, chiefly on prosodic constraints of *worser* in EModE) and, in broader diachronic terms, by González-Díaz (2006a, 2008: Chapters 6 and 7).

The diachronic perspective

While it is undisputed in diachronic accounts of adjective gradation that inflection was the default gradation strategy in Old English (Lass 1994: 149-150; Mitchell & Robinson 2007: 33, *inter alia*), two aspects concerning the analytic strategy have been subject to scholarly contention, namely its roots and the date at which it evolved in competition to the synthetic strategy.

The position held by a handful of authors is that the analytic mode developed in the 13th century, and hence in Middle English times, in analogy with Latin and possibly also French (Mustanoja 1960: 278-279; Strang 1970; Kytö & Romaine 2000: 172). The incursion of the analytic strategy

into the exclusively synthetic system of adjective gradation is generally considered a change from above. The innovative analytic variant steadily gained ground until the late 14th century (D'Arcy 2014: 220) and reached its peak in the Late Middle English period (Pound 1901: 19). A reversal of this trend, i.e. an increase in the use of synthetic forms, is assumed to have taken place in the Late Modern English period (post-1710, Kytö & Romaine 1997: 335), at which point the competing forms had attained levels of prominence comparable to the present-day distribution. Mondorf (2009b: 127) interprets the ensuing stabilization of the system as a functional specialization along the lines of a cognitive support strategy to which I will return in greater detail in Section 3.2.3.

The alternative view, held by Faiß (1977), González-Díaz (2006b, 2008) and Mondorf (2009b), *inter alia*, rests on the assumption that the gradual diffusion of analytic forms may as well be "native in origin and arose from the natural desire for greater emphasis and clarity which seems to lie behind so many other periphrastic constructions of the English language" (Mustanoja 1960: 279-280). This view of a change from below (see also Faiß 1977: 167-168 on this matter) is extended by González-Díaz (2006b: 731-733; 2008), who hypothesizes that the use of OE degree markers and adverbial intensifiers in conjunction with participles paved the way for the emergence of the analytic strategy. The establishment of this innovation is consequently regarded to have taken place as early as the second half of the 9th century, possibly facilitated by French influence (Mondorf 2009b: 121; D'Arcy 2014: 221).

Especially for the latter half of the 20th century, an increase in analytic comparative and superlative forms – and a simultaneous decline of inflections across the board – is reported by Potter (1969: 146-147) and Leech et al. 2009 (264-265). In the latter study, this trend of *analyticization* (Leech et al. 2009: 264) is interpreted as a typological development that has been going on since the earliest stages of the English language.

The synchronic perspective

The comparative and superlative alternations have become the subject of increasing scholarly interest since the final decade of the 20th century, with Bauer (1994) providing the first empirical study of this showcase of grammatical variation. Earlier accounts of adjective comparison – chiefly

comprehensive reference guides on English grammar – take an exclusively descriptive approach, observing that the choice of gradation strategy is very much determined by the morphophonological properties of the adjectival base, predominantly the number of syllables and the final segment(s) (Curme 1931: 500-501; Jespersen 1949: 347-356; Quirk et al. 1985: 461; Biber et al. 1999: 522; Huddleston & Pullum 2002: 1582-1588; further Fries 1993 on monosyllabic adjectives). By the end of the 20th century, the number of syllables has been promoted as a rule of thumb predicting the choice of one variant over the other, as advocated by König (1994: 540), who states that "[m]onosyllabic adjectives take the inflectional form [...]; adjectives with three syllables take the analytic form [...]. Disyllabic adjectives accept both strategies." For the primary locus of variation, namely disyllables, the final segment of the adjective is considered to turn the scales in favor of one of the two variants. In this context, the divergent preferences of disyllabic adjectives ending in *-y* and *-ly* have exhaustively been investigated (Bauer 1994: 51-60; Lindquist 1998; Matsui 2010: 195), with Chua (2018) arguing that considerations related to the lexical spread of comparative constructions may constitute a viable supplement to previous accounts that have exclusively focused on the morphophonological properties of adjectives ending in *-y*.

While still foregrounding type-specific features, Leech & Culpeper (1997) are the first to acknowledge the influence of contextual factors such as syntactic position (predicative vs. attributive), following *than*-phrases and degree modifiers (more comprehensively, González-Díaz 2008). In the first study that regards the comparative and superlative alternations as distinct phenomena, Peters (2000) identifies a range of "crossover" adjectives, i.e. lexical types that have diverging strategy preferences in the comparative and superlative. Strikingly, all seven of these types tend to favor the analytic comparative but the synthetic superlative (*risky*, *costly*, *deadly*, *friendly*, *cruel*, *polite* and *remote*). These findings give weight to the assumption that collocational and register-specific effects may at least partly condition the choice of gradation strategy, with Peters (2000: 311) concluding that the "interplay of these factors is intricate, suggesting a paradigm which is splintered rather than simply split". Pending further empirical validation, we should be skeptical of the crossover concept.

Couched within the paradigm of cognitive linguistics, Mondorf's (2002, 2003, 2009a, 2009b) corpus-based studies propose a functional division of labor between the two comparative variants along the lines of processing complexity. The general hypothesis underlying her research draws on Rohdenburg's (1996: 151) *Complexity Principle* and suggests a link that was essentially conjectured as early as Wilhelm von Humboldt (1836: 284-285) in broader terms[4]:

> In cognitively more demanding environments which require an increased processing load, language users tend to make up for the additional effort by resorting to the analytic (more) rather than the synthetic (-er) comparative. (Mondorf 2009b: 6)

This hypothesis advances a causal relationship between linguistic complexity and processing costs. According to Mondorf (forthc.), the compensatory mechanism of *more*-support can be illustrated by envisioning a hill that can be climbed along two different routes. One route is short and steep, representing the less explicit synthetic form (e.g. *cleverer*). The other is longer, winding, and less demanding, corresponding to the more explicit analytic variant (e.g. *more clever*). Under normal circumstances, we would take the direct, steep route (*cleverer*), as it offers the fastest ascent. Yet when we carry an additional burden – comparable to an increased cognitive load – the gentler, more circuitous route (*more clever*) becomes the more efficient choice. The corpus evidence presented in Mondorf's studies for 24 determinants of eight linguistic domains provides ample

---

[4] See the following quote from Humboldt's (1836) treatise *Über die Verschiedenheit des menschlichen Sprachbaues und ihren Einfluss auf die geistige Entwicklung des Menschengeschlechts* [On the diversity of human language construction and its influence on the mental development of the human species] (translation by Peter Heath in Losonsky [ed., 1999: 206], bold print added for emphasis):

The imaginative pleasure at a clever combination of indicative signs with a sonorous march of syllables is **replaced by convenience of understanding**, whereby **forms are dissolved** into auxiliary verbs and prepositions. The aim of **a readier clarity** is thus elevated at once above the other merits of language, since **this analytic method does indeed diminish the effort of comprehension**, and even in some cases **increases accuracy, where the synthetic method has more difficulty in attaining it**. But by the use of these **grammatical *auxiliary terms*** the inflections become more dispensable, and gradually lose their importance for the vigilance of the linguistic sense.

support for the notion of *more*-support, suggesting that functional motivations on multiple levels (e.g. semantics, pragmatics etc.) indeed appear to supplement the lexical partitioning of the paradigm according to the morphophonological properties of the adjectival base. Hilpert (2008: 413) objects that claiming the universality of this principle carries some risk of circularity, which, as (Lohmann 2010: 308) specifies, "may arise if complexity for a given variable is claimed when we find an effect towards the analytic variant, without providing independent evidence for this assumed complexity". Most importantly, the complexity-based argument as brought forward by Mondorf is too general in that it is indifferent to the distinction between type-specific factors and contextual constraints. I will argue that while the proposed cognitive support strategy has explanatory potential for the latter kind of variables, particularly those on the syntactic and pragmatic levels, the link of complexity and the alleged processing advantage afforded by the analytic form is far from transparent for the former group of factors.

Further, Mondorf (2009a, 2009b: 171-193) provides the first crossvarietal analysis of comparative strategy choice, focusing on the two main reference varieties BrE and AmE. Her findings indicate that AmE more readily deploys *more*-support as a strategy of mitigating processing complexity than BrE; this perceived greater sensitivity to complexity effects in AmE, however, can at least partly be accounted for by differences in style/register and attested gradability among the adjectives under investigation (Mondorf 2009b: 190-192).[5] In a corpus study of the comparative in the New Zealand and Canadian vernaculars, D'Arcy (2014: 240) concludes that the "comparative alternation does not appear to be particularly variable in casual, unscripted, and unreflective discourse", acknowledging, however, that informal speech is not the most suitable register in which to observe this phenomenon. A comprehensive account of varietal differences in this domain of grammatical variation remains a desideratum.

Boyd (2007) and Kunter (2015, 2017) use experimental designs to test the validity of Mondorf's (2009b) central hypothesis in perception and production. Triangulating psycholinguistic experiments by two corpus studies of BNC and COCA material, Kunter (2017) suggests that *more*-support

[5] According to Mondorf (2009b: 179), attested gradability is defined as "frequency of both comparative forms (synthetic plus analytic) per amount of text".

is indeed operative as a compensatory mechanism for cognitive complexity in production (Kunter 2017: 139-140); the evidence gleaned from his perception experiment, however, in which reaction times to the presented stimuli are considered proxies of cognitive processing complexity, prompts Kunter (2017: 91) to dismiss the applicability of this compensation strategy in perception. The distinction between perception and production clearly demonstrates the shortcomings of an exclusively corpus-based approach to variation in gradation strategy choice. Further, the results can be taken to imply that synthetic and analytic comparatives are processed by distinct cognitive mechanisms and stored separately in the mental lexicon (Kunter 2017: 96). In addition, Kunter's (2017) account of the comparative and possessive alternations concludes on the note that the complexity associated with the adjective type in the comparative alternation has significantly greater predictive power than noun complexity in the genitive alternation. According to Kunter (2017: 8), these findings indicate that the former alternation is morphological in nature, while the latter one is rather a syntactic phenomenon. This key insight prompts Kunter (2017: 225) to reformulate the principle of *analytic support* under the heading of *periphrastic support*:

> If the inflectional paradigm of a lexeme allows the choice between periphrastic and synthetic exponence, the probability to use the periphrastic exponent increases in cognitively more demanding environments which require an increased processing load, or if the lexeme itself is cognitively demanding.

Crucially, this principle accounts for the suggested cognitive mechanism underlying morphological alternations by considering the choice between the competing variants as a probabilistic function of the processing complexity inherent in the type properties and the constraints of the linguistic environment.

In his study of morphosyntactic persistence in spoken English, Szmrecsanyi (2006: 63-85) finds that speakers are more likely to use the analytic comparative if they were previously exposed to the degree marker *more* within a certain time interval while no such effect obtains for the synthetic comparative. The odds of an analytic prime-target match declines as a logarithmic function of the textual distance between prime and

target (Szmrecsanyi 2006: 84). Strikingly, analytic persistence is also observed if a comparative form is preceded by quantifier *more*. According to (Mondorf 2009b: 115), this finding can be taken to suggest that "even formal identity creates persistence effects in the absence of functional identity". Informal registers generally seem to be more persistent in the choice of comparative strategy than formal registers.

The multifactorial corpus studies by Hilpert (2008) and Cheung & Zhang (2016) rank several type-specific (i.e. morphophonological), contextual (i.e. syntactic) and frequency factors that have been found to influence the choice of comparative strategy according to their predictive power. Both accounts conclude that the morphophonological properties of the adjectival base are the strongest predictors of the comparative alternation, with the number of syllables by far outweighing all other factors under investigation (Hilpert 2008: 408). The frequency variables included in Cheung & Zhang's (2016: 573) model, i.e. the frequency of the positive form and the comparative-to-positive ratio as a measure of gradability, are found to even surpass the morphophonological ones in terms of effect size. Among the syntactic variables, the presence or absence of an infinitival complement weighs most heavily, as previously suggested by Mondorf (2003). In general, however, syntactic factors have the smallest impact of the three factor groups under scrutiny. Further, the complex interplay of factors revealed in Hilpert's analysis prompts him to challenge the scope of Mondorf's *more*-support hypothesis, arguing that postulating a universal compensation strategy is reductive and simplistic (Hilpert 2008: 412). Comparing the effects of the predictors in their comparative and superlative models, Cheung & Zhang (2016: 579) conclude that "'*most*-support' does not seem to fare as well as *more*-support". This indication motivates the empirical analyses that will be presented in Chapter 3 of this study. More generally, these multifactorial studies highlight the conceptual premise that gradation strategy choice is indeed a probabilistic and multicausal phenomenon.

Although speaker variables have received increased attention in recent corpus-based investigations of the alternations in focus, our knowledge of this factor group is still limited. In a study situated at the intersection of computer-mediated communication and variationist sociolinguistics, LaFave (2016) analyses corpora of different formats (written, spoken, instant messaging) to measure the effect of a speaker's age and education

on the choice of comparative and superlative variant. Higher levels of educational attainment are found to correlate with a more frequent use of analytic forms and females are more prone to use synthetic forms than their male counterparts. The latter observation ties in with most recent findings by Säily, González-Díaz & Suomela (2018) for English working-class females but appears at odds with the account by Grofulović & Jovanović (2016), who do not report any significant divergence in the strategy preferences of women and men. While LaFave's (2016) study is to be credited with identifying a set of variables that have largely been neglected in previous studies, his findings must be interpreted with caution for two reasons: His analysis is indifferent to the distinction between comparative and superlative forms; moreover, the impact of speaker-related variables is not quantified in relation to linguistic predictors. Empirical support has further been provided for the hypothesis that textual informality is correlated with analytic gradation (Mondorf 2009b: 188-190; Watanabe & Iyeiri 2020).

Psycholinguistic studies have further investigated the acquisition of the degree forms. Graziano-King & Smith Cairns (2005) propose a three-stage L1 acquisition model for the comparative: In stage 1, approximately up to the age of four or five, L1 learners list comparatives in the mental lexicon without any awareness of their morphological makeup. Since the comparative forms children encounter in the early stages of the acquisition process are mostly those of high-frequency monosyllabic types (e.g. *small, big, nice*), a phonological group with a virtually deterministic preference for *-er*, their mental records almost exclusively contain synthetic tokens. By the time children have amassed a sufficiently large repertoire of comparative types and tokens, distributional analyses allow them to abstract a suffixation rule for comparative formation; this insight marks the onset of stage 2. Consequently, overgeneralization of this rule occurs: Even low-frequency and polysyllabic adjectives will be graded by suffixation. As the learner's internal tallies are fed with input that appears at odds with the posited morphological rule, its applicability will henceforth be confined to adjectives for which they can retrieve sufficient positive evidence. The rule will be abandoned for types that are inconsistent with the regularities inferred from these mental records. This process of hypothesis testing eventually results in a regularized system of comparative formation which is partitioned according to frequency and phonological type,

with the analytic form acting as a default last resort if previous experience is too scarce to draw analogies with types that have an entry in the mental records. In this final stage, commencing approximately at the age of eight, L1 learners have developed an adultlike system of comparative formation. In the individual case, however, the proposed stages appear to be somewhat fluid (Graziano-King & Smith Cairns 2005: 371).

These insights have sparked interest in the differences between L1 and L2 processing of comparatives forms (Kırkıcı 2012; Hohaus, Tiemann & Beck 2014). Kırkıcı's (2012) elicitation study involving L1 Turkish learners of L2 English and an English L1 control group supports the view taken by Ullman (2001, 2005) that L2 processing relies more heavily on the storage of single-word units in the declarative memory system; L1 processing, on the other hand, draws on the procedural memory system and decomposes comparative forms into their component morphemes according to the rule-based generalizations described in the previous paragraph. This account ties in with the findings presented by Clahsen & Felser (2006) and Silva & Clahsen (2008) who report that L2 grammatical processing is characterized by more shallow parsing than L1 processing. An L2 learner's processing further seems to be subject to developmental change: The more proficient learners become, the more sensitive they are to morphological parsing and the less they rely on lexical storage (Kırkıcı 2012: 13).

1.2 Scope and structure of this study

The present work is structured as follows: Chapter 2 treats the English superlative from a theoretical perspective. A contrastive analysis of the comparative and superlative reveals systematic differences between the two degree constructions on the syntactic, semantic and idiomatic levels. These insights imply that the superlative alternation should be taken as a distinct instance of morphosyntactic variation rather than assuming a priori that the phenomenon in focus can sufficiently be accounted for by the mechanisms that have been shown to govern comparative strategy choice. Further, my considerations suggest that the English superlative is by itself far from clear-cut on the conceptual plane: A survey of the three superlative readings identified by Claridge (2007a), namely *relative*, *absolute* and *intensifying*, demonstrates that the distinction between these interpretations hinges primarily on semantic and pragmatic criteria while syntactic

clues such as the absence of an explicit set of reference or the definite/indefinite article can merely serve as a vague guideline. I will argue that the high level of ambiguity between these interpretations must be carefully considered in any investigation of superlative strategy choice.

The two studies that will be reported in Chapters 3 and 4 differ from each other along three major dimensions, namely in terms of (i) the predictors under analysis, (ii) the overall methodological setup and (iii) the theoretical frameworks that will be invoked to account for the findings.

The questionnaire study in Chapter 3 quantifies the relative effects of selected contextual constraints and speaker variables on the alternation. As for the former group of factors, the analysis zooms in on phonological, syntactic and pragmatic constraints whose effects have previously been interpreted along the lines of analytic support (Mondorf 2003, 2009b). The investigation will further shed light on effects of morphosyntactic persistence in superlative strategy choice. In addition, I will address the question of whether AmE makes more extensive use of the analytic superlative than BrE to compensate for effects of contextual complexity, as can be hypothesized on the basis of Mondorf (2009b: 171-193). To conclude this chapter, I will argue from a methodological perspective that elicitation studies constitute a valuable complement to the corpus studies that have been conducted on the alternation in focus in that they allow for the operationalization of linguistic and extralinguistic variables in a controlled setting. In interpreting the results of this study, I will rely on the theoretical frameworks of cognitive-functional linguistics and variationist sociolinguistics.

The corpus study that will be presented in Chapter 4 investigates the interplay of superlative frequency and the formal characteristics of the adjective (i.e. the number of syllables and final segments). The findings will be interpreted as evidence in favor of a usage-based connectionist view of superlative strategy choice. In this perspective, adjectives are hypothesized to be represented in an exemplar network whose structure can be defined in quantitative terms by the predictors covered in this study. Crucially, the structure of this network persistently changes upon experience with comparative and superlative forms. The results of this study will further allow an assessment of the role of frequency in the processing and storage of synthetic and analytic superlatives. To continue the crossvarietal approach pursued in Chapter 3, I will also examine whether my corpus

data hint at systematic differences between British and American English. For a comparison of superlative and comparative strategy choice according to gradation frequency, I will rely on the comparative data provided by Sönning & Hartmann (2019) as reference points.

Chapter 5 summarizes the findings of the present work and highlights questions and problems that should be considered in future studies of the comparative and superlative alternations. Discussing the merits and challenges of the methods used in the current investigation of superlative strategy choice, I will further assess the opportunities of methodological triangulation in the research on grammatical variation.

2 A conceptual approach to the English superlative

If we set out to account for the observation that the comparative alternation has been investigated both in wider scope and in greater detail than its superlative counterpart, a contrastive analysis of the syntactic and semantic idiosyncrasies of both alternation phenomena appears worthwhile. Huddleston & Pullum (2002: 1161-1162) point out that comparatives are exclusively used to form *term comparisons*, i.e. comparisons between a primary term and a secondary term, with the former superior[6] to the latter regarding the semantic property denoted by the adjective.

(1) Some of the countries that receive funding have much **larger** economies than most of the donor countries. [GloWbE US 410]
(2) The rapidly rising NPLs may soon be **larger** than the announced $158 billion of "planned" local government stimulus. [GloWbE US 2159]
(3) If he looks at the **bigger** picture, he will do what's best and step aside. [GloWbE US 38422]

Regardless of the syntactic function of the comparative, both attributive (1) and predicative (2) constitute term comparisons in which *than* is followed by the secondary term. These two examples further differ in that the secondary term in (2) is specified in a unit of measurement (a unit of currency, to be more precise), while in (1), the primary term is exclusively defined as relative to the secondary term, without any reference to an external scale; we do not know how large the respective economies are in absolute terms (e.g. judging by their GDP). At first glance, (3) does not represent an instance of term comparison, since it does not include a primary and secondary term; formulaic comparative + N combinations of this kind nevertheless allow for the comparative to be interpreted along the lines of a term comparison: The secondary term, i.e. the context or situation that is thought of as more specific than the one labelled as "bigger", has to be inferred from the co-text. These subtle differences between

[6] This chapter is not concerned with inferior comparatives (Andersen 1983: 100; Heine 1997: 109), i.e. comparatives involving the degree marker *less*, for which there is no synthetic equivalent.

comparative constructions do not have an impact on the syntactic and semantic categories by which they are classified: The exclusive function of the comparative is to determine the superiority of any given term to another with regard to a particular property in the absence of an external frame of reference.[7] As to their syntactic function, the vast majority of comparatives are either attributive or predicative.[8] It is this semantic and syntactic transparency and uniformity of the comparative that has allowed for the relatively straightforward operationalization and binary coding of syntactic predictor variables in previous studies.

The superlative, in contrast, is conceptually less clear-cut despite its formal similarity to the comparative (cf. Heine 1997: 109, 124-125).[9] This impression of subjectivity is supported by the fact that Claridge (2007a), in the most comprehensive study of this degree of comparison to date, distinguishes three 'interpretations' of the superlative: relative, absolute and intensifying superlatives. The present section aims to provide a differentiated account of these interpretations, starting with the fairly unambiguous relative use (Section 2.1). Sections 2.2 and 2.3 are concerned with the semantically and syntactically overlapping absolute and intensifying

---

[7] This theoretical sketch of the comparative does not cover the comparative correlative construction (Fillmore, Kay & O'Connor 1988; Culicover & Jackendorff 1999; Hoffmann, Horsch & Brunner 2018; Hoffmann 2019, *inter alia*), e.g. *The greater the risk, the greater the return*. The use of the comparative in this construction is special in that it does not fit into the syntactic and semantic mould of term comparisons as defined in this section. Previous studies of the comparative alternation, notably Mondorf (2009b: 12), exclude comparatives in this syntactic setting due to assumed effects of morphosyntactic persistence (cf. Section 3.2.3).

[8] It shall be noted that comparatives can also occur in postnominal position, in which case they behave similarly to predicative comparatives (cf. Mondorf 2009b: 63-70). This syntactic function is excluded from analysis in many studies on comparative variation (e.g. Hilpert 2008, Cheung & Zhang 2016). Lindquist (2000: 129), although presenting an example in which the comparative occurs postnominally, also restricts his analysis to the positional dichotomy attributive vs. predicative. This appears expedient due to Quirk et al.'s (1985: 418) observation that a postnominal adjective "can usually be regarded as a reduced relative clause" in which the relative pronoun and copula are missing: something *useful* ~ something *that is useful*.

[9] The semantic vagueness and multidimensionality of the superlative is reflected in the different labels that have been applied to its uses in theoretical treatises. Rusiecki (1985), for instance, subsumes Claridge's (2007a) absolute and intensifying uses under the unitary notion of "absolute". The present thesis adopts the tripartite distinction advanced by Claridge (2007a).

interpretations of the superlative, respectively. In the final subsection (Section 2.4), I present the implications of the divergent interpretations for the operationalization of predictor variables in the empirical part of the present study. Throughout this section, it will become evident that adjective gradation is a semantic process in the first place, with significant implications for the syntactic structures in which the superlative is embedded. I am not concerned with quantifier uses of *most*; Gergel (2015) provides a thorough examination of the diachronic trajectory the quantifier has taken from its Old English status as a lexical gradable adjective meaning 'largest (in degree)'.

2.1 Relative superlatives

Jespersen (1949: 392) notes that "in ordinary usage the superlative does not indicate a higher degree than the comparative, but really states the same degree, only looked at from a different point of view." This observation points to the use of the superlative as a means of set comparison (Huddleston & Pullum 2002: 1161-1162), i.e., in (4), a comparison between a referent and a set of at least three individual members, with the referent *she* ranked at the top of the scale of friendliness for the set of reference comprised by everyone in school.

(4) She had no friends, yet she was the friendliest person **in school**. [GloWbE US 16898]

(5) The food is excellent and the staff is the friendliest and most pleasant **I've ever experienced around here**, seriously. [GloWbE US 324006]

(6) **My** most interesting **patient** was a retired executive in his late seventies who had agonizing arthritis. [GloWbE JM 1824932]

The presence of restrictive modification is key to the relative interpretation of a superlative. The intended set within which the comparison takes place can syntactically be realized in different ways. In Claridge's (2007a: 141, Table 5) data, postmodifying prepositional phrases (as in (4)) and relative clauses (as in (5)) are about equally frequent in this function (approximately 35 % each), with determiners (such as possessive *my* combined with the nominal head *patient* in (6)) making up approximately

23 % of restrictive modifiers. Furthermore, relative superlatives are frequently premodified by modal adverbs such as *probably* or *perhaps*, which adds a subjective tone of doubt about the set comparison and slightly blurs the boundary between relative and absolute interpretations (cf. Claridge 2007a: 132).

From a syntactic point of view, superlatives can occur in three functions. More than two thirds of superlatives in Claridge's (2007a: 129) data are used in attribution, which ties in with Rusiecki's (1985: 135-136) findings;[10] 19 % appear predicatively, i.e. as a subject complement following a copular verb and without a determiner, as in (7); 11 % are fused-head noun phrases (cf. Huddleston & Pullum 2002: 415-416) and therefore nominal, as in (7), where the superlative is head of the noun phrase and preceded by the definite article.

(7) So, which city emerged as **the most polite** and which as **the rudest**? [GloWbE CA 673121]

(8) your mind's **most absorbent** when you're up to the age of five (Claridge 2007a: 129)

It is important to note that this breakdown of syntactic functions does not take into account semantic criteria. Instances like (8), where the subordinate clause introduced by *when* provides restrictive modification for the predicative superlative, are fairly unusual for set comparisons. The predicative position with zero determiner is assumed to be the marked domain of intensifying superlatives (cf. Rusiecki 1985: 140; Quirk et al. 1985: 466), particularly with analytic ones. This elative use of the superlative form is further elucidated in Section 2.3. Huddleston & Pullum (2002: 410) define "[F]used-head NPs [as] those where the head is combined with a dependent function that in ordinary NPs is adjacent to the head, usually determiner or internal modifier"; an alternative to the fused-head analysis of instances like (7) is advanced by Günther (2018), who, in her contrastive account of German and English adjectives used as nouns, argues in favor of ellipsis (see also Olsen 1988 on this issue). I do not intend to argue in

[10] In reference to Croft (1991) and Taylor (2002), Claridge (2007a: 130) argues that attribution is the primary function of adjectives, as the Latin etymon *adicere* implies. If we accept that the primary use is also the more frequent one, this certainly holds true. See Matthews (2014: 1-36) for a structuralist investigation of adjectival functions.

favor of either position at this stage; it should be noted, however, that previous studies on the superlative alternation have arguably failed to pay due attention to these intricacies at the intersection of syntax and semantics in their classifications. In one of the earliest corpus analyses of the comparative and superlative alternations, Lindquist (2000: 127) does not consider the possibility of fused-head NPs and refers to the superlative in the example sentence *The car bomb was* ***the deadliest*** *this year in Madrid* as predicative. Others have followed suit on this classification of syntactic functions (e.g. Cheung & Zhang 2016). In the only multivariate analysis of the superlative alternation to date, they adopt most predictor variables that have been operationalized in Hilpert's (2008) analysis of the comparative, for which the binary conceptualization of the syntactic function is adequate, as demonstrated at the beginning of this chapter. Despite the ambiguities that may arise in the classification of predicative and fused-head nominal uses of the superlative, I argue that the distinction between the three categories set forth above is crucial to the operationalization of the syntactic function as a predictor variable in any study of the superlative.

2.2 Absolute superlatives

Based on the finding that just about half of the superlatives (49 %) in Claridge's (2007a: 140) data are found to be relative, she suggests that factual comparison is not the principal function of the superlative form. This subsection addresses absolute interpretations of the superlative, which make up 42 % of superlatives in her data.

In Claridge's (2007a: 133) definition, the notion of universality is the defining feature of the absolute interpretation:

> Absolute [...] means that no restriction is present in the cotext, producing an unqualified and thus potentially universal superlative. If the frame is the whole world/the speaker's or hearer's world knowledge (i.e. the speaker's epistemic world), the limits are either extremely high [...] and/or vague, as well as often unverifiable.

This definition appears intuitive at first sight; the vagueness of the notion of universality, however, requires clarification. It is hard to determine objectively whether the scope of reference is universal or restricted. Consider the following examples:

(9) This is the most gorgeous place **in the world**.
(10) Lisa is the loveliest girl **in school**.
(11) Mount Everest is the **highest mountain**.

In (9), the set of reference is syntactically realized by a prepositional phrase and it is clearly the greatest possible, therefore universal (the most gorgeous of all places imaginable). While identical from a syntactic point of view, (10) differs from (9) in that other potential points of reference may lie outside the scope defined by the (restrictive) prepositional phrase; in other words, Lisa is ranked top on the scale of loveliness relative to the set of reference comprised of all girls in school, but other girls outside school may be lovelier than her. Both (9) and (10) clearly have an emotive or evaluative component. Rusiecki (1985: 137-139) does not distinguish between instances like (9) and (10), subsuming both under the relative sense; he applies the term 'absolute' to instances that Claridge (2007a: 133) considers intensifying, as will be discussed shortly. Although the frame of reference in (9) is the whole world, which would, on a semantic level of analysis, warrant the label 'absolute' in Claridge's terms, it would seem advisable to refer to any superlative with a syntactically explicit scope-qualifying expression as a relative superlative, regardless of the aspect of universality. The superlative in (11) is absolute by Claridge's (2007a: 133) definition in that it draws on the recipient's world knowledge to determine the universal set of reference, which consists of all mountains on planet Earth. From an exclusively semantic point of view, Sharvit & Stateva (2002: 454) are to be credited with the observation that relative and absolute interpretations only differ in the way in which the scope of reference is created. Example (12), drawn from Sharvit & Stateva (2002: 454), is ambiguous between the relative and absolute interpretation.

(12) John climbed the **highest** mountain.

The absolute reading implies that the mountain John climbed is higher than any other mountain; the relative reading focusses on the climbing achievement of John, i.e. he climbed a higher mountain than any other climber. The distinction between relative and absolute interpretations is therefore strongly determined by the context, and syntactic criteria can merely serve as a vague guideline here. Claridge (2007a: 137, 141-

142) points out that some syntactic structures, particularly prepositional phrases, do not restrict the scope of reference, but rather have the opposite effect of extending it to its extremities, e.g. *in the world*, *in memory*, *of all*; thus, restrictive phrases of this kind push the superlative towards intensifying readings.

"[U]niversal or modified existential quantified readings" (Fauconnier 1975: 356) similarly blur the boundary between relative and absolute interpretations. In cases like (13), in which the root adjective, due to its semantic properties, is located towards one end of a scale, negation can imply quantificational *any*-like readings ("People talking about Clinton don't have *any* clue…"). Such superlative + N bigrams are likely to be entrenched as idioms, in which case the superlative alternation is significantly restricted.

(13) People talking about Clinton **don't have the slightest clue** as to the agreement that the prosecutors had with Clinton. [GloWbE US 29454]

(14) My thoughts are with you and I offer you **my sincerest condolences** for this loss. [GloWbE GB 3030781]

(15) Look at it this way: most people know that Brown is **not the most pleasant** person to be around. [GloWbE GB 3050994]

The formulaic character of superlative + N bigrams is also apparent in the non-negated absolute reading (14), where the superlative reinforces the emotive and emphatic load of the utterance. Interpretations of this kind shade into intensifying uses, since the notion of extremity clearly does not involve any comparative scale here, but merely underscores the emotionality of the statement. (15) further highlights the fact that the negated superlative can also be employed for stylistic and rhetorical purposes. A negated adjective with a positive denotation is used to understate a point that could be made using an adjective with a negative denotation (e.g. *awful*, *terrible*, in this context). This stylistic device of *litotes* provides a more polite and less explicit alternative.

The semantics of adjectives deserve a provisional examination at this point. Claridge (2007a: 127) observes that regardless of their morphological realization, adjectives in the superlative form "are either inherently and primarily evaluative, have a second(ary), but established evaluative

meaning component or can fairly easily be contextually filled with attitudinal meaning." This observation, previously advanced by Leech & Culpeper (1997: 368-370) in more intuitive terms, holds particularly, though not exclusively true for absolute superlatives. The adjective *blunt* is a case in point: In the Corpus of Contemporary American English (COCA, Davies 2008–), a simple query for *bluntest* + N renders 12 hits, 11 of which can clearly be classified as evaluative in the sense of 'dull, insensitive', e.g. combined with *language*, *statement* and *assessment*. The odd one out here is *bluntest angle* in a mechanical context. The equivalent query involving the analytic superlative yields only a single hit in a distinctly non-evaluative context; combined with *tool*, the adjective has the meaning 'not sharp, obtuse'. An observation of this kind, though by no means conclusive, of course, licenses the tentative hypothesis that evaluative superlative + N bigrams can be accounted for along the lines of Sinclair's (1991) *Idiom Principle*; they are acquired and represented as a single lexical chunk whose meaning cannot be compositionally derived, as a construction in the conceptual framework of Construction Grammar (Goldberg 1996: 68, in particular). This assumption of idiomaticity has profound implications for the superlative alternation. Idiomatic superlative + N bigrams with evaluative connotations may exhibit a categorical preference for one of the two morphological variants, possibly the synthetic superlative, as hypothesized by Leech & Culpeper (1997: 369-371).

The final type of absolute superlative to be mentioned here is the superlative-*of*-N expression (cf. Claridge 2007a: 128). The syntactic analysis of this string is far from straightforward. Leech & Culpeper (1997: 369) use the term (*the*) Superlative + *of* + Noun Phrase construction and commit to the noun as the phrasal head. In (16), the superlative is part of a fused modifier-head in a Noun Phrase (cf. Huddleston & Pullum 2002: 538) and thus structurally similar to relative interpretations realized by fused-head noun phrases postmodified by a restrictive prepositional phrase as defined in the previous subsection. (18) is a superlative-*of*-N expression with no determiner immediately preceding the noun *people*.

(16) She has answered **the most important of your criticisms**.
(Huddleston & Pullum 2002: 538)

(17) *? **Of your criticisms** she has answered **the most important**.

(18) I met **the loveliest of people** there, made some good friends, [...] [GloWbE IN 812580]
(19) * **Of people** I met **the loveliest** there.

In both (16) and (18), the fusion of superlative and noun is strong. Extraction of the prepositional component is not possible in (19) and at least questionable in (17). Although the use of the preposition *of* implies a restriction of the scope of reference in (18), which would warrant a relative interpretation, the scope of reference is clearly universal, i.e. comprising everyone. The semantic classification as a fused modifier-head in an NP and concomitant ellipsis of the determiner *all* appears to offer a valid description of this construction. The syntactic ambiguities encountered here are similar to those raised in the context of (nominal) pseudo-partitive constructions (two nominal elements separated by the element *of*, e.g. *a pint of beer*; cf. Selkirk 1977; Keizer 2007: Chapter 6).

In Leech & Culpeper's (1997: 369-370) data, superlatives in the superlative-*of*-N expression are predominantly synthetic (77 %). At first glance, it appears inconceivable why these results support the authors' "hunch that the inflectional superlative is more likely to have [an] emphatic role than the periphrastic superlative". This runs counter to pragmatic accounts which conceive of grammatical weight simply as a matter of string length: The lexically more bulky analytic variant is more suitable for creating emphasis than the synthetic superlative, since it comes with the additional lexeme *most* (cf. Lindquist 2000: 126; Mondorf 2009b: 99-107). A simple query of superlative-*of*-N strings in the British National Corpus (BNC Consortium 2007) and subsequent analysis of the frequency breakdown yields the following tentative observations:

- The adjective types used in this expression tend to be rather short and frequent; these morpholexical properties have been shown to correlate with increased odds of the synthetic comparative and superlative (Hilpert 2008; Cheung & Zhang 2016).
- Semantically, most adjective types used in this expression are inherently emphatic and subjective. 13 of the 20 most frequent collocations contain the superlative *best* (cf. example 20). The nouns following *of* are mainly monosyllabic, as in (20), or disyllabic with initial stress, as in (21).

(20) Well [pause] we're not the **best of friends**, she just takes the piss. [laugh] [SP:PS52C]
(21) Mrs Thatcher gives Mr Gorbachev the **warmest of welcomes** as they start their talks at Number Ten. [SP:PS5XR]

Despite the lack of adequate empirical support, I advance the assumption that superlative-*of*-N strings are a showcase of a construction in the sense of usage-based Construction Grammar (Langacker 1987; Fillmore 1988; Goldberg 1995, 2003; Ellis & Ferreira-Junior 2009; Tomasello 2003); therefore, I will henceforth refer to this expression as a *construction*. A Superlative-*of*-N string is a construction with the variable slots N and Superlative. These slots can be filled with lexical material which is to some extent formally and functionally conditioned by the construction itself (cf. Hilpert 2014: 7). The filler material is further influenced by its degree of similarity to a prototype on which the construction hinges (cf. verb-island constructions in the usage-based sense of Tomasello 1992, 2008). The above observations warrant the tentative hypothesis that this prototype may be the superlative *best*. A prosodic effect presumably plays a role here, too: The buffer syllable *of* prevents a stress clash between *best* and the following initially stressed noun and thereby maintains the "ideal" alternating rhythm (cf. Schlüter 2005). It appears legitimate to object that all synthetic superlative types other than the suppletives *best* and *worst* are at least disyllabic and thus do not require a prosodic buffer before the following noun. In order for the construction itself to become entrenched (cf. Langacker 1987: 59 for the cognitive notion of *entrenchment*), however, only the token frequency of the prototypical superlative matters (cf. Tomasello 2008 in the context of the ditransitive construction). It is irrelevant at this stage that other synthetic superlative types are not subject to this prosodic effect. Once the construction is sufficiently entrenched, its productivity depends on the degree to which the variable slots allow for variation. The greater the formal and functional similarity of a prospective type to the prototypical superlative, the greater are the odds that it qualifies for the Superlative slot.

A closer look at the noun slot seems to support the speculative view that the productivity of the construction under investigation is indeed subject to constraints posed by the indecomposable unit of form and function (cf. Goldberg 1995: 4) inherent in the Superlative-*of*-N string. The nouns that qualify for this slot are predominantly used in the plural and have an

abstract meaning. Singular marking is possible but limited to abstract nouns that do not allow plural marking (e.g. the best of *luck*, the greatest of *respect*). It is therefore noteworthy that there is not only variation in the lexical items that can appear in both variable slots, but also in the morphological units that instantiate them (singular vs. plural noun in the N slot, synthetic vs. analytic marking in the superlative slot). The frequency of an adjective-noun combination in the positive form also appears to be correlated with the odds with which it is used in this construction (*good luck*, *great respect*). Hilpert (2014: 14-22) establishes four defining criteria for a construction. To conclude the line of argumentation pursued in this chapter, the Superlative-*of*-N construction shall be examined with regard to these criteria.

(i) *Deviation from canonical patterns*: The construction under investigation deviates from canonical grammatical patterns in that an adjectival form functions as head of a noun phrase postmodified by the preposition *of* and a nominal scope qualifier. Moreover, ellipsis of the determiner *all* immediately before the N slot imposes the notion of universality on the noun in this slot.

(ii) *Non-compositional meaning*: The meaning of the Superlative-*of*-N construction is more than the sum of its parts, as denoted by the principle of *coercion* (Michaelis 2004: 25). The expression *the loveliest of people* is not equivalent to *the loveliest people*, nor does the preposition *of* serve its default functions as a marker of possessive or partitive relations. The construction coerces the denotations of *lovely* and *people* as well as the grammatical functions of the preposition *of* into a meaning that is hard to approximate by periphrasis, as its compositional semantics are, among other aspects, subject to pragmatic constraints of the communicative situation.

(iii) *Idiosyncratic constraints*: Superlative-*of*-N constructions are idiosyncratically constrained on several levels of linguistic analysis. The noun slot can, for instance, only be filled by plural nouns or mass nouns with an abstract meaning. It also appears plausible to assume that the morphological realization of the superlative within this construction is disproportionately biased towards either strategy. This question, however, is not empirically addressed in the present study.

(iv) *Collocational preferences*: The explorative corpus queries described earlier in this section suggest that certain adjectives and nouns are used more frequently than others in the construction examined here.

The tentative considerations presented in this section entail the conclusion that the formal and functional features of the superlative in this construction merit further empirical engagement. In this context, aspects of frequency and phraseology certainly deserve as detailed an examination as differences between spoken and written language. If formal (written) genres are indeed found to be the primary domain of the construction under scrutiny, as hypothesized by Claridge (2007a: 137), this may at least partly explain the assumed predominance of the synthetic superlative. Szmrecsanyi (2009: 336-337) shows that various grammatical markers are significantly more synthetic in writing than in speech, invoking functional constraints and communicative demands to account for these findings. While written texts maximize output economy and thereby possibly increase processing complexity on behalf of the reader, spoken texts incur output diseconomies, or, as Chafe (1982: 37) puts it rather vividly: "In writing we have time to mold a succession of ideas into a more complex, coherent, integrated whole, making use of devices we seldom use in speaking." In analogy to Brems (2007: 297), who posits that the grammaticalization of size nouns (e.g. *a bunch of*, *loads of*) is spearheaded by the highly frequent prototype *a bit of*, it remains to be empirically tested whether the assumed prototype *the best of* N may be the center of gravity of this construction, analogically casting more infrequent, formally similar Superlative-*of*-N expressions in the same pragmatico-semantic mold. Her reassessment of Hoffmann's (2004) study of complex prepositions yields an important methodological caveat for the future operationalization of frequency variables in an empirical engagement with this construction: It is imperative that any such analysis be sensitive to the distinction between the frequency of the Superlative-*of*-N pattern as a whole and the frequency of the lexical types that instantiate the variable slots (cf. Brems 2007: 297).

The constructionist analysis set forth above highlights the fact that no single domain of linguistic investigation can solely account for the distributional mechanisms that underlie the formation of the Superlative-*of*-N construction – an idea that challenges the dictionary-and-grammar model

(Taylor 2012), which conceives of idioms as separate entities stored in an 'appendix' to the mental dictionary.[11] Ultimately, the Superlative-*of*-N construction exhibits traits of both absolute and intensifying interpretations: It sets up a universal scope of reference against which the superiority of the nominal referent regarding a particular adjectival quality is expressed and serves as a vehicle for emotion and emphasis in specific communicative settings.

2.3 Intensifying superlatives

It has already been noted that previous scholarship is inconsistent as to the conceptual and terminological distinction between absolute and intensifying interpretations of the superlative. Schibsbye (1970: 139), Rusiecki (1985: 137-139) and Leech & Culpeper (1997: 369) merge these two notions into the term "absolute" and thereby disregard the cognitive and semantic intricacies pertaining to the scope of reference elucidated above. Farkas & Kiss (2000: 437) proceed similarly, yet in conceiving of instances like (4c) (the Mount Everest example in the previous section) as "'absolute' absolute superlatives", they demonstrate sensitivity to the aforementioned problem; this label implies that along the lines of the terminology adopted in the present work, absolute superlatives are conceptually close to intensifying ones.

German philology, deeply rooted in the traditional grammar of the classical languages, has established the term *elative* for the main type of intensifying superlative to be discussed in this section. Since the elative is formally and functionally distinct from the superlative in settings involving comparisons, accounts of the German elative commonly treat this interpretation as a grammatical concept in its own right. This may be partly because the German elative is morphologically more diverse and firmly established in language use. Claridge (2007a) refers to this phenomenon as the intensifying function of the superlative, Schibsbye (1970: 139) uses the label "superlative used without the idea of comparison". Elatives are markedly different from the interpretations previously explored in that they do not require the definiteness of relative and absolute superlatives

[11] See Hilpert (2014: 3-8) for a summary of constructionist arguments against the dictionary-and-grammar model.

(cf. Koller 2007: 8-9; Fuhrhop & Vogel 2010: 90).[12] They can occur both in attributive and predicative positions. The predicative example (22) lacks a determiner and is therefore ambiguous between absolute and intensifying interpretations ('the loveliest of all' vs. 'very lovely'). Preceded by the definite and indefinite article, respectively, (23) and (24) can each be conclusively assigned to one of the two categories.

(22) Lisa is **most lovely**.
(23) Lisa is **the most lovely** girl.
(24) Lisa is **a most lovely** girl.

The indefinite article in premodification provides a knockout criterion for the synthetic superlative (* Lisa is **a loveliest** girl.). This renders *most* semantically equivalent to intensifying adverbs like *very* and thereby imposes a limit on the alternation investigated in this work (cf. Quirk et al. 1985: 466);[13] intensifying superlatives are thus categorically analytic in the sense of Rosenbach's (2002: 27) distinction of categorical and choice contexts. Diachronically, the intensifying capacity of *most* may be a remnant of its exclusively lexical status as a gradable adjective in Old English, when it grammaticalized as a degree marker in analytic superlatives (cf. Gergel 2015). In the previous section, I have shown that most absolute interpretations of the superlative have an emphatic or affective force, too. The situation is particularly intricate in cases where an implied or syntactically explicit universal scope of reference is used as an emotive and emphatic device. In line with Claridge (2007a), I consider only those instances intensifying uses that can be paraphrased by degree-modifying adverbs.

2.4 Implications for an empirical investigation of the superlative alternation

The insights gleaned in the present chapter have clear implications for any empirical engagement with the superlative. Its uses are primarily de-

[12] Koller (2007: 8) points out that morphosyntactically, the superlative in its comparative function consists in the superlative form predetermined by the definite article.
[13] Note, however, the following idiosyncratic example found by Claridge (2007a: 139): *That was rather a cheekiest thing to do.*

termined by the semantics of the root adjective and the pragmatic concerns of the communicative situation, i.e. its inherently emphatic and emotive load. Further, the superlative is conceptually more diverse than the comparative. The binary classification of syntactic functions (attributive vs. predicative) applied in previous studies of the superlative falls short of addressing the semantic and pragmatic intricacies that have been highlighted. The superlative in a fused-head noun phrase in the presence of restrictive modification is the default opposition to the attributive function in relative readings, while the predicative function is frequently ambiguous between different semantic interpretations. This is especially so if the context fails to provide an explicit set of reference. The presence of a syntactically realized scope of comparison, however, is not an infallible criterion for assigning a superlative to one of the two comparative categories (relative or absolute): Instead, the (non-)universality of a referential frame must be determined.

What sets the elative functionally apart from the superlative involved in comparisons is its indefiniteness. Syntactically, elatives are thus either determined by the indefinite article or realized by the predicative superlative without determiner. Notably, the alternation between synthetic and analytic forms is assumed to be significantly restricted due to the synonymity of *most* and default degree modifiers. These functional and morphological idiosyncrasies strongly suggest that research in this domain of variation must be attentive to the distinction between relative and absolute readings on the one side and intensifying interpretations on the other. Any empirical investigation of superlatives must further consider the pragmatico-semantic and stylistic limitations posed by the genre in which they prevail. Finally, phraseological effects seem to condition the uses of the superlative to a greater extent than those of the functionally rather uniform comparative.

3 Contextual and social predictors of superlative strategy choice: Questionnaire-based insights

The questionnaire study that will be presented in this chapter sheds light on the influence of seven contextual constraints and four speaker variables on the superlative alternation. After a consideration of the opportunities and challenges of a questionnaire-based approach to the superlative alternation (Section 3.1), I will provide a detailed account of the method used in this study (Section 3.2), explaining the structure of the questionnaire and the logistics of data collection. In this context, I will also introduce the individual predictors on the basis of previous research and highlight implications for the design of questionnaire items. Findings are presented in Section 3.3 and Section 3.4 concludes with a discussion.

3.1 Opportunities and challenges of a questionnaire-based approach to the superlative alternation

It is evident from the literature review in Section 1.1 that research into the comparative and superlative alternations has to date exclusively relied on corpus data, with the exception of Boyd (2007) and Kunter (2017), both of which have used experimental designs. The methodological preponderance of corpora in the study of gradation strategy choice springs from the prime interest in the morphophonological properties of the adjective type. The possibility to cover a wide range of adjective types is undoubtedly the greatest virtue of corpus approaches to this phenomenon of morphosyntactic variation. Further, the advantages inherent in the observational character of corpus data deserve a mention here. In the face of the obvious benefits of a corpus-based examination of the superlative alternation, it appears all the more necessary to consider the merits of questionnaire-based elicitation data in this research context. I will argue that the potential of this tool in the study of alternation phenomena has hitherto largely been underestimated.

The factors that have been identified as playing a role in gradation strategy choice can be grouped into four categories: (i) morphophonological properties of the adjective, (ii) frequency measures, (iii) constraints of the linguistic context and (iv) speaker variables. The present questionnaire study is concerned with the impact of factors that belong to the latter two

categories. First, I will throw into relief the shortcomings of corpus approaches to the research questions in focus and explain why a questionnaire-based elicitation study is a suitable method for the current empirical investigation. In doing so, I am guided by Dollinger's (2015: Section 3.1) comparative account of corpus and questionnaire methodologies. The first issue to be raised in this context is the widely discussed problem of corpus coverage. The validity and comparability of data extracted from a corpus depends on how it has been compiled, both qualitatively and quantitatively. It is obvious that a corpus analysis can only operationalize what the corpus manages to capture, and even the largest collection of linguistic material is limited in scope. Linguistic features may therefore not occur as frequently as would be necessary to allow for a thorough and reliable investigation. Indeed, previous studies of the comparative alternation have acknowledged this problem. Mondorf (2009b: 64, fn. 49) notes that the scarcity of her data does not allow for a reliable comparison of the effects of infinitival adjuncts and infinitival complements. Similarly, Säily, González-Díaz & Suomela (2018: 170) report that the number of adjective types that occur with an infinitival complement is very low in their Spoken BNC2014 dataset.

Questionnaires enable the researcher to design a prompt at their scientifically motivated convenience and consequently generate a great number of individual judgements on linguistic features that may be too infrequent in a corpus. Albeit one of its greatest assets, this research-tailored design option constitutes a conceptual drawback of elicited data: Informant judgements are always somewhat unnatural, regardless of how successfully the researcher manages to alleviate the effects of monitoring while the questionnaire is processed by the respondent. The *Observer's Paradox* (Labov 1972) addresses the problem that the logistics of data collection compromise the authenticity of elicitation data,[14] and hence their reliability; however, informants are aware that they are being monitored, or at least that their judgements will be evaluated, and this awareness may

[14] The problem of (lacking) authenticity can also be identified for corpus data. Stefanowitsch (2020: 24-25) argues convincingly that language ceases to be authentic as soon as it is sampled for inclusion in a corpus. In this process of decontextualization, a text loses features on various levels of linguistic analysis, most evidently those pertaining to the discourse situation (e.g. gestures, intonation). Therefore, even corpus material can at best be an approximation of authentic language use.

induce them to respond in a way that is socially desirable. The *social desirability bias* is arguably even stronger if a researcher is physically present when the questionnaire is administered. Furthermore, it is often impossible to determine whether a judgement represents a linguistic form that an informant merely passively accepts, uses in actual speech production or claims to use (Schütze 2016: 62). While the researcher can harness the visual clues of item design and explicit verbal instructions to control the cognitive influence of the setting in which data collection takes place, this scaffolding can barely eliminate the differences between elicited data and natural production data.

If we cannot rule out the eventuality that judgements are influenced by effects of social desirability, it remains to be discussed towards which superlative form the elicited data are likely biased in the present research context. Dollinger's (2015: Chapter 2) sketch of the history of questionnaire methodology suggests that for a long period of time, from the early adopters of this tool in the field of social dialectology in the late 19th century up to the turn of the millennium, questionnaires were primarily used to examine the geographic distribution of lexical variants and the variation of pronunciation across different locales. In both research domains, even laypeople are often consciously aware that for certain features that may be targeted in a highly monitored methodological setup there are standard and non-standard dialectal variants. Certain non-standard forms are socially stigmatized, and informants may therefore deliberately opt in favor of the socially accepted (prescriptive) standard even though this choice may not necessarily reflect their actual language use. Informants could, for instance, suppress *t*-glottalization in an interview with a field researcher or indicate a preference for *toilet* over *bog* when given the choice between these lexical variants in order to appear compliant with social conventions vis-à-vis the researcher.

For matters of morphosyntactic variation, and particularly the alternation phenomenon under study, the situation is somewhat different. There is, at least on the face of it, no standard that would render one superlative variant less acceptable than the other, nor is gradation strategy choice well-known to vary according to geographical locale or social categories. However, we have good reason to assume that L1 speakers of English are exposed to the same prescriptive rules on comparative and superlative for-

mation as L2 learners (cf. Section 1.1.2). In fact, some informants commented upon completion of the present questionnaire that they were explicitly taught morphophonological criteria for gradation strategy choice in school, while acknowledging that the linguistic context decisively influenced their judgements, too. This impression is in keeping with Schmidt & McCreary (1977), who suggest that many people are aware of prescriptive guidelines although they may not comply with them in actual language production. Couched within the generative paradigm, Chomsky's (1964) distinction between competence and performance addresses precisely this issue: What we do is often very different from what we know. Schütze (2016: 83) concludes that in some linguistic contexts, informants are aware of the difference between prescriptive standards and their intuitions; however, it remains a matter of debate how informants can successfully be prevented from drawing on prescriptive knowledge when making their judgements. Thus, it is imperative that a researcher deploying a questionnaire be aware of and try to alleviate this "irreconcilable tension between comparability and reliability" (Dollinger 2015: 56). In Section 3.2.1, I will explain how this methodological caveat affects the design of the present questionnaire in more practical terms.

A second major advantage of questionnaire studies over corpus analyses is that they enable a more reliable examination of speaker variables. Any study that intends to investigate the social correlates of linguistic phenomena must find a viable method of exhaustively documenting these details of interest. Even though most standard reference corpora used in linguistic research now come with some demographic information on the contributor of individual texts, this information is often too scarce, both in scope and quality, to draw empirically backed conclusions. The studies in Brezina, Love & Aijmer (eds., 2018) demonstrate the opportunities of the Spoken BNC2014 (Love et al. 2017), a single-variety corpus tagged for the categories age, gender and socio-economic status, in sociolinguistic research. Based on an early-access version of this corpus, Säily, González-Díaz & Suomela's (2018) account of the social correlates of the productivity of comparatives suggests that this group of variables indeed plays a role in the comparative alternation; however, they issue a caveat that again arises from corpus compilation (pp. 177-178): sampling imbalances across social categories as a consequence of an "opportunistic" sampling procedure (Love et al. 2017: 326-327). This is not to say that elicitation data

gleaned from a questionnaire are *per se* balanced across these categories, but the researcher can deliberately influence the recruitment process to adjust perceived imbalances. Put simply, if we are interested in differences between BrE and AmE, and our dataset contains considerably more British than American responses, we should get some more Americans to fill in our questionnaire. If properly informed, statistical models can cope with these imbalances; nevertheless, the reliability of the estimates will inevitably be constrained for variables that are unevenly covered in the dataset. In sum, questionnaires afford a high degree of control over the sampling process, and hence over the structure of the dataset; the researcher can harness these assets to avoid the problems posed by the potential scarcity and imbalance of corpus data.

3.2 Method

Now that the general strengths and weaknesses of my methodological approach have been considered, it is time to zoom in on the practical realization of the questionnaire method used in this study. This section walks the reader through the chronological steps of the elicitation method. Section 3.2.1 presents the design of the questionnaire and discusses the implications for the logistics of data collection. The account of the data collection method offered in Section 3.2.2 compares the written and web-based survey formats of the questionnaire, and discusses the challenges posed by mixed-mode data collection. The predictor variables are described in Section 3.2.3. Section 3.2.4 concludes the methodological considerations of this elicitation study with an outline of the techniques and tools used for data analysis.

3.2.1 Questionnaire design

The linguistic main part of the questionnaire consists of 120 items that were designed as a forced-choice judgement task. First, I defined a set of adjective types that were subsequently used for the wording of item prompts. Both the number of determinants and the methodological outline conditioned the choice of adjective types in previous synchronic studies of the comparative and superlative alternations. The studies by Hilpert (2008: 404) and Cheung & Zhang (2016: 568) allow for a large number of

comparative types and tokens to be included in the analysis. Drawing on the BNC, these studies apply different criteria to their selection of adjective types. Hilpert (2008: 404) includes all adjectives with at least a single token in both variants, acknowledging, however, that many of these adjectives "could be thought of as occurring in only one variant". Cheung & Zhang (2016: 567) eliminate adjectives that behave categorically in their data, i.e. types for which one variant accounts for more than 99.5 % or less than 0.5 % of all tokens. D'Arcy (2014: 225) recommends an even narrower range, stating that "standard practice within the variationist paradigm is to exclude forms with nearly categorical distributions (i.e., values higher than 95 percent and lower than 5 percent) when considering the operation of the variable grammar". A thoroughly devised method of adjective selection was all the more important in the present research setting since type-specific predictors are not the primary focus of this study. It was therefore advisable to ensure that (i) the selected types are not categorical by virtue of their morphophonological properties and (ii) no single phonological type (e.g. disyllabic adjectives ending in /-i/) will be overrepresented in the data.

Both quantitative and qualitative criteria served as a guideline for adjective selection. Cheung & Zhang's (2016) superlative dataset[15] gleaned from the BNC contains 161 adjective types with an analytic-synthetic ratio (ASR) between 0.005 and 0.995 after exclusion of elatives and quantifier uses of *most*. In the current study, this range was narrowed down to $0.05 < x < 0.95$ to zoom in on types which are still less prone to gravitate towards the categorical extremes; restricting the set of adjectives to types with at least two occurrences in both superlative variants yielded 68 types. The ASR of these types was determined in the COCA (Davies 2008–) using the interface provided at *english-corpora.org*, and types with fewer than two tokens in either variant were discarded; elatives and quantifier uses were not weeded out. This resulted in an overlap of 48 adjective types. Subsequently, qualitative criteria were applied to further reduce the number of types. Adjectives had to allow both attributive and nominal use in the superlative. In addition, care was taken that the set of adjectives exhibited some diversity with regard to syllable number and final segment to rule out that a particular phonological type would unduly influence the

[15] Thanks are due to Lawrence Cheung for sharing the dataset used in Cheung & Zhang (2016).

data. I further ensured that the selected adjectives were evenly spread across the range of overall superlative frequency (i.e. synthetic and analytic tokens combined). 19 adjectives (4 monosyllabic, 15 disyllabic) were ultimately selected according to these criteria. Another four adjectives (1 monosyllabic, 3 disyllabic) were added to the pool of adjectives for item formation even though their token counts and/or ASR violated the previously defined quantitative criteria for inclusion (*clever*: only one single analytic token and ASR = 0.04 in the BNC; *fierce*: ASR = 0.048 in COCA; *scary*: ASR = 0.027 in COCA; *tricky*: ASR = 0.029 in COCA). This decision was motivated by previous accounts of the comparative alternation that suggest a considerable degree of variation for these adjectives. Table 1 classes the selected adjectives according to their phonological type, i.e. the combination of syllable number and final segment(s).

Table 1. Morphophonological properties of the adjectives used for item construction

number of syllables	final segment(s)[16]	adjectives
monosyllabic	miscellaneous	*foul, blunt, scarce, fierce, cruel*[17]
disyllabic	/-i/	*tricky, risky, healthy, scary, happy*
	/-li/	*friendly, costly, lovely, deadly, lively*
	/-ə(r)/, /-ɪə(r)/, /-ɪr/	*clever, severe, sincere*
	/-t/	*polite, remote*
	miscellaneous	*handsome, profound, pleasant*

For practical reasons, adjectives ending in /-ə(r)/, /-ɪə(r)/, and /-ɪr/ are collectively referred to in the following as /-ə(r)/ adjectives. This group is admittedly heterogeneous, yet its members share the presence of a final /r/, which in non-rhotic varieties such as RP is only latent but may reappear in liaison contexts. In some transcription systems of GA, this /r/ is treated as r-coloring that permeates the preceding vowel (e.g. /ˈklevɚ/),

[16] The literature on the degree alternations is remarkably inconsistent as regards the notation of the adjectival ending. The different notational conventions, as well as their implications for the classification of individual adjectives, will be considered in greater detail in Section 4.5.2. In this chapter, I predominantly use phonological notation.

[17] The OED lists both /kruːl/ and /ˈkruːəl/ as phonological variants of *cruel*. In the current study, this adjective is considered monosyllabic.

rather than as an independent segment. Adjectives in this set also vary in rhythmic structure and stress placement, but are treated as one class here.

It is purely coincidental that seven of the 23 chosen adjective types constitute "crossovers" in Peters' (2000: 308-309) data, i.e. adjectives for which the comparative tends to be formed analytically, while the synthetic variant is preferred in the superlative, namely *risky*, *costly*, *deadly*, *friendly*, *cruel*, *polite* and *remote*. Further, four of the types included in the questionnaire were identified as adjectives for which analytic forms seem to be gaining ground in present-day English (Quirk et al. 1985: 462), namely *cruel*, *polite*, *handsome* and *pleasant*.

Informants saw sentences and were asked to make a choice between the two superlative variants. The visual presentation of the items is exemplified in Figure 1. Three response options were vertically aligned to the right of the sentence prompt, two of which represented the synthetic and analytic superlatives of the respective adjective. Informants could further indicate that both superlative variants appeared equally valid in the respective sentence by selecting the intermediate answer category of "no preference". A visually separate opt-out category was provided to the right of the main response options.

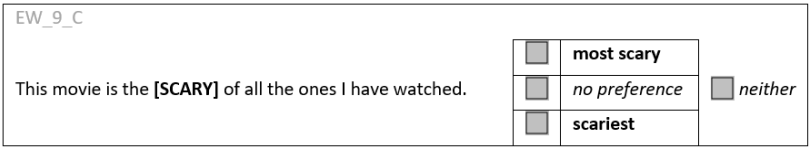

Figure 1. Control item EW_9_C in the written questionnaire

The inclusion of a neutral response option and an opt-out category is controversially discussed in accounts of survey and questionnaire methodology (cf. Bethlehem & Biffignandi 2012: 113-115). Kalton, Roberts & Holt (1980: 77) report that the availability of an intermediate response category increases the probability of informants expressing a neutral view. A similar conclusion is drawn by Sudman & Bradburn (1982) for the opt-out category. Efforts to discourage respondents from selecting this option by reducing its visual prominence have further been shown to be ineffective (DeRouvray & Couper 2002; Tourangeau, Couper & Conrad 2004). Although the instructions provided at the top of the questionnaire form familiarized informants with the item format and urged them to go by

their initial feeling, they still may have felt indifferent to the linguistic choice they were asked to make. Décieux et al. (2015) suggest that forcing respondents to opt for one of the two superlative forms increases the rate of premature break-off and decreases answer quality; I therefore decided to offer both the intermediate option and the opt-out category. The opt-out category was designed for the case that respondents consider a sentence prompt ungrammatical regardless of which superlative variant is inserted.

The pool of items includes 120 prompt sentences as illustrated in Figure 1. These items were organized into 60 pairs (56 of which were used in the present study). Each item pair included a "control" and a "test" item. The test and control items of a given pair contained the same target adjective. The control item was constructed to exhibit a relatively low level of complexity. It serves as a neutral baseline. Test items were derived by manipulating the control sentence with regard to a single contextual constraint, while making no other changes. This is illustrated in Figure 2, which shows the test item whose control counterpart was shown in Figure 1. The manipulated variable is *end weight*: In this test condition, the superlative is rendered in sentence-final position. This constraint is assumed to increase the probability of the *most*-superlative as compared to the control item EW_6_C, in which the superlative is presented sentence-medially (cf. Section 3.2.3 for a discussion of the end weight constraint).

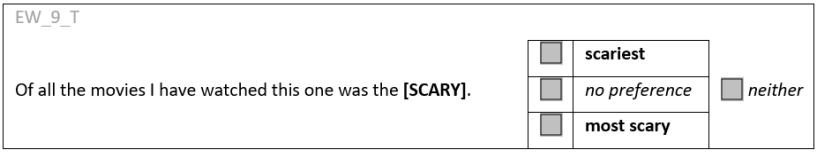

Figure 2. Test item EW_9_T in the pen-and-paper version

Since the only linguistically relevant difference between test and control item pertains to the end weight factor, the difference between the proportions of *most* in the test and control conditions can be attributed to the effect of the constraint in focus.

Table 2 specifies the contextual constraints operationalized in the test conditions.[18] These will be explained in greater detail in Section 3.2.3.

Table 2. Contextual constraints operationalized in the questionnaire study

Constraint	Expected tendency in test condition	Item pairs	Adjectives
Phonological			
Rhythmic harmony	synthetic	9	*fierce, clever, polite, friendly, sincere, severe, happy, profound, foul*
Segmental *horror aequi*	analytic	9	*lively, remote, clever, fierce, deadly, healthy, blunt, polite, scary*
Syntactic			
Syntactic position	analytic	9	*lovely, healthy, severe, deadly, foul, handsome, lively, pleasant, tricky*
Infinitival complementation	analytic	4	*pleasant, tricky, costly, risky*
Synthetic persistence	synthetic	8	*handsome, sincere, risky, costly, scarce, lovely, remote, cruel*
Analytic persistence	analytic	8	*handsome, sincere, risky, costly, scarce, lovely, remote, cruel*
Pragmatic			
End weight	analytic	9	*deadly, profound, scarce, remote, cruel, happy, friendly, blunt, scary*

Each test item served to activate one of the seven constraints, and we are interested in whether (and how) this affects the choice of variant. Each informant was confronted with both members of a given pair. The questionnaire was designed to approximate an ideal scenario in which item

[18] The online appendix at https://osf.io/9hf7v lists all 120 item prompts by contextual constraint.

pairs differ only along a single dimension, e.g. with/without infinitival complementation. Since all other factors were kept as equal as possible, differences between test and control items can be taken to reflect the effect of the manipulated variable.[19] The main response categories (synthetic form, no preference, and analytic form) were treated as an ordinal response variable. Responses where informants chose "neither" were not considered in the analysis.

A short sociodemographic section preceded the linguistic part, inquiring personal details of the informants such as age, gender, educational attainment and native variety of English.[20] Educational attainment was operationalized as an ordinal variable with four categories: primary education, secondary education, undergraduate and postgraduate degree. Respondents were asked to specify the highest educational level they had attained or were in the process of attaining. Details about their occupation and the place in which they had spent their childhood were not considered in the analysis. The number of years spent outside the country in which their native variety is spoken was excluded from analysis, too, since none of the participants mentioned any significant stays of this kind in connection with this item.

3.2.2 Data collection

A mixed-mode recruitment strategy (Dörnyei & Taguchi 2010: 60-62) appeared necessary to obtain a sufficiently large dataset and cater for the social heterogeneity of respondents that was required for an analysis of speaker variables. The combined sampling method involved two different questionnaire formats: a written pen-and-paper questionnaire and an online version.

The recruitment logistics of the written format required that the 60 item pairs be subdivided into four versions of 15 item pairs each. Each of the four versions was in turn set up in three sub-versions in which the

[19] It is self-evident that even in a highly controlled methodological setup such as the present questionnaire the effect of certain linguistic constraints cannot be completely neutralized, e.g. collocational preferences.

[20] For the sociodemographic section of the questionnaire see the respective document in the online appendix at https://osf.io/nemcb.

sequence of items was manually rearranged to balance serial-position effects.[21] Each form therefore elicited a subset of 30 items. I ensured that items with the same adjective did not appear in immediate succession; further, care was taken that the control and corresponding test item were not presented within a sequence of at least ten items. The order in which the two superlative options were presented in the response panel on the right (cf. Figure 2) was manually randomized to balance possible effects of *straightlining* (Schonlau & Toepoel 2015), i.e. respondents giving identical answers to items in a battery of questions with the same response scale; the position of the *no preference* and *opt-out* categories was kept consistent across items. Instructions emphasized the need to read the item prompts out loud or silently to oneself prior to making a judgement in order to activate the impact of phonological constraints. In addition, informants were urged to make their self-reporting judgements individually and refrain from drawing analogies between items in the questionnaire. A pilot study with 20 native speakers of English, ten British and ten American, was conducted to assess the usability of the instrument. The qualitative feedback I received from the pilot informants pertained mainly to the wording of individual item prompts and did not raise any objections to the overall conceptual makeup of the questionnaire.[22] The mean survey completion time as reported by the pilot informants was at 5 minutes. This is in keeping with Maniaci & Rogge (2014: 457) who establish a link between dropout rates, survey length and recruitment incentives:

> [...] meta-analyses have indicated that survey length is one of the strongest contributors to response rates [...]. Based on rates of recruitment, study completion, and feedback received from participants, we find that individuals are generally willing to complete 5–10 minute studies without any sizeable monetary recruitment incentive.

Convenience sampling was employed to recruit respondents for the pen-and-paper format. River cruise passengers from English-speaking countries proved a convenient informant group. As a local tour guide in

---

[21] One of the twelve subversions of the written questionnaire is available at https://osf.io/nemcb/.

[22] Further, thanks are due to the participants of the English Linguistics Research Seminar at the University of Bamberg for their invaluable comments on an earlier draft of the questionnaire items.

the city of Bamberg, Germany, I had access to this specific group of informants, which, by virtue of the exclusive way of travelling, tends to be very homogeneous as regards the social parameters of interest.

The online format was designed to remedy this imbalance. An online version of the questionnaire was set up in *LimeSurvey* (LimeSurvey 2018). Care was taken that the visual arrangement of the online version was as similar as possible to the written format (cf. Figure 3).

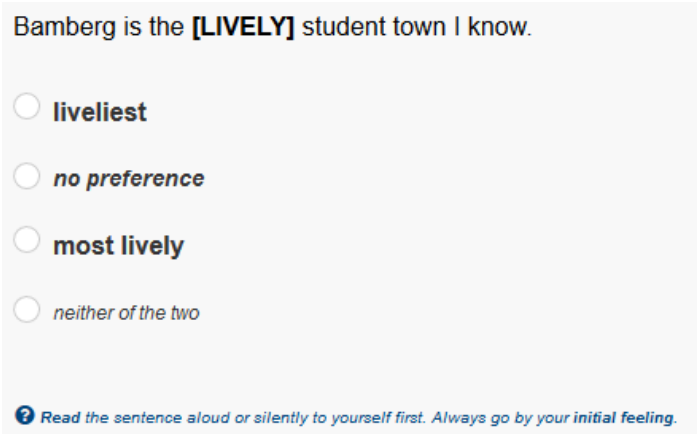

Figure 3. Test item HA_1_T in the web version of the questionnaire

The four response categories were vertically aligned below the sentence prompt to ensure that the visual presentation is compatible with the dimensions of a smartphone screen. Text formatting was used to make the opt-out category less prominent than the main response options. Radio buttons prevented informants from selecting more than one response. Informants could not return to a previous question or item once the 'Next' button at the bottom of the webpage had been clicked. As with the written questionnaire, the sociodemographic details of interest were inquired at the beginning of the survey. After that, respondents were randomly assigned numbers from 1 to 4; these numbers corresponded to the four subversions of 15 item pairs each. Within each subversion, randomization was applied to both the sequence of items and the order in which the two superlative options were presented while the positions of the intermediate response option and the opt-out category were kept consistent (cf. Figure 3). A progress indicator was provided at the top of the web interface.

Respondent-driven "snowball" sampling (Heckathorn 2002; Salganik & Heckathorn 2004; Dewaele 2018: 279-280) was used as a recruitment strategy for the online format. The URL of the web questionnaire was sent to friends and acquaintances living in English-speaking countries, who were asked to distribute the link in their own personal environment. In addition, the link was distributed via different social media channels, most prominently in Facebook groups designed to offer and ask for help in particular English-speaking regions and cities. These local "share-and-care" communities were strategically targeted to balance the varietal makeup of the combined sample towards BrE; further, these groups were assumed to be relatively heterogeneous as regards the other speaker variables operationalized in the present study. I also shared the link with Linguistics Departments in the Anglosphere, whose staff proved very supportive by encouraging their students to take part in the study. Due to the recruitment efforts of fellow linguists, participants with a linguistic background may be overrepresented in the sample; it remains debatable, however, whether a significant number of informants with considerable metalinguistic knowledge will distort the results, which were designed to be based on intuition in the first place. The reduced reliability that may arise from these circumstances is outweighed by the diversity and size of the dataset. 528 informants completed the web questionnaire, 6,524 trials resulted in a break-off. These incomplete data rows were altogether excluded from further analysis. Further, 41 online informants were weeded out since they had indicated that they did not consider themselves native speakers of English.

The possibility to glean paradata, i.e. client-side and server-side information on the questionnaire completion process, is a substantial advantage of web survey applications.[23] Figure 4 plots the web format completion time as a percentage of all complete trials, rounded to full minutes.

[23] The contributions in Kreuter (ed., 2013) demonstrate how paradata can be harnessed to understand response behavior and optimize the survey process.

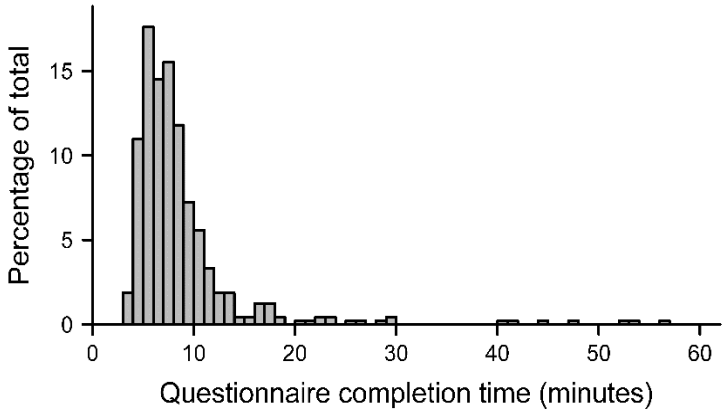

Figure 4. Web questionnaire completion time as a percentage of the total

Just under two-thirds of trials took between 5 and 8 minutes, while 75 % of all informants completed the survey within the range of 5 to 10 minutes. This distribution turns out as expected if we consider that each item in the web format was presented on a separate screen and the next item was only displayed approximately two seconds after the informant had confirmed their response by clicking 'Next'. The mean interview time for all 528 trials was 8:26 minutes; this factors in the few respondents who paused and resumed survey completion at a later point. Three informants completed the questionnaire in less than 3:30. These timings gave rise to the suspicion that the respective informants were prone to *satisficing*, i.e. they tried to give satisfactory responses with minimal effort (cf. Krosnick 1991 for a definition of the problem, and Roßmann 2017 for a comprehensive methodological account). This tendency has been shown to be more prevalent in web surveys than in face-to-face formats, where the physical presence of an interviewer possibly prompts the respondents to exercise greater diligence and patience while completing the questionnaire (Heerwegh & Loosveldt 2002). A qualitative analysis of the suspicious response sets indeed revealed patterns that were indicative of satisficing; these respondents proved susceptible to response-order effects by straightlining one of the response categories, either the one presented first or the intermediate option (nondifferentiation). In addition, two of these response sets contained a significant number of unanswered items, many of them in sequence. As my concerns about the quality of these

responses turned out to be legitimate, all three informants were discarded from the analysis. The final dataset comprised responses from 675 informants, 193 of which used the written format and 482 the web format.

Given the differences between the two modes of data collection that have so far been discussed, it appears reasonable to object that simply combining the two datasets will be detrimental to data quality. Heerwegh & Loosveldt (2002, 2008) report that web survey respondents generally produce data of lower quality as compared to respondents in an interviewer-assisted setting. At odds with this account, Bethlehem & Biffignandi (2012: 190) note that differences in data collection mode do not always translate into different survey results. In any case, researchers employing mixed-mode surveys must be aware of the methodological challenges posed by both the survey design and the logistics of data collection. Post-hoc quality checks on the basis of web survey paradata serve as viable diagnostic criteria in this regard. The two questionnaire formats, self-administered and interviewer-assisted (although to a limited degree here), presuppose different cognitive frameworks in which informants are exposed to the linguistic problem in focus: The persuasive power of the researcher in the manual mode is assumed to reduce nonresponse and breakoff rates, and prevent satisficing; it remains a matter of debate, however, whether the high degree of intrusion by the researcher promotes more sizeable social desirability effects as compared to the seemingly more anonymous self-administered web format.

3.2.3 Linguistic constraints and hypotheses

This section introduces the seven linguistic constraints that were operationalized in the current questionnaire. In describing the individual variables, I will proceed as follows: First, I will examine the respective variable from a theoretical perspective and review previous work on its role in gradation strategy choice. Based on these insights, I will then explain how the variable in focus was operationalized in the current study and provide exemplary questionnaire items for illustration. The survey of each variable concludes with a discussion of the outcomes that should be predicted on the basis of previous research. The constraints are grouped according to their linguistic subdiscipline. The first group comprises the phonological constraints *rhythmic harmony* and *segmental* horror aequi, followed by the

syntactic predictors *syntactic position*, *infinitival complementation* as well as *synthetic* and *analytic persistence*. Finally, the pragmatic-syntactic variable *end weight* is addressed.

Phonological constraints

The extent to which linguistic subsystems can be assumed to interact has controversially been discussed in the literature.[24] Formal theories of grammar in the Chomskyan tradition (e.g. Chomsky 1964; Chomsky & Halle 1968) as well as traditional cognitive accounts (e.g. Fodor 1983) advance the strict autonomy of individual linguistic modules and deny any interaction between them. Under the umbrella of *phonology-free syntax*, Zwicky (1969) and Pullum & Zwicky (1988), *inter alia*, propose that phonological representations cannot have the slightest influence on syntax. This radical notion of *modularity* has subsequently been revised to permit varying degrees of interaction. Vogel & Kenesei (1990) argue in favor of a unidirectional relation between syntax and phonology, negating any influence of phonological constraints on syntactic processes. At odds with this conception, Zec & Inkelas (1990: 378) and Schlüter (2005: 257-306, 2015: 199-202) propose models that allow for the bidirectionality of this interaction. It is particularly the latter account that provides the theoretical framework for the present investigation of phonological constraints on the superlative alternation. The phonological factors that have been assumed to affect syntactic realizations operate both on the segmental and supra-segmental levels. The current study operationalizes one factor on each level, namely *rhythmic harmony* and *segmental horror aequi*, i.e. the avoidance of identical sounds in (near-)adjacency.

Rhythmic harmony

The prosodic structure of languages is organized in a systematic way. According to the *Principle of Rhythmic Alternation*, the ideal rhythm in the stress-timed English language consists in the alternating sequence of stressed and unstressed syllables (cf. Couper-Kuhlen 1986: 60; Kager 1989: 2; Schlüter 2005: Chapter 2, *inter alia*). Violations of this pattern can take two forms: stress clashes and stress lapses (Nespor & Vogel 1989). A stress clash is defined as a sequence of two stressed syllables without an

[24] Schlüter (2005: 2-6) provides a comprehensive account of this discussion.

intervening unstressed syllable, as in *severe storm* [σ'σ'σ]; a stress lapse renders two unstressed syllables in immediate succession, as in *clever apprentice* ['σ σ σ'σ σ]. Clashes constitute a more serious infringement than lapses. Drawing on psycholinguistic evidence (McClelland & Rumelhart 1981; Levelt 1989, *inter alia*), Schlüter (2015: 202) argues that rhythm is neurophysiologically anchored in the human brain. Let us assume that there exists a prosodic node which responds to the condition '+stress'. Every time a stressed syllable is processed, this node receives activation and fires, immediately followed by a recovery phase; on the neurosensory level, the ideal rhythm thus manifests itself as the regularly alternating sequence of activation and recovery of the node which is sensitive to '+stress' (cf. also MacKay 1987: particularly Chapters 5 and 8). Since the reactivation of the node during this recovery phase is assumed to incur additional processing costs, language users employ compensation strategies to avert this premature reactivation. The great majority of rhythmic problems are resolved by mechanisms that operate on the same linguistic level, i.e. *beat addition*, *beat deletion* and *stress shift* (Kelly & Bock 1988; Nespor & Vogel 1989; Schlüter 2005: 28-31). If compensation strategies on the rhythmic level are incapable of averting an imminent overload of processing effort, the next higher level is penetrated to remedy the situation. This concept thus posits that levels of linguistic analysis are hierarchically structured and compensatory tasks are passed on in a bottom-up fashion, from the rhythmic subsystem via morphology and syntax to the highest level of semantics.

Ample empirical support has been provided for the claim that prosodic problems can be negotiated on the levels of morphology and syntax. Schlüter (2005, 2015: 181-199) identifies such effects in an investigation of a range of morphological and syntactic variation phenomena such as the suffixation of past participles, adverbial and infinitival marking, and redundant comparative marking in a diachronic perspective. Studies by McDonald, Bock & Kelly (1993), Benor & Levy (2006) and Lohmann (2014: 112-114) indicate that rhythmic constraints play a role in the constituent order of binomials. Investigating the interaction of rhythm and frequency, Lohmann (2014: 112-114) further invokes psycholinguistic evidence to account for the finding that high-frequency binomials are sensitive to prosodic effects while no such effect is obtained for binomials of low fre-

quency. The recall-based production study by Lee & Gibbons (2007) suggests that language users include or omit the complementizer *that* in order to balance rhythmic irregularities.

Mondorf's (2009b: 17-24) analysis of prosodic constraints on the comparative alternation focusses on the avoidance of stress clashes at the junction of the comparative form and the initial syllable of the following word. Positional differences also play a role in this context: In attributive use, monosyllables and finally-stressed disyllables exhibit a more sizeable preference for the synthetic comparative than initially-stressed disyllables, as the former can avert a stress clash by inserting the *-er*-suffix as a buffer before an initially-stressed noun (cf. examples 25 and 26). This prosodic pattern is the default for attributive adjectives, since 94 % of disyllabic and 72 % of trisyllabic nouns are initially-stressed (cf. Berg 2000: 274; also see Mondorf 2009b: 18); evidently, a stress clash also occurs with a monosyllabic noun.

(25) *a remoter island* [σσˈσσˈσσ]
stress clash averted by suffixation

(26) *a more remote island* [σˈσσˈσˈσσ]
periphrasis induces stress clash

The situation is different for initially-stressed disyllabic adjectives that are used attributively with an initially-stressed noun. The analytic comparative maintains the alternating rhythm in the prosodic sequence under investigation, while the synthetic *-er*-buffer entails a stress lapse involving the two final syllables of the adjective, as in examples (27) and (28).

(27) *a cleverer student* [σˈσσσˈσσ]
suffixation induces stress lapse

(28) *a more clever student* [σˈσˈσσˈσσ]
stress clash averted by periphrasis

Two observations are worthy of note: First, if a stress lapse is indeed more tolerable than a stress clash, as hypothesized by several authors (see Schlüter 2005: 20 for a summary), it can be inferred from examples (25) through (28) that the synthetic comparative generally affords a slight rhythmic advantage when used attributively. Second, the averted stress clash in (28) comes at the expense of a stress clash at the junction of *more*

and the initial syllable of the adjective if we posit that the analytic comparative marker is generally stressed in connected speech.

When constructing items for rhythmic harmony, I was guided by Schlüter & Knappe's (2018: 70) advice to focus on attributive structures. Prenominal contexts afford the advantage that the superlative and the following noun are closely tied to each other, both syntactically and prosodically. Further, attributive cases enable a more clear-cut rhythmic analysis than non-attributive ones (superlatives in predicative and nominal contexts) in that lexical stress can reliably be identified in both components (i.e. adjective and noun); this is all the more so because English nouns predominantly have initial stress (cf. Berg 2000: 274). Fused-head NPs (cf. Section 3.2.3), in contrast, tend to contain short function words whose (secondary) stress pattern in connected speech may vary according to speaker.

In item pairs that were designed to test *rhythmic harmony*, I manipulated the stress patterns of the nouns and the syllable numbers and stress patterns of the adjectives. This was aimed at creating a scenario in which the synthetic superlative consistently offers a prosodic advantage in the test condition while the analytic variant is rhythmically preferable in the control condition. The rhythmic analysis starts with the initial syllable of the superlative and ends on the first *stressed* syllable of the following noun. This design is illustrated in Table 3: In the test condition RH_3_T, the synthetic variant maintains an alternating sequence of stressed and unstressed syllables; the analytic variant, by contrast, causes a stress clash at the junction of superlative and noun. In the control condition RH_3_C, it is the analytic superlative that comes with the "ideal" alternating rhythm while the synthetic superlative entails a stress lapse.

Table 3. Exemplary item pair for rhythmic harmony

Code	Item	Rhythmic pattern	
		synthetic	analytic
RH_3_T	This was probably the **[POLITE]** answer he has ever given.	σˈσσˈσ	ˈσσˈσˈσ
RH_3_C	That was probably the **[POLITE]** response he has ever given.	σˈσσσˈσ	ˈσσˈσσˈσ

As illustrated in reference to Table 3. Exemplary item pair for rhythmic harmony, the wording of items that were designed to operationalize the predictor *rhythmic harmony* rests on the premise that the synthetic variant consistently offers a sizeable prosodic advantage over its analytic competitor in the test condition while the reverse applies in the control condition. A critical analysis of my items conducted after launching the questionnaire revealed that a clear-cut binary analysis of prosodic patterns as pursued here is valid for the item pairs RH_3, RH_5, RH_6 and RH_8, in the sense that in each item, one of the variants produces the ideal alternating rhythm while the other variant entails an infringement thereof. All other RH items, however, do not afford such a clear-cut opposition on prosodic grounds, as shall be illustrated in reference to the item pair RH_4 (Table 4).

Table 4. Prosodically ill-conceived item pair RH_4

Code	Item	Rhythmic pattern	
		synthetic	analytic
RH_4_T	This school has the **[FRIENDLY]** atmosphere of all.	ˈσσσˈσ	ˈσˈσσˈσ
RH_4_C	This is the **[FRIENDLY]** community of all.	ˈσσσσˈσ	ˈσˈσσσˈσ

Recall that the rhythmic analysis starts with the initial syllable of the superlative and ends on the first *stressed* syllable of the following noun. In the test condition RH_4_T, the synthetic superlative comes with a stress lapse, and the analytic superlative induces a stress clash. In the control condition RH_4_C, the synthetic variant provokes a double stress lapse, and the analytic variant entails a sequence of stress clash and stress lapse.

It is visible therefore that in the control and test conditions of this item pair, either superlative variant results in a violation of the ideal alternating rhythm. Although previous studies posit that clashes are more strongly avoided than lapses (cf. Nespor and Vogel 1989: 87; Kager 1995: 372), it does not appear expedient in the present methodological setup to quantify against each other the assumed degrees of prosodic (dis)harmony afforded by each superlative variant. We should legitimately assume that subtle differences of this kind would have no measurable impact on the choice of superlative form. Thus, I decided to exclude from the analysis RH item pairs whose test and control conditions did not yield a clear-cut binary opposition of the ideal alternating rhythm and an infringement thereof. I consequently restricted my analysis to the unproblematic item pairs RH_3, RH_5, RH_6 and RH_8.[25] With due regard to the above considerations, I advance the following working hypothesis for the variable *rhythmic harmony*:

Hypothesis: When facing the choice between the synthetic and analytic superlative form, language users tend to opt for the rhythmically more harmonious one. In the present context, we expect a preference for the synthetic superlative in the test condition and a preference for the analytic superlative in the control condition.

Segmental horror aequi

The avoidance of identical morphophonological units in adjacency has predominantly been investigated from an optimality-theoretic point of view. Studies in this paradigm have typically drawn on the *Obligatory Contour Principle* (Leben 1973; McCarthy 1986; Yip 1988) as a general output constraint which prohibits phonologically identical morphs in immediate sequence. In their account of the *Repeated Morph Constraint*, Menn & MacWhinney (1984) mention *lily*-adverbs as instances of this phenomenon, i.e. adverbs like *uglily*, for which periphrasis by *in an ugly way* is a valid remedy (cf. Ackema & Neeleman 2005: 90); however, they acknowledge that tolerance of repetition is the default in such cases

[25] Note, however, that the problematic item pairs RH_1, RH_2, RH_4, RH_7 and RH_9 were only excluded from the analysis of the predictor *rhythmic harmony* but were included in all other analyses described in Section 3.2.5.

(Menn & MacWhinney 1984: 528). In other derivational contexts, haplology is deployed to avert this repetition: The verbal suffix *-ate* is truncated if adjectivization by *-able* occurs, e.g. *demonstrate – demonstrable* (Plag 1998: 204).

Segmental identity avoidance has further been operationalized as a predictor in studies of morphosyntactic variation. For the genitive alternation, Hinrichs & Szmrecsanyi (2007) and Szmrecsanyi et al. (2017) find that a final sibilant in the possessor decreases the likelihood of the *s*-genitive. Interestingly, the results presented in Szmrecsanyi et al. (2017: 7, 21) indicate that the effect of this constraint is subject to a sizeable degree of cross-varietal variation. As early as Sweet ([1892] 2014: 327) and Jespersen (1949: 344), identity avoidance for reasons of euphony was identified as a factor conditioning the choice of comparative and superlative variant (see also Kytö & Romaine 2000: 185). As for adjectives ending in a sibilant or /-st/, the near-categorical preference for *most* has been related to the prospective adjacency of two word-internal sibilants in the synthetic superlative; similarly, adjectives in /-ə(r)/ can be assumed to favor *more* to avoid /-ərə(r)/ sequences in the synthetic comparative. The corpus results presented by Mondorf (2009b: 24-30) tentatively support these hypothesized preferences for the comparative of adjectives in /-ə(r)/ and the superlative of adjectives in a sibilant or /-st/; despite the fact that written corpus data have successfully been used to study phonological constraints on grammatical variation (particularly since Schlüter 2005), the validity of these findings is yet to be ascertained using alternative methods. Two problems have been addressed insufficiently in previous accounts of this constraint: The first one pertains to the phonological structure of the allegedly more "euphonious" analytic superlative of an adjective in a sibilant or /-st/: the superlative marker *most* also contains the consonant cluster /st/, whose immediate repetition is, by hypothesis, discouraged in the synthetic variant, even though the identical segments do not cooccur word-internally, as in (29):

(29) If the Sharia law is followed in letter and in spirit, it is perhaps the **most just** system available, but any wrong doing [sic] makes it look very bad. [GloWbE US 331037]

Thus, it remains debatable whether the *most*-superlative really affords as great an advantage over its synthetic competitor as has been assumed. The second problem directly emanates from the first one, yet it more generally applies to the limits of identity avoidance: Does this effect require that the identical strings are morphs? And further, does identity avoidance operate across word boundaries? As to the former question, Menn & MacWhinney (1984: 538) conclude, albeit cautiously, that haplology and dissimilation do not presuppose the morphological status of identical elements and can also occur morpheme-internally. Plag (1998: 204) argues along the same lines; his account, however, is likewise confined to derivational processes. The question whether segmental identity avoidance effects can be observed across word boundaries has, to my knowledge, only been considered by Bresnan (2011), in the context of the dative alternation. Her data suggest that speakers do not avoid the ditransitive dative when a final sibilant in the recipient immediately precedes an initial sibilant in the theme, as in example (30):

(30) The Jew said, "By Him Who gave **Moses superiority** over all the people!" [ˈməʊzɪ**z s**uːˌpɪərɪˈɒrɪti] [GloWbE GB 3698912]

Prior to formulating a hypothesis for the constraint presently under theoretical scrutiny, it appears imperative to explore the cognitive mechanisms that may underlie the avoidance of segmental identity in adjacency. Rohdenburg's (2003: 236) *horror aequi*[26] principle denotes "the widespread (and presumably universal) tendency to avoid the repetition of identical and adjacent grammatical elements or structures". This principle has predominantly been put to the test in the context of various *grammatical* phenomena (Rohdenburg & Schlüter 2000; Mair 2002; Vosberg 2002, *inter alia*). Mondorf (2003: 278-279) advances a phonological conception of *horror aequi*, assuming that "the neurological motivations are the same in grammar and phonology"; these motivations consist in the tendencies (i) to inhibit reactivation of neurons within a given timespan, and (ii) to create sufficiently distinct adjacent elements to facilitate recog-

[26] The term was first used by Karl Brugmann in his treatise *Das Wesen der lautlichen Dissimilationen* [The nature of sound dissimilation] (1909: 146) to refer to the cognitive aversion to similar or identical sounds in adjacency.

nition and processing. Along these lines, final /-st/ in the synthetic superlative form is notionally akin to a stressed syllable in Schlüter's (2005) nodal-activation account of rhythm: Too small a buffer between two identical segments incurs an increased processing load and induces speakers to avoid *earnestest* in favor of *most earnest*, as hypothesized by Mondorf (2009b: 24-26), whose analysis is restricted to word-internal *horror aequi*. In his review of Mondorf (2009b), Lohmann (2010: 308) legitimately challenges the explanatory power of *more*-support in the context of segmental identity avoidance. Indeed, it is hard to understand how a speaker's preference for *most earnest* over *earnestest* should be related to the proposed trade-off between explicitness and economy. Instead, Lohmann (2010: 308) argues that segmental *horror aequi* should be accounted for along the lines of nodal activation in speech production, as suggested by Schlüter (2005) for rhythmic constraints on grammar: The choice of the analytic superlative avoids the repeated activation of a node during a refractory phase in which reactivation is linked to increased processing effort.

The present study is concerned with segmental identity avoidance across word boundaries, i.e. contexts in which attributive synthetic superlatives are collocated with nouns starting in /st-/. While the phonological buffer consists of a schwa in word-internal instances of repetition like *earnestest*, no phonological material intervenes in the scenario I intend to investigate; however, speakers may insert a pause at the word boundary. A simple query in the GloWbE corpus using the search string `st*_jjt st*_n`[27] (CLAWS 7 tagset, *english-corpora.org* interface) shows that this phonological pattern is by no means limited to single cases; the query yields 34,445 tokens of 1,900 types and includes highly frequent collocations such as *highest standards* (2,013 tokens, 1.06 pmw), *brightest star* (483 tokens, 0.25 pmw) and *earliest stages* (367 tokens, 0.19 pmw).

The wording of items that were designed to operationalize the predictor *segmental horror aequi* is exemplified by the item pair HA_3. In the test item (31), the adjective prompt is followed by an initially-stressed disyllabic noun starting in /st-/; the noun in the control condition (32) is synonymous or semantically close to the one used in the test condition and identical on prosodic grounds. Crucially, however, the noun following the

[27] This search string must be read as follows: any noun starting in <st> preceded by any superlative adjective starting in <st>.

target superlative in the control condition does not start in /st-/ or a sibilant.

(31) Sam is the **[CLEVER]** student in class. [HA_3_T]
(32) Jamie is the **[CLEVER]** pupil in class. [HA_3_C]

In the present methodological design, in which a clash of identical segments is induced, informants are assumed to seek a bypass on the next higher level of linguistic representation, namely morphology: I expect a preference for the analytic superlative since this variant averts the repetition of the /st/-cluster in immediate adjacency. The following hypothesis for the constraint *segmental horror aequi* suggests itself:

Hypothesis: In collocation with nouns starting in /st-/, language users tend to prefer the *most*-superlative in order to avoid the prospective repetition of the /st/-cluster in the coda of the synthetic superlative and the onset of the noun.

Eight item pairs were designed to test this hypothesis.

Syntactic constraints

It is in the domain of syntax that the link of linguistic complexity and processing efficiency – as proposed by Mondorf (2009b) – appears most transparent. Mondorf (2009b: 7) argues that *more* is a more reliable cue for the comparative than the *-er*-suffix and refers to competition with the productive derivational suffix in agent nouns like *painter* (cf. Köpcke 1998: 308-309). However, it has been discussed in Chapter 2 that *more* and *most* also occur as quantifiers; these contexts clearly constitute competitors to analytic comparatives and superlatives since similar phrase structures often complicate disambiguation. This applies even more to the superlative, as the corpus analyses in Chapter 4 will demonstrate: It is often impossible to draw a sharp distinction between quantifier *most* (which we can assume to be no less frequent than quantifier *more*), a genuine superlative degree phrase and an elative. Further, there is no derivational suffix in English that competes with the synthetic superlative suffix *-est* in terms of cue reliability. Thus, it appears legitimate to weaken Mondorf's argument in the

domain of the superlative: As for cue reliability, analytic *most* does not come with a substantial advantage over synthetic *-est*. The second argument adduced by Mondorf (2009b: 7) to account for the processing advantage of *more*, however, also holds for the superlative: The synthetic alternant is more difficult to parse since it merges two functions into a single form (*cleverest* ADJ SUP); in the more explicit analytic superlative, there is a one-to-one correspondence of form and function, which obviates the need for mental decomposition (*most* SUP *clever* ADJ). As yet, this hypothesis has not been empirically tested in the domain of the superlative alternation. Closely related to this aspect, another argument in favor of the processing advantage afforded by the analytic form is specified by Boyd (2007) as the *buying time hypothesis*: In the left-to-right processing of a sentence, the cue to the degree phrase is rendered prior to the lexical cue instantiated by the adjective. The fact that these two functions are each assigned a separate lexeme allows the speaker to buy themselves additional processing time that can be used to select a suitable adjective after the degree marker by inserting filler material between the lexemes, as in the comparative example (33):

(33) toy story is a kids movie, we are both 18, how about we try and watch something **more** you know... **Romantic**?
[GloWbE US 627197]

The current questionnaire study investigates two syntactic constraints whose operational mechanisms are firmly grounded in the cognitive-functional processing account set forth in the previous paragraph, namely *syntactic position* (attributive vs. nominal) and the presence of an *infinitival complement*. The third constraint to be covered in this section, namely *persistence* in superlative strategy choice, will be explained along the lines of structural priming.

Syntactic position

There is ample empirical evidence to suggest that attributive comparatives favor *-er* while predicative comparatives come with higher probability of *more* (Braun 1982; Leech & Culpeper 1997: 366; Szmrecsanyi 2005: 74; Claridge 2007a: 129; Hilpert 2008: 407; Mondorf 2009b: 80-88; Cheung & Zhang 2016: 576). In reference to Bolinger (1967: 9-11), Mondorf (2009b:

88-90) argues that predicative adjectives tend to give temporary or unexpected information about their referents while attributive ones generally denote permanent characteristics of the noun they premodify: A *responsible man* is a man with a universal sense of responsibility, the string *she is responsible (for)* puts the respective person in charge in a particular situation, e.g. a temporary project. Mondorf's line of argument subsequently shifts to iconicity: The second maxim of Givón's (1991: 87) *Quantity Principle* posits that "[l]ess predictable information will be given more coding material". Since predicative adjectives provide less expected information, the lexically bulkier *more*-comparative is more likely to occur in this function. In addition, due to the additional lexeme bearing primary stress, the analytic variant is prosodically heavier, too.

The theoretical considerations in Chapter 2 emphasized that attribution is the default function of the superlative; more than two thirds of superlatives in Claridge's (2007a: 129) data occur in this position. Comparatives, on the contrary, are fairly evenly split between attributive and predicative uses (cf. Lindquist 2000: 126). It has further been noted that nominal superlatives in fused-head NPs represent the default opposition to attributive superlatives while the predicative function frequently shades into absolute interpretations, particularly if an explicit set of reference is absent. Payne & Huddleston (2002: 416) note that

> [s]uperlatives are ranking expressions: they inherently pick out a unique entity as of higher rank than the rest, as do comparatives when comparison is made between two entities. They are therefore more like definite determiners than are ordinary adjectives, which serve only to constrain the denotation of the nouns they modify. This is reflected in the position of superlatives [...]: they are typically closer to the determiner than other adjectives [...].

The crucial difference between predicative comparatives and nominal superlatives in a fused-head NP hinges on the opposition of *term* and *set comparison*. This has syntactic implications. In the comparative example (34), the subject complement slot is clearly instantiated by an adjective; the superlative in (35), however, has nominal traits: It comes with the definite article and is notionally fused with the subject in a determiner-like fashion; further, it is postmodified by an *of*-prepositional phrase denoting a quasi-partitive reference. From a generative point of view, there is a

filler-gap dependency (cf. Hawkins 1999) between the NP in subject position and the vacant nominal slot following the superlative. The noun leaves a trace in this slot (*t*1).

(34) This issue was **more important** than all others that were raised in the discussion.
(35) **[This issue]**1 was **the most important** ***t*1** of all those that were raised in the discussion.

What does this conceptual difference imply for the study of positional constraints on the superlative alternation? Two aspects have so far been integral to the line of argument pursued in this section: First, positional differences entail differences in processing complexity; second, due to a filler-gap dependency, the superlative in a fused-head noun phrase has noun-like characteristics.[28] To combine these insights, it appears logical to investigate attributive and nominal superlatives with regard to noun phrase complexity. Berlage (2014: 1-18) proposes two measures of NP complexity, namely length (i.e. counting words) and internal structure (operationalized as the number of phrasal nodes or the degree to which an NP is sentential). The item pair POS_7 is used to compare the syntactic complexity of attributive superlatives (37) and nominal superlatives in a fused-head NP (36) according to these metrics.

(36) [**The debate about nuclear energy]**1 was [the **liveliest / most lively** ***t*1** I have seen in a while]NP. [POS_7_T]
(37) [The **liveliest / most lively** debate I have seen on this channel]NP was the one about abortion last night. [POS_7_C]

Let us first consider the criterion of NP length. A comparison of (36), which contains a nominal superlative in a fused-head NP, and (37), in

[28] Adjectives and nouns are notionally very similar. In his structuralist account, Matthews (2014: 21) points out that grammarians up to the early 20th century have summarised adjectives and nouns under the category *common noun* (also see Radatz 2001: 33-36; Rauh 2010: 11-30). The functional distinction of these word classes is more transparent in Modern English than in other, particularly older Indo-European languages. Morphologically, the crucial difference between nouns and adjectives is that the latter lost their inflection for number in the transition from Old to Middle English.

which the superlative occurs attributively to the noun *debate*, suggests that if the same superlative variant is used in both sentences, (37) is longer by a single word only. This marginal numerical difference is assumed to have a negligible impact on processing complexity in the contrastive analysis of attributive and nominal superlatives. A noun phrase's degree of *sententiality* is a measure of its structural complexity. The nominal superlative in (36) can be considered sentential in the sense that it serves as the direct object of the following contact clause by which it is postmodified in a filler-gap construction. Postmodification by a relative clause is a genuinely nominal property; superlatives with a type of postmodification generic to nouns blur the boundaries between word classes and can thus be considered structurally more complex than superlatives in attributive use (i.e. the most frequent adjectival function). By hypothesis, superlatives in a fused-head NP also require greater effort than attributive ones from a processing perspective. While parsing (36) from left to right, recipients cognitively reactivate the nominal referent *debate* when encountering the extraposed superlative *liveliest*/*most lively*: A nominal trace t_1 follows the superlative. The more words intervene between the superlative and its referent, the higher are the processing costs incurred by the fused-head NP. By hypothesis, as the distance between the superlative and its nominal referent increases, so does the urge on behalf of the speaker to reduce processing complexity by repeating the noun in the slot to the right of the "dangling" superlative. Thus, it appears reasonable to argue that a nominal superlative in a fused-head NP is syntactically more complex than an attributive superlative. This insight is hypothesized to translate into a processing advantage of the analytic superlative: In a fused-head NP, *most* comes with a slightly shorter Constituent Recognition Domain (CRD) than the *-est*-suffix, i.e. recipients need to parse less linguistic material in order to recognize the immediate constituents of a phrase (cf. Hawkins 2001: 5 for the *Early Immediate Constituents Principle*; Berlage 2014: 9). When faced with the choice between two forms that differ in structural transparency, language users systematically prefer the more transparent option (Hawkins 1999: 215-216). The *most*-variant allows for an earlier recognition of phrase structure since the degree marker precedes the adjective. According to this line of argument, we should expect to find support for *most*-support in our investigation of the predictor *syntactic position*.

At the beginning of this section, I noted that the argument advanced by Mondorf (2009b: 88-90) to account for differences in processing complexity between attributive and predicative comparatives is essentially a discourse-pragmatic one. These syntactic positions vary with regard to the information status of the adjective presented in the respective slot (predictable vs. unpredictable). As for the superlative, a similar iconic argument can be made in favor of the expectation that nominal superlatives in a fused-head NP have a higher probability of *most* than attributive ones. In (36), the superlative instantiates the subject complement and thus represents a constituent. While certainly not dispensable from a syntactic point of view, the attributive superlative in (37) is governed by the noun it premodifies and does not instantiate a constituent of its own. Consequently, a nominal superlative in a fused-head NP is more integral to the structure of a sentence than an attributive one. Further, in (36), the determiner must follow the copular verb and the noun it references; thus, example (38), in which the nominal superlative occurs preverbally, is ungrammatical:

(38) * The **liveliest / most lively** I have seen in a while was the debate about nuclear energy.

Restricted to the slot of subject complement, the superlative of a fused-head NP necessarily appears in a sentence-medial (or sentence-final) position, in either case following the subject and verb. The position of attributive superlatives in a sentence is more flexible: In their function as premodifiers, they can be part of any constituent that contains a noun (i.e. also the subject) and can thus be found in a pre- or postverbal position.

According to the principle of *End Focus* (Quirk et al. 1985: 1357), information in an utterance is commonly presented in a linear fashion, from given to new. As nominal superlatives generally appear in a late sentential position (as opposed to attributive ones, which come with greater positional flexibility), they are more likely to provide information that is new and unexpected (and hence more relevant) to the recipient. If the superlative in a fused-head NP is (i) more pivotal to the syntax of a sentence and (ii) more momentous as regards the serial presentation of information as compared to an attributive superlative, it can be argued in reference to Givón's (1991: 87) *Quantity Principle* that nominal superlatives

should receive more coding material than attributive ones. In keeping with the analytic support principle (Mondorf 2009b: 6), therefore, nominal superlatives in a fused-head NP are hypothesized to come with a greater share of the lexically bulkier *most*-variant as compared to superlatives in attributive function. This prediction also appears reasonable from a phonological point of view: The analytic variant has an additional stressed syllable, thereby adding greater weight to the stand-alone nominal superlative.

The arguments presented in this section can be boiled down to the following hypothesis for the variable *syntactic position*:

> ***Hypothesis***: Nominal superlatives in a fused-head noun phrase are more likely to be formed by means of *most* than superlatives in attributive contexts.

Nine item pairs were designed to operationalize the functional opposition of attributive superlative (control condition) and nominal superlative in a fused-head NP (test condition).

Infinitival complement

Of the six types of adjectival complementation mentioned by Quirk et al. (1985: 1220), three have been tested for their role in the comparative alternation: *than*-phrases, prepositional complements and infinitival complements. Mondorf (2009b: 57-80) suggests that prepositional and infinitival complements increase the probability of the *more*-comparative and thus warrant an argumentation along the lines of *more*-support. Infinitival complementation has been shown to significantly increase the proportion of analytic comparatives (Mondorf 2003: 268, 2009b: 64-72; Szmrecsanyi 2006: 63-85; Hilpert 2008: 408); notably, however, these studies have relied on corpus data containing a relatively small number of infinitival complements. Mondorf (2003) was criticized by González-Díaz (2004: 369) for this methodological weakness: Her results are based on a set of 28 adjectives, of which only 17 occur with infinitival complements, with eight being attested fewer than five times with a *to*-infinitive. Säily, González-Díaz & Suomela (2018: 170) also acknowledge the problem of data scarcity for the predictor in focus, reporting that only few adjectives in

their study occur with an infinitival complement. Boyd's (2007: 30-32) experimental results confirm the insights that have been gained on the basis of corpus studies: His findings that analytic comparatives are read faster than synthetic ones in the presence of an infinitival complement can be taken to support the hypothesis that *more* has a processing advantage over *-er*.

As for the superlative, Cheung & Zhang's (2016) corpus study suggests that infinitival complementation only marginally increases the probability of *most*. Prior to formulating a hypothesis for this syntactic constraint, let us consider in greater detail the syntactic structure of superlatives complemented by a *to*-infinitive. Mondorf (2009b: 70) distinguishes between three types of infinitival constructions following a comparative: tough movement, *it*-extraposition and infinitival complement.

Examples (39) to (41) represent an instance of each type as a complement to a superlative.

(39) *Tough Movement*
[Your upper-back muscles]$_1$ are the **most tricky to** target ***t***$_1$ in any exercise. [COINF_2_T]

(40) It-*Extraposition*
When it comes to placing furniture in a room, it is **easiest to** do if your room has a focal point like a fireplace.

(41) *Infinitival Complement*
Of all contenders for the throne, he emerged as the man **most fit to** be king.

First, it is noteworthy that *to*-infinitive complements are excluded from non-attributive positions, i.e. they can only complement nominal superlatives in a fused-head NP (39 and 40) and postnominal superlatives (41).[29] *It*-extraposition (40) is syntactically different from the other two constructional types in that there is a lower degree of dependency between the superlative and the infinitive; syntactic rearrangement of the respective constituents assigns each of them a separate function: *Of all the options they had, the easiest was to sell the house*. [ASVCs] This may account for Mondorf's (2009b: 71) findings that *it*-extraposition is the only constructional type that does not come with the increased sensitivity to complexity effects postulated along the lines of *more*-support. Of the remaining two types of infinitival complementation, tough movement constructions are assumed to entail the highest degree of fusion of superlative and infinitive. In (39), the superlative *the most tricky* does not specify a characteristic of the nominal referent but rather describes the verbal action denoted by the infinitive. The notional object of the infinitive occurs in the subject slot. From a transformational point of view, tough movement leaves a trace (t_1) after the transitive infinitive, resulting in the cognitive reactivation of the nominal referent *your upper-back muscles*; this incurs higher processing costs on behalf of the recipient and, to argue along the lines of Mondorf (2009b), requires *most*-support. For methodological reasons, all four test items that have been designed to operationalize the constraint *infinitival complement* contain the superlative in a tough movement context. A questionnaire study that aims to compare types (39) to (41) would have to include a sufficiently large number of test items for each of the three test conditions, as well as an equal number of control items. The ensuing extension of the item pool would pose serious challenges for the logistics of data collection. The semantic domain of tough movement adjectives is restricted: Mair (1990: 26; 58-59) points out that while infinitival subject clauses can combine with five semantic classes of adjectival pred-

[29] Strictly speaking, predicative superlatives with zero article are non-attributive, too, e.g. *I find it most difficult to pay attention when I'm around people who talk a lot*. The theoretical considerations in Chapter 2 suggest that superlatives of this kind are frequently ambiguous between different interpretations (absolute – relative – elative). Zero-determined predicative superlatives are not included in the present analysis to rule out that the syntactic and semantic idiosyncrasies of these cases affect the analysis.

icates, namely expressing potential, frequency, necessity/desirability, degrees of ease/difficulty and value judgements, tough movement constructions are confined to the latter two. The four item pairs designed to operationalize the predictor *infinitival complement* were constructed using four different adjectives which have been attested to exhibit varying degrees of alternation in the comparative; *tricky*, *costly* and *risky* denote degrees of difficulty, *pleasant* expresses a value judgement. The selection of adjective types was motivated by Mondorf (2009b: 71).

An item pair that operationalizes the contextual constraint in focus is structured as follows: In the test condition (42), the superlative occurs in a fused-head NP and is followed by a *to*-infinitive. The notional object of the infinitive functions as the syntactic subject of the sentence. The control condition (43) renders the superlative as part of a fused-head NP, too; crucially, however, there is no infinitival complement. This syntactic difference has semantic implications: The superlative in (42) describes the verbal action denoted by the infinitive. In (43), by contrast, the superlative specifies a property of the nominal referent.

(42) **[Your upper-back muscles]**$_1$ are the **[TRICKY]** to target t_1 in any exercise. [COINF_2_T]

(43) The question about his love life was certainly the **[TRICKY]** of all. [COINF_2_C]

So far, I have discussed why fused-head NPs followed by an infinitival complement should be considered more complex and thus more difficult to process than those without such a complement. What remains to be fleshed out in the current context is this: Why is *most* more advantageous than *-est* when followed by a *to*-infinitive in a tough movement construction? As noted in considering the predictor *position*, the superlative variants differ with regard to their Constituent Recognition Domains (Hawkins 2001). In general, early identification of phrase structure is associated with low processing complexity, while late identification incurs higher processing costs. At an early stage of the parsing process in (44), *most* alerts the recipient to the fact that a degree phrase follows; in (45), it is only after parsing the *-est*-suffix that the phrase structure can be determined. Essentially, the sequence in which linguistic information is pre-

sented at the onset of the degree phrase is reversed: While the morpho-syntactic cue to the following degree phrase comes first in (44), lexical information in the form of the base adjective is supplied initially in (45). Upon parsing the degree-marking suffix, recipients have to correct their hypothesis that an adjective phrase headed by *tricky* follows immediately after the determiner – a hypothesis that appeared plausible prior to encountering *-est*. Thus, drawing on Mondorf (2009b: 58-59), we can assume that *most* provides a processing advantage over *-est*.

(44) the most tricky to target NP[the DegP[**most** AP[tricky IP[**to** target]]]]
(45) the trickiest to target NP[the DegP[AP[tricky **-est**] IP[**to** target]]]

Mondorf's (2009b: 58-59) second argument zooms in on the sequence of adjective phrase and infinitive phrase. Analytic (44) renders the heads of these phrases in adjacency; head-adjacent orderings are assumed to facilitate phrase recognition since they entail a strong degree of fusion between the adjective and dependent infinitival complement. In synthetic (45), the adjacency of these phrasal heads is prevented by the intervening *-est*-suffix. In line with Boyd's (2007: 32) *warning hypothesis*, we can thus assume that the *most*-superlative offers a crucial cognitive advantage: The phrase-initial degree marker serves as a signal to recipients that the following adjective phrase is complex and urges them to devote increased processing effort to this phrase.

The above considerations entail the following hypothesis for the predictor *infinitival complement*:

Hypothesis: Infinitival complements in a tough movement construction increase the complexity of a degree phrase. Language users tend to compensate for the concomitant increase in processing effort by resorting to the cognitively more advantageous *most*-superlative.

The questionnaire contains four item pairs to gauge the impact of this constraint.

The variable *infinitival complement* was originally designed to be one of two manifestations of a superordinate variable labelled *complementation*. The theoretical considerations in Chapter 2 revealed that the default interpretation of the superlative is the relative one: The noun attributed or

referenced by the superlative is superior to all other entities comprised by a set of reference with regard to the characteristic denoted by the adjective. Quirk et al. (1985: 463) report that 52 % of superlatives (as opposed to only 25 % of comparatives) come with an explicit basis of comparison; this basis is often realized by a prepositional phrase, as in the control condition (47). In (46), there is no explicit set of reference; it is up to the recipient to infer this information from the previous discourse.

(46) This project was certainly the **[RISKY]** one. [CONO_4_T]
(47) This was the **[RISKY]** of all the projects he has undertaken so far. [CONO_4_C]

When constructing the present questionnaire, I assumed that this act of inferring the referential set would translate into higher processing costs and thus, along the lines of *most*-support, increase the odds of the analytic superlative. In the process of data collection, several arguments induced me to dismiss this line of reasoning. First, if a reference set can be inferred from previous discourse, it necessarily represents discourse-given information. In the context of positional differences, Mondorf (2009b: 88-90) argues that it is discourse-new material that is commonly given more coding material. This is in line with Wasow (1997: 85), who discusses givenness/newness as a measure of grammatical weight. Since predicative comparatives tend to give unexpected, discourse-new information, they are more likely to be realized by the lexically bulkier analytic form. It is thus hard to justify why a superlative without an explicit referential frame should require *most*-support. Further, it has been noted that the number of words in a phrase can be considered an indicator of its complexity (cf. Berlage 2014: 1-14). According to this measure, then, a degree phrase containing an overt set of reference is more complex than a degree phrase without such scope-restricting material. Further, conflating prepositional phrases of the present kind and infinitival complements into a single variable contravenes the distinction between modifiers and complements. The syntactic dependency of restrictive modifiers and the superlative form is relatively weak: The postmodifying prepositional phrase in (47) can be separated from the superlative by topicalization. Infinitival complements, on the contrary, are strongly fused with their adjectival

head and do not allow relocation within the sentence. Ultimately, the operationalization of the variable in focus is at odds with the overall methodological framework of the present study; every constraint is operationalized as a comparison of two clearly defined conditions, i.e. a test and control condition. The variable *complementation*, however, was originally devised as a triple comparison with two test conditions (*infinitival complementation* and *implicit restrictive modification*) mapped to the same control condition. A design of this kind, however, compromises the interpretation and comparability of my data; it is fair to say that this variable was ill-conceived from the very start, both conceptually and methodologically. Consequently, I excluded the four item pairs that operationalized the constraint *implicit restrictive modification* from further analysis and thereby reduced the variable *complementation* to *infinitival complement*, defined as the presence (test condition) and absence of an infinitival complement (control condition).

Morphosyntactic persistence in superlative strategy choice

It is a linguistic commonplace that language is inertial and repetitive to some degree. Production of linguistic material is significantly influenced by exposure to identical or similar material in previous discourse, which may constrain the creativity applied by language users in subsequent production. In his work on morphosyntactic persistence in spoken English, Szmrecsanyi (2006: 41-42) identifies three factors that lie at the root of persistence effects, namely (i) properties of the human speech production and processing system, (ii) accomplishment of discourse management functions and (iii) maintaining speaker-hearer economy.

From a psycholinguistic perspective, linguistic persistence has chiefly been investigated along the lines of *priming*. As one of the earliest and arguably most influential experimental studies on syntactic priming in production, Bock (1986) investigated sentence formulation processes with prompts that operationalized the transitive (active vs. passive) and dative (prepositional vs. double object construction) alternations. Her results imply that "formulation processes are somewhat inertial and subject to such probabilistic factors as the frequency or recency of use of particular structural forms" (Bock 1986: 355). As to the latter factor, a sizeable wealth of psycholinguistic experiments yields results of notable inconsistency: While some studies have suggested that the likelihood of a prime-target

match significantly decreases with the length of the intervening input (e.g. Levelt & Kelter 1982; Branigan, Pickering & Cleland 1999), others have provided evidence for the considerable longevity of a prime, thus arguing in favor of the view of priming as a form of implicit procedural learning (Bock & Kroch 1989; Bock & Griffin 2000; Branigan et al. 2000; Bock et al. 2007, *inter alia*).

Priming effects have been demonstrated to operate on various levels of linguistic analysis (see Szmrecsanyi 2006: 11-22 for a summary). Previous accounts have pivoted on several recurrent questions, three of which I will discuss in the following. In doing so, I will highlight implications for the study of persistence in gradation strategy choice. The first of these questions appears most urgent in the light of the methodology adopted in the present study: Do priming effects operate across modalities? In other words, is priming confined to the condition that the prime and target units are produced by the same person or is there evidence to assume the possibility of *comprehension-to-production priming*? Bock et al. (2007) use an experimental method to investigate cross-modal priming effects in the transitive and dative alternations. Participants were exposed to a spoken prime instantiated by one of the syntactic variants; they were subsequently asked to describe an unrelated event in a picture. Crucially, the content of the picture could be described using a single sentence that could be realized by one of the constructional alternants. Previous perceptive exposure to one of the constructions in the prime sentence significantly increased the likelihood of participants producing the same syntactic variant in the target sentence. This effect was even found to endure over longer stretches of linguistic input that intervened between prime and target. Comparing these findings to the results presented by Bock & Griffin (2000) in a production-to-production setting using identical primes and targets, the authors conclude that "structural persistence from comprehension to production [...] was comparable in magnitude to persistence from production to production. This adds to the evidence of crossmodality persistence." (Bock et al. 2007: 452) Similar conclusions are drawn by Potter & Lombardi (1998), Pickering & Branigan (1998) and Cleland & Pickering (2003) in similar experimental settings. In one of the first corpus-based approaches to syntactic priming, Gries (2005) identifies both intra-modal and cross-modal priming effects; he notes, however, that the latter are weaker than the former (Gries 2005: 374). Further, Litcofsky & van

Hell (2019) provide support for bi-directional cross-modal priming, hence confirming the hypotheses that (i) previous production of a syntactic construction facilitates the comprehension of the same construction in subsequent discourse, and that (ii) previous perception of a syntactic construction facilitates the production of the same construction in the target.

The design of my questionnaire items rests on the premise of cross-modal priming, for which there is adequate empirical support, as I believe to have demonstrated. The prime sentence in the test condition (48) immediately precedes the target sentence and renders an analytic prime in attributive function; this precondition is assumed to increase the odds of the analytic superlative in the target slot. The control condition (49) comes without a sentence containing a superlative prime.

(48) Crude oil is among the **most expensive** liquids on the market. That is because it is one of the **[SCARCE]** resources on earth. [PERANA_5_T]

(49) Gold is one of the **[SCARCE]** resources on the planet. [PERANA_5_C]

In order to activate priming effects in both modalities, informants were instructed to read the item prompts out loud or silently to themselves before intuitively selecting the preferred superlative form.

A second issue that features prominently in the literature on priming pertains to the time course of effect: How are priming effects influenced by the length of the linguistic material intervening between prime and target? In their study of syntactic priming in written production, Branigan, Pickering & Cleland (1999) report that priming effects diminished significantly when a single sentence fragment intervened, and vanished entirely when prime and target were separated by four sentence fragments. Branigan et al. (2000), on the contrary, find that syntactic priming in spoken production does not decay as rapidly as in the written mode; this prompts the authors to conclude that the durability of a prime depends on the mode of production. The corpus results presented by Szmrecsanyi (2006: 75-83) for persistence in the choice of comparative forms imply that the proportion of prime-target matches decreases logarithmically as the textual distance between the two forms in focus increases; Szmrecsanyi (2006: 79) terms this relationship the "forgetting

function". These insights have profound implications for the design of persistence items in the current questionnaire: The textual distance between prime and target should be kept as short as possible. As to the item design elucidated above, one may validly object that prime and target could be rendered in even closer proximity by coordinating them in attribution to a noun, as in (50):

(50) He built a house and trading post and became one of **the wealthiest and most influential** men of the period. [GloWbE US 1408997]

Instances like these, however, are problematic for pragmatic reasons. While the literature on syntactic coordination would predict persistence on the grounds that processing of the second conjoin is faster if it is structurally equivalent to the first (Leech & Culpeper 1997: 367-368; Frazier, Munn & Clifton 2000), the BNC data presented by Cheung & Zhang (2014) point in a different direction: In line with the pragmatic principle of *end weight*, speakers are inclined to render the "heavier" superlative in the second slot. This tendency was observed as early as Rohr (1929: 111): "In the case of two superlatives, one of which is formed by means of a suffix and the other periphrastically, the one with suffix was usually preposed." Therefore, an item configuration modelled on (24), containing a synthetic prime conjoined with the target, would come with the potentially confounding factor of *end weight*, possibly entailing a higher probability of *most* in the target and thereby overriding the persistence effects the present variable is designed to measure. To circumvent this problem, prime and target were rendered in two separate but syntactically similar sentences; importantly, cohesion was ensured by means of deictic reference (cf. (23a), where *it* references *crude oil* and *that* references the entire preceding prime sentence).

A third and final question to be addressed in this section on persistence in superlative strategy choice pertains to the linguistic description of the primed phenomenon itself: Can we expect the two superlative alternants to induce priming effects of equal strength? For the comparative, Szmrecsanyi (2006: 84) concludes that

> [w]hile we were unable to obtain evidence that the presence of a generic adjective taking synthetic comparison in the discourse increases the odds for synthetic comparison in

> [the target], the presence of the token more in the discourse immediately preceding [the target] can clearly increase the odds for analytic comparison in [the target] significantly. A possible reason why the token more triggers analytic comparison, but the affix -er does not trigger synthetic comparison, is that for one thing, more is a lexical item of its own; second, it has more phonological substance than -er, which is phonetically often just realized as [ə]. In psycholinguistic terms, then, it should not be surprising that more is a better prime than -er.

This explanation for the advantage of *more* over *-er* with regard to priming centers on differences in phonological salience. Notably, the analytic comparative is also reliably primed by quantifier *more*; thus, "even formal identity creates persistence effects in the absence of functional identity" (Mondorf 2009b: 115). This effect is what Pickering & Ferreira (2008: 437) consider the "lexical boost" to structural priming. Previous studies have provided ample empirical support in favor of lexically boosted priming (e.g. Pickering & Branigan 1998; Cleland & Pickering 2006); however, these studies concur in the view that this effect is stronger for content words than for function words. These insights suggest that what has been labelled morphosyntactic priming can at least partly be accounted for on the phonological/lexical level. Ultimately, it remains debatable whether morphosyntactic persistence in gradation strategy choice as defined by Szmrecsanyi (2006) should more accurately be conceived of as lexical priming of *more*/*most*.

Szmrecsanyi (2006: 39-40) further mentions Rohdenburg's (2003: 236) *horror aequi* principle as "a rival empirical phenomenon" to persistence. As is apparent from the considerations of phonological constraints in gradation strategy choice, this principle can be taken to suggest the opposite of structural priming, namely the avoidance of structural repetition in adjacency; however, persistence in gradation strategy choice clearly differs from the phenomena that have been studied with regard to *horror aequi* (e.g. Rohdenburg & Schlüter 2000 on a range of grammatical identity effects from a diachronic and synchronic perspective; Mair 2003 on Doubl*ing* in the context of gerundial complementation; Vosberg 2003 on gerundial vs. infinitival complementation in diachrony). The *horror aequi* principle premises that identical grammatical elements occur in a short textual distance from each other; in some contexts that have been tested for

horror aequi effects, the linguistic items in focus are even rendered in immediate adjacency, as in example (25) taken from Rohdenburg & Schlüter (2000: 466, bold print for emphasis in the original).

(51) ?She looked upon this solution **as as** good **as** that one.

In the previous paragraph, I proposed that the current operationalization of the variable *persistence* necessitates that the prime and target superlatives be rendered in separate phrases but in the closest possible proximity. The resulting textual distance between prime and target is inevitably greater than in contexts that have been analyzed along the lines of *horror aequi*, in which sufficiently distinct elements in adjacency are assumed to have a facilitatory effect on processing. Despite the conceptual differences between identity avoidance in the sense of *horror aequi* and persistence as operationalized in the current study, the predictor in focus comes with interpretative challenges: It runs the danger of inviting post-hoc interpretations that fit the purpose: If we find the effect, this can be taken in support of morphosyntactic priming; if we do not find the effect, an explanation along the lines of structural identity avoidance seems appropriate. Since both scenarios appear equally plausible, building and testing hypotheses for morphosyntactic persistence in general is an intricate undertaking.

The present questionnaire measures the effects of two persistence variables, namely *synthetic persistence* and *analytic persistence*. In the test condition, the prime – i.e. the sentence preceding the superlative target – contains a *most-* or *-est*-prime, respectively. Eight item pairs were constructed for each persistence variable. The above considerations entail the following hypotheses for the constraint in focus:

Hypothesis: An analytic superlative in the prime sentence results in a greater proportion of *most* in the target structure. A synthetic prime, on the contrary, does not significantly increase the probability of observing *-est* in the target.

End weight

Ample empirical support has been provided for the role of end weight as a constraint on grammatical variation, both on the clausal and phrasal level. In the genitive alternation, 'heavier', i.e. longer possessors tend to favor the *of*-genitive, while 'lighter', i.e. shorter possessors are more likely to take the *s*-genitive (Altenberg 1982; Rosenbach 2005; Hinrichs & Szmrecsanyi 2007: 453; Szmrecsanyi et al. 2017: 2), resulting in a constituent order that can be described as 'short before long'. A similar relation between constituent length and constituent order has been identified as a predictor in a number of other contexts of grammatical variation, e.g. the dative alternation (Bresnan et al. 2007; Gerwin 2014: 48), the alternation of extraposed and non-extraposed *that*-clauses and cleft constructions (Erdmann 1988; cf. Berlage 2014: 24-25 for a summary) as well as coordinate constructions (Lohmann 2014: 117-118).

Conceptually, end weight is firmly grounded in Behaghel's (1909) *Law of Growing Constituents*, which states that the order of syntactic units is partly determined by their length. Eitelmann (2016: 396) points out, however, that the scope of end weight is not limited to constituent ordering along the lines of 'short before long'. According to the definition by Quirk et al. (1985: 323), the *End Weight Principle* "encourages the placing of more complex and communicatively more important units towards the end of the noun phrase". Quirk et al.'s (1985: 323) definition of end weight involves a second dimension which allows for this constraint to be classified as a pragmatic (rather than purely syntactic) phenomenon. This aspect is expressed by the principle of *End Focus*, according to which "it is common to process the information in a message so as to achieve a linear presentation from low to high information value" (Quirk et al. 1985: 1357). It has been a linguistic universal, not only since Grice's (1967) *Cooperative Principle*, that human conversation rests upon the cooperation of interlocutors. A pragmatic guideline that conceptualizes this conventionalized tendency to present discourse-given before discourse-new information is

couched in terms of the *Given-New Contract* (Clark & Haviland 1977).[30] Hence, if (i) recipients expect new information to occur towards the end of a clause and (ii) new information receives more coding material, lexically bulkier units should more likely be found towards the clausal end. Eitelmann (2016: 396) objects that along these lines, end weight – simply conceived of in terms of length – could be regarded as epiphenomenal to end focus, referring to studies which suggest a relation between information value and structural choice regardless of syntactic weight. Thus, while the conceptual mystery surrounding the interplay of syntactic and pragmatic factors must remain an issue for future studies to resolve, we nevertheless adopt the term *end weight* for the constraint under consideration, simply because the notion is intuitive and amenable to operationalization in the present methodological framework.

End weight is by no means a new arrival among the constraints on gradation strategy choice. Rohr (1929: 21) remarks that the analytic comparative is used in the final constituent to add emphasis. On purely phonological grounds, Lindquist (2000: 126) argues that the *most*-superlative is used to create end weight due to an additional stress on the degree marker and an additional consonant (/**m**əʊ**st**/ vs. /ɪ**st**/). In the context of the comparative alternation, Mondorf (2009b: 100, 106-107) suggests that end weight can be expected to have a facilitatory effect in production and comprehension. Eitelmann's (2016: 416) account is confined to advantages on behalf of the producer.[31] He argues that the compensatory effect of the more explicit variant (i.e. the analytic superlative, in the present context) is twofold:

[30] It is worthwhile noting that the terminology for the concepts of 'given' and 'new' varies according to the theoretical background of the respective authors. In this context, I use Halliday's (1967) terminology. Proponents of generative grammar adopt the terms 'presupposition' and 'focus' (cf. Jackendorff 1972, for instance). See Clark & Haviland (1977: 3-4) for a summary of this terminological discussion.

[31] Eitelmann's (2016) study is concerned with two phenomena of syntactic variation: semi-reflexive verbs with the explicit *self*-pronoun vs. the zero variant (*to hide oneself* vs. *to hide* ∅) and finite verb forms vs. *do*-supported ones in EModE affirmative declaratives (*he did love* vs. *he loved*).

> On the one hand, language users can be assumed to save extra costs by the choice of a variant that is easier to process, and on the other hand, the bulkier variant buys language users additional time at the end of a sentence and gives them a head start for planning the next one.

These motivations are in line with psycholinguistic accounts that conceive of end weight primarily as a speaker-centered notion (e.g. Wasow 1997). The following caveat must therefore be issued with regard to the method adopted in this study: If the facilitatory effect of end weight exclusively pertains to production, it remains to be seen if elicitation data gleaned through a questionnaire can adequately gauge the impact of this constraint. Here, it appears important to reiterate a point raised at the beginning of this section: Conversation is a cooperative process. Speakers can therefore be assumed to draw on their communicative experience when planning utterances: They are aware that some constructions incur higher processing costs than others and endeavor to facilitate comprehension on behalf of the recipient. Hence, end weight should more accurately be regarded as an interactive rather than a purely speaker-centered effect (cf. Arnold et al. 2000: 51).

My questionnaire includes eight item pairs that operationalize the predictor *end weight*. For the target form to be subject to end weight, the superlative must occur in clause-final position (cf. 52); the superlative necessarily constitutes the head of a fused-head NP, with the restrictive modifier (instantiated here by an *of*-phrase) rendered clause-initially. In the control item (53), the superlative represents the head of a fused-head NP in clause-medial position. The scope-restricting prepositional phrase follows the superlative.

(52) Of all students in class Chris was the **[FRIENDLY]**. [EW_7_T]
(53) Nigel was the **[FRIENDLY]** in the group. [EW_7_C]

The clause-final superlative in (52) entails topicalization of the restrictive modifier, resulting in a discontinuity. Thus, end weight has different implications for the comparative and the superlative: Comparatives do not require scope-defining modification; their referential term is introduced by means of a *than*-phrase. *Than*-phrases are classified as complements rather than modifiers since they cannot be dislocated from the degree form:

(54) *Than ice water is heavier.

The syntactic complexity that arises from the discontinuity triggered by a superlative in clause-final position is hypothesized to increase the proportion of *most* in such an environment. We may tentatively assume that the end weight factor is a stronger constraint on the superlative than on the comparative alternation. Two factors add up to this assumption: (i) the pragmatic motivations underlying the planning, production and processing of utterances, and (ii) the syntactic complexity that arises from the topicalization of the restrictive modifier.[32] The following working hypothesis emanates from the above considerations:

Hypothesis: A superlative subject to end weight is more likely to be realized analytically than a clause-medial superlative.

3.2.4 Speaker variables

In addition to the methodological difficulties of obtaining satisfactory corpus data for an examination of speaker variables, adjective gradation does not appear to be a particularly promising subject for sociolinguistic analysis in the first place. Both superlative alternants are equally established in the English language, highly frequent and their respective use is, at least on the face of it, largely determined by the morphophonological characteristics of the adjective. Further, there seems to be no overt standard against which one variant can generally be judged more prestigious than the other. The same holds true for most other morphosyntactic alternations. In recent years, however, advancements in corpus methodology have paved the way for more engagement with the social correlates of such phenomena. Jenset, McGillivray & Rundell (2018), for instance, examine social predictors of the dative alternation on the basis of data gleaned from the Spoken BNC2014 (Love et al. 2017). They find that age has no effect on the preference for either dative variant, while postgraduates and male speakers are slightly biased towards the prepositional dative. Crucially,

[32] The terms *topicalization* and *left-dislocation* are often used synonymously. This is inaccurate, however, as can be inferred from Rivero (1980: 365-366): While topicalization comes with a gap in the slot from where an item is moved, left-dislocation leaves a pronominal copy in the original slot. See also Gregory & Michaelis (2001) on this conceptual distinction.

these findings imply that speaker variables play a role in morphosyntactic alternations, although they may not weigh as heavily as linguistic ones. The present section introduces the four social variables investigated in the questionnaire study at hand, namely *age*, *gender*, *education* and *variety*.

Age

The present study operationalizes the variable *age* in an apparent-time setting. In this sense, age differences per se do not have any explanatory power; they rather serve as a surrogate for real-time differences, i.e. data collected at different points in time (cf. Bailey et al. 1991: 241; Bailey 2013: 312). Linguistic differences between age groups are interpreted as reflecting diachronic changes in the target population that occurred in the period the respective individuals lived through. This allows for "the history of a linguistic process to be viewed from the perspective of the present" (Tagliamonte 2012: 43).

English is assumed to have undergone a typological shift since Old English times from a largely synthetic to a rather analytic language (cf. Baugh & Cable 1993: 60; König 1994: 539; Tagliamonte 2012: 241). This comprehensive development manifests itself in the loss of inflections and the concomitant need for a more rigid word order to mark grammatical relations. For the 20th and 21st centuries, scarce empirical support has so far been furnished in favor of this claim in the domain of adjective gradation. Leech et al. (2009: 264-267) find a slight but not statistically significant gain in the analytic comparative on the basis of BrE and AmE corpus data covering the period from 1961 to 1992. Similarly, Szmrecsanyi's (2006: 85) corpus-based analysis of the comparative concludes on the note that "there is apparent-time evidence that analytic comparison is indeed spreading". These findings further dovetail with the tentative conclusions drawn by Leech & Culpeper (1997: 370-371) from their analysis of the same time span as well as those by Grofulović & Jovanović (2016: 206). Covering a greater diachronic range, Kytö & Romaine (2000) suggest the opposite: They find that in the big picture, *-er* and *-est* have reasserted themselves since the EModE period and consider the BrE lead over AmE in the change towards synthetic comparison an instance of colonial lag (Kytö & Romaine 2000: 190). This trend is in turn consistent with the observation by Szmrecsanyi (2012: 655) that syntheticity has been on the rise

since the 17th century across a diverse range of morphosyntactic alternation phenomena after a trough in the previous two centuries.[33] Which implications do these rather inconsistent findings have for the present study of the superlative? The informants' age in my questionnaire data ranges from approximately 20 to 80 years; this range can therefore be considered a temporal analogue for the period from the mid-1950s until the mid-2010s if we draw on two fundamental prerequisites of the apparent-time hypothesis:

(i) An individual's idiolect has been solidified by late adolescence, i.e. approximately age 17 (cf. Tagliamonte 2012: 45).
(ii) The system of adjective gradation remains relatively stable across an individual's lifetime. The key motivations of superlative strategy choice are not subject to linguistic changes that result from changes of an individual's role in society.

The present analysis of the variable *age* thus rests on the premise that the judgements gleaned from my oldest informants represent the status of English spoken in the mid-1950s (i.e. when their idiolects can be taken to have been largely solid), and the data points from my youngest informants are representative of the English language spoken in the mid-2010s. The second assumption thus casts doubt on the influence of *age-graded variation* (cf. Labov 1994: 84; Tagliamonte 2012: 46-47; Bailey 2013: 324; Krug & Sell 2013: 74-75). Age-grading is generally assumed to be limited to linguistic features that can be consciously controlled. Arguably, this is not the case for variation in gradation strategy choice, where both forms coexist as prescriptively valid alternants.

Still the question remains as to which trajectory the superlative may have taken in the period under study. Two scenarios have been proposed in the literature. Following Kytö & Romaine (2000) and Szmrecsanyi (2012, 2016), we should expect an increase in the use of the synthetic superlative and a concomitant decrease in the use of the analytic one. Neither study, however, specifically zooms in on the time period covered in the present questionnaire-based account. Kytö & Romaine (2000) examine

[33] The hypothesised reversal of a typological shift from synthetic to analytic, and back to synthetic, has been couched in terms of the *linguistic cycle* (cf. van Gelderen 2006: 284).

the comparative and superlative alternations in a broader diachronic perspective from 1620 to 1990; crucially, they pool together corpus data from 1950 to 1990 into a single diachronic category. Szmrecsanyi's (2012, 2016) studies investigate the development of syntheticity and analyticity from the 12th to the early 20th century across a wide range of morphosyntactic phenomena; his analysis is not limited to degree forms and, most notably, does not cover the time frame of interest.

The alternative scenario, as advocated by Leech & Culpeper (1997: 370-371) and Leech et al. (2009: 264-267), suggests that *most* has taken away market share from *-est* in the latter half of the 20th century. Both studies compare corpus material from the early 1960s and the early 1990s (gleaned from corpora of the Brown family). This trend further receives tentative support from Säily, González-Díaz & Suomela (2018: 179), whose study of the BNC and the Spoken BNC2014 (Love et al. 2017) yields a mild increase in the overall token frequency of the analytic comparative in the twenty-odd-year-period under investigation.

There are valid reasons to subscribe to the second account. Crystal (2004: 523-525, 2012: 507-510) notes that institutionalized prescriptivism has declined towards the end of the 20th century. As of the 1970s, educational practice in the UK, and simultaneously in other English-speaking countries, has moved away from a rule-based, dogmatic notion of language to embrace a more pragmatic concept thereof, which involves "the ability to adapt knowledge to meet the needs of differing circumstances and a readiness to judge cases on their merits" (Crystal 2004: 524). This fundamental shift of educational paradigms was accompanied by similar linguistic tendencies in other domains of public life that challenge the primacy of an established standard and accentuate the merits of linguistic variation. The BBC, for instance, has abandoned the hegemony of a Southern British Standard to mirror the increasing social diversity of its audience. In the light of this trend, it is reasonable to hypothesize that older speakers are more invested in the prescriptive tradition than younger ones, who have grown up in a linguistically less dogmatic environment.

As for the superlative alternation, prescriptive standards dictate that the morphophonological characteristics of the adjective type determine the choice of gradation strategy. The alternative approach, characterized

by communicative pragmatism, aims at minimizing production and processing costs. From a cognitive perspective, it has been argued that this facilitatory effect is predominantly achieved through the more explicit analytic variant (Mondorf 2009b); this tendency possibly carries over to contexts of low complexity. Indeed, the feedback I received from several informants implies that *most* is intuitively more accessible than *-est*. One of the participants couched this impression in terms of a workaround in the present questionnaire study: "When in doubt, opt for *most*." In lockstep with the decline of prescriptivism, the adoption of the digital media and the rapid dissemination of the linguistic innovations they promote may also play a part in the assumed tendency towards more *most*. Although the characteristics of Internet language can at best be approximated as "writing which has been pulled some way in the direction of speech" (Crystal 2011: 21), it stands to reason that particularly the digital surrogates of face-to-face-conversation (e.g. instant messaging, blogs, chats) share certain communicative needs with the spoken medium. Szmrecsanyi (2009: 335) finds that spoken texts exhibit a significantly higher degree of analyticity than written texts and concludes that "spoken English places a premium on explicitness, transparency, and the minimization of comprehension complexities". LaFave (2016: 322) reports that synthetic gradation is significantly less frequent in instant messaging than in speaking or writing. It appears valid to assume that younger speakers embrace these tendencies more strongly than older speakers. These insights warrant the following hypothesis for the variable *age*:

Hypothesis: The apparent-time perspective suggests a linear trend towards analyticity. The younger the speaker, the greater the proportion of *most*.

Gender

Gender differences in the choice of gradation strategy have, to my knowledge, only marginally been tapped into in previous studies. Drawing on data from the spoken parts of the BNC and the Spoken BNC2014 (Love et al. 2017), Säily, González-Díaz & Suomela (2018) are concerned with the productivity of the two comparative formation strategies. They report a combined effect of gender and age, with the lowest degree of synthetic productivity found for female speakers under the age of 30. In gen-

eral, men appear to use *-er* more productively than women (Säily, González-Díaz & Suomela 2018: 176-177). As the examination of productivity involves zooming in on the type rather than token frequencies of both strategies (cf. Bybee 1985; Dąbrowska 2008, *inter alia*), this study may only indirectly inform the present investigation of the superlative. In a corpus study of gender differences in comparative and superlative strategy choice, Grofulović & Jovanović (2016) find that while subtle differences in gradation strategy choice arise between men and women if the analysis zooms in on material pertaining to the same register (fiction, academic or journalistic), their results imply that if the predictor variable *register* is disregarded, women and men do not exhibit systematically diverging strategy preferences. It is worthy of note, however, that graded forms (synthetic and analytic combined) are more frequent across the board in texts by female as compared to male authors; this observation is taken to indicate a higher level of expressiveness in women's language (cf. Grofulović & Jovanović 2016: 205).

LaFave's (2016) analysis of written vs. spoken vs. instant messaging data implies that women use proportionately more synthetic forms while men show a stronger preference for analytic gradation. These findings, however, can at best be taken as tentative in the light of the points raised earlier, namely that (i) his examination is confined to AmE data, (ii) his data disproportionately come from younger language users and, most strikingly, (iii) no distinction is made between comparatives and superlatives.

In her examination of NZE corpus data, D'Arcy (2014: 229) observes that the analytic comparative is proportionally more prevalent among women as opposed to men; her data further seem to imply that higher socioeconomic groups use more analytic forms than lower ones. She goes on to argue that these findings tie in rather nicely with sociolinguistic theories that relate the use of linguistically complex structures to females and higher socioeconomic status (D'Arcy 2014: 228-229). This view is not in keeping with Mondorf's (2009b) *more*-support hypothesis, which considers the analytic comparative a means of *mitigating complexity*. D'Arcy's (2014: 228-229) conclusion appears all the more dubious in the light of Mondorf's (2009b: 190-192) finding that informal style tends to come with a reduction of complexity afforded by *more*. It is not immediately apparent why a preference for the analytic strategy should be shared by informal

style and higher socioeconomic status. In sum, the current state of research on gender differences in gradation strategy choice is rather inconsistent.

Since much of the evidence available to date has only indirect relevance for the research question addressed here, what is it then that we should expect for the variable *gender* in superlative strategy choice? As with *age*, it is worthwhile to revisit established sociolinguistic theories. Regardless of which scenario best accounts for the assumed apparent-time development towards more *most* (the decline of institutionalized prescriptivism or the rise of digital communication, as remarked in the previous section on age), it is clear that such a diachronic trend is a change from below. Labov (2001: 292) states that "[i]n linguistic change from below, women use higher frequencies of innovative forms than men do". Notably, Labov refers to innovative *forms*. In the present research context, however, neither of the superlative forms is per se new in diachronic terms; it is a change in gradation strategy *preferences* I intend to predict, conditional on gender.

Nevertheless, the theory of a female-led increase in the proportion of analytic superlatives has some appeal: Tagliamonte (2012: 63) acknowledges women's role as vanguards in 90 % of linguistic changes. Further, if the female propensity to adopt linguistic features with a wide regional distribution, as advocated by Cheshire (2002: 430) in rather general terms, holds for the superlative, women should be more inclined to more explicit, easier-to-process forms. This development can be thought of as universal and supra-local in nature, potentially reinforced by the communicative changes of the digital era. Further, this hypothesis would accommodate Woods' (1997: 117) assumption that in spoken interaction, women tend to be more cooperative and willing to facilitate processing on behalf of the recipient. It is imperative, however, to alert the reader to the extremely tentative character of the following hypothesis:

Hypothesis: Women are slightly more inclined to the analytic superlative than men.

After all, it is important to reiterate that the key principles of variationist sociolinguistics have been identified chiefly in the study of *phonological* features. Thus, it appears unlikely that the patterns of (morpho)syntactic

variation mirror those identified for phonology all too closely, as Cheshire (2002: 439) notes. This, however, should not discourage empirical engagement with speaker variables in linguistic domains other than phonology. In the present study, gender is treated as a binary variable.

Education

The two social variables that have so far been discussed appear relatively intuitive in terms of their operationalization in the current questionnaire. Age is customarily measured on a metric scale; gender is typically treated as a two-way categorical factor. Methodological approaches to education, however, have been remarkably heterogeneous, both in sociolinguistic research and beyond. This situation can largely be attributed to two interrelated circumstances: First, there is disagreement as to the explanatory power of education as a social variable. Second, education is essentially a social construct; as societies are diverse and subject to permanent change, so too are the educational profiles they produce. I will consider both aspects in turn before outlining the coding practice adopted in the present study and contemplating the expected outcomes.

It is important to note at the outset that there is no immediate causal relationship between education and language use (cf. Al Wer 2002: 42). Rather, the link between an informant's educational background and their language use rests upon two basic assumptions: (i) The higher an individual's degree of education, the longer and more intensely they have been exposed to institutionalized instruction. This may have linguistic implications: Since rule-based language teaching is retreating rather slowly in favor of a more pragmatic and variation-oriented approach, as has previously been argued, speakers with high educational attainment can be assumed to be more aware of and comply more strictly with prescriptive guidelines. (ii) From a sociological perspective, education is among the array of factors that are highly correlated with an individual's life chances and thus represents an important co-indicator of *social class* as a more comprehensive speaker-centered predictor. In this context, Labov's (2001: 60-61) Philadelphia neighborhood study serves as a prime example, for its socioeconomic status index is designed as a composite of three factors (education, residence value and occupation) measured on an ordinal scale with six levels each. Ash (2002: 408) objects, however, that in Labov's study, education inadequately contributes to the social stratification of his

sample since 'high school graduate' is the modal level of education in all neighborhoods under investigation, and advocates that education be operationalized as a distinct variable, independently of social class. In any case, education and the composite variable of socioeconomic status can be interpreted as a proxy for the nature and intensity of an individual's social contacts (cf. Krug & Sell 2013: 74), both in the professional and private domain, and consequently serves as an indicator of the linguistic norms to which they adhere.

Further, the increasing heterogeneity of career tracks and socioeconomic life histories, both within and across societies, has methodological implications. Employing a questionnaire in the study of education as a social variable often requires that researchers strike a balance between oversimplifying and overspecifying educational profiles. The *Maltese English Questionnaire* introduced by Krug & Sell (2013: 94) asks informants to provide detailed information about their careers, including the type of school attended at primary and secondary levels, qualifications attained or ongoing, current occupation as well as their parents' occupations and highest qualifications. While a fine-grained survey of this kind would certainly do justice to the considerable educational diversity of the target population, logistic constraints on data collection warrant that education be operationalized in a simpler and less specific fashion in the present study. The largely uncontrolled setting in which sampling and questionnaire processing takes place presupposes that (i) response categories be general and culture-unspecific, i.e. category labels should not be tied to a single national educational system, and (ii) providing information on education should not require too much time on behalf of the informants to prevent breakoff. The present methodological design thus focusses on the highest educational qualification as an indicator of *educational attainment*. This sociological concept has two subcomponents, namely duration of education and educational success, as Schneider (2016: 11) points out. Commonly interpreted as a proxy for competencies, educational attainment zooms in on learning outcomes rather than underlying developmental processes. In the current questionnaire, education is treated as an ordinal variable with four factor levels, labelled 1 to 4: Primary/Elementary Education, Secondary Education (High School), Undergraduate (Bachelor's), Postgraduate (Master's/PhD).

LaFave (2016) is the only study to have considered the effect of educational differences in gradation strategy choice. The finding that more highly educated speakers use more analytic forms leads him to conclude, albeit tentatively, that *more* and *most* can be considered prescriptively more favorable than their synthetic equivalents by virtue of their being more explicit. Since higher levels of education are associated with greater susceptibility to the prescriptive ideology, it would seem reasonable to assume that as educational attainment increases, so does the proportion of *more/most* (cf. LaFave 2016: 321). I have already commented on the link between education and prescriptivism; while the tendency towards using overtly prescribed forms should be marked among older highly educated speakers, differences between higher and lower levels of education are expected to be less pronounced in the younger generations, whose instructional experience with grammar coincided with the decline of prescriptivism, as previously discussed in reference to Crystal (2004: 523-525). I do not concur with LaFave (2016: 321), however, on the claim that analytic forms are *prescriptively* preferable. It has been noted that a more pragmatic approach to linguistic usage has been making inroads into the practice of grammar teaching, with a concomitant retreat of the rule-centered doctrine of language instruction prevalent prior to the 1990s. "Pragmatic", in this sense, should be taken to imply a commitment to forms that are easier to process and less complex; with regard to Mondorf's (2009b) *more*-support hypothesis, then, we should assume that speakers drawn to linguistic pragmatism make more use of the more explicit *most*-superlative. On the whole, it is hard to determine how educational attainment should relate to linguistic prescriptivism or pragmatism. On the one hand, highly educated speakers can be assumed to have been exposed to institutionalized prescriptivism for a longer time than their less well-educated peers; on the other, however, more intense exposure to linguistic instruction increases an individual's experience with and sensitivity to language, which enables them to adapt more successfully to communicative situations that require a reduction of processing complexity, e.g. when conversing with an L2 learner of English. In the light of this ambiguity, it appears most sensible not to commit to either scenario at this point and remind the reader that the present investigation of the variable *education* is exploratory.

Variety

Varietal differences have recently come to the fore in the variationist research on morphosyntactic alternations. These studies have typically been situated at the crossroads of probabilistic grammar (cf. Klavan & Divjak 2016 and Divjak 2019: Chapter 3 for a critical survey), World Englishes and variationist sociolinguistics, and conceive of grammar as a quantitative and usage-based notion, constituting "the cognitive organization of one's experience with language" (Bybee 2006: 711): Differences between varieties with regard to alternation phenomena are considered the product of their speakers' varying degrees of (lifelong) exposure to the alternants under study. The studies in this paradigm particularly probe into differences in constraint strengths across varieties to account for the observed variation in morphosyntactic choice. As for the genitive alternation, Hinrichs & Szmrecsanyi (2007) find that the *s*-genitive gained ground from the 1960s to the 1990s. This trend, they suggest, is more marked in AmE than BrE and can be accounted for in terms of a range of constraints that appear to be stronger in American press material due to greater textual economy, among other factors. Likewise invested in the explanatory power of cognitive complexity and communicative efficiency, Levshina (2018) reports a sizeable degree of cross-varietal homogeneity in the use of the bare vs. *to*-infinitive following *help*. In her analysis, all but one variety pattern similarly. The relatively weak effects in her US model are attributed to the empirically backed impression that there is increasingly less competition between the alternants in AmE, with the use of the *to*-variant declining in this variety (Levshina 2018: 17-18). On the whole, these findings are in keeping with Szmrecsanyi et al. (2017), whose analysis of the genitive alternation concludes with the observation that while individual constraints may vary across linguistic communities, there is a grammatical "common core or nucleus [...] present in all the varieties" (Quirk et al. 1985: 16). For some alternation phenomena, cross-varietal variability is greater than for others. The language-internal constraints on the genitive alternation appear relatively stable across varieties (cf. Heller, Szmrecsanyi & Grafmiller 2017: 23); for the dative alternation, however, previous studies have drawn a picture of greater cross-lectal fluidity (Bresnan & Hay 2008; Szmrecsanyi et al. 2017: 19; Röthlisberger 2018). The present work investigates exactly this question on the basis of elicitation data: How do language-internal constraints on the superlative alternation

vary in a cross-varietal perspective? And further, what does this mean for the choice of gradation strategy? The standard reference varieties BrE and AmE will be in focus here.

In Mondorf's (2009a, 2009b: 171-193) analyses of the comparative alternation, differences between BrE and AmE are related to two generalizations that are widely acknowledged in the study of British-American contrasts, namely *regularization* and *colloquialization*. Colloquialization establishes a relation of diverging stylistic preferences and national variety, or rather the different societal structures underlying the varieties under study. Previous work is suggestive of the claim that AmE is more colloquial and less formal than BrE. Rohdenburg (2009: 180), for instance, shows that in the second half of the 20th century, the use of reflexive structures has declined more markedly in the former variety; this tendency may be ascribed to the assumed lead of AmE in avoiding cognitively complex and formal language (further see Biber et al. 1999: 513; Kövecses 2000: 235-246; Tottie 2002: 176). He points out, however, that BrE and AmE are sensitive to the same contextual constraints in this domain of variation (cf. Rohdenburg 2009: 180). As for the comparative, Mondorf (2009b: 171) argues that AmE and informal styles show greater susceptibility to complexity constraints than BrE and formal styles. Regularization denotes the alleged American trend of developing more regular forms where BrE is prone to hold on to irregular forms. This powerful generalization is well-established in the contrastive study of the two main reference varieties (e.g. Kövecses 2000: 189-199; Levin 2009). According to Mondorf (2009b: 190-192), AmE can be considered more regular than BrE in that the functional division of labor, which she shows to have emerged diachronically (Mondorf 2009b: 166-169), is more advanced in this variety. Along the lines of *more*-support, then, AmE should be expected to use more analytic superlatives than BrE in the presence of complexity constraints, which leads to reduced variability in the American system of comparative strategy choice. In Mondorf's (2009b: 171-193) study of British-American differences, complexity constraints are as such not empirically operationalized. Her analysis suggests a higher proportion of analytic comparatives in AmE for three of the four morphophonological groups investigated and regardless of syntactic position. This proportional lead in analytic forms is taken to indicate a more pronounced use of *more* to compensate for cognitive complexity. This observation, however, remains to be addressed

more systematically in a cross-lectal comparison of contextual complexity constraints.[34] My contrastive analysis of BrE and AmE aims to follow up on these indications by comparing contextual constraints across varieties. A separate model was designed for this purpose. Section 3.2.5 specifies the differences between this *variety model* and the *overall model*. Crucially, my elicitation-based analysis does not take into account the potential intermediary variable of style as discussed by Mondorf (2009b: 171-193). Thus, can we reasonably expect that AmE is more sensitive to effects of contextual complexity than BrE without considering the factor of style? While Mondorf's argumentation along the lines of colloquialization is conditional on diverging stylistic preferences, no such premise applies to the assumption that AmE is more advanced with regard to the functional partitioning of the degree formation system couched in terms of *more*-support, which she considers an instance of regularization. Nonetheless, I have good reason to issue a caveat at this point: Schlüter's (2009: 128) cross-lectal study of the interplay of phonology and grammar concludes on the note that "it is impossible to argue that either BrE or AmE is more sensitive to rhythmic alternation", since functional universals are, by their very nature, stable across varieties. It remains to be seen, after all, whether *most*-support can likewise be considered a universal feature of English. Despite these reservations, however, it is operationally expedient to align my working hypothesis for the syntactic and pragmatic constraints under scrutiny in a cross-varietal perspective with Mondorf (2009a, 2009b: 171-193):

[34] The only variables for which Mondorf (2009b: 178-183) provides empirical data to support her hypothesis of the American lead in *more*-support are frequency measures, namely positive frequency and attested gradability, defined as the sum of synthetic and analytic comparatives of an adjective type per amount of text. As with all other factors investigated in her work on the comparative, frequency effects are summarised under the umbrella of *more*-support. The role of frequency in the superlative alternation will come under more intense scrutiny in Chapter 4.

Hypothesis A: AmE is more prone than BrE to mitigate contextual complexity by making greater use of the easier-to-process analytic superlative.

As for persistence in superlative strategy choice, my cross-varietal analysis is exploratory. Intuitively, there is no convincing argument that would lend plausibility to the assumption that varieties exhibit different degrees of sensitivity to morphosyntactic priming effects. Thus, it appears reasonable to formulate the following hypothesis:

Hypothesis B: AmE and BrE are equally sensitive to the effects of structural priming in superlative strategy choice.

As with the contextual constraints and speaker variables previously discussed, the predictor *variety* enables us to compare the proportion of *most* for each variety, with the effect of the respective variety gauged independently of all other predictors in the model.

It appears worthwhile in the present context to revisit previous studies which have examined comparative strategy choice across varieties. Kytö & Romaine (2000) can be credited with pioneering British-American contrasts in the comparative and superlative alternations. Drawing on corpus data from 1750 to 1990, they identify a moderate American lead in analytic comparatives for disyllabic adjectives in the earlier two of the three periods under study (1750-1800; 1850-1900; 1950-1990); interestingly, it is in the most recent period that they find a reversal of this trend, with the two variants on par with each other in AmE (50 % synthetic vs. 50 % analytic), while BrE exhibits a preference for *more* (38 % synthetic vs. 62 % analytic; cf. Kytö & Romaine 2000: 183). As for the superlative of disyllabic adjectives, they draw a slightly different picture for the latter half of the 20th century, i.e. the only period directly relevant for the research question in focus. While the analytic lead is only marginal in BrE (48 % synthetic vs. 52 % analytic), AmE shows a noticeable bias towards the synthetic superlative (61 % synthetic vs. 39 % analytic; cf. Kytö & Romaine 2000: 184). Aside from the diachronic dimension of this account, these results suggest that in the time period under scrutiny in the current study, American disyllables are more synthetic than British ones and that this tendency is

even more marked in the superlative. Mondorf's (2009b: 172-173) comprehensive treatment of Kytö & Romaine's (2000: 172-173) study raises two objections that merit attention. First, highly frequent types may have skewed the tally towards syntheticity in their dataset; this scenario would imply a more or less universal positive relation of frequency and syntheticity, which is impossible to maintain without the support of frequency data. I will return to this assumed relation in greater detail in Chapter 4. Second, the American lead in synthetic comparatives and superlatives may be type-specific. Indeed, Mondorf (2009b: 176, 192-193) reports that disyllables ending in /-i/ deviate from all other morphophonological categories in her analysis in two ways: They are less sensitive to contextual constraints and make up the only category for which AmE has a greater share of synthetic comparatives than BrE. The prospect that characteristics of the adjective type lie at the heart of cross-varietal variation in the choice of superlative form seems tempting, especially since this factor group is customarily regarded as having the greatest predictive power for gradation strategy choice, as demonstrated earlier.

The pool of adjectives used in my questionnaire comprises five disyllables in /-i/, two of which are subject to analysis in Mondorf (2009b: 176), namely *tricky* and *risky*. Caution is warranted, however, when generating hypotheses upon the findings of both Kytö & Romaine (2000) and Mondorf (2009b) in the context of the present research question. Neither study controls for contextual constraints. Thus, it might be the case that at least some of the British-American variation that has been ascribed to diverging strategy preferences for individual morphophonological categories in these accounts is in fact due to factors of contextual complexity: In line with *more*-support, then, the American lead in sensitivity to constraints of contextual complexity would be expected to translate into a higher proportion of *most* as compared to BrE. This assumption is also valid from a statistical perspective: As will be fleshed out in the following Section 3.2.5, the overall model does not contain interactions of the predictor *variety* and the contextual predictors operationalized in the questionnaire. It is possible, therefore, that varietal differences may at least in part be attributed to effects that can be related to processing complexity. The above considerations yield the following hypothesis:

Hypothesis C: American respondents use a greater proportion of *most* than British respondents.

As for the two other varietal categories that qualified for inclusion in the model (cf. Figure 5), AusE and CanE, there has as yet been no empirical engagement with the comparative and superlative alternations. Thus, my analysis of these national varieties is necessarily exploratory. Figure 5 provides descriptive statistics for the four speaker variables operationalized in the questionnaire.

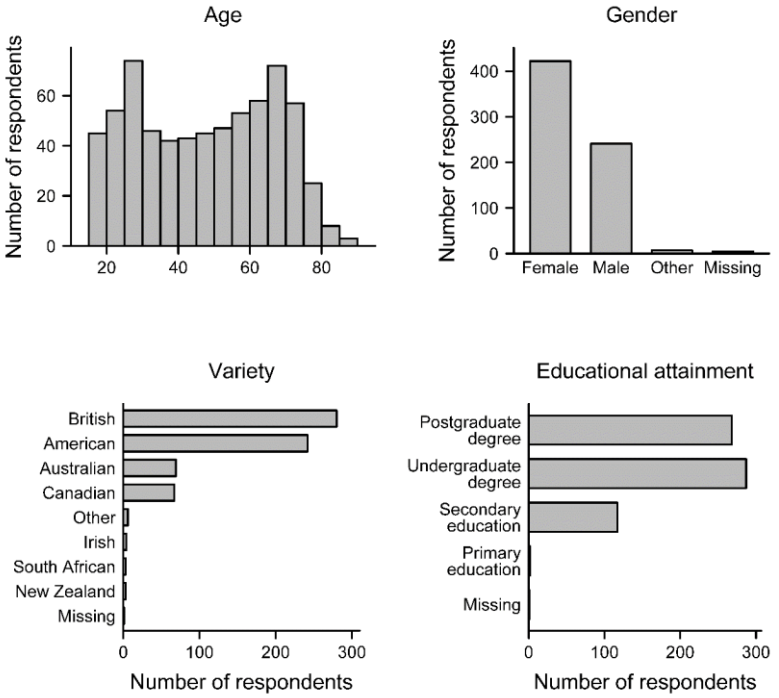

Figure 5. Descriptive statistics for speaker variables (n = 675 informants)

The age distribution covers the range from 16 to 85 years. 77 % of informants represent one of the two main references varieties, BrE (n = 280) and AmE (n = 242). AustrE (n = 69) and CanE (n = 67) are large enough to be considered varietal categories of their own. The residual category "Other" is further included in the analysis. This category comprises (i) all native varieties for which the sample was too small to be considered a category

of its own, namely Irish English (n = 4), New Zealand English (n = 3) and South African English), and (ii) n = 6 informants who specified that their native variety was not listed as a response option in the questionnaire. Interestingly, all these informants stated that their native variety was English as spoken in Trinidad and Tobago. As for gender, my sample comprises considerably more female (n = 422) than male informants (n = 241). N = 12 informants were excluded since they either did not specify their gender (n = 5) or chose the gender category "other" (n = 7). As for education, 82 % of informants have obtained (or are in the process of obtaining) an academic degree (n = 287 undergraduate, n = 268 postgraduate). N = 119 informants stated that secondary education was their highest qualification; for n = 2 informants, primary school was the highest level of education attained. A single informant did not specify their highest educational qualification.

3.2.5 Data analysis

Since the first methodological treatise of multilevel modelling in linguistic research (Baayen, Davidson & Bates 2008), the research community has become increasingly sensitive to the hierarchical structure of linguistic data (cf. Tagliamonte & Baayen 2012; Gries 2015). Emphasizing this structural prerequisite, Sönning (2018: 6) argues that multilevel models should be the default statistical tool in (linguistic) research (cf. also McElreath 2020: 14-15) and "linguists should be justifying why they are not using a multilevel model rather than the other way around". My questionnaire data are no exception in this regard.

The dataset generated from the present questionnaire contains n = 20,039 data points, each representing a single rating. In the data hierarchy, these will be referred to as level-1 units. These individual observations are grouped in meaningful ways, i.e. there is a logical connection between the data points that the model needs to be informed about in order to draw legitimate inferences (***assumption of non-independence***). The logical connection between level-1 units is conceptually realized by three types of ***clustering***; these are the level-2 units in the model. First, my data points are clustered by source, as is common in most linguistic research contexts. Thus, each informant (n = 675) contributed n = 30 ratings to this study. Informants may vary in their preference for a particular gradation

strategy, whether this tendency be due to variation by social variables or a conscious urge to be internally consistent across items. I therefore expect similarities between data points from the same informant. Second, the ratings are clustered on linguistic grounds, too. As the morphophonological properties of the root adjective have been demonstrated to systematically influence the choice of gradation strategy, my data are clustered by adjective (n = 23). Finally, it has been noted that the questionnaire items were designed in such a way as to reflect variation along a single contextual constraint in order to operationalize the difference between test and control items as a measure of contextual complexity. Observations are therefore also clustered by item (n = 120) and item pair (n = 60). The relation between level-1 and level-2 units is one of ***partial crossing***. Thus, not every informant got to rate every item and adjective in the study. Similarly, not every adjective is combined with every contextual constraint. Clustering enables me to assess the relative importance of the three sources of variation covered in this study. Every unit within each cluster can in turn be considered a data point, and comparisons can be drawn between them to gauge ***between-cluster similarity***, i.e. the degree of variation among informants, adjectives and items.

The data structure is represented in the regression models I apply to the data. Thus, each level-2 unit (informants, adjectives, items) is assigned a ***varying intercept*** (also called a 'random intercept'), i.e. a unique average that describes the overall tendency of a given cluster (e.g. a particular adjective, informant or questionnaire item) on the outcome scale. The variation among clusters can then be modelled by including cluster-level predictors, which express relevant attributes of the clusters (e.g. the varietal background of a given informant, or the status as a "control" or "test" condition for a certain questionnaire item). Every data point therefore contributes to three varying intercepts, one for each cluster. In addition, ***varying slopes*** (also called 'random slopes') on item pair (control and test item) and informant were added for each contextual predictor (*synthetic persistence*, *analytic persistence*, *syntactic position*, *infinitival complement*, *segmental horror aequi*, *end weight*, *rhythmic harmony*). This allows the effect of the contextual variables to vary across subjects and questionnaire items.

I used a set of regression models to address my research questions. My first model[35] addresses the social dimension of variation (see Section 3.2.4), and thus focusses on the variation I observe among the informants in our study. This model includes, as predictors, the variables *age*, *gender*, *education* and *variety*. This model included random intercepts for informants (informing the model that we are essentially dealing with n = 659 observations[36] – i.e. speakers – when studying socio-demographic patterns of variation), adjective and item pair.

For the contextual predictors (see Section 3.2.3), I proceeded by fitting a separate regression model[37] for each of the seven constraints. These models were structured identically, including, as a fixed effect, the focal predictor (e.g. *end weight*) that was manipulated via the design of our item pairs (test vs. control item). Each constraint model also included random intercepts for informants and adjectives, as well as by-informant and by-adjective random slopes for the focal constraint.

Finally, another set of seven models[38] were fit to assess whether the effects of the contextual constraints differ between BrE and AmE. Essentially, each model estimates the constraint of interest separately for BrE and AmE, thus revealing potential differences between the varieties. In statistical terms, the model included, as a fixed effect, an interaction between variety and each of the seven constraints. Apart from this, the models were structured identically to those not including this interaction (see previous paragraph).

[35] The R script for fitting the model as well as the regression table can be found in the web appendix (https://osf.io/hjeq2/). Thanks are due to Lukas Sönning for his invaluable help with the statistical analysis.

[36] This number deviates from the original n = 675 since informants with missing information ("NA") for at least one of the social predictors were excluded from the analysis.

[37] For details on the model specification and output, see https://osf.io/hjeq2/.

[38] See https://osf.io/5896e/ for details.

The statistical analyses that have been conducted in this study use methods of Bayesian inference.[39] All analyses[40] were carried out in R (R Core Team 2020), using the facilities provided in the packages *brms* (Bürkner 2018) and *RStan* (Stan Development Team 2020). For data visualization, I relied on the packages *lattice* (Sarkar 2008) and *ggplot2* (Wickham 2016). To make the results as interpretable as possible, the presentation strategy that will be adopted in the following draws on two key features: First, intuitive measures such as proportions (or differences between proportions) will be given priority over log-odds or odds ratios and communicated in the form of graphs, and second, posterior uncertainty estimates will be visually displayed by 50 % and 90 % error bars and bands. For the present purposes, these uncertainty intervals can be interpreted along similar lines as frequentist confidence intervals.[41]

3.3 Results and discussion

The current section is structured as follows: The bird's eye view of results presented in Section 3.3.1 enables me to determine the relative effect strengths of all predictors operationalized in the questionnaire study. The following Section 3.3.2 affords a more detailed investigation of results obtained for the linguistic constraints. On this basis, I will evaluate the explanatory power of the assumed cognitive compensatory mechanism of *most*-support in the superlative alternation. The role of speaker variables in this domain of morphosyntactic variation is assessed in Section 3.3.3. Section 3.3.4 addresses differences between British and American English with regard to gradation strategy choice. Drawing on selected statisti-

[39] Bayesian data analysis offers some crucial advantages over the frequentist methods that are widely used in quantitative work. Two of these advantages shall be briefly mentioned: First, Bayesian analysis provides great flexibility for the summary and presentation of results. Second, it enables the researcher to include into the analysis pre-data information about the phenomenon under scrutiny (in the form of prior distributions). For a more detailed survey of the advantages of the Bayesian paradigm see Sönning (2020: 46-50).

[40] The R script written for the statistical analysis is provided in the online appendix at https://osf.io/5896e/. The questionnaire dataset is openly available via the *Tromsø Repository of Language and Linguistics* (Beland 2021).

[41] See Sönning (2022: 28-30) for an account of the different interpretations of frequentist confidence intervals and Bayesian uncertainty intervals.

cal metrics, I will address different sources of variation in my data (Section 3.3.5) and determine the degrees of variability among informants and item pairs (Section 3.3.6). Finally, I will critically reflect on the quality of the current elicitation instrument (Section 3.3.7). In this context, I will highlight selected strengths and weaknesses of corpus and elicitation data in the study of grammatical alternations.

3.3.1 Overall model: A bird's eye view of linguistic constraints and speaker variables

The present section provides a summary of results for two dimensions of systematic variation in the data:

- external (social) predictors: *age*, *gender*, *education* and *variety*
- internal (linguistic) predictors: the seven contextual constraints

Let me first focus on the linguistic predictors. Recall that each of the models that were fit to the data included

- the predictor of interest as a fixed effect: This estimate indicates whether the preference for *most* is sensitive to the respective predictor. A positive value indicates an increase in the probability of observing *most*.
- by-subject random slopes for the predictor of interest: I am interested in the distribution of these random slopes, since this distribution indicates how much the effect of a constraint varies among informants. This is worthwhile to know, since it tells me how consistent the informants were with regard to how they responded to manipulations of the contextual constraint in focus.
- by-adjective/item-pair random slopes: My main interest here is again to see whether there was variation among the item pairs that were designed to elicit the effect of interest. I am hoping for consistent patterns, that is, I would like all item pairs to point into the same (or at least a similar) direction. Note that, for a given constraint, each item pair includes a different adjective (cf. Section 3.2.1). This means that, in case we are confronted with a deviant pair, this may be attributable to either (i) the wording of the items,

(ii) the adjective itself, or (iii) both. It is impossible to disambiguate between these potential explanations for idiosyncratic item pairs.

I will start with a visual summary of the patterns that can be observed in the data. To this end, I will construct what are often called marginal (or partial) effects plots. These displays are drawn for each predictor of interest. When interpreting these graphs, it is best to think of the estimates as showing isolated differences between different values or levels of a certain predictor. These differences are "isolated" (or marginal/partial) in the sense that the model is asked to hold constant ("control for") the other predictors in the model and only allow the focal variable to vary. Varying the focal predictor (e.g. *age* from 20 to 80 years, or *infinitival complement* from absent to present), I can ask the model to produce the predicted proportion of analytic superlatives at these predictor values. Comparing these proportions allows me to quantify how large (or small) a difference we observe between representative predictor values. I can attach an uncertainty interval to these differences and compare the differences across the predictors in my models.

Figure 6 illustrates the implied predictions of the overall model on the proportion scale.[42] It shows how the predicted proportions of analyticity (on the y-axis in the bottom row of panels) change as the value of one predictor (on the x-axis) is changed while all others are held constant. Note that in the charts for the contextual constraints, the predictor setting that is expected to result in a greater proportion of *most* is always on the right. The labels in these charts must be read as follows (from left to right):

- *Rhythmic harmony*: SH stands for synthetic harmony, AH for analytic harmony. Thus, if the *-est*-variant is prosodically more harmonious, the proportion of *-est* is expected to be greater. The reverse applies in the case of analytic harmony.
- *Horror aequi*: H = segmental *horror aequi*
- *Syntactic position:* A = attributive, N = nominal (fused-head NP)
- *Infinitival complement:* C = infinitival complement

[42] See the online appendix at https://osf.io/d6c29 for a summary of the posterior predictions for the internal and external predictors.

- *Persistence:* P = superlative prime in the previous sentence (synthetic or analytic, respectively), Ø = absence of prime/contextual constraint
- *End weight*: E = sentence-final position

The top row of panels displays the difference between selected values for each variable. For non-binary variables, the values chosen to calculate this difference are marked with an arrow on the x-axis of the respective proportion chart below. This difference value enables us to determine, in a bird's eye fashion, whether the change of a predictor value results in the hypothesized change in proportional analyticity. With the exception of the predictor *education* (for which we had no clear expectation), the arrangement of predictor levels and differences is such that we should observe an upward trend from left to right, and therefore a positive difference in the top row of panels. Since I predict that the test condition will result in higher odds of *most* for all contextual constraints except *synthetic persistence* and *rhythmic harmony*, the difference for these six constraints is calculated as test value *minus* control value. Conversely, in the case of *synthetic persistence*, the test value is subtracted from the control value since I expect an increased proportion of *-est* in the presence of a synthetic prime. As for *rhythmic harmony*, we subtract the value for synthetic harmony (i.e. items in which the synthetic superlative affords a prosodic advantage) from the value for analytic harmony.

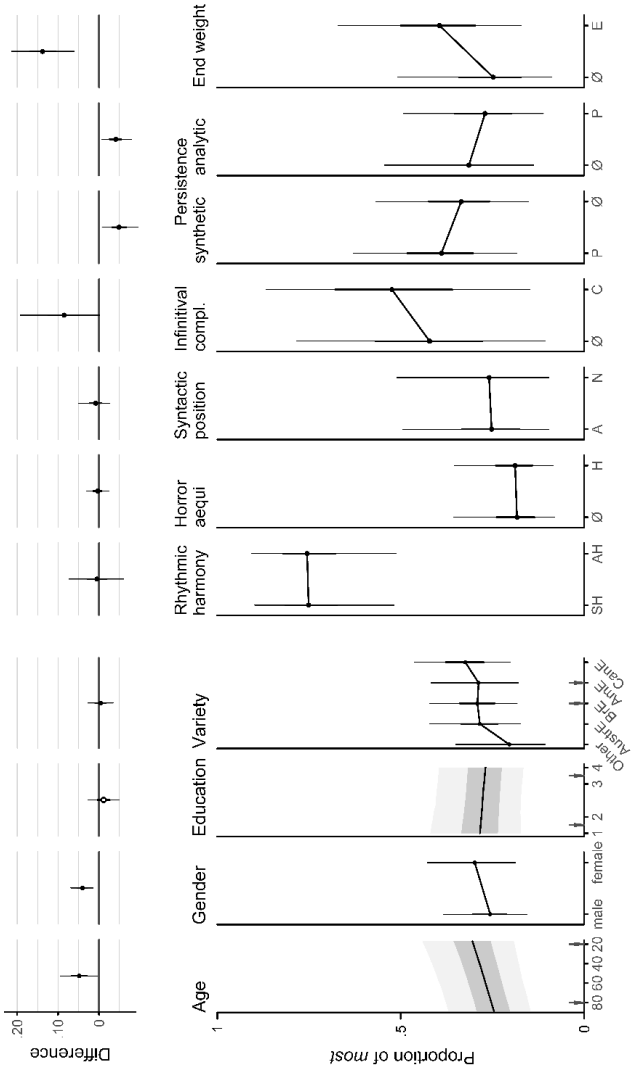

Figure 6. Proportions of the analytic superlative for the four speaker variables and seven contextual constraints in the overall model (bottom chart). The chart at the top renders the differences of the reference values marked on the x-axis. The error bars/bands display the 50 % and 90 % posterior uncertainty estimates.

I will start with the results for the four speaker variables. For informant *age*, the model suggests a mild increase in the proportion of *most* from 20- to 80-year-olds ($\Delta_{20\text{-}80}$ = .05 [.00; .10])[43]. Females use more analytic superlatives than males ($\Delta_{f\text{-}m}$ = .04 [.01; .07]). The flat trend for education implies that there are no sizeable changes in proportional analyticity as we move along the four-point ordinal scale. Accordingly, the difference value $\Delta_{edu\ 3.5\text{-}1.5}$ = –.01 [–.05; .03] renders no effect in either direction. It is important to note that the ordinal scale for *education* was slightly adapted in the statistical analysis: As the focal comparison for this variable was designed to be university graduate (undergraduate and postgraduate) versus primary/secondary education only, the difference between the ordinal categories 2 (secondary) and 3 (undergraduate) was considered twice as large as the remaining differences on the scale. The means between the respective adjacent categories (1.5 and 3.5) were consequently used for calculating the difference in the chart at the top.

The results for *variety* paint a rather inconsistent picture: While the proportional analyticity values are virtually identical for AmE (.28 [.17; .42]), AustrE (.28 [.17; .42]) and BrE (.29 [.18; .42]), that for CanE is at .32 [.20; .46], and for the residual category 'Other' it is at .20 [.10; .35]. Our comparison of AmE and BrE shows that the difference we observe is negligible and estimated at –.005 [–.04; .03] (i.e. 0.5 percentage points).

Let us now turn to the results for the seven contextual predictors. We find no support for the predicted tendency of language users to alternate between the superlative forms to maintain the ideal alternating rhythm in the overall data (Δ_{rhy} = .00 [–.06; .07]). The assumed avoidance of /st/ at the junction of superlative coda and noun onset along the lines of segmental *horror aequi* is also not borne out by the data: Contexts in which the synthetic superlative would cause the repetition of /st/ in immediate adjacency do not come with a significantly higher proportion of *most* (Δ_{hor} = .00 [–.02; .03]).

The functional opposition of attributive and nominal superlatives in a fused-head NP does not translate into an appreciable difference in the proportion of analyticity (Δ_{pos} = .01 [–.03; .05]). *Infinitival complement* (Δ_{inf} = .09 [.00; .20]) and *end weight* (Δ_{endw} = .14 [.05; .22]) have the strongest effects on the choice of superlative form of all constraints under analysis.

[43] The values in square brackets indicate the [lower and upper] limits of the 90 % posterior uncertainty estimates.

The presence of an infinitival complement and the end weight factor appear to increase the probability of *most* by 9 and 14 percentage points, respectively.

Contrary to expectations, the predicted proportions of *most* for the two persistence variables speak in favor of identity avoidance, possibly as a means of stylistic variation: While a synthetic prime in the previous sentence comes with a decrease of 5 percentage points in the share of *-est* in the superlative target ($\Delta_{p\_syn} = -.05$ [−.10; −.01]), an analytic prime reduces the probability of *most* by 4 points ($\Delta_{p\_ana} = -.04$ [−.08; .00]).

3.3.2 Linguistic constraints in the overall model

Phonological constraints

The findings presented in the previous section suggest that neither of the two phonological predictors under analysis has a noticeable effect on superlative strategy choice: The expected tendency of informants to prefer the rhythmically more harmonious superlative is not borne out by the data at hand. Similarly, we find no support for the hypothesized inclination to segmental *horror aequi*, i.e. the urge to avert the repetition of /st/ in immediate adjacency at the junction of an attributive synthetic superlative and the noun it premodifies.

The results for *rhythmic harmony* thus fail to provide unequivocal support for Schlüter's (2005: 267-271, 2015: 200-201) non-modular but layered and interpenetrative model of linguistic subsystems that allows for prosodic conflicts to be bypassed on levels of linguistic analysis other than phonology. According to this model, violations of the Principle of Rhythmic Alternation can be prevented on the next higher level of morphology, that is by opting for the prosodically less problematic superlative variant. Both phonological constraints under study can theoretically be accounted for in terms of the *Obligatory Contour Principle* (McCarthy 1986; Yip 1988), which generally prohibits the adjacency of identical units, whichever form they may take (cf. Section 3.2.3). The neurological mechanisms underlying this principle as hypothesized by Schlüter (2005: 257-306) can be sketched as follows: Levels of linguistic processing are connected via nodes which undergo a cycle of activation and recovery. In perception, activation spreads from the lowest level, the phonetic one, to levels of

higher order, i.e. morphology, syntax and semantics. In the present context, the connection of the phonetic and morphological levels through nodal activation paves the way for a view of language processing as alternating between a planning and execution phase, whereby a loop of feedforward and feedback simultaneously pre-activates both superlative competitors on the morphological level and feeds this information into a phonetic node. Looking ahead to subsequent phonetic elements in the planned utterance, the phonetic node feeds back into the morphological ones; ultimately, in the execution stage, the variant with the highest level of activation wins out over its competitor (cf. Schlüter 2005: 279). In practice, this would imply the following scenario for segmental *horror aequi* in the default attributive example *the* [CLEVER] *student*: The morphological nodes for *cleverest* and *most clever* are preactivated; this information is fed into the phonetic node, which subsequently estimates the activation levels of both morphological nodes in anticipation of the following /st/ cluster in *student*. Ultimately, the node for *most clever* knocks out its synthetic competitor and is activated in the execution stage due to a sufficiently large phonetic buffer that intervenes between the identical consonant clusters in *most* and *student*.

Why, then, do the phonological predictors *rhythmic harmony* and *segmental horror aequi* yield no effect in my analysis? We may legitimately raise doubts as to whether the design and logistics of the present questionnaire study, relying, after all, on elicited ratings of forced-choice items, can reliably operationalize the constraints in focus. Although explicitly instructed to do so, some informants can be assumed not to have read the questionnaire items out aloud, thereby failing to reinforce the assumed phonetic conflict in auditory perception. Methodological modifications are required to gauge the effect of phonological constraints more accurately in future studies, ideally adopting a design that requires informants to produce the form of interest independently (cf. the production task in Kunter 2017: 108-109) instead of selecting the preferred form from a given range of options. As was pointed out in the theoretical description of *segmental horror aequi* (cf. Section 3.2.3), corpus studies are rather ill-suited to examine this constraint, since natural production data contain only a very limited number of instances in which the superlative of non-categorical types is followed by word-initial /st-/. The results at hand add to the body of inconsistent evidence regarding the effects of segmental identity

avoidance on morphosyntactic alternations: They are in keeping with Bresnan's (2011) finding that the imminent adjacency of recipient-final and theme-initial sibilants in the ditransitive dative (e.g. *to give the kids sweets*) fails to increase the odds of the prepositional variant, but contradict Szmrecsanyi et al.'s (2017: 20-21) conclusion that speakers resort to the *of*-genitive when facing the *s*-genitive of a noun ending in a sibilant. However, the tentative findings I have provided should not be taken to invalidate Schlüter's (2005: 257-306) nodal activation account sketched in the previous paragraph.

Syntactic constraints

My findings for the predictor *syntactic position* do not lend support to the hypothesis that nominal superlatives have a greater probability of *most* than attributive ones. This observation raises doubts about the complexity-based line of argument pursued in Section 3.2.3, which rested upon the assumption that fused-head NPs involve a high degree of fusion between the nominal head and the extraposed superlative. The ensuing increase in noun phrase complexity was linked to higher processing costs as compared to an attributive superlative. In generative terms, it was argued that a trace of the nominal head in the slot to the right of the extraposed superlative prompts language users to cognitively reactivate the head noun; by hypothesis, this reactivation incurs additional processing expenses that, by analogy with Mondorf (2009b), can be compensated by the *most*-superlative, since this variant affords earlier and thus easier recognition of phrase structure. In the light of the findings for the predictor *syntactic position*, it is conceivable that a comparison of attributive and genuinely predicative superlatives possibly would have yielded more sizeable effects in the expected direction. While the results for this functional opposition could more suitably be compared to those for the comparative, predicative superlatives are often ambiguous between different semantic readings, as I have argued in Chapter 2 in reference to Claridge (2007a). Future studies should be attentive to this distinction.

In line with expectations, an infinitival complement results in a robust increase in the probability of the analytic superlative. It is important to remind the reader that both the control and test conditions for this variable rendered the superlative in a fused-head NP in order to eliminate positional effects. The results can be taken as empirical evidence in favor of

the hypothesis that a *to*-infinitive entails increased processing costs. The assumed cognitive mechanisms underlying this effect will be sketched in the following, again drawing on the example used in Section 3.2.3.

(55) **[Your upper-back muscles]**$_1$ are the **most tricky to** target t_1 in any exercise. [COINF_2_T]

First, topicalization of the [object]$_1$ in (55) leaves a trace (t_1) in the slot from where it has been dislocated; when parsing the utterance, the cognitive reactivation of the object in this slot requires additional processing effort. The link of phrasal complexity and processing costs has received ample empirical support in previous work (e.g. Hawkins 1992; Rohdenburg 1995, 1996; Berlage 2014, *inter alia*). As for the three criteria of phrasal complexity established by Berlage (2014: 2), it is obvious without further analysis that a nominal superlative complemented by a *to*-infinitive wins out over a superlative lacking such complementation by having (i) more linguistic material, (ii) more phrasal nodes and (iii) a higher degree of sententiality.

Second, in reference to Mondorf (2009b: 59), the processing advantage of *most* in a tough movement construction can be accounted for in terms of Hawkins' (2001) *Early Immediate Constituents Principle*: Head-adjacent orderings enable early phrase structure recognition and are thus preferred over orderings in which linguistic material intervenes between phrasal heads. The analytic superlative in (55) warrants a logical presentation of syntactic relations, whereby the degree marker *most* primes the degree phrase, which in turn is completed by the lexical information contained in *tricky*; the phrasal cue to the infinitival complement (*to*) occurs in immediate adjacency to the adjective. The synthetic superlative, by contrast, produces the following sequence of syntactico-semantic information: adjective > degree > complement. Successful parsing of this sequence requires that recipients establish the dependency of adjective and complement. The intervening degree-marking suffix increases the distance between the constituents in focus, while the analytic form maintains their adjacency.

Gibson's (2000) *dependency locality theory* (DLT) affords a more comprehensive psycholinguistic account of this assumed relationship between the distance of syntactically co-dependent items and the processing

resources required on behalf of the recipient. According to Gibson (2000: 95), two processes incur computational costs during sentence parsing:

1. Performing structural integrations: connecting a word into the structure for the input thus far.
2. Keeping the structure in memory, which includes keeping track of incomplete dependencies.

In order to demonstrate the role of these component processes in the parsing of superlatives complemented by a *to*-infinitive, I will again draw on the opposition *trickiest* vs. *most tricky to target* in example (29) above. Parsing the analytic variant keeps both storage and integration costs at a minimal level since the dependencies between the adjectival head and its co-dependent elements are established immediately; as for the synthetic variant, by contrast, the lexical information contained in *tricky* must be stored in memory over a sequence of two linguistic items, as the degree-marking suffix prevents the parser from immediately resolving the dependency between the adjective and its complement. This situation can further be explained along the lines of neural activation: The successful integration of the infinitival complement and its adjectival head requires that linguistic characteristics of the adjective, e.g. the semantics and collocational preferences of *tricky*, be retrieved from memory. Since the overall scope of activation available for sentence parsing is limited, the activation level of the adjective decreases "as intervening words are processed and integrated into the structure for the input" (Gibson 2000: 103). Thus, in the case of *trickiest to target*, two aspects can be assumed to increase processing costs: First, the linguistic information in *tricky* must be stored and retrieved upon encountering *to target*. Second, other structural integrations take place in the interim: the dependency between the degree suffix and the antecedent adjective is established; the likelihood of upcoming linguistic units is computed. In terms of the overall activation in the parsing system, these interim integrations take away market share from the primary integration that resolves the dependency between head and complement. It is important to remark that Gibson (2000), in order to substantiate his DLT, exclusively draws on examples of nesting complexity, i.e. relative clauses nested between a NP and a verb, as in (56):

(56) The reporter [$_{S'}$ who the senator attacked] disliked the editor. (Gibson 2000: 96)

According to this line of argument, then, the linguistic elements that interfere with the integration in focus are words. There is, however, no evident reason why this theory should not apply to smaller units of linguistic code (e.g. inflectional suffixes) in the same way. Further, Gibson (2000: 104-105) proposes a method of quantifying the structural integration costs incurred by a dependency as the linear function of the distance between the co-dependent items. As every intervening item is assumed to correspond to 1 energy unit (EU), the processing advantage afforded by the analytic over the synthetic superlative in a tough movement construction is minimal according to this measure. Nevertheless, this difference in processing complexity might translate into a noticeable preference for *most*, as suggested by the questionnaire data at hand.

The observation that infinitival complementation has a relatively sizeable impact on the superlative alternation ties in with the results presented by Hilpert (2008: 407) and Cheung & Zhang (2016: 573); notably, for the comparative, this constraint is the strongest syntactic predictor in both studies. In the light of the present findings, however, it is difficult to account for the observation reported in Cheung & Zhang (2016: 576-577) that for the superlative, adverbial intensifiers and position score higher in the constraint ranking than infinitival complementation. This discrepancy is possibly due to two factors: First, Cheung & Zhang (2016) did not control for the end weight factor, which has been shown to interact with the variable *syntactic position* and is the strongest contextual predictor of the analytic superlative in both models that were fit to the data at hand. This may have unduly increased the effect of *syntactic position* in their superlative model. Second, the authors did not confine their analysis of this variable to obligatory *to*-infinitive complements in tough movement constructions but may have also included optional infinitival complements and adjuncts (see Mondorf 2009b: 60-61 for the distinction between the three types), which, by hypothesis, come with a lower degree of syntactic and conceptual fusion between adjective and complement.

The results for both persistence predictors turn out contrary to expectations in that they consistently suggest the avoidance of structural paral-

lelism: A synthetic superlative prime in attribution to a noun in the sentence preceding the target entails a decrease in the probability of the *-est*-superlative; an analytic prime in the previous sentence results in a reduced proportion of *most*. However, the observation that the present results do not tie in with the previous literature on priming does by no means call into question the efficacy of this cognitive mechanism. Instead, a note on the operationalization of morphological persistence is in order: In the questionnaire study, informants were instructed to read the prime sentence and subsequently choose their preferred superlative variant from a given range of options. While a forced-choice setting differs in significant ways from naturalistic language production, the present design parallels comprehension-to-production, i.e. cross-modal priming effects as attested in psycholinguistic accounts (Bock et al. 2007; Litcofsky & van Hell 2019, *inter alia*). However, there is corpus-based evidence to suggest that cross-modal priming is weaker than intra-modal priming (Gries 2005: 374). Elicited data of the kind presented here may thus have little predictive power for effects of production-to-production priming in the sense of Szmrecsanyi (2006: 63-86).

Two further methodological concerns with regard to eliciting persistence effects must be addressed: First, as has been noted, the logistics of data collection required that each version of the questionnaire contained 30 items, all of which were designed to test superlative strategy choice; there were no distractor items. It is thus not far-fetched to assume that the superlative forms encountered in previous items might have influenced my informants' choice of superlative variant in a persistence item, in the sense that they either deliberately avoided or aimed for structural repetition on stylistic grounds.

Second, sensitivity to priming effects can be thought of as varying considerably across adjective types. Gries (2005: 386-387) notes that the collostruction strength of a linguistic item may influence its "primability" and advises that future studies control for this variable. My statistical analysis included the cluster-level predictor *adjective* and was thus sensitive to variation by adjective type. Collocation statistics, e.g. the type and token frequencies of superlative + noun bigrams for the adjectives used in the questionnaire, however, were not taken into account. Since the test and control conditions for both persistence variables rendered the superlative in attribution to a noun, this aspect may have affected the results for this

constraint. This caveat appears all the more relevant since two out of three superlatives occur in attribution, as noted in Section 3.2.3. Collocational effects are, by hypothesis, weaker in the comparative alternation, as the proportions of attributive and predicative uses of the comparative differ only marginally (Lindquist 2000: 126).

In sum, my questionnaire method poses methodological and conceptual challenges for the operationalization of persistence. Despite their respective shortcomings in the study of priming effects (Gries 2005: 386-387), both experimental and usage data appear more suitable in this research context than the elicitation data used in the present investigation.

End weight

The end weight factor comes with the greatest proportional increase in analyticity of all contextual constraints under study. In the theoretical discussion of this constraint (cf. Section 3.2.3), both pragmatic and syntactic motivations were identified as underlying this effect. In reference to Eitelmann (2016: 416), I argued that the end weight effect can be assumed to apply both in production and comprehension. In production, the extra time provided by the lexically bulkier analytic superlative at the end of a sentence can be invested by the producer in planning the following one. In addition, the one-to-one correspondence of form and function, with the code for the degree preceding the code for the adjectival root, grants additional time for adjective selection. I will again draw on the exemplary test item EW_7_T to illustrate the implications of my questionnaire results for sentence parsing.

(57) Of all students in class Chris was the **most friendly**. [EW_7_T]

The fused-head NP in (57) renders the nominal superlative in clause-final position, which requires that the restrictive modifier (instantiated by an *of*-prepositional phrase) be topicalized. This dislocation results in the maximum distance between the head of the NP and its modification. As with the predictor *infinitival complement*, Gibson's (2000) *dependency locality theory* can be adduced to account for the assumedly higher processing costs of clause-final as compared to clause-medial superlatives. In (57), it is not until the clause-final superlative is parsed that the "dangling" prepositional modifier can be integrated into the overall syntactic structure.

The linguistic material intervening between the co-referents requires that (i) the linguistic information contained in the modifier be stored until the dependency in focus can be established, and (ii) the parser conduct structural integrations in the interim that in turn require neural activation. Along these lines, then, it is obvious that a nominal superlative in clause-final position comes with considerably greater parsing effort than a nominal superlative in clause-medial position. It is not yet clear, however, why the analytic superlative should be more advantageous from a parsing perspective than its synthetic competitor. According to Eitelmann (2016: 415), the preference for explicit variants in clause-final position can be attributed to the *sentence wrap-up effect*, for which ample empirical support has been provided in psycholinguistic studies (Just & Carpenter 1980; Hirotani, Frazier & Rayner 2006; Tiffin-Richards & Schroeder 2018, *inter alia*). In the clause-final position, it is assumed, the human parser runs a review program that resolves dangling dependencies, checks the validity of dependencies established so far and corrects false semantic and pragmatic inferences if necessary. This process of 'wrapping up' the sentence consumes additional processing energy. It has been argued earlier that the analytic superlative comes with an ideal mapping of form and function, while the synthetic variant merges two functions in a single lexeme. Thus, the energy saved when parsing the analytic superlative can be allotted to the wrap-up process.

As the end weight factor has not yet been quantified in multifactorial analyses of the comparative alternation, the relative strength of this effect in the comparative vs. superlative alternation is presently a matter of speculation. In the line of argument set forth in this section, however, we should reasonably expect that end weight is among the stronger constraints on superlative strategy choice. Neither of the two complement types of the comparative (*to*-infinitive and prepositional phrase) can be dislocated from the adjectival head;[44] therefore, since predicative comparatives in clause-final position cannot have a complement, no processing costs can be incurred that correspond to those required for establishing the dependency between a clause-final superlative and its topicalized

[44] Mondorf (2009b: 78-80) shows that *than*-phrases constitute a postmodification rather than a complement; this is due to a weaker degree of fusion, both syntactic and semantic, between the adjective and the complement as compared to the other two complementation types (*to*-infinitive and prepositional phrase).

modifier in a fused-head NP. It would be worthwhile for future studies to compare the strength of this predictor in multivariate analyses of the comparative and superlative alternations on the basis of identical data sources. Ultimately, it can be concluded that the processing advantage afforded by the analytic variant arises from the syntactic structure in which the superlative is embedded; thus, the end weight effect has both a syntactic and pragmatic dimension.

Five of the seven contextual constraints operationalized in the present questionnaire study were adapted from Mondorf's (2009b) analysis of the comparative. In her account, the effects of these linguistic predictors, as well as a wide array of other determinants, were subsumed under the cognitive compensatory mechanism of *more*-support (Mondorf 2009b: 6), which premises a link of processing complexity and explicitness on the level of the linguistic code. I have noted that according to Boyd (2007: 32), the processing benefit of the more explicit analytic variant can be ascribed to two hypothesized advantages: The *warning hypothesis* holds that the preposed degree marker serves as a signal to the recipient that a degree phrase follows; thus, the cue to phrase structure is rendered prior to the lexical information denoted by the adjective, the knowledge of which is secondary to phrase structure recognition. The *parsing hypothesis* proposes that the one-to-one correspondence of form and function afforded by the analytic form obviates the need to mentally decompose a dimorphemic unit. In the following section, I will more comprehensively assess the explanatory power of this hypothesized cognitive compensation strategy in the superlative alternation.

Support for most-support?

The results of the present elicitation study empirically corroborate the hypothesis that the cognitive benefits couched in terms of *more*-support (Mondorf 2009b) can also be identified for the superlative alternation; however, the applicability of this compensation strategy appears to be limited to syntactic constraints. While it has at length been elucidated how the two processing advantages of the *most*-superlative relate to syntax, this connection is less conclusive for constraints pertaining to other linguistic domains, e.g. identity avoidance in phonology. Thus, the notion of analytic support remains a problematic umbrella term, as has been noted by Hilpert (2008: 413). Notwithstanding this conceptual criticism, the results

at hand support the assumption that the analytic variant is indeed advantageous from a processing perspective: This claim has legitimately been challenged by several scholars before me (Szmrecsanyi 2009: 322; Bauer, Lieber & Plag 2013: 119-120; Kunter 2017: 50). Kunter (2017: 50) notes that the analytic variant requires that the human parser (i) retrieve the lexical information contained in two separate lemmas in the mental lexicon (the degree marker and the adjectival root) and (ii) establish the syntactic link between the two units. This can be taken to suggest that, after all, the parsing effort associated with *more* and *most* may not be as small as has previously been assumed. While this objection appears valid on theoretical grounds, the empirical evidence I have furnished supports the view that the analytic superlative is indeed easier to process than its synthetic competitor in contexts of increased syntactic complexity.

Another point of contention in the discussion of analytic support in gradation strategy choice revolves around the question of whether this mechanism operates both in production and perception. Mondorf (2009b: 68-69, fn. 51) explicitly decides not to comment on this question. Kunter (2017: 226), whose study of the comparative and genitive alternations specifically zooms in on this opposition, concludes that while the analytic comparative has a facilitatory effect in production, no such effect can be observed in perception. Since the elicitation data I have provided clearly are closer to production than perception data, the present findings support Kunter's verdict on production. After all, however, it appears unclear why speakers and listeners should prefer different gradation strategies. I noted earlier, albeit in a different context, that communication is essentially a cooperative activity: Producers can be assumed to draw on their own experience as recipients and consequently choose the variant they consider advantageous in perception. Thus, due to this process of persistent adaptation, we should expect that, all other things equal, producers and recipients prefer the same superlative form, i.e. the analytic one, in contexts of high complexity.

While the results I have presented speak in favor of *most*-support, it is hard to determine on the current data basis how this compensation strategy relates to *more*-support in terms of effect size. Cheung & Zhang (2016: 579) provide support for their conclusion that analytic support operates more strongly in the comparative than the superlative alternation. The discussion of the individual constraints in focus, however, consolidates

the insights that have been gleaned in rather theoretical terms in Chapter 2: Due to the fundamental differences between the two degree forms on various levels of linguistic analysis, it remains debatable whether the effects of *more-* and *most*-support can be reliably quantified against each other at all. Future studies that intend to compare the degree alternations with regard to contextual complexity may need to proceed with caution.

In an attempt to assess the effect of contextual constraints on superlative strategy choice, the present study zoomed in on a small subset of adjectives for which previous studies suggest alternating behavior. As such, the results at hand have limited predictive power for categorical types. Nonetheless, it can be assumed that the functional overlap, or "division of labour" (Mondorf 2003: 298), between the alternants is less pronounced for types whose strategy preference is largely determined by morphophonological properties. By hypothesis, it is a combination of these properties and frequency variables that determine the scope of *most*-support. The interplay of these factor groups will be examined in Chapter 4.

3.3.3 Speaker variables in the overall model

In this section, the results for the four speaker variables included in the overall model (*age*, *gender*, *education*, *variety*) will be addressed in turn. For ease of reference, the model-based trends for these speaker variables as shown in Section 3.3.1 are quoted again in Figure 7.

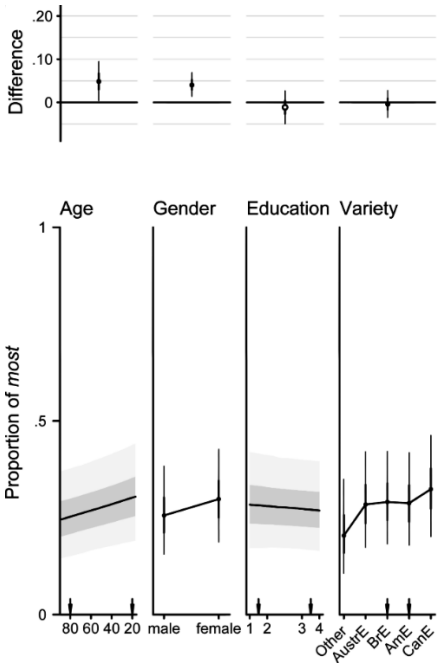

Figure 7. Proportions of the analytic superlative for the four speaker variables in the overall model (bottom chart). The chart at the top renders the differences of the reference values marked on the x-axis. The error bars/bands display the 50 % and 90 % posterior uncertainty estimates.

Age

The results for *age* suggest a mild but statistically robust apparent-time trend towards more *most* in the period from the mid-1950s until the mid-2010s. These findings are in line with those presented by Leech & Culpeper (1997: 370-371), Szmrecsanyi (2006: 85) and Leech et al. (2009: 264-267), who concur in the view that the analytic comparative has gained ground in the latter half of the 20th century. In Section 3.2.4, I attributed this instance of language change to the decline of institutionalized prescriptivism and the concomitant rise of communicative pragmatism that has been in progress across the Anglosphere since the 1970s (Crystal 2004: 523-525, 2012: 507-510). At the heart of a pragmatic use of language lies processing efficiency and thus, on the level of the linguistic code, explicitness. In line with the argumentation set forth in the previous subsection, it is plausible that this cognitively advantageous quality is afforded exclusively by the analytic superlative. However, since the overall model

does not consider potential interactions between contextual constraints and the social predictor *age*, it remains to be elucidated whether the empirically indicated rise in analyticity holds irrespective of the complexity of the linguistic environment or mirrors increased sensitivity to contextual constraints along the lines of *most*-support. If the use of cognitive compensation strategies indeed varies according to degrees of formality, as Mondorf (2009b: 171-188) implies, it is not unlikely that younger and older speakers employ *most*-support to different extents. In the following, I will argue that the increase in analytic superlatives is independent of contextual complexity but can be taken to reflect a more general preference for explicit morphosyntactic variants. The time period under study saw fundamental changes in various domains of society, many of which can be subsumed under the term *globalization*. The growing importance of English as a lingua franca in a globalized world gives rise to systematic language change, as can be inferred from Carlier, De Mulder & Lamiroy (2012: 292):

> The more languages spread over large populations and involve frequent language contact between individuals who are related to each other by weak ties, the faster languages may evolve by regularizing mechanisms, ultimately also reducing their morphological and grammatical systems.

When native speakers communicate with non-native speakers or native speakers from a different varietal background, they may aim to reduce the complexity of their utterances in an effort to facilitate comprehension. Due to the parsing advantage of the analytic over the synthetic superlative, which has at length been accounted for along the lines of a one-to-one mapping of form and function, the system of degree formation, by hypothesis, undergoes regularization towards the more explicit analytic variant. In degree strategy choice, *more* and *most* do not only come with the benefit of being processed prior to the lexical information contained in the adjective; importantly, owing to their high frequency, these lexemes are also well-entrenched in a learner's memory as the default degree marker *and* adverbial quantifier, so that the activation of this item has become a highly automated routine from the early stages of the acquisition process. Native speakers of English can be assumed to be intuitively aware of this effect and regularize towards *more*/*most*, possibly even when con-

versing with fellow native speakers. Sensitivity to this practice is hypothesized to have increased throughout the second half of the 20th century due to the advance of English as a lingua franca. This account receives support from Michaelis & Haspelmath (2017: 15), who find that grammars generally become more analytic in situations of language contact. Future studies of the degree alternations should bring into focus the interactions between contextual constraints and the variable *age* to tackle the question as to whether the observed apparent-time trend towards analyticity is dependent on or applies regardless of contextual complexity.

Gender

As for the variable *gender*, the results of the overall model support the tentative hypothesis that women use slightly more *most* than men. It is imperative to reiterate that my analysis of this social predictor was largely exploratory in nature. In Section 3.2.4, I related the assumed female-led increase in analytic superlatives to rather general sociolinguistic principles, maintaining that women as primary linguistic innovators should be expected to spearhead the adoption of features with a wide regional distribution. The line of argument advanced in the previous paragraph suggested that the growing importance of ELF in the period under scrutiny fueled a tendency towards greater linguistic explicitness instantiated by a rise in analyticity in the domains of syntax and morphology. It is possible that women are more advanced than men as regards this change in progress. However, I decidedly caution that, pending further empirical investigation, this finding should not be taken as evidence in support of the vague stereotype that women are more cooperative communicators than men, as couched in equally tentative terms by Cameron (1985: 51).

Education

The descriptive account of the variable *education* in Section 3.2.3 concluded on the admittedly unsatisfactory note that formulating a hypothesis for this predictor is a rather intricate undertaking. The results imply that the proportion of *most* does not vary significantly according to education level. Two alternative scenarios were advanced for this predictor, both of which revolve around the link between educational attainment and the opposition of linguistic dogmatism and pragmatism. In the light of the

evidence I have furnished, it is valid to challenge this assumed relationship and further raise doubts about the conceptual and methodological validity of the variable *education* as operationalized in the present questionnaire. A four-point ordinal scale modelled on the conventional and culture-unspecific breaks of the education system is not fine-grained enough to accommodate the growing diversity and permeability of present-day career trajectories. It is imperative that future work on the social correlates of morphosyntactic alternations investigate this predictor on a higher level of granularity. Owing to the limitations posed by the logistics of data collection, this was not possible in the study at hand.

Variety

For the predictor *variety*, the overall model charts a picture of homogeneity: AmE, AusE and BrE score almost identically on the analyticity scale. CanE, the odd one out, has a marginally greater proportion of *most* than the other varietal categories. With the main reference varieties AmE and BrE so similar to each other in terms of proportional analyticity, it is hard to argue that the perceived difference between the AmE and CanE analyticity scores is due to BrE influence, which, according to Boberg (2012), is traceable at least in CanE vocabulary and pronunciation. Moreover, based on the data I have presented, it is indefensible that CanE should be generally more sensitive to contextual complexity than AmE. An additional variety model based on the AmE and CanE sub-datasets would be necessary to shed light on this scenario; due to the small size of my CanE sample, however, an analysis of this kind would not yield reliable inferences. The CanE lead over AmE in terms of analyticity, as suggested by the overall model, should not be taken as indicative of a systematic trend. Across the board, the present analysis implies that the four main reference varieties under study are rather uniform as regards the overall proportions of analytic superlatives, even though it is likely that the influence of individual constraints may vary according to variety. Drawing on the results of the variety model, I will explore this possibility for BrE and AmE in the following section. The hypothesis that American informants generally use more *most* than their British peers is not borne out by the data at hand, however.

The insights provided in this section lend plausibility to the view that the English superlative is indeed a sociolinguistic variable, since its formal

realization varies according to selected social characteristics of language users. While *most* is used slightly more often by younger (as compared to older) speakers and females (as compared to males), superlative strategy choice seems to be largely independent of education and national variety. In line with expectations, however, speaker variables govern the alternation in focus to a considerably lesser degree than contextual constraints. This verdict dovetails with the findings presented by Jenset, McGillivray & Rundell (2018: 206) on the dative alternation.

3.3.4 British-American differences

The results presented in this section rely on a model that will henceforth be referred to as the *variety* model. This model differs from the overall model in one additional feature: It contains interactions between the linguistic predictors and the varietal categories BrE and AmE and thereby allows us to compare these varieties with regard to their sensitivity to contextual constraints.[45] We will start out with our results for these constraints (Section 3.3.4.1) and then take a more detailed look at the individual adjectives in the two varieties (Section 3.3.4.2).

Linguistic constraints in a crossvarietal perspective

The presentation of results proceeds from the bottom to the top of the grid in Figure 8. This figure again renders posterior predictions on the proportion scale. The bottom panel displays the proportion of the *most*-superlative in the control versus test settings for each constraint, distinguished by variety.[46] Note that in the bottom panel, again, the predictor setting that is expected to result in a higher probability of *most* appears on the right. The labels in the bottom panel are the same as those used in Figure 6 (Section 3.3.1).

The middle row of panels shows the difference between the test and control values, i.e. a difference between two proportions. As in Figure 7, positive values indicate an effect in the predicted direction.

[45] This section is a slightly extended version of the results section in Beland (2022).

[46] The tables in the online appendix at https://osf.io/psqta list the predicted proportions of *most* for each constraint in the BrE and AmE data.

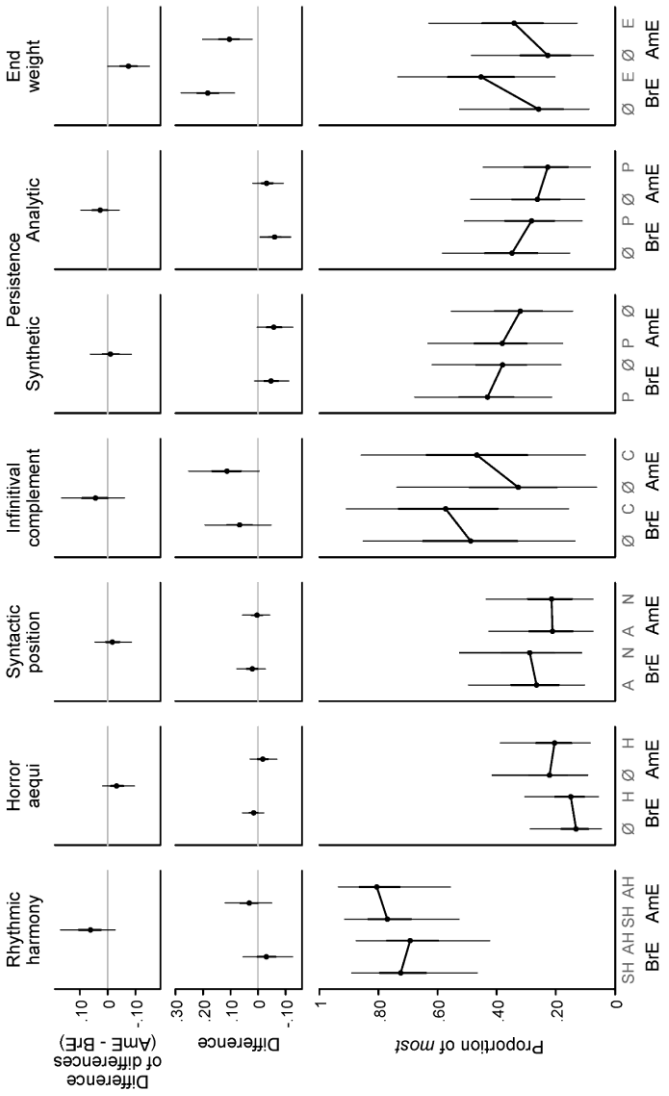

Figure 8. Proportions of analyticity for test and control conditions of the seven contextual constraints distinguished by variety (bottom panel). Difference of analyticity values in test and control conditions for BrE and AmE (central panel). Difference of differences, i.e. varietal difference of AmE and BrE test-control differences (top panel). The bold and fine lines visualize the 50 % and 90 % posterior uncertainty intervals.

The top row of panels shows whether and how BrE and AmE informants differed in their sensitivity to a given contextual constraint. It renders the difference between the values shown in the middle row and it is therefore a difference of differences. Positive values imply that American English is more sensitive to the proposed effect, in the sense that we see a greater tendency to opt for *most*-support in environments in which the analytic variant is assumed to be cognitively more advantageous along the lines of Mondorf (2009b).

The predicted tendency of language users to alternate between the superlative forms to maintain the ideal alternating rhythm receives tentative support in my AmE data ($\Delta_{\text{rhy AmE}} = .03$ [–.05; .12]) while no such effect is found for BrE ($\Delta_{\text{rhy BrE}} = –.03$ [–.13; .06]). The assumed avoidance of /st/ at the junction of superlative coda and noun onset along the lines of segmental *horror aequi* is not borne out by the data: Contexts in which the synthetic superlative would cause the repetition of /st/ in immediate adjacency do not come with a significantly higher proportion of *most* in either variety under scrutiny ($\Delta_{\text{rhy BrE}} = .02$, $\Delta_{\text{rhy AmE}} = –.02$).

The comparison of attributive and nominal uses reveals mildly diverging varietal trends: BrE exhibits a slight effect in the expected direction, with the fused-head NP scoring higher in analyticity than the attributive superlative ($\Delta_{\text{pos BrE}} = .02$ [–.02; .08]); my results for AmE, however, neither support nor invalidate the hypothesis of an increase in *most* along the lines of processing complexity ($\Delta_{\text{pos AmE}} = –.02$ [–.08; .04]).

Infinitival complementation comes with the second-strongest effect of all predictors, raising the likelihood of observing *most* by 7 and 11 percentage points ($\Delta_{\text{inf BrE}} = .07$ [–.04; .19]; $\Delta_{\text{inf AmE}} = .11$ [.00; .25]). The results for this constraint are statistically robust although the estimates come with relatively sizeable uncertainty; this is partly due to the fact that the questionnaire contained only four item pairs for this variable. In terms of effect strength, infinitival complementation is only surpassed by end weight, which leads to stable preferences for the analytic superlative in BrE ($\Delta_{\text{endw BrE}} = .18$ [.08; .28]) and AmE ($\Delta_{\text{endw AmE}} = .10$ [.02; .20]). Strikingly, BrE reacts more strongly to the end weight factor than AmE ($\Delta_{\text{AmE-BrE}} = –.08$ [–.16; –.01]), the opposite applies for infinitival complementation ($\Delta_{\text{AmE-BrE}} = .04$ [–.06; .17]).

As for persistence in superlative strategy choice, the presence of a synthetic prime in the preceding sentence comes with a moderate increase of

the *most*-superlative, while an analytic prime appears to decrease the proportion of *most*, consistently suggesting the avoidance of structural parallelism. We observe only minor differences between BrE and AmE with regard to the strength of these effects (synthetic persistence: $\Delta_{\text{AmE-BrE}} = -.01$ [−.09; .06], analytic persistence: $\Delta_{\text{AmE-BrE}} = .02$ [−.04; .09]). These findings seem to shed critical light on the predicted prime-target matches derived from Szmrecsanyi's (2006) corpus-based account of the comparative.

Overall, my results imply that neither variety under study is systematically more sensitive to the hypothesized complexity and priming effects. The main reference varieties are on equal footing with regard to the two persistence variables, syntactic position and segmental *horror aequi*. Varietal differences can be observed for the remaining three constraints: End weight turns out stronger in BrE, infinitival complementation and rhythmic harmony have greater effects in AmE.

The variety model was designed to test a hypothesis that was derived from Mondorf's (2009b: 192) account of the comparative, namely that the functional division along the lines of *more*-support is more advanced in AmE as compared to BrE. At odds with this assumption, the results I have presented for the superlative (cf. Figure 8) unambiguously suggest that neither variety is systematically more inclined than the other to compensate contextual complexity by resorting to the analytic variant, which, by hypothesis, is cognitively advantageous. This general impression obtains despite the observation that sensitivity to individual constraints varies moderately according to variety. These findings are consistent with previous studies which report an overall high degree of cross-varietal homogeneity in grammatical alternations (Szmrecsanyi et al. 2017; Heller, Szmrecsanyi & Grafmiller 2017: 21, *inter alia*), with subtle probabilistic differences for particular linguistic predictors. Nothing else was expected: It is hard to envisage that the scope of cognitive mechanisms underlying the processing of a language should be subject to variation across space. While it appears plausible that the functional specialization of the gradation strategies has developed in diachrony (Mondorf 2009b: 117-169), with the strength of this effect increasing in apparent time, as can be inferred from the discussion of the variable *age* in Section 3.3.3, it is unlikely that similar differences exist between the two main reference varieties in the 21st century. Instead, the American lead in analytic support reported by

Mondorf (2009b: 171) may at least in part be accounted for by stylistic differences between her American and British corpus material; however, correlations of style and linguistic complexity,[47] or rather the compensation/avoidance thereof, do not permit inferences about the link between complexity and variety.

Zooming in on individual adjectives

Next, I turn my attention to the adjectives that form part of my analysis. Figure 9 displays the BrE and AmE means of the 23 adjectives in the questionnaire on the proportion scale from 0 (synthetic) to 1 (analytic). These model-based predicted proportions of *most* for each adjective (out of the three response categories) were determined for each variety, exclusively based on the control items to minimize the influence of contextual factors. All adjectives were used in at least three item pairs (cf. the online appendix at https://osf.io/9hf7v); the analyticity values are therefore averages across the values obtained for all control items in which the respective adjective occurred.

[47] This assumed correlation is tentatively supported by Watanabe & Iyeiri (2020), who report a higher share of analytic comparatives and superlatives in British newspapers with a lower as compared to a higher level of textual formality.

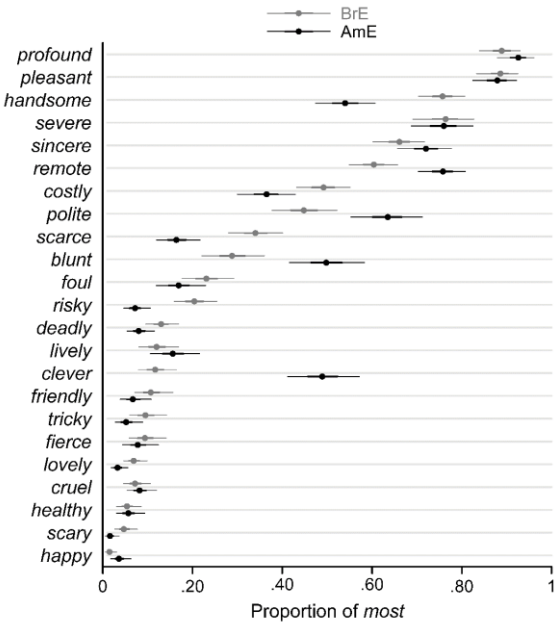

Figure 9. BrE and AmE analyticity values for the 23 adjectives used in the questionnaire (proportion scale). The adjective values are based on the control items only. The error bars display the 50 % and 90 % posterior uncertainty estimates. See the online appendix at https://osf.io/3rkmj for all analyticity values and posterior uncertainty estimates.

The values broadly cover the range from (near-)categorical syntheticity to analyticity, slightly inclined towards the former extreme. In great part, this is owing to the large group of disyllables ending in /-i/ and /-li/, which appear to be predominantly synthetic with analyticity values consistently below .25. *Costly* is the odd one out with values of .37 (AmE) and .49 (BrE). Among the disyllables in /-ə(r)/, /-ɪə(r)/ and /-ɪr/, *severe* and *sincere* are drawn towards *most* with values ≥ .66 in both varieties. Notably, both adjectives are more analytic in AmE than in BrE. *Clever* stands out by the greatest intervarietal difference of all adjectives under study ($\Delta_{\text{AmE-BrE}}$ = .37). This adjective prefers *-est* in BrE (.12) and exhibits considerable variation in AmE (.49). Disyllables with a final consonant cluster are almost categorically analytic, with *pleasant* and *profound* scoring ≥ .88. In the group of disyllables with miscellaneous final segments, *handsome* shows sizeable alternation in AmE (.54) but is inclined to *most* in BrE (.76). Despite their rhythmic similarity, *polite* and *remote* diverge by > .10 in both

varieties, with *polite* exhibiting greater alternation than *remote*. Both adjectives are more analytic in AmE than in BrE. The five monosyllables score rather heterogeneously on the analyticity scale. *Cruel* and *fierce* consistently prefer the synthetic superlative with values ≤ .10. For *foul*, both varieties display similar scores (.23 for BrE and .17 for AmE); the 90 % posterior uncertainty intervals, however, extend further to the right than to the left. The data hint at substantial variation in the superlative strategy preferences of *blunt* and *scarce*. These types receive values between .16 and .50, with *scarce* (BrE = .34, AmE = .16) significantly more synthetic in AmE than in BrE, while the reverse obtains for *blunt* (BrE = .29, AmE = .50). Generally, intervarietal variation is greatest for adjectives ranging in the middle of the analyticity scale, between .20 and .80. Notably, the adjectives in this section come with relatively wide uncertainty intervals, too. Adjectives outside this range pattern similarly in BrE and AmE, have narrow uncertainty intervals and gravitate towards the extremes.

In general, intervarietal variation is particularly pronounced for adjectives within a window of variation around the center of the analyticity scale, whereas (near-)categorical types pattern rather uniformly in both varieties. Towards the extremes, it is assumed, the morphophonological properties of the adjectival base knock out the effects of any constraints that are not type-specific. Thus, the influence of contextual constraints on the alternation is likewise limited to a window in which functional-cognitive negotiations take place, shielded from the gravitational forces exerted by type-specific factors. This verdict is in keeping with multivariate accounts of the degree alternations (Hilpert 2008; Cheung & Zhang 2016) which empirically substantiate the intuitive expectation that the influence of speaker variables and contextual constraints on the alternation is subordinate to the impact of the morphophonological characteristics of the adjectival base.

It is important to reiterate that the questionnaire operationalized 23 adjectives for which previous studies and preliminary corpus analyses found alternating gradation strategy preferences. Further, neither the overall nor the varietal model contained predictors characterizing the adjectival base. Therefore, the following insights into different morphophonological subgroups are tentative at best. The groups that have attracted most attention in the scholarship are disyllables in /-li/ and /-i/; this is particularly because previous studies have suggested diverging

tendencies for these two types of final segment in the comparative (cf. Hilpert 2008: 407; Chua 2018: 466): While final /-li/ prefers *more*, final /-i/ comes with a higher probability of *-er*. Further, according to Mondorf (2009b: 192-193), disyllables in /-i/ show deviant behavior in that they are less sensitive to contextual complexity and are not markedly more analytic in AmE as compared to BrE. The questionnaire data yielded the proportions of *most* for ten adjectives in this category, with five each ending in /-li/ and /-i/. On these admittedly shaky empirical grounds, there is little support for the hypothesized differences between the two formal subcategories in focus: Except for *costly*, the adjectives with these final segments are predominantly synthetic. Varietal differences are subtle across the board; only *risky* and *costly* exhibit sizeable differences between the analyticity values in AmE and BrE, notably with a greater proportion of *most* in the latter variety.

Albeit only statistically robust for *clever*, the finding that all three disyllables in /-ə(r)/ have a higher proportion of *most* in AmE than BrE might be phonologically motivated: The synthetic *comparative* of these adjectives comes with the repetition of the root-final segment in immediate adjacency. This represents an instance of word-internal *horror aequi*. This form is possibly avoided more strongly in AmE as compared to BrE due to an overall higher degree of rhoticity in this variety (Mondorf 2009b: 27). The subsequent preference for the *more*-comparative is paralleled by a bias towards *most* in the superlative. The same effect may also partly explain the analytic preference of *costly*: Word-internal *horror aequi* triggers the avoidance of the synthetic superlative (*cost<u>li</u>est*). As for the disyllables in my data, two further type-specific variables deserve to be commented on in qualitative terms, namely lexical stress and etymology: In line with Hilpert's (2008: 407) results for the comparative, iambic (i.e. end-stressed) adjectives apparently incline towards *most*, the reverse seems to hold for trochaic ones. It remains to be resolved in future studies of prosodic effects on gradation strategy choice whether the measurement of rhythm should be confined to the form in focus or encompass the context to the left and right thereof, as advocated in the present study.[48] Germanic origin

[48] See also Schlüter (2015: 202) who states that "dealing with rhythm beyond the word also improves our understanding of other parts of linguistic structure, e.g. questions such as why a certain morphological or syntactic form is chosen rather than a competing one".

favors the synthetic superlative, whereas Latinate adjectives are predisposed towards the analytic variant, as suggested by LaFave (2015: 2, 6-7). These factors may further contribute to our understanding of the observation that among disyllables in /-ə(r)/, /-ɪə(r)/ and /-ɪr/, *sincere* and *severe* perform categorically while *clever* displays considerable variation.

Apart from identifying the linguistic and extralinguistic factors that govern the superlative alternation, the analysis at hand also allows a comparison of corpus and questionnaire data. The four plots in Figure 10 compare the proportions of *most* for the 23 adjectives in four different datasets, namely in the BrE and AmE questionnaire data, the BNC and COCA.[49] I manually identified elatives and quantifier uses of *most* in the COCA data and discarded these instances.

[49] The analyticity values for the adjectives in the four datasets are provided in the online appendix at https://osf.io/3rkmj. Thanks are due to Lawrence Cheung for sharing the BNC superlative dataset used in Cheung & Zhang (2016). Note that elatives and quantifier uses of *most* were discarded by the authors in the pre-processing stage.

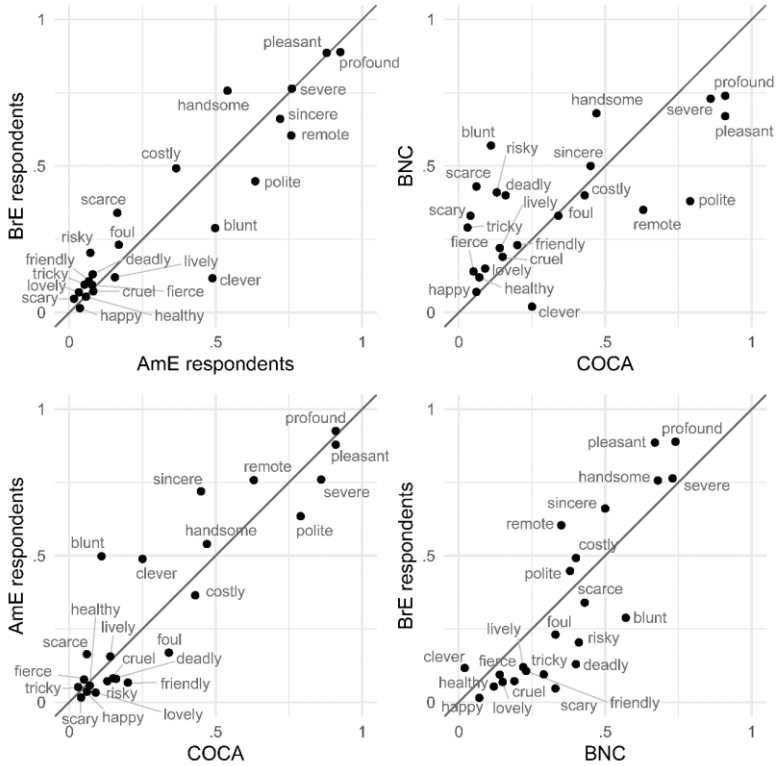

Figure 10. Analyticity values for the 23 adjectives used in the questionnaire (proportion scale). Top left: BrE and AmE questionnaire respondents. Top right: BNC and COCA. Bottom left: AmE questionnaire respondents and COCA. Bottom right: BrE questionnaire respondents and BNC. The online appendix at https://osf.io/3rkmj provides the analyticity values for all four sub-datasets.

Values above the diagonal line in each plot indicate that an adjective is more analytic in the dataset specified on the y-axis, the converse applies to values below the diagonal line. The top-left panel plots the BrE and AmE proportions of analyticity. Recall that these proportions are exclusively based on the control items. As such, this is simply a different way of illustrating the information contained in Figure 9. While the proportions of *most* in the varietal subsections of the questionnaire data converge towards the categorical extremes, there is considerable variation around

the center of the spectrum; on the whole, neither variety appears more analytic than the other.

The top-right panel compares the analyticity values gleaned from the two standard reference corpora. These values were computed from raw counts. The corpus data yield results of remarkable inconsistency. The BNC and COCA values converge towards the synthetic extreme, albeit not as markedly as in the questionnaire data; the values are rather dissimilar, however, towards analytic categoricity. Across the board, the degree of varietal divergence is significantly higher in the corpus data than in the questionnaire data. For eleven types, the intervarietal difference $\Delta_{\text{COCA-BNC}}$ is > .20, with peaks on both sides, e.g. *polite* ($\Delta_{\text{COCA-BNC}} = .41$) and *blunt* ($\Delta_{\text{COCA-BNC}} = -.46$). Strikingly, while the analyticity values for three adjectives (*foul*, *costly*, *sincere*) are similar in the middle of the spectrum, COCA exhibits sizeable variation for types that range between .30 and .60 in the BNC. Most of these adjectives are near-categorically synthetic in COCA but show alternating behavior in the BNC. The divergent varietal preferences observed for selected adjectives in the questionnaire data (top-left panel) are mirrored in the corpus data: *Scarce*, *handsome* and *risky* are more analytic in the BrE data; *remote*, *polite* and *clever* have more *most* in the AmE data. *Blunt* is exceptional in this regard: While the questionnaire data suggest an American lead in analyticity, the proportions gleaned from the corpora indicate a greater proportion of *most* in BrE.

The panels at the bottom left and bottom right draw comparisons between the questionnaire and corpus data pertaining to the same national variety. In the bottom-left panel, the analyticity values are largely similar towards the extremes. This trend is also noticeable in the bottom-right panel, albeit less pronounced. The divergence between the American questionnaire and corpus data is systematic: Adjectives with analyticity values between .25 and .65 in COCA are generally more analytic in the questionnaire. The S-pattern that can be identified in the plot at the bottom right implies that there is a tendency towards both extremes in the questionnaire: A preference for *-est* in the corpus is even more pronounced in the questionnaire, predominantly analytic types in the BNC are still more analytic in the corresponding elicitation data. Remarkably, the identified tendency towards the extremes also seems to be systematic with regard to type-specific properties: Adjectives that turn out more synthetic in the BrE questionnaire data are predominantly disyllables ending

in /-i/ and /-li/, e.g. *tricky* ($\Delta_{\text{BrE- BNC}} = -.19$), *scary* ($\Delta_{\text{BrE- BNC}} = -.28$), *deadly* ($\Delta_{\text{BrE- BNC}} = -.27$) and *risky* ($\Delta_{\text{BrE- BNC}} = -.21$); the group of adjectives for which the BrE and AmE questionnaire data contain more *most* than the respective corpus material includes disyllables with a final consonant (*remote, handsome*) as well as final /-ə(r)/, /-ɪə(r)/ and /-ɪr/ (*sincere, clever*).

3.3.5 Sources of variation

This section sheds light on four sources of variation in the data, namely

(i) *adjective*: This is the variation among the $n = 23$ adjectives that were used to construct the questionnaire items.
(ii) *informant*: This reflects the variation that is due to informants, i.e. the degree to which the n = 675 respondents differ in their overall preference for the analytic superlative.
(iii) *control item*: This is the variation between the n = 56 control items.
(iv) *difference (test - control)*: This is the variation among the differences we observe between the test and control conditions of the n = 56 item pairs.

To explore the variation that can be attributed to these factors, a model with none of the predictors we have discussed previously was fit to the data. This null model only contains the cluster-level predictors *adjective*, *informant* and *item*.

The horizontal axes in Figure 11 show the latent preference for *most*. This is the scale on which our ordinal regression models operate. For the summary of our results so far, we have been looking at proportions, which are back-transformed from this underlying latent scale. To get a sense of the sources of variation in our data, it is more convenient to consider this latent scale. Before we go further, we will briefly explain the connection between this abstract scale and the outcome proportions.

The bell-shaped curve is the heart of the model. We can think of it as representing a distribution of inclinations towards *most*. We have three ordered outcome categories (*-est*, 'no preference' and *most*), and the model establishes, based on the distribution of observations in the data, two thresholds. These are marked as vertical grey lines in Figure 11. With these thresholds, the bell-shaped curve allocates probability mass to our three response options. The probability mass is equal to the area under

the curve that falls in the three intervals. We see that, overall, the most frequently observed response category was *-est*.

In general, the way this regression model works is as follows: If a predictor is associated with a change in the outcome, the magnitude of this association will be reflected in its regression coefficient. For *end weight*, for instance, this coefficient was 0.41. This quantity refers to the latent scale shown in Figure 11. It denotes by how much the bell-shaped curve shifts, and into which direction. The model tells us that if we change *end weight* from 0 to 1, i.e. switch from a neutral context to a superlative occupying the clause-final position, the curve shifts rightwards by 0.41. Looking at Figure 11, we have to imagine this shift. Note that the thresholds remain in place. As a result, the three areas under the curve, i.e. the response probabilities of our three outcomes, change: The proportion of *most* increases, and it was this proportion and this increase that has been at the center of our attention so far.

In what follows, we will look directly at the latent scale. Our interest will be in the amount of variation that stems from the different dimensions of variability in our data. Each point in Figure 11 represents either a single adjective (top), single speaker, a control item, or a difference between test and control item (bottom). Each point shows the "effect" of a certain unit, or cluster. Thus, for the adjective at the far left (*happy*), the bell-shaped curve shifts leftwards to about –1.5. This means that almost all of the probability mass is allocated to *–est*: *happy* is decisively synthetic. While we have to imagine these shifts, our primary focus when studying Figure 11 is on the amount of variation we observe among (i) adjectives, (ii) subjects, (iii) control items, and (iv) differences between test and control items. The amount of variation can be condensed and expressed as a measure of dispersion; the standard deviation of each distribution has been added at the left margin of the display.

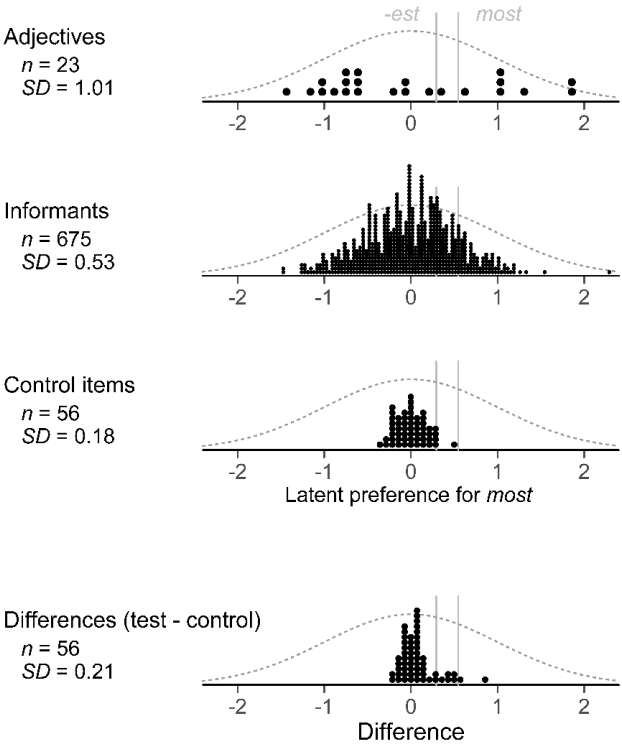

Figure 11. Variation among adjectives, informants, control items and test-control differences on the latent scale

The n = 23 adjectives in the graph at the top cover the whole range of the latent scale, indicating a considerable degree of variation among adjectives (SD = 1.01). This is despite the fact that our choice of adjectives was aimed at a set of actively alternating forms. Still, this great variability according to adjective is consistent with our previous knowledge of the comparative and superlative alternations: There is general consensus in the literature that the morphophonological characteristics of the adjective are the strongest predictors of degree strategy choice. In a global perspective, we observe that adjectives with a preference for *-est* are slightly overrepresented in the set of adjectives chosen for item construction. This is especially due to disyllabic adjectives in /-i/ and /-li/: Ten of the 23 adjectives in the questionnaire fall within these morphophonological categories.

The second graph from the top displays the variability among informants (n = 675), thereby giving a reflection of the variation that could, in principle, be due to social factors. It is considerably smaller than that between adjectives, but still sizeable (SD = 0.53). Thus, for a small share of respondents, *most* is in fact the preferred strategy on the whole.

We observe a relatively low degree of variation among the mean analyticity values for the n = 56 control items (SD = 0.18). This is a positive sign with regard to the construction of the questionnaire: It seems that, once the effects of adjective and informant were partialled out, the wording of my control items did not produce an undue bias towards one of the superlative variants that could be ascribed to contextual constraints. They may therefore be considered fairly close approximations of the baseline, i.e. the hypothetical condition in which contextual factors are assumed to exert the least possible influence on the choice of superlative form.

Finally, we turn to the distribution of differences between test and control items, that is, the differences between the analyticity values of the test and control conditions of each item pair (n = 56). The interpretation of this batch of points differs from the other ones. What we are interested in here is how much variation we were able to generate through our deliberate manipulation of the linguistic constraints. Points that hover around zero represent item pairs which saw very similar responses for the control and test conditions. In such item pairs, the manipulation of the contextual constraint did not produce the expected effect: It led to no (or only minor) observable changes in the distribution of the responses. In a hypothetical setting, where all constraints were operative in the expected direction, all of these points would be dislocated to the right, signaling the expected increase in analyticity. This, however, is not the case. As our findings illustrate, there are constraints that seem to be inert, and some appear to operate in the other direction. The added value of including these differences into Figure 11 is that it allows us to appreciate the magnitude of the variation we were able to generate via the contextual constraints to that which we observed among control items, informants, and adjectives. It is evident that the amount of variation due to our deliberate manipulations of item pairs is rather small. This is coherent with the findings of corpus studies which have consistently ascribed a minor role to non-phonological constraints (Hilpert 2008; Cheung & Zhang 2016).

3.3.6 Contextual constraints: Consistency among item pairs and informants

A further type of variation that our model documents, and which is of diagnostic interest for the present investigation, is the variation among item pairs and among informants with regard to the direction and magnitude of the seven contextual constraints. So far, we have simply averaged over these, but now we will take a closer look. This source of variation is of interest for two reasons: First, the variation among item pairs gives us insight into the quality of our instrument. Thus, we have designed nine item pairs to assess the effect of *end weight*. Ideally, we would like all item pairs to tell a similar story – we would like to see a consistent effect of *end weight*. For informants, on the other hand, we are interested in whether a certain constraint is operative in similar ways across informants. Some may respond more to a certain manipulation, others may be less sensitive. If we find little variation among informants, for instance, we would consider the average trends, or differences, we have reported to represent the behavior of our sample well.

Figure 12 shows the amount of variability that we observe, for each of the seven constraints, among informants (grey histogram) and item pairs (filled circles). For the three constraints whose effect is estimated to be of negligible size – *segmental horror aequi*, *rhythmic harmony* and *syntactic position* – both items and informants show consistent responses. For the persistence settings, the same is true: Items and informants show very similar behavior, suggesting that, while small in magnitude, the average difference we observe between the test and control items is stable across informants and questionnaire item pairs. For *infinitival complement* and end weight, we observe more variability, especially among informants. This tells us that the overall average difference masks a considerable extent of between-speaker variability with regard to the magnitude and even direction of these constraints. Still, all item pairs point into the same direction, albeit with varying strength. For infinitival complement, we observe a curious bimodal distribution of informants.

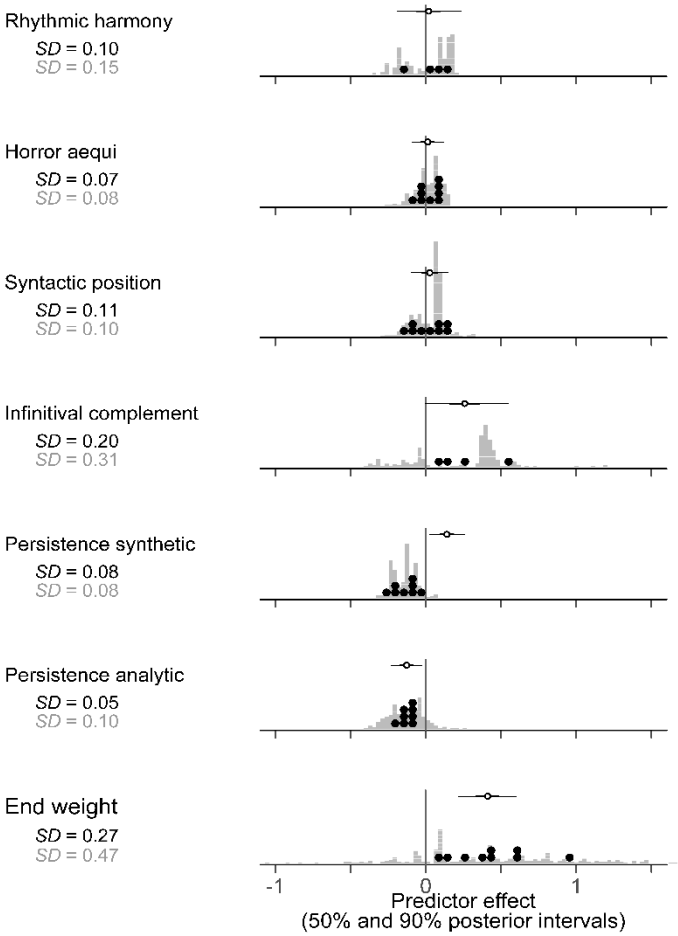

Figure 12. Variation among informants and item pairs for the seven contextual constraints. The white dot at the top of each panel indicates the mean effect of all item pairs that operationalize the respective constraint.

Overall, then, Figure 12 can be taken to suggest that although the degree of sensitivity to the set of contextual constraints does vary between item pairs, this variation merely concerns the magnitude of the difference, not its sign. On the whole, therefore, the questionnaire items offer consistent estimates of the effect of the constraint in focus, and this consistency suggests that, for the persistence settings, *infinitival complement*

and *end weight*, the differences we observe between test and control conditions may be attributed to the operationalized constraint.

However, we identify a single outlier among the item pairs whose test-control difference is considerably larger than the mean. This outlier is the largest observed item pair difference for the end weight constraint. Its unusually large value (close to 1) is particularly apparent from the bottom panel in Figure 12. An inspection of this item pair (EW_5) is necessary to determine whether this seemingly large effect is in fact due to the contextual constraint or can be taken to hint at the possibility that the end weight effect was not successfully isolated.

Let us take a closer look at the wording of the test item EW_5_T and the corresponding control item EW_5_C:

(58) Of all fates one can suffer being blind is certainly the **[CRUEL]**. [EW_5_T]

(59) A brain tumour made him suffer the **[CRUEL]** fate of all. [EW_5_C]

The qualitative feedback I received from the participants of my pilot study did not raise any objections as to the phraseological and grammatical acceptability of either item in focus. On the level of spelling, it is worthy of note that the BrE and AmE conventions differ in that the former variety favors reduplication of root-final <l> in the synthetic superlative. In both questionnaire modes, the BrE convention was adopted throughout. It appears altogether implausible, however, that this detail should substantially influence American respondents in their choice of superlative form.

Further, rhythmic considerations merit attention here: It was noted that while the OED lists both /kruːl/ and /ˈkruːəl/ as phonological variants of *cruel*, this adjective is treated as monosyllabic in the present study (cf. Section 3.3.4). In this case, the analytic superlative would induce a double stress clash in the control condition EW_5_C (*most cruel fate* [ˈσˈσˈσ]); the synthetic superlative, however, produces an alternating sequence of stressed and unstressed syllables (*cruellest fate* [ˈσσˈσ]). This prosodic difference can be assumed to translate into a preference for the synthetic variant.

Another aspect weighs even more heavily here: It is integral to the conceptual setup of this study that the test and control items of a given item

pair differ only along a single contextual dimension. The items EW_5_T and EW_5_C, however, differ not only with regard to the constraint whose effect they were designed to capture (i.e. *end weight*), but also with regard to a second contextual factor for which previous studies have suggested differences in processing complexity, namely *syntactic position* (cf. Section 3.2.3): While the control item renders the superlative in attribution to a noun, the end weight factor in the test item requires the superlative to be nominal and dislocated from its nominal referent *fates*. It is possible, therefore, that the effects of two constraints, namely *syntactic position* and *end weight*, add up to an exceptionally great difference between the predicted proportions of *most* for the test and control conditions of the item pair EW_5.[50]

Though my exemplary consideration of the item pair EW_5 has revealed that manipulating an item along a single contextual dimension is an intricate undertaking, our investigation of the sources of variation in the data warrant the conclusion that on the whole, the questionnaire items appear to add up to an adequate instrument for the purposes of the present investigation.

3.3.7 Questionnaire data in the study of grammatical alternations: A critical evaluation

Questionnaire data are a fairly new arrival in the research literature on grammatical alternations. The present study is, to my knowledge, the first to have supplied data of this kind for gradation strategy choice. Although innovative in their own right, the experimental designs employed in psycholinguistic accounts of the comparative alternation (Boyd 2007; Kunter 2017) differ widely from the elicitation method introduced in Section 3.2. Since the great majority of studies in the domain of morphosyntactic variation have typically relied on corpus data, it appears all the more necessary to reflect on the assets and drawbacks of the method and data type at hand. In this section, I aim at such a critical assessment, notably also drawing comparisons, where appropriate, to corpus data.

It has been argued in Section 3.2.5 that linguists should be aware of the hierarchical structure of their data. As a consequence of the elicitation

[50] Recall, however, that the current data do not lend support to the hypothesized differences in contextual complexity between attributive and nominal position. (cf. Section 3.3.1).

instrument used in this study, we can expect similarities between data points that are nested within the same adjective, item and informant. In Section 3.3.3, we examined the degree of variability in the current data according to these parameters. We found a high degree of variation among the n = 23 adjectives in the questionnaire. This was ascribed to the observation that the morphophonological properties of the adjective are the most powerful predictors of gradation strategy choice. The intermediate degree of variation among informants was taken to imply that speaker variables in fact play a role in the superlative alternation, although their effects are clearly subordinate to those pertaining to the adjective. In our analysis of the variation among item pairs, we identified a single outlier with an exceptionally large difference between the control and test values (EW_5). A close inspection of the item pair in focus revealed that the test and control item differed along two contextual constraints that can be linked to differences in processing complexity and, perhaps, rhythm. It emerged that while my questionnaire may be assumed to constitute a sufficiently valid instrument, the conceptual premise of isolating a baseline condition of lowest complexity comes with challenges. This must be acknowledged as a clear drawback of the method, at least in the current research setting.

Further, it has at length been explained that my questionnaire dataset is a composite of data gleaned through two different collection modes (cf. Section 3.2.2). Thus, we may legitimately inquire: How do the results of the online respondents compare to those obtained through the pen-and-paper format? An interaction term between the questionnaire mode and each predictor was included in the overall model.[51] The analysis reveals no statistically significant differences between the questionnaire modes for all constraints except for *end weight*: Informants who used the online version were more sensitive to the end weight constraint in the sense that they favored *most* more strongly in clause-final position. Despite this, the model diagnostics suggest that only a minor share of the variation observed in the data can be ascribed to differences in data collection mode. This cross-modal convergence ties in with methodological accounts in psychological (Gosling et al. 2004) and second language research (Wilson & Dewaele 2010).

[51] See https://osf.io/hjeq2/.

A particularly urgent methodological caveat was that the lack of researcher control and the anonymity of the collection logistics could compromise the validity and integrity of the online results. Evidently, this concern is unjustified in the present context (see Gosling et al. 2004: 98-99 for a similar verdict). Sprouse (2011) conducted two identical acceptability judgement tasks in two different methodological settings, namely via *Amazon Mechanical Turk* (AMT), a commercial platform for human intelligence tasks, and in the laboratory. He concludes that "AMT data are almost indistinguishable from laboratory data" (Sprouse 2011: 155) and thereby provides further support for my findings on the cross-modal comparability of elicitation data.

Let us finally reflect on how questionnaire data relate to corpus data in terms of data quality. Figure 10 compared the 23 adjectives in the questionnaire study with regard to their analyticity values in the BrE and AmE questionnaire data and the respective standard reference corpora BNC and COCA. In general, the current elicitation data are more extreme than corresponding production data gleaned from the corpora. Adjectives for which the BNC yields a synthetic or analytic preference are even more markedly inclined towards the respective extreme in the BrE questionnaire data. The picture is slightly different for AmE: While the analyticity values from the two sources largely converge in proximity to the categorical extremes, adjectives for which the COCA implies an alternating character are noticeably drawn towards the analytic pole in the questionnaire data. We can assume that the situation in the BrE data would be similar to what we have observed in AmE if the BNC and the COCA were of equal size: I have already noted that the BNC values may have been thrown askew, at least in part, by the small number of superlative tokens for some of the adjectives under study. This artefact of the corpus design may also to some extent explain the divergent varietal trends for several adjectives in the proportional values derived from the corpora, even though it cannot be ruled out that a considerable share of the variation for individual types genuinely reflects different superlative strategy preferences in BrE and AmE. The overall verdict that elicitation data are more extreme than comparable corpus data is indeed not surprising: Although the informants were instructed to base their judgements on intuition, effects of social desirability and the logistic circumstance that respondents had the time to

ponder their choices may have unduly reinforced the application of prescriptive rules. Along these lines, then, the marked analytic preference in the AmE questionnaire data for types that are drawn towards neither pole in the COCA material can possibly be accounted for by the assumption that *most* is the default for types that are not categorical by virtue of their morphophonological properties.

In the light of these findings, it is hard to argue that either type of data is objectively 'better' or more reliable than the other. Both come with their own benefits and drawbacks, some of which have been addressed in a contrastive fashion in Section 3.1. One of the most urgent caveats about questionnaire data was raised in the previous paragraph, namely the assumed tendency of informants to adhere to prescriptive rules. A similar effect, however, might also compromise the reliability of corpus data that have been gleaned from written material: The grammar checkers that are customarily implemented in word processing software nowadays use language-specific algorithms for morphology. In view of the evidence I have presented for individual adjectives, it is dubious on what grounds *Microsoft Word* recommends that *most clever* and *most costly*, for instance, be replaced by their synthetic competitors, notably regardless of the variety of English selected in the language options.[52] For none of the 23 adjectives operationalized in the present questionnaire, *Microsoft Word* accepts both superlative variants; this contravenes the empirically informed insight that "the realization of [this] inflectional category is probabilistic, instead of being grammatically or lexically strictly determined" (Bauer, Lieber & Plag 2013: 119). Therefore, we may legitimately be concerned that reliance on the suggestions of grammar checking software could reduce the variation in gradation strategy choice. Surprisingly, this major drawback of corpus data has not been acknowledged in previous studies of the comparative and superlative degrees.

Ultimately, both corpus and elicitation data can provide valuable insights into the probabilistic mechanisms underlying the alternation in focus. It is imperative that researchers in the domain of morphosyntactic variation be sensitive to the respective strengths and weaknesses of either

[52] Under the heading *Comparative use*, the Microsoft Word grammar checker provides an explanation for these suggested corrections; this explanation, however, is completely unfounded in linguistic terms, even in the eyes of the most adamant prescriptivist: "Use 'more' and 'most' with adjectives with no comparative form, but not with adjectives that have one."

type of data as well as the methodological and interpretative challenges they pose.

4 A usage-based connectionist view of superlative strategy choice

The study that will be presented in this chapter relies on corpus data to investigate the interplay of superlative frequency, i.e. the number of superlative tokens (synthetic plus analytic) per amount of text, and the formal, i.e. morphophonological characteristics of the adjective. In interpreting the trends that emerge from the data, I will adopt a usage-based view of lexical representation. The current study thus differs from the questionnaire study outlined in Chapter 3 with regard to three main aspects, namely in terms of (i) the predictors in focus, (ii) the method that will be used (i.e corpus instead of questionnaire), and (iii) the theoretical framework within which the findings will be interpreted (i.e. usage-based instead of cognitive-functional and sociolinguistic theory).

The chapter is structured as follows: After outlining the fundamentals of usage-based Construction Grammar (Section 4.1), Section 4.2 gives a survey of the literature on morphological processing and lexical categorization and sketches the principles of connectionist category learning. In Section 4.3, a connectionist exemplar model of superlative strategy choice will be introduced. After an outline of the research questions this study intends to address (Section 4.4), Section 4.5 provides a description of methods and data. Results are presented in Section 4.6 and Section 4.7 concludes with a discussion.

4.1 Linguistic structure in usage-based Construction Grammar

The usage-based approach abandons the conceptual bedrocks of Generative Grammar, namely the assumptions that language is to some extent innate and that there is a clear distinction between competence and performance. Instead, linguistic structure is seen as emergent from events of usage upon which the mental records of language are persistently updated throughout an individual's lifetime (Langacker 1987; Bybee 2006). Crucially, there is no explicit a priori knowledge of linguistic rules: It is by analyzing the distributional properties of linguistic input that language users derive structural patterns. These patterns are conceived of as experience-based regularities rather than rules, as probabilistic rather than

strictly determined. Usage-based theory rests upon the premise that linguistic representations are lexically specific (Diessel 2017) and is therefore at odds with the structuralist view that grammatical categories are independent of their lexical instantiations, as advocated in the dictionary-and-grammar model (cf. Taylor 2012: 8). The linguistic units abstracted from the input are referred to as *constructions*. According to Bybee (2010: 9), a construction is "a direct form-meaning pairing that has sequential structure and may include positions that are fixed as well as positions that are open". Therefore, all linguistic information about a construction is contained in the construction itself, with no intermediate level of representation; this includes information on all layers of linguistic analysis, notably also collocational preferences and pragmatic implications (Hilpert 2014: 1-2). Constructions come in a variety of shapes and sizes, ranging from individual morphemes and words to multi-word sequences such as idioms and syntactic patterns with highly variable slots (e.g. the ditransitive construction). These slots are the "open positions" in Bybee's definition. The degree of schematicity, i.e. the productivity of the open slots, varies considerably among constructions (Blumenthal-Dramé 2012: 7-8; Diessel & Hilpert 2016: 8): While idioms like *all of a sudden* mark the lower end of the schematic cline, with the instantiation of all lexical slots fixed, the two constructions under scrutiny in the current study, the synthetic X-*est* and the analytic *most* X construction, come with great variability in the lexical realization of the adjectival slots.

How do language users acquire constructions? Construction grammarians commonly provide a rather intuitive answer to this question: by means of domain-general cognitive abilities that operate below the level of consciousness and are integral to other human activities such as sensory perception and physical motion (Tomasello 2008: 27-32; Bybee & Beckner 2009: 829-832; Cordes 2017; Diessel 2019: 23-39; Divjak 2019: 16). These basic capacities include categorization, i.e. abstracting regularities from previously encountered instances of a construction, statistical learning and *chunking*, which denotes the process that previously separate linguistic units, if repeated frequently, are eventually processed and stored as a single unit with idiosyncratic meaning (Ellis 1996, 2002: 172; Blumenthal-Dramé 2012: 4-26; Pfänder & Behrens 2016: 4-5); these sequences become automatized or entrenched (Diessel 2016: 226). There is ample empirical evidence to suggest that the abstraction of schemas, a

process referred to as "schematization" (Langacker 2008: 17), begins in early infancy and rests upon two types of cues, namely phonetic cues such as stress and pauses, and distributional cues that predict the order of units within speech strings (Jusczyk 1997: chiefly Chapters 4 to 6). The experimental study by Saffran, Aslin & Newport (1996) implies that eight-month-old children successfully segment speech input on the basis of these cues (cf. also Aslin & Newport 2012). Similar findings have been obtained for children at seven (Marcus et al. 1999) and twelve months of age (Gómez & Gerken 1999).

Frequency and similarity

In usage-based theory, the emergence of grammatical categories depends on two key characteristics of the input, namely frequency and similarity. Two frequency metrics are commonly distinguished: token and type frequency. Token frequency is the frequency with which a particular construction occurs in total. High token frequency leads to entrenchment of a form-meaning pairing, e.g. the sequence X-*est* as the synthetic superlative construction. Type frequency, on the other hand, denotes the number of distinct linguistic units – adjectives, in the case of the superlative constructions – that can occur in the schematic slot. As such, type frequency measures the productivity of a schema and indicates the degree to which it may extend to novel items (Diessel 2019: 16-17). The higher the type frequency of a construction, the less likely it is to become associated with a single highly frequent item (Bybee & Thompson 1997: 384). As for the two competing superlative constructions, however, it can be inferred from previous studies that the productivity of the respective schemas is further restricted by formal, morphophonological characteristics of the adjectives. A simplistic rule of thumb proposes that the X-*est* construction prefers mono- and disyllabic adjectives with selected final segments, while *most* X appears to favor types with three or more syllables. The observation that trisyllabic types like *unhappy* are more likely to form the synthetic than the analytic comparative (Mondorf 2009b: 37) highlights the fact that formal similarity (to disyllabic *happy*, in this case) is fluid and that membership of schematic categories is ideally conceived of as probabilistic. These insights are particularly relevant for the network models that will be addressed from Section 4.2.2 onwards.

4.2 Morphological processing and lexical categorization

The insights that have been offered in the previous subsection have implications for a lively debate in the literature on morphological processing that has been going on since the 1980s. This debate is commonly referred to as the past-tense debate, since the primary but by no means exclusive focus of empirical and theoretical scrutiny has been the English past tense (see Pinker & Ullman 2002 and McClelland & Patterson 2002 for critical reviews). Under this heading, cognitive linguists and psychologists have been concerned with a host of interrelated questions that can be boiled down to a single one: How are inflected forms processed, stored and represented in memory? The contributions to this discussion have traditionally relied on two different conceptual architectures: Rule-based accounts, which invoke two different processing mechanisms for irregular and regular inflections, and associative network accounts, which propose that regularly and irregularly inflected forms are handled by a single mechanism. The present subsection reviews both strands of research, drawing on studies of the processing and representation of both inflectional and derivational morphology. While rule-based models foreground the working mechanisms of morphological processing, associative network accounts make additional assumptions about the organization of the mental lexicon. Both approaches offer valuable perspectives for the current study of the superlative alternation. In this respect, it will be particularly rewarding to investigate how previous studies have dealt with frequency variables.

The review I will conduct in the following is selective in that it merely intends to capture the key assumptions of the models under consideration that can be harnessed in the present research context. Sonnenstuhl-Henning (2003: 29-43) offers a more comprehensive survey of the models that have been proposed in the literature on morphological processing. Similarly to the account at hand, she makes a binary distinction between rule-based, or rather "morphology-based" (Sonnenstuhl-Henning 2003: 42-43) models, and associative network models; however, her categories are more fine-grained in that they also pay regard to the neurophysiological underpinnings invoked in the respective studies, an analysis of which does not fall within the remit of this study.

4.2.1 Rule-based dual-route architectures

Rule-based models of morphological processing have originated in the words-and-rules approach (Pinker & Prince 1988; Pinker 1999), which proposes, contrary to the domain-general capacities at the heart of usage-based theory, that the human brain has two distinct language-specific modules: a dictionary containing lexical entries and a system of combinatorial grammatical rules that processes (novel) linguistic input using symbolic computation. Models in the rule-based paradigm customarily have a dual-route architecture, i.e. they assume that while irregular inflectional forms have separate lexical entries that are retrieved each time the form is encountered, processing of regular inflectional forms involves morphological decomposition by fully predictable rules (cf. Sereno & Jongman 1997: 434). The main point of contention in this family of models is the extent to which regular morphology requires rule-driven parsing and whole-word storage. Dual-route architectures are therefore typically situated along a scale between full listing and full parsing, with various intermediate positions. Despite the conceptual differences that will be exposed in the following, these models share a crucial commonality: The speed of lexical access is frequency-sensitive. There is disagreement, however, as to the operationalization of this factor.

Dual-route models on the scale between full listing and full parsing

The accounts by Butterworth (1983) and Taft & Forster (1975) mark the respective extreme positions on the scale between full listing and full parsing. Butterworth (1983)[53] has been credited as the pioneer of the full-listing approach, which starts from the premise that both morphologically simple, i.e. monomorphemic, and complex, i.e. polymorphemic, words are stored as single units in the mental lexicon. The model only vaguely acknowledges the availability of rules for morphological decomposition as

[53] Sonnenstuhl-Henning (2003: 41) lists Butterworth's model among associative network models. I follow Abel (2003: 181), however, in classifying this model as rule-based, since my focus is on demonstrating the continuum between full parsing and full listing rather than the assumed principles of lexical organization in the mental lexicon, which are only vaguely sketched in Butterworth (1983). Similarly, Bybee (1995: 432) criticizes the fact that the literature on dual-route models appears to invoke the existence of associative networks, especially for irregular morphology, without specifying the operative and structural principles thereof.

a fallback if direct lexical access fails; the nature and operational mechanisms of these rules, however, remain unclear (cf. Abel 2003: 181). Postulating such a fallback, however, is crucial to accounting for the processing of novel forms for which lexical records are unavailable (Frauenfelder & Schreuder 1992: 169), as we will see shortly. Butterworth's design has legitimately been criticized for methodological reasons, too: The decision to define the speed of word recognition solely as a function of surface frequency, i.e. the frequency of a word regardless of the frequency of its constituents, falls short of addressing potential effects of morphological root frequency (cf. Baayen, Dijkstra & Schreuder 1997: 95-96).

In diametrical opposition to this design, the full-parsing model (Taft & Forster 1975) proposes that morphologically complex regular words are always decomposed before the root morpheme is lexically accessed.[54] Hence, processing of *unlucky* involves two serial processes: First, the nominal root is identified and the affixes are "stripped off"; second, the lexical entry for *luck* is accessed (Taft & Forster 1975: 638). There is no separate lexical entry for *unlucky*. Methodologically, this approach abandons surface frequency and instead places a premium on cumulative stem frequency, i.e. the combined frequency of all words in which the respective unit can be identified as the stem. As for the current example, this involves summing the frequencies of *luck* and all lexical items that can be derived from this root, including *lucky*, *unlucky* and *luckless* as well as the degree inflections and adverb forms (*-ly*) of these adjectives. It is evident that even this model cannot entirely do without storage: At least the stem must have a lexical entry that can be accessed after morphological decomposition. The conceptual opposition between the two models that have been reviewed in this paragraph is thus more accurately described as one of pervasive and minimal storage rather than full listing and full parsing (cf. Baayen, Dijkstra & Schreuder 1997: 95).

An intermediate position between these extremes is taken by two influential dual-route models that are concerned with the processing of derived and regularly inflected forms: The Augmented Addressed Morphology Model (AAMM) (Caramazza, Laudanna & Romani 1988; Laudanna, Badecker & Caramazza 1992, *inter alia*) and the Morphological Race

[54] Only instances of derivational affixation are subject to empirical scrutiny in Butterworth (1983) and Taft & Forster (1975) but the insights gleaned from the review of these studies carry over to inflectional morphology.

Model (MRM) by Frauenfelder & Schreuder (1992).[55] The AAMM assumes that both routes, the direct lexical access route and the decomposition route, which invokes access representations of stems and affixes, are activated for previously encountered items and that the former is fast-tracked by default, with decomposition merely serving as a fallback – a possibility that was rather cautiously implied by Butterworth (1983). Only in the case of novel forms, for which direct lexical access inevitably fails, will the parsing route be employed: There is no race between the two routes, since direct access is always given priority. Relying on experience with processed forms as the key determinant of processing route, the AAMM makes a central contribution to the dual-route debate; the authors remain non-committal, however, as to how much experience with a form is required before it receives a permanent lexical entry (Caramazza, Laudanna & Romani 1988: 300). The possibility of a frequency threshold for whole-word access representations will be explored in due course.

By contrast to the AAMM, the MRM assumes that the two access routes are activated simultaneously and compete in a "race" for completion. The speed of the whole-word route is positively correlated with the surface frequency of the form to be processed: Irrespective of their morphological complexity, high-frequency forms are more likely to be stored as a lexical unit (cf. Stemberger & MacWhinney 1986: 25). The speed of the parsing route hinges on two formal factors of the unit to be parsed: phonological transparency, i.e. the phonological similarity of the affixed form and its stem, and semantic coherence, i.e. the degree to which the meaning of an affixed form can be predicted by combining the meanings of stem and affix. These factors are related to the productivity of the affix, as Abel (2003: 184, fn. 26) remarks: Words containing productive affixes are more likely to be phonologically transparent and semantically coherent than words containing unproductive affixes. Therefore, it is the frequency with which an affix has been involved in the successful completion of the parsing route rather than the surface frequency of the affixed word to be parsed that is positively correlated with the activation level of the affix, and hence the speed of the parsing route (Frauenfelder & Schreuder 1992: 178). Operationally, then, the outcome of the morphological race is determined by the interplay of surface frequency and the transparency of the parsing process.

[55] The present review of these models is largely based on Abel (2003: 182-185).

What counts? Surface frequency, root frequency and relative frequency

It is apparent that the dual-route architectures reviewed above raise questions about the role of frequency. Previous studies have commonly linked whole-word representation to high surface frequency, while high root frequency, i.e. the lemma frequency of the root of a complex form, has been taken as supportive of decomposition. A central question emanates from this vague guideline: Is there a frequency threshold for whole-word representation? In other words, assuming that morphological processing is efficiency-driven, can we empirically determine the point at which storage and parsing break even in terms of processing costs? Alegre & Gordon (1999) addressed this question in a series of lexical decision experiments involving regularly inflected forms.[56] While they were able to identify whole-word frequency effects for morphologically simple, i.e. monomorphemic, forms across the board, regularly inflected forms only exhibited such effects above a threshold of 6 per million. Forms with surface frequencies below this threshold may thus be handled by the parsing route. The findings presented by Baayen, Wurm & Aycock (2007), however, imply that whole-word storage already occurs at a much lower surface frequency, significantly below the frequency threshold suggested by Alegre & Gordon (1999). They further refer to studies of Dutch inflectional morphology to support this idea (cf. Baayen, Wurm & Aycock 2007: 448). Regardless of the frequency measure in focus, we should be skeptical about a general threshold for such effects. This caveat appears particularly urgent in the light of the remark by Sereno & Jongman (1997: 425) that differences in the inflectional structures of English nouns and verbs should

[56] There seems to be general consensus in the lexical access literature that irregular inflections never undergo decomposition, thereby ruling out dual-route lexical processing, but are most likely stored in an associative network of the kind that will be discussed in Section 4.2.2 (Bybee & Slobin 1982; Bybee & Moder 1983; Prasada & Pinker 1993; Bybee 1995, *inter alia*). Therefore, regular inflections have been the primary subject of empirical scrutiny in studies of dual-route morphological processing.

not be ignored: For nouns, the relative frequency of roots and their respective inflected forms is higher than for verbs.[57] It is reasonable to assume, therefore, that the insights that have been provided for nominal and verbal inflectional morphology do not straightforwardly carry over to adjectives.

The main insight gleaned from my review of dual-route architectures is that the models in this family rest upon two assumptions about the role of frequency in lexical processing: First, the higher the surface frequency of a form, the more likely it is to be stored as a whole since decomposition and reassembly of a highly frequent inflected form would incur higher processing costs than rote storage. Second, cumulative stem frequency, i.e. in inflectional morphology, the frequency of the stem and its inflectional variants, is positively correlated with the odds of morphological decomposition. Acknowledging these fundamentals, Hay (2001, 2003: 71-95) argues that the relation of these frequency measures is essential to determining which forms will be handled by which processing strategy: Her *relative frequency hypothesis* states that "even low frequency words can display low levels of decomposability, if they are more frequent than their base words [...] [a]nd even high frequency words can display high levels of decomposability, if their base words are more frequent still" (Hay 2003: 73). Crucially, the higher the ratio of a word's root-to-surface frequency, the more likely it will undergo decomposition. Hay's (2003: 102-103) analyses, which are exclusively based on empirical insights into *derivational* morphology, further imply that due to a lower correlation between the stem and surface frequencies of suffixed as compared to prefixed forms, effects of whole-word storage for suffixed forms occur only at higher relative frequencies as compared to prefixed forms. Suffixed forms can thus be considered "more resistant" to whole-word representation than prefixed forms.

Even though most of the studies that have been referenced in the previous paragraphs are empirically concerned with derivational affixation,

[57] Alegre & Gordon (1999) appear to be indifferent to this aspect. Further, the authors altogether ignore the existence of adjectival inflections, listing the comparative and superlative suffixes *-er* and *-est* among derivational suffixes that trigger nominalization (e.g. *-ity* and *-ness*, cf. Alegre & Gordon 1999: 51). Although there are valid arguments to suggest that the boundary between inflection and derivation is not clear-cut in the case of the English degree-marking suffixes (Fábregas 2014: 286-290), it is undisputed in the literature on the alternation in focus that suffixation for degree is an inflectional process.

frequency can be assumed to play an equally integral role in the processing of inflected forms. It has emerged that defining the role of frequency in dual-route morphological processing is an intricate undertaking. The models that will be discussed in the following section consider frequency as interacting with a second explanatory variable, namely (semantic and phonological) similarity.

4.2.2 Bybee's dynamic network model of lexical representation

According to Bybee's (1985, 1988, 1995) schema-based network model, morphological structure of any kind – derived, regularly and irregularly inflected forms alike – emerges from individual traces of linguistic experience which have separate lexical representations of varying strengths in a lexicon whose structure is constantly adapting upon exposure to linguistic input. Bybee (1988: 131) emphasizes that this lexicon is not to be interpreted as "the mental counterpart of a dictionary" containing entries carved in stone; rather, the representations in this lexicon are written in sand: their strength increases with use and fades with disuse. Crucially, these representations are associated in a network whose architecture is shaped by two interacting organizational principles, *lexical connections* and *lexical strength*, both of which will be addressed in turn.

In the usage-based perspective, lexical representations are persistently derived and adapted upon individual instances of linguistic experience. In the mental lexicon, these representations are thought to be connected by formal similarity. Lexical items can be similar on semantic and/or phonological grounds. A network architecture arises as similar representations map onto each other to varying degrees, with the strength of the relations commonly visualized by lines of different shapes. Figure 13, adapted from Bybee (1985: 130, 1988: 134-135), illustrates how adjectival forms may be organized through lexical connections in an associative network. This figure exemplifies connections between four disyllabic adjectives in /-i/ and their graded forms. Fine solid lines mark phonological identity and broken lines indicate phonological similarity (e.g. [ɪ] and [ʌ] are both vowel sounds). Thicker solid lines represent relations that are both phonological

and morphosemantic[58] in nature (e.g. the phoneme /ɪ/ as a morphological marker of denominal disyllabic adjectives). Only where these two types of connections run in parallel can morphological paradigms be established, as will shortly be demonstrated.

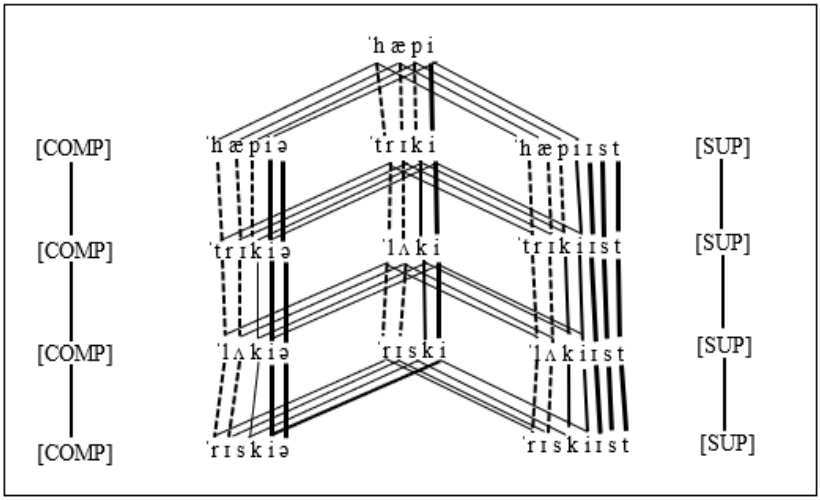

Figure 13. Network of phonological and morphosemantic connections adapted from Bybee (1985: 130, 1988: 134-135)

Lexical strength, the second organizational principle of Bybee's associative network, captures the effect of lexical frequency on morphology. It is widely acknowledged that morphological irregularity correlates with high token frequency. The Conserving Effect of high token frequency (cf. Bybee & Thompson 1997) predicts that persistent experience with forms contributes to the autonomy of their representations, i.e. they have only few and weak connections to other representations in the network. Highly frequent forms commonly instantiate paradigms of low productivity. The suppletive paradigms of the comparative and superlative serve to demonstrate this relation: The high token frequency of *worse* and *worst* can be

[58] It is worthy of note that Bybee (1985: 132) calls connections of this kind *semantic* connections, arguing that "forms with the same semantic features indicating morphological categories are associated with each other" and "connections on the level of expression [...] have the same structure as natural semantic categories". Since the term *semantic* lacks terminological precision in this context, I refer to this kind of connection as *morphosemantic*.

observed to have a twofold conserving effect: First, these forms are prevented from being regularized as **badder* and **baddest*; second, the gradation strategy of less frequently graded adjectives is unlikely to be modelled on these forms due to the lack of phonological similarity. A word's lexical strength and the number and degree of its lexical connections are inversely related, as Bybee (1995: 238-239) points out: Low-frequency words rely on strong lexical connections to other members of a paradigm but have a low degree of lexical strength, while high-frequency words, which, by hypothesis, come with great lexical strength, have only weak connections and thus a high degree of autonomy.

How is morphological structure established according to the model? A morphological relation, the strongest possible kind of link in the network, can only arise where phonological and semantic relations run in parallel (cf. Bybee 2010: 118). The irregular past tense paradigm involving the stem-internal vowel change /ɪ/ → /ʌ/ has customarily been adduced as a case in point for the development of morphological paradigms (cf. Bybee & Moder 1983; Bybee 1988: 134-135): The group of verbs that form their past tense using this vowel change includes, among others, the types *string, sting, cling* and *stick*. These types are connected within a cluster due to their shared morphophonological features. The morphological rule for past tense formation is established upon the connections between the individual members of this cluster. On the basis of these members, or, to use a different term, *exemplar representations*, a *schema* is abstracted. A schema can be conceived of as a structural blueprint for the paradigm in focus, containing slots of varying degrees of fixedness. In the current example of the irregular past tense, the phonological makeup of the schema can be described as

$$\text{C (C) (C) } \text{ʌ} \left[\left[\begin{array}{c} \text{velar} \\ \\ \text{nasal} \end{array} \right] \right] \quad \text{(cf. Bybee 1988: 135)}$$

According to Bybee (1995: 240), schema formation requires (i) a minimum of six or eight exemplars and (ii) strong connections among exemplars. Importantly, a schema can be extended to novel forms if they are found to be sufficiently similar to the exemplar representations in the cluster from which the respective schema was abstracted. The productivity

of a schema varies considerably. Since the slots of the irregular past tense schema involving stem-internal vowel change are relatively fixed, its productivity is significantly lower as compared to schemas that allow for greater phonological variation in their slots, such as the regular past tense schema involving suffixation by *-ed*.

How do these insights carry over to gradation strategy choice? By analogy to the classical example cited in the previous paragraph, the synthetic comparative and superlative schemas can be derived from the connections in Figure 13, the latter of which takes the form [ˈ(C) (C) V C (C) i ɪ s t].[59] Evidently, similar schemas for the synthetic superlative emerge from other clusters of formally similar exemplars, e.g. disyllables in /-li/. One may object, however, that while the exemplar-based emergence of schemas as suggested by Bybee (1985, 1988, 1995) appears plausible for inflectional paradigms, her model cannot account for analytic gradation. Although the original version of Bybee's model does indeed not address the emergence of schemas that have a syntagmatic dimension, the assumption that language users draw similar analogical generalizations for the *most* X construction on the basis of stored exemplars is in accord with usage-based construction grammar. These exemplars can likewise be thought of as mutually associated in a cluster whose architecture is shaped by the interplay of lexical strength and lexical connections, as will be demonstrated in Section 4.3.

Building on this idea, Booij (2013) argues that both complex words and multiword expressions instantiate morphological constructions within a hierarchically organized constructicon. This means that frequently used analytic superlatives such as *most beautiful* can be stored as holistic exemplars at lower levels of abstraction, much like single complex words. Through repeated use, these *most* X combinations become entrenched as conventionalized form-meaning pairings, reflecting Booij's (2013) claim that morphology and syntax interact within the same network of constructions. Thus, even analytic gradation can be accounted for by exemplar-based mechanisms, as frequent exposure promotes storage and lexicalization alongside the schematic synthetic pattern X-*est*.

[59] Exemplars such as *iffy, itchy* and *achy* attest to the fact that the initial consonant of this schema is optional, even though this cannot be inferred from Figure 13.

4.2.3 Connectionist exemplar models

The models that have previously been reviewed focus on the associative organization of linguistic representations but remain relatively unspecific as to the cognitive operations that give rise to linguistic structure based on the associations of these representations. This section is concerned with connectionist models, a family of models that seek to account for exactly these operations. It is important, therefore, to point out right at the outset that "[c]onnectionist models [...] are not an alternative to [...] exemplar models, but merely a computational framework in which either type of model (or, something in between) can be instantiated", as Ambridge (2018: 34) puts it. Thus, a cognitive model can be both connectionist and exemplar-based at the same time (cf. Nosofsky, Kruschke & McKinley 1992; Kruschke 1992), drawing on the representational structure of the latter and the computational (learning) operations of the former, as Busemeyer & Diederich (2010) demonstrate. They compare prototype and exemplar models employing connectionist computations. In the context of the present study, it will suffice to sketch the basic characteristics of exemplar connectionist models and highlight implications for linguistic analysis. For a comprehensive review of the connectionist tradition in the cognitive sciences see Pospeschill (2004). The survey by Thomas & McClelland (2012) focusses on computational aspects of the connectionist approach.

Connectionist models, also known as *artificial neural network models* or *parallel distributed processing (PDP) models*, have been developed in the cognitive sciences to account for human category learning in general and have been trained on a wide range of cognitive phenomena such as language. Originally designed to explain how the brain executes genuinely human intelligence tasks of varying complexity, connectionist models have also intensely been adopted in the research on artificial intelligence to implement computer-driven technologies such as facial recognition and machine translation. The two volumes on *Parallel Distributed Processing* by Rumelhart, McClelland and the PDP Research Group (1986) are generally considered the foundations of the connectionist paradigm. Within this extensive research project, Rumelhart & McClelland's (1986) analysis of the acquisition of the English past tense can be credited with

pioneering connectionist models in the study of linguistics, more precisely in morphological processing (see Kruschke 1992: 37 for a summary). The connectionist approach views cognitive processing as neurophysiologically anchored in the architecture of the human brain (cf. Schlüter 2005: 266); the computational mechanisms invoked in the connectionist tradition are thus modeled on those assumed to operate in the human neural architecture, even though the anatomical plausibility of these mechanisms is a matter of debate (cf. Thomas & McClelland 2012: 28-30).[60]

In what follows, I will sketch a simple connectionist exemplar model inspired by Kruschke (1992: 23) and Busemeyer & Diederich (2010: 23) to illustrate the basic design and operating principles of such architectures (cf. Figure 14).[61] Among the wide array of connectionist models, an important, although often far from clear-cut distinction is to be drawn between systems that rely on *local* representations plus spreading activation as an operational mechanism and those that draw on *distributed* representations. Hinton, McClelland & Rumelhart (1986: 85) point out that

> [t]he main difference is that in [the former] case there is a particular individual hardware unit that acts as a "handle" which makes it easy to attach purely conventional properties like the name of the concept and easier for the theorist who constructed the network to know what the individual parts of the network stand for.

Thus, a system that uses local representations invokes a one-to-one correspondence of exemplar representations and neural nodes; as a result, "the structure of the physical network mirrors the structure of the knowledge it contains" (Hinton, McClelland & Rumelhart 1986: 77). An exemplar network of local representations is thus similar to a semantic network in which each concept is assumed to have a physical correlate in

[60] The present survey will not be concerned with the neuroanatomical correlates of connectionist architectures. See Pospeschill (2004: 30-39) and Schlüter (2005: 260-265) for summaries of these aspects.

[61] Connectionist models come in a variety of designs. While there is general agreement that nodes and weighted connections are the integral parts of the architecture, the main point of divergence seems to be the function and structure of the exemplar layer. Even the models proposed by Kruschke (1992) and Busemeyer & Diederich (2010), from which the present model draws inspiration, differ in several crucial aspects.

the neural architecture. While conceptually intuitive, local representations lack neurobiological plausibility (cf. Peschl 1996: 579). In a system that uses distributed representations, by contrast, each exemplar is represented by multiple nodes and each node is involved in the representation of multiple exemplars; consequently, it is impossible to determine an exemplar's physical location in memory (cf. Hinton, McClelland & Rumelhart 1986: 80). Distributed representations have a number of advantages, two of which are addressed in Schade (1999: 32-33): First, a defective node does not immediately result in the loss of a representation; second, distributed representations are generated by learning operations that are amenable to computational modelling. At the same time, Schade (1999: 32-33) argues convincingly that neither advantage is relevant for a connectionist model of speech production and reception. This also holds for the connectionist exemplar model of superlative strategy choice that will be developed in the following. Such a model must be descriptively convenient in the first place and does not necessarily have to claim computational adequacy. The fact that local representations enable a one-to-one mapping of linguistic units of any kind (e.g. phonemes, syllables or words) and nodes appears particularly appealing in this context (cf. Schade 1999: 32-33): Crucially, this design feature can be used to combine connectionist principles of category assignment and Bybee's (1985, 1988, 2010) usage-based network architecture of lexical representations into a single model. The model that will be introduced in Section 4.3 therefore invokes local representations of adjectival exemplars since these will prove expedient for illustrating the structural principles of lexical connection and lexical strength. It is a central matter of debate, however, whether these exemplar representations are to be interpreted as *nodes* (as in Kruschke 1992: 23, for instance) or *activation patterns/values* (as in Busemeyer & Diederich 2010: 23-25). I have noted above that the difference between the two terms is essentially one of descriptive convenience versus neuroanatomical and computational adequacy. Notwithstanding this terminological and conceptual dilemma,[62] I will follow Kruschke (1992: 23) in referring to the entities in focus as *exemplar units* and *exemplar nodes*.

[62] Helm (1991: 100-101) argues that the distinction between local and distributed representations, and thus the question as to the nature of the underlying entities (nodes or activation patterns), is essentially pointless.

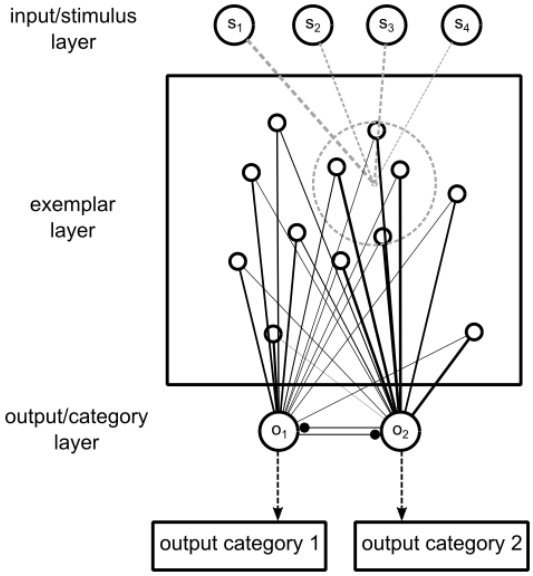

Figure 14. Basic architecture of a connectionist exemplar model adapted from Kruschke (1992: 23) and Busemeyer & Diederich (2010: 23)

The model illustrated in Figure 14 has two main components, namely (i) *nodes*, or simply *units*, which correspond to neurons on the neurophysiological level, and (ii) multiple connections of varying strengths between them. The nodes are arranged on three layers. The exemplar layer in the center of the model contains exemplar nodes which correspond to individual exemplars relevant to the category assignment task in focus. The arrangement of exemplar nodes is not random; in Section 4.3, I will specify how the systematic relations among individual exemplars can be conceived of on the basis of similarity. On the stimulus layer, stimulus nodes correspond to individual stimuli of different kinds (visual, auditory or other), all of which are relevant for the category decision at hand. Activation propagates from the stimulus nodes into the exemplar layer and focusses on a point that marks the centroid of a circular receptive field. The signal spreads outwards from this point, with the activation of an exemplar node declining as a function of its distance from the centroid. The degree to which an exemplar will be adduced to compute the output category depends on the amount of activation its node receives. The connections from the stimulus layer to the exemplar layer (the grey dashed lines

in Figure 14) are not equally strong. That is to say that the activation strengths of stimulus nodes vary; strong connections come with a higher degree of activation than weak ones. This provision accounts for the assumption that some types of stimuli are more reliable indicators in a particular categorization task than others, and that humans are capable of ranking stimuli according to their utility on the basis of their experience. Exemplar nodes are further associated with the output nodes by connections whose weights are constantly adjusted upon the outcomes of previous category decisions. Connection weights correspond to the activation levels of neurons: In line with the Hebbian axiom "what fires together, wires together", the more often a connection is activated, the "weightier" it gets and the higher are the odds that this connection will be reactivated in the future. The thickness of the lines connecting exemplar nodes and output nodes in Figure 14 indicates the weight of a connection. The winning output node (i.e. the node that receives most activation in a particular run) fires, and the corresponding output category is assigned as a result. At the same time, there are inhibitory connections holding between the competing output nodes that prevent the "defeated" node from firing and thereby interfering with the winner in the process of generating the output response (cf. Schlüter 2005: 270). While activation must spread through the system in a top-down fashion for the output category to be determined, the signal can be assumed to propagate in the opposite direction once an output node has fired. This mechanism enables that the weights of connections holding between nodes on all layers be adjusted upon the outcome of every run.

It is apparent from the model architecture sketched above that in the connectionist view, learning can be defined as the product of the persistent frequency-governed adaptation of connection weights or, to put it less technically, "change as a result of experience" (Goldblum 2001: 53). This key insight makes a strong case for statistical learning as a central cognitive capacity and, as will become evident in the next section, assigns "frequency of occurrence [...] its rightful place as a fundamental shaper of a lexical system always dynamically responsive to experience" (Monsell 1991: 150).

4.3 A connectionist exemplar view of superlative strategy choice

To begin with, I will address an issue that was set aside in the previous section: How are adjectival representations organized on the exemplar layer? Figure 15 showcases an associative exemplar network of root adjectives as proposed by Sönning & Hartmann (2019). The architecture is shaped by the two organizational principles that give rise to linguistic structure according to Bybee (1985, 1988), namely lexical connections and lexical strength.

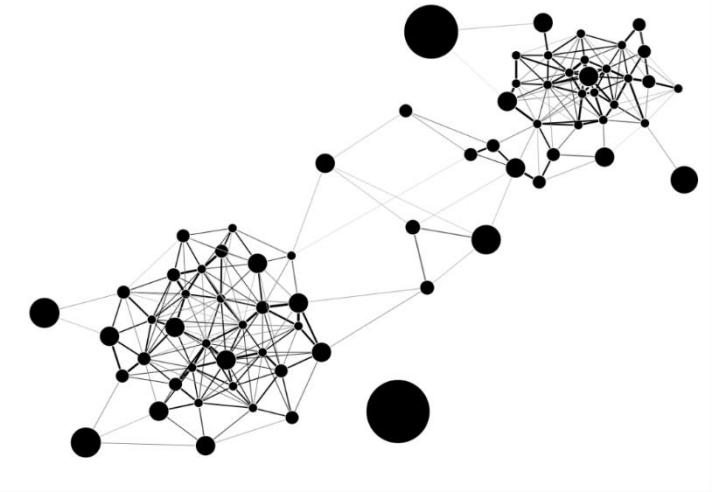

Figure 15. Exemplar-based network of adjective types (adapted from Sönning & Hartmann 2019)

The dots of different sizes represent individual exemplars. The larger a dot is, the greater its lexical strength due to its high token frequency. Lexical connections among exemplars are indicated as lines of different strengths: Finer lines are usually also longer than bolder ones. The closer two exemplars are in the grid, the more similar they are on morphophonological grounds and the more strongly they are interconnected. Figure 15 further illustrates how lexical strength and lexical connections interact. Exemplars of low and intermediate lexical strength are integrated into clusters due to their many and strong lexical connections to formally similar exemplars (cf. the clusters at bottom left and top right). Types of higher frequency form only few and weak connections in the network, the

most frequent ones are completely autonomous. This network of exemplars, which is generated and persistently updated as linguistic input is processed, provides the representational basis for the connectionist computation of gradation strategy choice.

Now that the basic operational mechanisms of connectionist models and the hypothesized representational structure of exemplars have been specified, it is time to integrate these aspects with each other. Figure 16 illustrates how the superlative output category is determined in the model.[63] A distinction is made between three different values of gradation frequency,[64] namely "new" adjectives, i.e. types with which a language user does not have any previous gradation experience[65] (top left), adjectives of intermediate (top right) and high frequency (bottom).

In the case of a new adjective, stimulus nodes that each represent individual phonological characteristics such as number of syllables, final segment, and lexical stress transmit activation into the exemplar layer. This activation focusses on a point (marked by a blue dot in the figure) from where the signal propagates outwards in a circular fashion. Since the exemplar space does not yet contain a representation of the adjective under scrutiny at this point, all exemplars that receive activation are adduced for superlative category assignment. Each exemplar node has weighted connections to the output nodes o_{syn} and o_{ana}, which produce the synthetic and analytic superlative forms, respectively; the weights of these connections are adapted upon experience with graded forms. The degree to which activated exemplars serve as the basis of analogy depends on their distance to the centroid of the circular field of activation, i.e. on their formal similarity to the adjective to be graded. In the present example, therefore, all exemplars located within the blue circle receive activation. As the activated exemplar nodes have stronger connections on aggregate to the synthetic than the analytic output node, the former node first reaches threshold activation, fires and sends an inhibitory signal to its competitor. The new adjective is thus attracted to the superlative

[63] The model that will be presented here does not take into account the effects of contextual constraints as investigated in Chapter 3.

[64] In the following Section 4.4, I will explain why quantifying experience with *graded* forms is most appropriate from the usage-based perspective adopted in the present study.

[65] In the current model, experience with graded forms is assumed to encompass both the perceptive and productive dimensions, i.e. a language user's (passive) exposure to and (active) use of graded forms.

schema instantiated by the cluster of exemplars with which it integrates due to formal similarity (cf. Langacker 1987: 371-372).

As for adjectives in the ranges of intermediate and high gradation frequency, the stimulus nodes directly activate adjectival representations on the exemplar layer. Since the exemplars already have weighted connections to both output nodes, superlative category assignment does not involve summing the connection weights of formally similar exemplars as hypothesized for new adjective types. The high-frequency exemplar in Figure 16 has a strong connection to the analytic output node and a significantly weaker connection to the synthetic output node. The adjective in focus is therefore highly likely to form the analytic superlative. The outcome probabilities for the intermediate-frequency exemplar are distributed more evenly; however, the adjective has been integrated into a cluster that instantiates the X-*est* schema and is thus more likely to be graded synthetically than analytically.

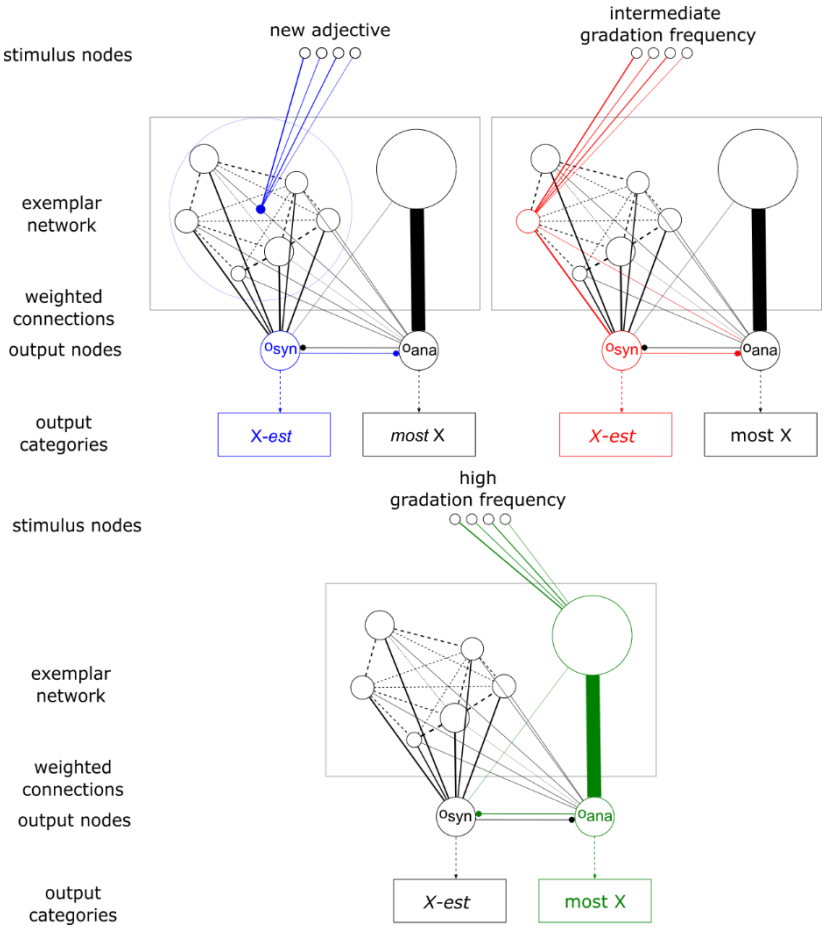

Figure 16. Connectionist exemplar model of superlative strategy choice exemplifying category assignment for adjectives without gradation record (top left), adjectives of intermediate (top right) and high gradation frequency (bottom)

4.4 Research questions

The connectionist exemplar view of superlative strategy choice that was presented in the previous section calls for a joint investigation of two factor groups, namely frequency and the type-specific properties of the adjective.

Every study that intends to investigate frequency effects must initially be concerned with the question of *what* to count – and *why*. Previous studies of comparative strategy choice have almost exclusively relied on two frequency metrics, namely positive frequency (i.e. the token frequency of the root adjective) and the comparative-to-positive ratio of an adjective (Hilpert 2008; Cheung & Zhang 2016). As for positive frequency, Hilpert (2008: 395) observes that *easy* – highly frequent in the positive – is near-categorically synthetic in the comparative while low-frequency *queasy* is not; his multifactorial analysis, however, treats positive frequency and type-specific properties (number of syllables, final segment etc.) as separate variables, thereby allowing no inferences about a possible connection between positive frequency and morphophonological similarity. The comparative-to-positive ratio can be related to two aspects: First, if Hay's (2001, 2003) relative frequency hypothesis also holds for inflectional morphology, an assumption that has not yet been subject to extensive empirical scrutiny, this ratio could be taken to determine whether a comparative form is more likely to be handled by morphological decomposition or to have a whole-word representation in a dual-route architecture, as noted in Section 4.2.1. Second, this ratio has been linked to the semantic notion of gradability, with weakly gradable types more likely to form the analytic comparative (Hilpert 2008: 397; Kunter 2017: 16). Mondorf's (2009b: 179) concept of *attested gradability* measures gradability as comparative tokens (synthetic plus analytic) per amount of text

The present study quantifies experience with *graded* forms, which should occupy center stage in the usage-based view of gradation strategy choice. In the connectionist exemplar model presented in 4.3, the weights of the connections between the adjectival exemplar nodes and the category nodes producing the superlative outcome (as illustrated in reference to Figure 16) are persistently adjusted upon the mental records of this experience; equally important, this experience governs the relations between

exemplar representations in the network by adjusting their lexical connections and lexical strengths. The present study therefore zooms in on superlative frequency, i.e. the combined token counts of an adjective type's synthetic and analytic superlative forms. It is legitimate to ask, however, why the present investigation is confined to the superlative and does not extend to the comparative. Indeed, if every token of experience with *livelier* reinforces the connection between the lexical representation of *lively* and the output node producing *liveliest*, as is likely the case, would it not be more accurate to combine the comparative and superlative token counts into a joint frequency measure? While conceptually appropriate, there are methodological limits to this proposal, to the effect that the ensuing extension of the token set to be extracted from the current corpus would considerably increase the time required for manual disambiguation (cf. Section 4.5.2). The study at hand thus exclusively focusses on superlative frequency and leaves an investigation of *gradation frequency* as a composite of comparative and superlative frequency for future work.

As remarked earlier, the present analysis draws methodological and conceptual inspiration from Sönning & Hartmann's (2019) investigation of the comparative in that it zooms in on differences in superlative frequency among formally defined clusters of adjective types. The authors argue that research on *contextual constraints* should exclude categorical types and only rely on alternating types. If the analysis is aimed at identifying *type-specific patterns*, however, types with categorical strategy preferences provide highly informative data points. The study at hand follows this advice and probes into frequency ranges that have not yet been subject to empirical examination, both in terms of type and token frequency. Table 5 compares previous multifactorial analyses of gradation strategy choice with regard to selected methodological parameters.

Table 5. Overview of multifactorial studies quantifying type-specific factors and frequency metrics in gradation strategy choice [66]

study	corpus	focus	type restriction	types	tokens
Hilpert (2008)	BNC	comparative	alternating types only	247	79,918
Cheung & Zhang (2016)	BNC	comparative superlative	alternating types only alternating types only	308 162	22,320 9,469
Sönning & Hartmann (2019)	BNC	comparative	categorical and alternating types; suppletives, participles and compounds excluded	2,921	207,171
current study	GloWbE US + GB	superlative	categorical and alternating types; suppletives, participles and compounds excluded	4,093	775,283

My analysis differs from Sönning & Hartmann (2019) not only regarding the three most salient criteria that can be derived from Table 5, namely the data source, the investigative focus (comparative/superlative) and the frequency dimensions under scrutiny, but also with respect to other methodological aspects (cf. Section 4.5). Despite these differences, the connectionist exemplar view at the heart of both studies gives rise to two common central research questions to be addressed in the following:

(1) What is the relation between *gradation frequency* and *analyticity* (measured as the proportion of analytically graded forms)? This question aims at identifying global patterns and disregards cluster effects based on formal similarity.
(2) How do these two metrics relate to each other among groups of morphophonologically similar adjectives? This question specifically zooms in on formal differences between adjective types.

The idea that the frequency of graded forms might at least partly account for different gradation strategy preferences, and that diverging distributional patterns of adjectives within the same morphophonological

[66] The type and token columns list the counts that were considered in the statistical analyses of the respective study. This does not include tokens and types that were discarded in the pre-processing stages. Section 4.5.2 sketches the pre-processing stages of the present study.

group might be governed by this frequency metric, is not new: Mondorf's (2003: 260-262, 2009b: 40-42) studies imply that for selected monosyllabic types with non-categorical behavior, analyticity is negatively correlated with comparative frequency. However, her analysis of disyllables in <-y> – likewise restricted to a small range of alternating types – does not hint at a systematic relation between the two variables. As we turn to the predictions that arise from the usage-based model discussed in Section 4.3, it will also be rewarding to consider trends that would be expected on the basis of alternative accounts of gradation strategy choice and, at least at first glance, do not fit the theoretical assumptions from which the present analysis originates. Due to a high degree of methodological and conceptual overlap, it should not be surprising that my predictions converge by and large with those made by Sönning & Hartmann (2019) for the comparative. It remains to be seen, however, to what degree the results will converge, and further, whether systematic deviations in the obtained distributions can be ascribed to structural differences between the comparative and the superlative.

As for the gradation strategy preferences of low-frequency types, three scenarios appear plausible: First, adjectives that have very rarely or not at all been graded before employ the globally more productive strategy, i.e. the strategy with the higher overall type frequency (see, for instance, MacWhinney 1978, particularly pp. 16-17, 82). The data provided by Sönning & Hartmann (2019) suggest that if the analysis is not limited to alternating types, the type frequency of the analytic comparative is 4.6 times as high as the type frequency of the synthetic comparative, while the token frequency of the synthetic comparative is 3.0 times the token frequency of the analytic comparative. Along these lines, therefore, types of low frequency should be expected to turn to *more* and *most*. The second explanation establishes an inverse link of frequency and processing complexity: Assuming that low-frequency types are less well-entrenched in any of the comparative alternants than high-frequency types, Mondorf (2009b: 40-41) argues that *more*-support serves as a fallback strategy for types of low frequency. The third scenario emanates from the usage-based view of gradation strategy choice presented in Section 4.3: The gradation schema that will be chosen for a particular low-frequency type is determined by analogy with morphophonologically similar exemplars on the basis of connec-

tionist computational mechanisms. These types have no lexical connections to other adjectives in the exemplar network and no or minimal lexical strength.

As for types with a high gradation frequency, the usage-based perspective holds that these exemplars have gained enough lexical strength to be autonomous from morphophonological clusters in the network. Their high degree of autonomy (cf. Bybee & Brewer 1980) results in a categorical preference for one of the two strategies. High frequency further has implications for dual-route morphological processing. Whole-word frequency effects of the kind reported by Alegre & Gordon (1999) are assumed to apply to types above a certain threshold frequency, at which rote access to the graded form is faster, by hypothesis, than morphological decomposition. Although the empirical evidence provided in Alegre & Gordon (1999) is limited to whole-word frequency effects in affixational morphology, thereby only accounting for rote access to instances of the X-*est* schema, the scenario that high-frequency types instantiating the *most* X schema appears no less compatible with usage-based construction grammar: As Divjak (2019: 141) puts it, the "loosening of paradigmatic bonds", i.e. cutting the ties to other exemplars in the cluster, to speak with the network model set out above, comes with a "strengthening of syntagmatic bonds", to the effect that multi-word sequences are processed and stored as units (see Arnon & Snider 2010, Tremblay & Baayen 2010 and McCauley & Christiansen 2014 for chunking of phrasal strings). A complexity-based line of argument in the sense of Mondorf (2009b: 41) would predict that high-frequency types, due to their being entrenched in one of the two strategies, incur low processing costs and thus do not require *more-*/*most*-support. If we accept that whole-word access consumes fewer resources in terms of processing than morphological decomposition, this complexity-driven prediction ties in nicely with the usage-based implications set out above.

Finally, what should we expect for types in the intermediate range of gradation frequency? While it appears reasonable to assume that the threshold value for whole-word-frequency effects proposed by Alegre & Gordon (1999), namely 6 pmw, marks the upper limit of this range, its lower limit is hard to determine on empirical grounds. On the conceptual level, this is the frequency at which types begin to establish lexical connections with formally similar exemplars in the network architecture,

gradually reinforcing their cluster membership and accumulating lexical strength in the process. It is in the ranges of low and intermediate frequency that most types with alternating behavior are predicted to be found; the higher a type's gradation frequency, the more likely it is to have categorical strategy preferences. This trend is assumed to obtain both in the global perspective and on the more granular level of those individual morphophonological subgroups for which previous studies have implied sizeable variation. As for dual-route morphological processing, types in the ranges of low and intermediate frequency are handled by the decomposition route, by hypothesis.

Probing into formally defined groups of adjectives will further allow for more conclusive comments on some of the type-specific issues that were raised on the basis of the questionnaire data in Section 3.4.4. In this context, I will revisit the hypothesis that disyllabic types in /-i/ and /-li/ have diverging strategy preferences (cf. Chua 2018: 466). Moreover, the analysis will address the alternation observed among disyllables in /-ə(r)/. Two integral comparisons will be drawn in the analysis: First, to continue the crossvarietal dimension pursued in the first empirical part of this work (Chapter 3), my corpus data will enable a comparison of the main reference varieties BrE and AmE. Second, drawing on the BNC data provided by Sönning & Hartmann (2019), my investigation will expose diverging and converging trends in the distributions of the comparative and the superlative.

4.5 Method

Reviewing the tradition of electronic corpora in English corpus linguistics, Mair (2006) identifies two methodological approaches that have developed independently from each other: In a spirit of philological accuracy, the "small-and-tidy" approach has foregrounded the fine-grained analysis of linguistic surface structure, thereby traditionally drawing on clean corpus material of a small size. The corpora of the Brown family and the ICE corpora have most typically been used in this tradition. Corpora located towards the "big-and-messy" end of the spectrum, by contrast, have provided the researcher with the large dimensions of data that were required for a computer-assisted analysis of frequency, thus uncovering the distributional regularities of linguistic variables. The original BNC can

be considered the first attempt to combine the benefits of both approaches in that it contains 100 million words and meets the characteristics that enable a thorough philological examination typical of the "small-and-tidy" tradition, i.e. a genre-balanced makeup, extensive part-of-speech tagging and rich metadata. While the "big-and-tidy" approach has enjoyed popularity upon the early success of the BNC, it is expected that "small-and-messy" corpora will be of little use to corpus linguists in the future (cf. Mair 2006: 356). More recently, corpus compilers have begun to harness the enormous potential of the World Wide Web to increase the size of corpora far beyond the benchmark set by the BNC. Drawing on usage-based theory, the present study views superlative strategy choice as firmly grounded in experience with graded adjective forms. In the light of the empirical focus on frequency that can be derived from this theoretical starting point, it appears advisable to use a corpus of large size; this will allow for an investigation of gradation frequency ranges that have not yet been subject to empirical scrutiny.

In 4.5.1, I will discuss why the GloWbE corpus is a suitable corpus for the present analysis; a critical assessment of GloWbE will further highlight that while this corpus is undoubtedly big, there are limits to its tidiness. The chronological outline of data collection and coding that will be offered in Section 4.5.2 illustrates that a great deal of manual cleaning was necessary to remedy this shortcoming. Section 4.5.3 explains the statistical model used for data analysis.

4.5.1 The GloWbE corpus

The *Corpus of Global Web-based English* (GloWbE) is comprised of 1.9 billion words drawn from 1.8 million webpages from 340,000 websites. Containing online material from six Inner and 14 Outer Circle varieties, the corpus has proved a valuable resource for research in the domains of web-based English and World Englishes. GloWbE incorporates two text types: Informal blogs account for about 60 per cent of the corpus, the remaining 40 per cent consist of more formal web-based text types such as online newspapers, magazines and company websites. Davies & Fuchs (2015: 4) note that in terms of stylistic makeup, GloWbE can be taken to be similar to the ICE corpora, which also roughly have a 60/40 split of informal and

formal language. Drawing such structural parallels between the two corpus projects, however, appears rather optimistic: While the compilation of GloWbE largely relied on automatic text retrieval, the compilers of ICE have expended great manual effort to ensure that the respective subcorpora be balanced according to stylistic criteria and representative of the national variety they intend to represent. Loureiro-Porto (2017) compares GloWbE and ICE with regard to several methodological aspects; her results reveal significant differences between the blog material contained in GloWbE and the spoken sections of ICE, and imply that the text categories of GloWbE (general websites and blogs) significantly converge with regard to several features of informal language (Loureiro-Porto 2017: 467). The latter finding lends support to the observation made by Röthlisberger (2018: 51) that the distinction between the two text categories covered in GloWbE is not always clear-cut.

Notwithstanding these general points of methodological criticism, there are three major reasons why GloWbE appears a suitable corpus for the study at hand, namely (i) its size, (ii) the fact that it enables a comparison of varieties and (iii) its large share of material from informal blogs, which has been claimed to mirror casual, unscripted speech. These three advantages of the GloWbE corpus will critically be addressed before turning to some of its inconveniences.

The current study taps into the US and GB sections of GloWbE, each of which contains about 387 million words. At its total, therefore, the present corpus is roughly 8 times as large as the BNC, which was used in the analyses of gradation strategy choice by Hilpert (2008) and Cheung & Zhang (2016). Since my prime focus is on the interplay of type-specific factors and gradation frequency, a corpus was needed that reliably captured superlative types in all frequency ranges, notably also encompassing a sizeable number of very rarely graded adjectives. Further, we can hope to draw valid comparisons between BrE and AmE, since the US and GB sections are of roughly identical size and comparable in terms of register and style. Mair (2015: 29) highlights exactly these advantages of GloWbE:[67]

[67] The special issue of *English World-Wide* 36 (1) provides a detailed discussion of GloWbE, also addressing methodological problems that arise from the corpus architecture. This issue contains a focus article by Davies & Fuchs (2015) and responses by Mair (2015), Mukherjee (2015), Nelson (2015) and Peters (2015).

> The researcher using GloWbE as a stand-alone corpus to investigate mid- and low-frequency phenomena in World Englishes can thus rest assured that, data-wise, they stand on firm ground. Most grammatical constructions and lexicogrammatical variables distinguishing varieties of English are in this frequency band, and currently only GloWbE provides the amount of data and the convenient and powerful search interface needed for systematic comparison across varieties.

Furthermore, with about 60 % of its material drawn from blogs, GloWbE can be considered as representing, in large part, the most speech-like and informal language of the Internet.[68] Formal texts such as (online) newspaper articles are customarily pre-written in a word-processing program (e.g. Microsoft Word), whose grammar checker may prompt users to reconsider their choice of superlative strategy, thereby potentially eliminating variation in the data.[69] The input boxes of blog interfaces, by contrast, usually come without a spell checker; hence, there are no visual clues that would urge the author of a blog post to review previously entered text, resulting in greater authenticity and spontaneity of the linguistic material.

Evidently, the architecture of GloWbE also comes with disadvantages that can be linked to the rather general aspects of corpus compilation addressed at the beginning of this section. First, as the present study is concerned with frequency, a particularly urgent problem pertains to the sizeable amount of duplicate text in the corpus. On the GloWbE website, the compilers acknowledge this issue and note that they regularly run scripts to identify and eliminate duplicates, thereby gradually "cleaning" the corpus. The majority of these cases seem to represent *intra-page* duplicates, i.e. strings of text that are repeated at least once on the same webpage.

---

[68] There is, of course, variation in the degree to which blog language resembles spoken language. Grieve et al. (2011) use the Biber Tagger to investigate register variation among blog texts along the dimensions proposed by Biber (1988, 1995). Their cluster analysis identifies two major blog types, namely personal blogs and thematic blogs, both of which exhibit what the authors consider "the standard blog voice" (p. 320). This voice is described as "a highly personal and conversational style".

[69] I have already hinted at the problem of grammar checkers in the study of grammatical variation in Section 3.4.5.

These mostly occur on webpages that have a section with reader comments; if contributors reply to a previous comment, its text is often automatically repeated prior to the new input. It appears mandatory in the light of the research questions at hand that such instances be manually identified and weeded out in order to prevent distorted tallies.

A second problem of GloWbE data springs from what has been framed as one of its greatest assets: Blogs can be seen as representing unrevised conversational language in a written format. This advantage, however, comes with an issue that has also been reported in Claridge's (2007b: 92) methodological account of internet forum language: Texts from this source often contain misspellings, typos and missing spaces between words. These errors might have affected the accuracy of the automatic tagger, resulting in (minor) perturbations of type and token counts.

A final problem to be addressed here can be boiled down to this question: How reliably can the text of a website represent a particular variety of English? Davies & Fuchs (2015: 4) point out that while country-code top-level domains such as *.ca* (Canada) and *.au* (Australia) provide a fairly transparent guideline for country assignment, Google uses a combination of three heuristics to associate generic top-level domains such as *.com* or *.org* with their respective countries of origin, namely (i) the IP address of the server hosting the website, (ii) the origin of those who link to the website and (iii) the origin of those visiting the website. Even if we can be reasonably certain that the website's country of origin has correctly been identified according to these criteria, we have good reason to object that the commentary board of a blog site which has been assigned to GloWbE US, for instance, may contain text written by visitors from the UK; this eventuality appears particularly urgent for websites covering topics of supralocal relevance (cf. Claridge 2007b: 94). It is impossible, of course, both for the corpus compiler relying on automated compilation mechanisms and the researcher tapping into GloWbE, to determine the degree to which such instances might blur varietal comparisons on the basis of GloWbE data. Cook & Hirst (2012) are concerned with the question whether a web corpus compiled from country-code top-level domains reliably represents the variety of the corresponding English-speaking country. They use the British and Canadian country-code top-level domains to compile national web corpora for the varieties in focus (i.e. BrE and CanE) and compare these corpora to the BNC, the COCA and the Strathy Corpus

along the lines of selected metrics of similarity. Their results imply that web corpora compiled from websites with a given national top-level domain are relatively similar to non-web-based standard reference corpora of the same variety but quite dissimilar to standard reference corpora of other varieties. This can be taken to suggest that web corpora built from such domains fairly reliably represent national varieties. Although two points of methodological criticism appear in order, namely (i) that Cook & Hirst's (2012) analysis does not consider that differences in text type and register may affect the comparability of websites, and (ii) that the similarity metrics applied in their study only cover linguistic features on the word level, thereby neglecting syntactic differences between varieties, we can be relatively confident that on the whole, the GB and US sections of GloWbE provide a representative data basis for the national varieties under scrutiny.

4.5.2 Data collection and coding

It is important to note that I used the POS-tagged full-text data of GloWbE for data collection. On the corpus website, the compilers state that in this version, which can be downloaded under a fee-based academic license, "[e]very 200 words, ten words are removed and are replaced with '@'" for reasons of copyright.[70] Due to their randomness, these omissions do not significantly affect the inferences that will be drawn from the data; however, the obtained frequencies may deviate slightly from those that would be gleaned from the online version, which uses the full data.

In the initial query, I retrieved the type and token frequencies for all superlative forms from the US and GB sections of GloWbE using the CLAWS7 search strings `*_JJT` (synthetic) and `most_RGT *_JJ` (analytic, i.e. *most* tagged as superlative degree adverb followed by the adjectival root) in CQPweb (version 3.2.31). The types obtained through the initial query were then converted to lower case to avoid the double-counting of types.[71] Table 6 displays the results of this query.

[70] Cf. <https://www.corpusdata.org/limitations.asp>.

[71] Particularly in blogs, superlatives were occasionally found entirely in upper case.

Table 6. Type and token counts retrieved through the initial query

	GB		US		Total	
	types	**tokens**	**types**	**tokens**	**types**	**tokens**
-est	2,885	449,763	3,175	403,026	4,685	852,789
most	6,692	173,321	6,443	157,265	9,376	330,586

Pre-processing was conducted in two stages. In the first stage, I screened the type list for false positives such as *Budapest*, *Oktoberfest*, which were mostly the product of incorrect tagging. These types were consequently discarded. In addition, lemmas of types whose formal characteristics constitute a definite knockout criterion for one of the superlative alternants were excluded from further analysis. This applied to superlatives of suppletive paradigms (*best*, *worst*) and participles (e.g. *interested*, *entertaining*, *chosen*). As for the latter, simply discarding all lemmas ending in *-ed* and *-ing* would have eliminated true positives such as *kindred* or *sterling*. A careful manual disambiguation of all types retrieved through the initial query was therefore necessary. Types such as *affortable* or *aweful* were retained in the dataset if these character strings unambiguously represented misspellings of existing adjectives.

The morphological class of compounds further required a differentiated approach. Mondorf's (2009b: 42-56) data comprise a considerable number of adjectival compounds of the type Adj + N that exhibit alternating behavior in the comparative (e.g. *low-key*, *high-profile*), with the superlative suffix attached to the adjectival element; assuming that for compounds of this type, gradation strategy choice is governed both by overall string length and the strategy preferences of the adjectival component, however, one is hard-pressed to explain why *high-risk* (78 synthetic, 7 analytic tokens) and *high-tech* (7 synthetic, 14 analytic tokens) exhibit diverging distributions in Mondorf's comparative data. Simple queries in the GB and US sections of GloWbE similarly hint at sizeable variation in the superlative among adjectival compounds of this type, e.g. *high-profile* (133 synthetic, 278 analytic tokens). The formal heterogeneity of the compound class constitutes a major problem for the analysis. This class in-

cludes instances of the types Adj + N (*high-class, long-term*), N + Adj (*environment-friendly*), N + [V + *ing*/*ed*] (*time-consuming*) and Adj + [V + *ing*/*ed*][72] (*wide-ranging, short-lived*) as well as phrase compounds that lack a genuinely adjectival component but have lexicalized as adjectives, e.g. *cutting-edge* and *state-of-the-art*. A study that aims to shed light on the gradation strategy distributions of compounds must distinguish between the different types of compounds. Additional type-specific categories would have to be established to give due consideration to this distinction. While beyond the scope of the present investigation, such an extended analysis might be a promising avenue for future research. Instances of the compound types Adj + [V + *ing*/*ed*] and N + [V + *ing*/*ed*] as well as phrase compounds were altogether discarded. As for the compound types Adj + N and N + Adj, orthography served as an expedient guideline, although legitimate objections may be raised to this approach.[73] Types with hyphenated spelling were automatically excluded without further inspection; for solidly spelled compound adjectives in these classes, the degree of lexicalization served as the main criterion for inclusion in the pool of adjectives for further analysis. Therefore, the guiding question in disambiguating these types was this: Does the item have (non-)compositional meaning, or in other words, to which degree is the meaning of the whole predictable from the meaning of its constituents? If this degree was determined to be low, as for *commonplace* and *steadfast*, for instance, the respective adjective qualified for further analysis. All other adjective types were retained for further

[72] Schmid (2005: 122, fn. 11) remarks that instances of the types N + [V + *ing*/*ed*] and Adj + [V + *ing*/*ed*] are borderline cases between compounding and derivations by suffixation since the V + *ing*/*ed* elements usually do not represent established lexemes. He labels these instances *synthetic compounds*.

[73] Referring to a noticeable degree of spelling variation in nominal compounds such as *girl*(*-*)*friend*, Schmid (2005: 132-133) argues convincingly that orthography is an unreliable indicator of compound status. Sanchez-Stockhammer (2018) provides an in-depth multivariate analysis of compound spelling in English, notably also encompassing adjectival compounds.

analysis if their morphological makeup complied with English word formation rules; this also applied to ad hoc formations such as the blend *terrorific*.[74]

The lemmas of the superlative types that qualified for further analysis were then coded for the following characteristics:

- Number of syllables: 1, 2, 3+
- Final segment(s): /i/, /li/, /ə(r)/, /l/, consonant cluster, other
- Superlative frequency, i.e. the sum of *-est-* and *most*-tokens as retrieved through the initial query, per million words (log-scaled, for ease of graphical presentation)

In Section 3.2.1, I remarked that previous studies have been remarkably inconsistent in their notation of the adjectival ending. Three notational patterns feature prominently in the literature, namely (i) phonological notation, e.g. /-li/, (ii) graphemic notation, e.g. <-ly> and (iii) the final grapheme(s) in italics, e.g. *-ly*, particularly in studies that intend to highlight the fact that the final segment categories /-li/ and /-i/ also represent morphemes. In the current study, I will use phonological notation and adopt one of the alternatives where the focus of analysis is on spelling or morphology.

The different notational conventions that can be identified in the previous literature also have consequences for the classification of individual types. Mondorf (2009b) does not distinguish between <-y> and <-ly>, listing *costly*, the only <-ly>-disyllable in her data, among adjectives ending in <-y>. In line with Chua (2018), the adjectives that were assigned to the /-i/- and /-li/-categories in the current study are uniform both in orthographic and phonological terms, with their endings spelled <-y> and <-ly>, and their final segments pronounced /-i/ and /-li/, respectively. Types such as *sly*, *dry* and *eerie*, which meet only one of these conditions, were consequently classified as "other". A similar problem arises for adjectives ending in <-l(e)>: Mondorf (2003: 283-284, 2009b: 35-36) assigns types as *subtle*, *noble* and *fickle* to this category. The current study follows

[74] The full list of adjective types included in the final dataset can be found in the web appendix at https://osf.io/vty9r. The varietal datasets containing all adjective types from the GB and US sections of GloWbE are available at https://osf.io/wf56d (GB) and https://osf.io/2bq4a (US).

Mondorf (2003, 2009b) in coding every adjective that ends in <-l(e)> as /l/, even if this sound is part of a consonant cluster. This decision is primarily motivated by the urge to be consistent with the classification adopted in Sönning & Hartmann's (2019) study of the comparative, which will serve as a main point of reference in Section 4.6.4.

In the second stage of pre-processing, the R package *tidyverse* (Wickham 2017) was used to extract the individual synthetic and analytic tokens of each adjective from the GB and US sections of GloWbE. To filter out duplicate tokens, the left and right context (20 words in total) were combined and identical cases (i.e. those with the same string of words in the left and right context) were eliminated. Table 7 specifies the number of duplicates that were eliminated in the GB and US subsets.

Table 7. Number of eliminated duplicate tokens by varietal subset

duplicates	GB	US
-est	15,436	14,599
most	5,305	5,481

The remaining tokens then needed to be screened to weed out adverbial superlatives, quantifier uses of *most* and elatives. To reduce the amount of manual work, I settled for a cut-off of n = 50 tokens: If a variant for a given adjective occurred more often, I disambiguated only 50 tokens. To this end, a randomized subsample of n = 50 was drawn from the combined dataset (GloWbE GB and US) and used for disambiguation. The percentage of true positives in this subsample was consequently extrapolated, resulting in an adjusted overall token count for each form. For expediency, the same scaling factor was used for both varieties. Table 8 illustrates this operation for *most able* and *ablest*; in this example, the subsample yielded 94% true positives for *ablest*, and 100% for *most able*. The scaling factors 0.94 and 1.00 were consequently used for extrapolating to the overall token counts, respectively. The subsample size of n = 50 was chosen to balance extrapolation accuracy and disambiguation effort (cf. Sönning & Hartmann 2019). Thus, doubling the subsample size (and hence the time required to identify false positives) would have rendered the estimate of the overall adjusted token count only marginally more accurate. The token counts of spelling variants, both widely recognized ones (e.g. *cosy*/*cozy*) and cases in which a misspelling could unambiguously be

matched with an existing type (e.g. *abun**d**ant*/*abun**t**ant*), were combined in the analysis.

Table 8. Extrapolation of true positives in the subsamples for *ablest* and *most able*

form	overall n_{tokens}	true positives in subsample	adjusted n_{tokens}
ablest	94	47/50 (94 %)	89
most able	145	50/50 (100 %)	145

Figure 17 illustrates the distribution of scaling factors for the *-est-* and *most*-types with n > 50 tokens. Subsamples were disambiguated for more than twice as many *most*-types (n = 509) as *-est*-types (n = 221); this indicates that in high ranges of gradation frequency (n > 50 tokens), *most* is the more productive strategy among the types that enter the present analysis.[75] The overall share of false positives was higher in the *most*-subsamples (*Mean* = 0.936) than in the *-est*-subsamples (*Mean* = 0.982), just under two thirds of which did not contain a single false positive (as compared to one third of *most*-subsamples). This was in line with expectations since tokens retrieved through the search string `most_RGT *_JJ` also comprise adverbial superlatives, elatives and instances of quantifier *most* while the *-est*-superlative does not have constructional competitors with orthographic identity. On the whole, the number of false positives in the subsamples of both *-est-* and *most*-types was lower than expected; thus, my assumption that elatives would account for a significant proportion of cases in the *most*-subsamples seems to have been unfounded.

[75] Note that the actual type frequency for *most* X is probably much higher, since participles were excluded from the analysis.

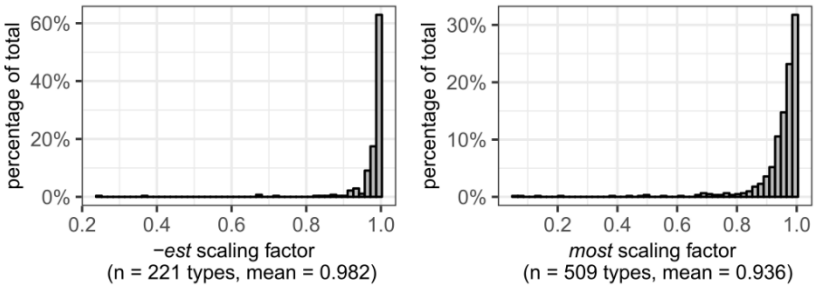

Figure 17. Distribution of the scaling factors for both superlative variants as percentage of total

Many false positives among *-est*-tokens represented misspellings of words that clearly do not represent a superlative, such as example (60), where the noun *elitist* is misspelled as the synthetic superlative of *elite*.

(60) Is there anyone left who does not feel Obama is an arrogant **elitest**? [GloWbE US 3258368]

Due to the existence of constructional competitors, the disambiguation of *most*-tokens was more cumbersome. CLAWS7 provides different tags for quantifier *most* (`_DAT`) and the superlative degree marker (`_RGT`). Although the initial query was restricted to the latter, the extracted data contained tokens in which *most* can be interpreted as clearly representing an instance of or shading into quantifier use.[76] Example (61) represents one such ambiguous case that required manual disambiguation.

(61) So after filling in the online questionair I could match my answers with the partyprograms and get an overview of parties I have the **most similar** views with. There was this new party I have a lot of similar views with. [GloWbE US 459613]

In this example, *most* is tagged as a superlative degree marker, thus suggesting genuine superlative use. The determiner *a lot of* in the right

[76] The query also yielded tokens of *most* that came with the ambiguity tag `_DAT_RGT`.

context, however, makes a quantifier interpretation of *most* more plausible. Instances of this kind were discarded after a thorough analysis of the left and right context.

In total, 67,678 (out of the original ~800,000) tokens were manually disambiguated (17,069 *-est*-tokens and 50,609 *most*-tokens). Token subsamples were screened for 689 of the 4,093 adjective types in the dataset (16.7 %); these are the types that have n > 50 tokens in one or both superlative variants. Figure 18 shows superlative frequency (on the log-scale) against proportion of *most* for all 4,093 types. It is noteworthy that considerable variation in superlative strategy choice occurs in the frequency range in which all tokens were manually disambiguated (blue), with the distribution of those types resembling an M rotated by 90 degrees. Despite some outliers with alternating behavior (especially between 0.1 and 1 pmw), types for which a subsample of tokens was disambiguated (red) exhibit rather categorical strategy preferences and are located in the right-hand regions of the frequency band.

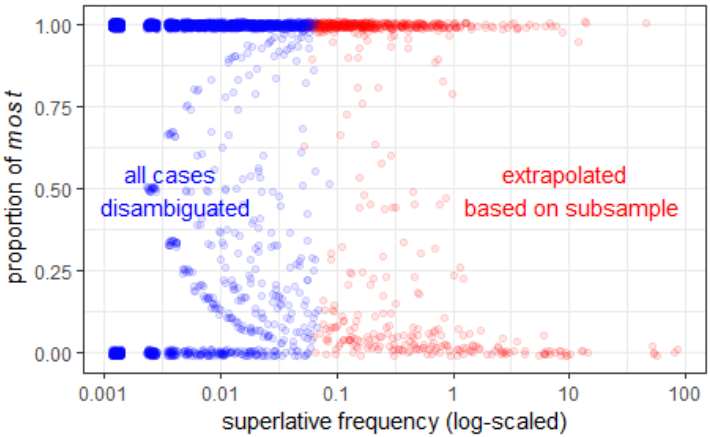

Figure 18. Types for which a subsample of tokens was screened (n = 683, red) and types for which all tokens were disambiguated (n = 3,410, blue)

Table 9 displays the type and token counts after the second pre-processing stage, i.e. the counts that were used for the analysis. Note that the tallies obtained from the initial query have been reduced considerably in the two pre-processing stages.

Table 9. Type and token counts after the second stage of pre-processing

	GB		US		Total	
	types	tokens	types	tokens	types	tokens[77]
-est	855	278,166	797	248,183	1,038	526,349
most	2,943	129,791	2,902	119,143	3,644	248,934

The summary of counts in Table 9 indicates tilted distributions: In terms of token count, X-*est* is about twice as frequent as its analytic competitor, both in the overall dataset (*most* 1 : 2.11 -*est*) and the varietal subsets (GB 1 : 2.14, US 1: 2.08). In terms of type frequency, *most* X exceeds that of X-*est* by a factor of 3.5; again, this trend is consistent across varieties (GB 3.4, US 3.6). These distributional patterns are further in line with those observed by Sönning & Hartmann (2019) for the comparative.

Before we go further, however, I need to mention a setback I encountered in the process of extracting, down-sampling and disambiguating hits from GloWbE. The local GloWbE version I initially relied on included curious tagging errors, which led to numerous graded adjectives not being classified as such. As a result, the set of graded forms returned by my first query did not exhaust the full set of superlatives in the corpus. I tried to correct this error to the best of my abilities. Since the unique corpus IDs of the individual instances ("match") changed after correcting this tagging error, this task turned out to be rather complex and error-prone. I proceeded by identifying, in the more accurate body of corpus hits, those superlative forms that were underrepresented or absent in the first query. I down-sampled these (where necessary) and then merged the individual files, to form a new (and more complete) basis for extrapolation. The figures reported in the current section are all based on the corrected counts. While I cannot guarantee the absence of errors in this patchwork, I should note that the results of my analyses did not change appreciably once I had corrected for the initial tagging errors.

4.5.3 Data analysis

In Section 3.2.5 on analyzing the data collected from my questionnaire informants, I argued that any statistical analysis of linguistic data should

[77] Note that the token counts provided in this table are the extrapolated counts.

consider the logical connections among individual observations. The data points in the current dataset are grouped by a single cluster variable, namely *adjective type*: The n = 775,283 observations (level-1 units) are nested within n = 4,093 adjectives (level-2 units), and we should reasonably expect data points contributed by the same adjective to be similar to each other. Disregarding this hierarchical structure would lead to overconfident uncertainty estimates (i.e. error bands that are too narrow) and thus invalid inferences (Sönning 2020: 44).

We adopt a beta-binomial regression model, which essentially treats each adjective as an individual data point (cf. McElreath 2020: 382-385). The rationale for my use of the beta-binomial model is basically the same aim as that for the multilevel model described in Section 3.2.5. My analysis respects the fact that we are not dealing with 775,283 independent data points. Rather, these are grouped by adjective and it is in fact the behavior of these adjectives that we are concerned with.[78] This means that we effectively have 4,093 observations that inform the data summaries. In my analyses, I foreground the role of phonological traits of the adjectives and their superlative frequency. These lexemes are then broken down into subgroups of interest, e.g. low-frequency monosyllabic adjectives ending in a cluster (e.g. *apt*). However, even upon sub-classifying types according to these characteristics, I expect to find a noticeable amount of residual variation between types due to the idiosyncratic behavior of lexemes and other systematic differences my analysis does not take into account. This remaining variation among adjectives that, from the viewpoint of our analysis, instantiate the same condition (i.e. number of syllables, final segment(s), and superlative frequency) is captured by the beta-binomial model using a beta distribution. This distribution describes the scatter of adjectives around the predicted average proportion for a subgroup (e.g. 15 % *most*). The model thereby sensitizes my estimates to the amount of between-adjective variation that is observed. If it is large, for instance, the statistical uncertainty of my estimates is adjusted, as the estimates then become less precise.

[78] In the present analysis, clustering by source was impossible due to the down-sampling, which was in turn due to the large number of tokens and the need for disambiguation. The analysis is thus a compromise between scale and feasibility.

Bayesian estimation was conducted in R (R Core Team 2020) using the package *rethinking* (McElreath 2016).[79] The packages *lattice* (Sarkar 2008) and *ggplot2* (Wickham 2016) were used for data visualization.

4.6 Results

Let us now turn to the findings of the present study. We will begin with a descriptive look at the superlative frequency distribution based on the overall dataset (Section 4.6.1).[80] Section 4.6.2 then zooms in on the superlative frequency trends by phonological subgroup. Section 4.6.3 sheds light on the superlative frequency trends of adjectives prefixed by *un-* and their non-negated counterparts (e.g. *unhappy* and *happy*). Section 4.6.4 adopts a crossvarietal perspective, comparing the superlative frequency trends in the GB and US subsections of the current dataset.[81] Section 4.6.5 draws a comparison between the current results for the superlative and those obtained by Sönning & Hartmann (2019) for the comparative.

The presentation strategy that will be adopted in this section relies on two metrics: The proportion of analytic superlatives, which takes values between 0 and 1, will be shown on the vertical axis. Superlative frequency, which I measure as the number of superlative tokens per million words in my data, will be displayed on the (log-scaled) horizontal axis. For individual adjectives discussed in the text, these metrics will be supplied in the form *adjective* ($prop_{most}$, $freq_{sup}$), with the latter metric provided either in absolute terms, i.e. as token counts (n_{sup}) or relative to the size of the corpus, i.e. per million words, depending on which format appears more informative in the respective context. For instance, if the focus is on an individual type with only a single superlative token in the current corpus, the absolute token count $n_{sup} = 1$ is significantly more informative than the relative metric $freq_{sup} = 0.001$ pmw. If we are concerned with global trends in particular ranges of the frequency spectrum, however, the relative notation will be more intuitive. For clarity of exposition, the frequency spectrum was divided into five bands, as illustrated in Figure 19.

[79] The R script written for the statistical analysis is available at https://osf.io/2sypm/. Thanks are due to Lukas Sönning for his help with the analysis.

[80] For the overall dataset, see https://osf.io/em4qy/.

[81] For this comparison, I relied on the varietal dataset available at https://osf.io/mgvdp/.

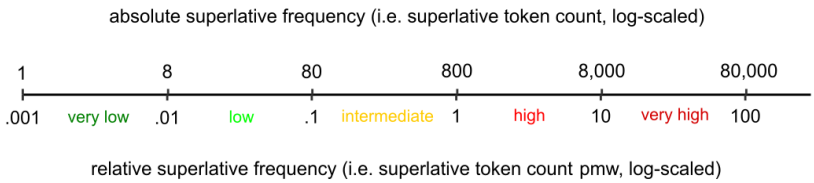

Figure 19. Superlative frequency spectrum (log-scaled): Relative frequency values are provided below the axis, roughly corresponding absolute values are displayed above the axis. The five frequency ranges and their labels were defined for ease of interpretation.

Thus, if we are interested in the patterns of strategy choice in the range of *very low* superlative frequency, for instance, this is to say that we focus on adjectives with an absolute superlative token count of 1 to 8 and a relative superlative frequency of about 0.001 to 0.01 pmw. It is important to note that the axis ticks as well as the labels for the respective superlative frequency ranges (e.g. *very high*) were chosen merely for ease of reference.

4.6.1 General trends

The graph at the top of Figure 20 maps the proportion of *most*-tokens of the n = 4,093 adjectives in the overall GloWbE dataset to the log-scaled superlative frequency pmw. The distributional pattern, which was introduced in Section 4.5.2 to highlight which types were affected by the extrapolation, resembles an M rotated clockwise by 90 degrees. The comparative distribution exhibits similar patterns, as can be inferred from the graph at the bottom of Figure 20. It is important to note, however, that this distribution is based on a corpus of a much smaller size; taken together, the present superlative corpus is about 8 times as large as the BNC material used by Sönning & Hartmann (2019)[82]. The current study thus extends to significantly lower levels of gradation frequency and includes higher type and token numbers.

[82] Thanks are due to Lukas Sönning and Stefan Hartmann for sharing their comparative dataset.

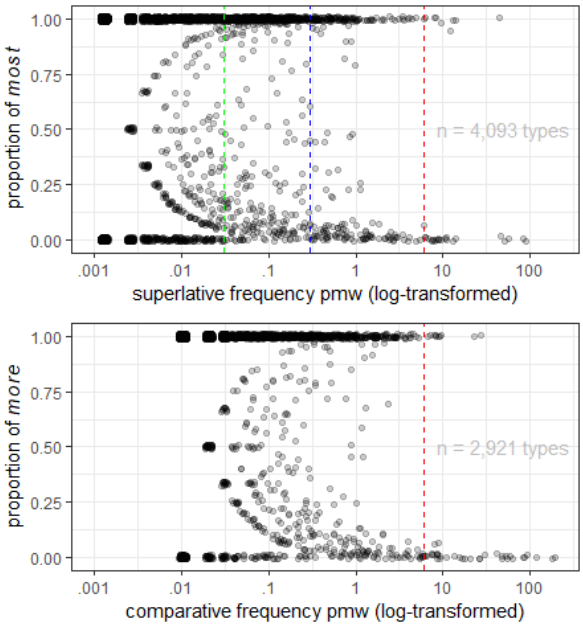

Figure 20. Gradation frequency distribution for all superlatives (top) and comparatives (bottom, BNC data from Sönning & Hartmann 2019) in the respective datasets

In the following, I will focus on general trends that can be identified in the superlative distribution. In the range of very low frequency, the overall shape is largely determined by the frequency measure: *Hapax legomena* can only take categorical values on the y-axis; as the superlative frequency increases, so too does the potential number of distinct analyticity scores an adjective can have. In the top panel (superlative data), the distribution of non-categorical types resembles the shape of a crescent whose tails converge with the categorical columns at a value of approximately 0.03 pmw (green line). Alternating behavior is particularly prevalent in the frequency bands below 0.3 pmw (n = 233 tokens, blue line). In this range, a large number of non-categorical types have a moderate or strong preference for the synthetic superlative, scoring between 0.50 and 0.00 on the analyticity scale. In the high frequency range, starting at approximately 1 pmw, adjectives almost exclusively show categorical trends; up to a superlative frequency of 6 pmw (red line), the overall density of

categorical types is in fact higher for *most* than for *-est*.[83] Above the 6-pmw-mark, we find more types that stick to the synthetic than to the analytic extreme; this applies to both clusters in this frequency band,[84] i.e. around 10 pmw (including types such as *popular*, *old* and *low*) and the highly frequent types with values between 40 and 100 pmw (e.g. *important*, *late*, *big*, *great*, *large* and *high*). The red line at 6 pmw further marks the threshold frequency above which, in a dual-route architecture of morphological processing, superlatives may be assumed to be processed by the direct lexical route (Alegre & Gordon 1999). In the dataset at hand, 32 adjective types have superlative frequencies above 6 pmw. Notably both in the current data and the comparative dataset provided by Sönning & Hartmann (2019), also types with a categorical preference for *most* score in this frequency range. With a growing body of evidence suggesting that multiword sequences can be processed and stored holistically (Arnon & Snider 2010; Tremblay & Baayen 2010; McCauley & Christiansen 2014, *inter alia*), we can assume, therefore, that analytic superlatives could have a lexical representation of their own provided that they are frequent enough.

4.6.2 Superlative frequency trends by phonological subgroup

This section analyses the superlative frequency trends for the phonological subgroups among the adjectives, i.e. combinations of syllable number and final segment, thereby also zooming in on individual types. In presenting the results, I will refer to two figures: Figure 21 illustrates the overall trends predicted by our regression model for each phonological subgroup, with the 50 % and 90 % uncertainty estimates provided in the form of error bands. These can be interpreted along similar lines as confidence intervals. Figure 22 specifies the position of individual adjectives on the grid. To facilitate cross-referencing, these graphs can also be viewed via the OSF repository.[85]

---

[83] All graphs in this and the following sections use the graphical device of *jittering*: Individual data points are minimally displaced from their actual position, particularly where they would overlap otherwise. This artificial noise helps us distinguish between clusters of high and low density but adds a tone of imprecision to some data points. The un-jittered version of the superlative frequency distribution in Figure 21 is available at https://osf.io/9axk2/.

[84] Comparing the graphs in, the "gap" between 16 and 70 pmw in the superlative distribution should not concern us unduly.

[85] See https://osf.io/hyctd/ (Figure 21) and https://osf.io/djv87/ (Figure 22).

Monosyllabic types

The monosyllabic adjectives can be grouped into four categories of final segments, with no data points for /-i/ and /-li/. These appear in the top row of panels in Figure 21. We observe a uniform trend across the four categories: Below a superlative frequency of roughly 0.1 pmw, adjectives exhibit considerable variation. Types that exclusively occur with *most* are limited to frequency values lower than 0.1 pmw, while categorical *-est* occurs in all frequency ranges. A general tendency towards *-est* develops as the superlative frequency increases, with highly frequent types (from approximately 1 pmw onwards) almost exclusively using the synthetic alternant.

The alternating trend in the low-frequency range appears to be weakest for the group of monosyllables in /-ə(r)/ (n = 19), which mainly comprises types with a (near-)categorical preference for *-est* across the frequency spectrum. Three outliers with a relatively high proportion of *most*-tokens can be identified, namely *sour* (0.29), *dour* (0.43) and *dire* (0.61); the deviant behavior of these adjectives is also manifest in the comparative data provided by Mondorf (2009b: 29-30) and Sönning & Hartmann (2019). Why should we consider *direr*, *dourer* and *sourer* more objectionable than *purer* or *sorer*, for instance? Mondorf (2009b: 29-30) ascribes the analytic preference of the three adjectives in focus to differences in the syllabification of the synthetic comparative forms: Suffixation by *-er* results in onset-coda identity within the final syllable of these three types (e.g. *di.**rer***) while coda-coda identity should be considered a valid option for the other adjectives in this phonological subgroup (e.g. *su**r.er***).[86] According to Mondorf (2009b: 29-30), the urge to avoid the synthetic comparative is stronger if suffixation by *-er* leads to onset-coda identity; we can reasonably expect language users to prefer *most* with *sour*, *dour* and *dire* by analogy, even though the hypothesized phonological effect does not directly carry over to the superlative (i.e. there is no risk of segmental identity in the synthetic superlative of these adjectives). A higher overall degree of rhoticity in American as compared to British dialects further gives reason to hypothesize that this effect is stronger in AmE than in BrE. The data at hand only

[86] Mondorf (2009b: 29-30) notes that the CEPD (Jones et al. 1997) is non-committal as to how the synthetic comparative of most adjectives in this subgroup should be syllabified, listing both *sur.er* and *su.rer* as variants, for instance; for the three adjectives in focus, however, only the variants that create onset-coda identity (e.g. *di.rer*) are provided in the CEPD.

partly support this assumption: *Dour* ($prop_{most\ BrE} = 0.25$, $prop_{most\ AmE} = 0.67$) and *dire* ($prop_{most\ BrE} = 0.44$, $prop_{most\ AmE} = 0.84$) exhibit a higher proportion of *most* in GloWbE US; the varietal analyticity scores of *sour* ($prop_{most\ BrE} = 1.00$, $prop_{most\ AmE} = 0.00$), however, are at odds with this hypothesis.[87] While Mondorf's (2009b: 29-30) line of argument focusses on the syllabification of the comparative, the phonological structure of the adjectival base may also help explain the special behavior of *sour*, *dour* and *dire*: Notably, the three adjectives in focus differ from the other monosyllabic ones in /-ə(r)/ in that they end in a sequence of three vowel sounds. This begs the question whether this sequence constitutes a triphthong or a diphthong + /ə^{r}/. The OED consistently advocates the former option, indicating that these types are monosyllabic (e.g. /daʊə/ and /saʊə/). However, the three adjectives form minimal pairs with English nouns which, according to the OED, constitute disyllables, e.g. *dire – fire* /ˈfʌɪ.ə/, *dour – sour – shower* /ˈʃaʊ.ə/.[88] Regardless of how the tricky issue of syllabification is solved,[89] this striking phonological similarity may induce language users to draw analogies to disyllables in /-ə(r)/, a group for which periphrasis appears to be the predominant gradation strategy both in the superlative (cf. Figure 21) and the comparative, as can be inferred from Sönning & Hartmann (2019) (cf. Figure 28, Section 4.6.5).

Among monosyllables ending in /-l/ (n = 18), only a single type stands out due to its relatively high analyticity value, namely *real* (0.72). On the one hand, this strategy preference may be a relic of the French etymon; on the other, the *most*-variant avoids the perceptual ambiguity between the synthetic superlative form and the phonetically similar noun *realist*.

The subgroup of monosyllables ending in a consonant cluster (n = 60) and the "other"-group comprising types with miscellaneous final segments (n = 186) exhibit considerable heterogeneity with regard to several type-specific criteria. In the low and intermediate ranges of superlative frequency (from approximately 0.01 to 1 pmw), we observe that, with

---

[87] Note, however, that the combined token numbers for *sour* ($n_{sup} = 8$) and *dour* ($n_{sup} = 7$) are very small. It will be worthwhile to examine in the future whether the suggested varietal trend bears scrutiny in a more extensive dataset, ideally comprising both the comparatives and superlatives of the three adjectives under investigation.

[88] For *tower*, the OED lists both the disyllabic (/ˈtaʊ.ə/) and monosyllabic (/taʊə/) options.

[89] Despite the rules that apply to the syllabification of English words (cf. Jones 2006: xiii), there is still a considerable degree of subjectivity inherent in any attempt to break a word down into syllables.

growing frequency, both groups develop a general trend from moderate alternation (although inclined towards *-est*) to (near-)categorical syntheticity. The group of monosyllables with "other" final segments further includes the five adjectives with the highest superlative frequency in the data, namely (in descending order) *late*, *big*, *great*, *large* and *high*.

Zooming in on individual types in these subgroups, it is interesting to examine whether effects of segmental *horror aequi* can be identified. The data comprise four monosyllabic adjectives ending in the /st/ cluster: Expectedly, the high-frequency type *fast* ($prop_{most}$ = 0.00, n_{sup} = 5,486) shows a categorical preference for *-est* and thus appears to be largely resistant to the hypothesized identity avoidance effect. As for the remaining three types, namely *just* (0.88; 57), *chaste* (0.75; 23) and *moist* (0.35; 12),[90] the former two predominantly take *most* while the latter seems moderately inclined towards *-est*. Although the relatively robust strategy preferences of *just* and *chaste* may at least in part be due to segmental *horror aequi*, we should not overestimate the strength of this effect. On phonological grounds, there is little reason to assume that *most* is significantly less objectionable than *-est*: By hypothesis, the urge to avoid identical segments decreases with the size of the buffer that intervenes between these segments. In the case of monosyllables, the buffer does not substantially increase in size if *most* is chosen (compare, for instance, *justest* /ˈdʒʌ**stɪst**/ and *most just* /ˈməʊ**st** ˈdʒʌ**st**/). Further, on the suprasegmental level, the analytic superlative inevitably comes with a stress clash while the synthetic superlative produces a sequence of unstressed and stressed syllables; since the former scenario constitutes an infringement of the ideal alternating rhythm, we could likewise expect a higher proportion of *-est*. Another factor that must be considered when comparing individual types – especially in the heterogeneous final segment categories "cluster" and "other" – is etymology. While *fast* is of Germanic origin, the OED indicates that *just*, *chaste* and *moist* were borrowed from French. It appears reasonable to suggest that these adjectives may have inherited their gradation strategy preferences from the respective etymons. From a usage-based perspective, the observation that *fast* is categorically synthetic and seems completely resistant to effects of segmental identity avoidance can ulti-

[90] Strikingly, my results for *moist* are at odds with those provided by Mondorf (2009b: 26): In her data, notably on the *superlative* of selected disyllabic adjectives, 4 of 5 tokens are analytic.

mately be ascribed to its high superlative frequency: In the associative exemplar network of adjectives (cf. Section 4.3), *fast* has gained a high degree of autonomy and is very weakly connected to phonologically similar types of low and intermediate frequency (such as *just*, *chaste* and *moist*). Further, with a superlative frequency of 7.08 pmw, this adjective scores above the threshold for whole-word storage proposed by Alegre & Gordon (1999), suggesting that *fastest* is holistically retrieved from the mental lexicon rather than morphologically decomposed and reassembled on demand.

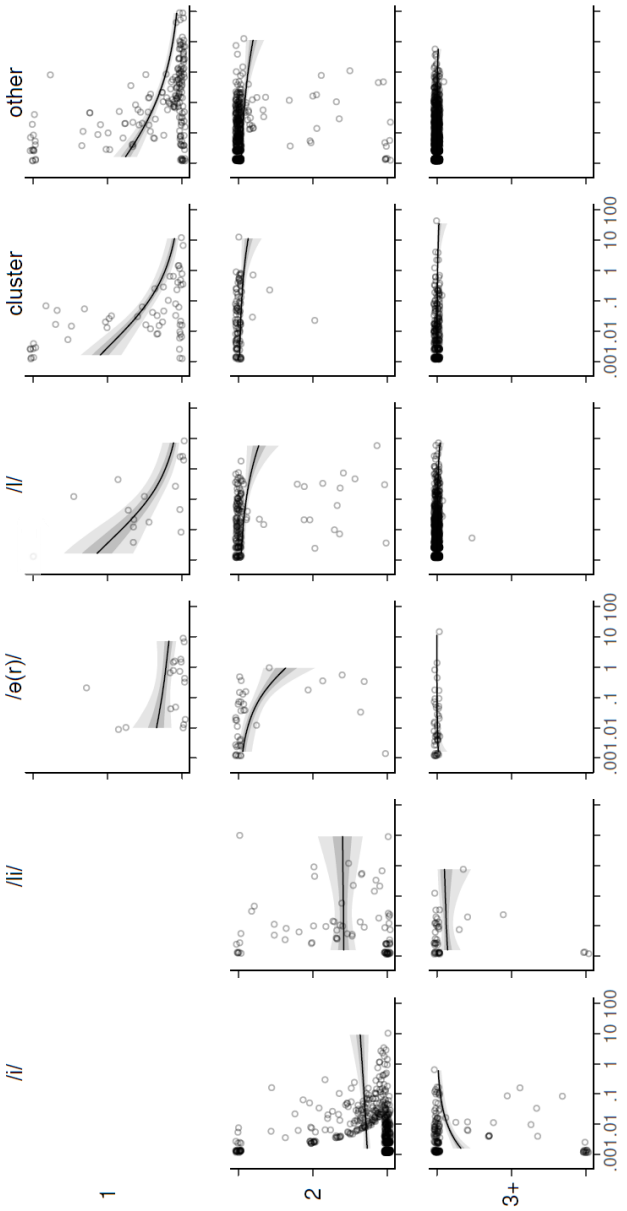

Figure 21. Superlative frequency trends for phonological subgroups. Error bands visualize the 50 % and 90 % uncertainty estimates.

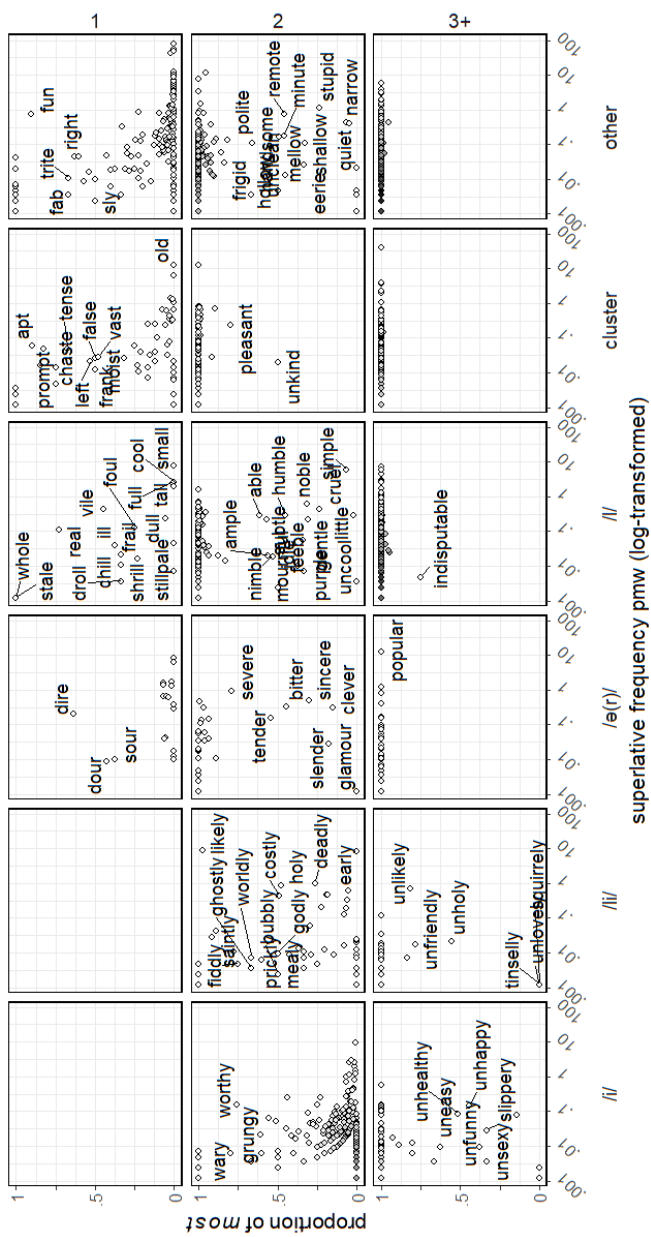

Figure 22. Adjective types in the individual phonological subgroups

Disyllabic types

As for disyllabic adjectives, we observe two fundamentally different superlative frequency trends across the phonological subgroups: Types ending in /-i/ and /-li/ appear to prefer the *-est*-superlative in all frequency ranges; this tendency is more marked in the /-i/-category. Notably, however, there is a great deal of variation in these two categories, particularly in the bands of low and intermediate frequency (below 0.1 pmw). In the remaining four groups of final segments, we identify a (near-)categorical preference for *most* across the entire frequency spectrum, although the observed categoricity seems to decrease slightly in the intermediate- and high-frequency ranges; this trend is most pronounced in the group of disyllables ending in /-ə(r)/, /-ɪə(r)/ and /-ɪr/, which includes alternating types with frequencies between 0.1 and 1 pmw such as *clever*, *sincere* and *bitter*.

Our investigation of the gradation frequency trends of individual phonological subgroups starts with the disyllabic types that feature most prominently in the literature on gradation strategy choice, namely those ending in /-i/ and /-li/. Importantly, these two groups differ from the others in that their members do not only have the same final segment but also share a common derivational suffix: We can thus treat /-i/- and /-li/-disyllables both as phonological and *morphological* subgroups. A comparison of the gradation frequency trends in Figure 21 exposes differences between the adjectival categories in focus: First, the preference for *-est* – notably consistent in all frequency ranges – is stronger among /-i/-disyllables. In this group, we find categorical *-est* across the entire frequency spectrum, with *easy* ($prop_{most} = 0.00$, 10 pmw) being the most frequent type, while for /-li/-disyllables, categorical syntheticity is limited to superlative frequencies below 0.03 pmw, with the exception of *early* (9 pmw) In the intermediate and high frequency ranges, /-li/-disyllables exhibit considerable variation (e.g. *holy*, *deadly* and *lovely*), although they appear to favor the synthetic superlative. Three outliers can be identified: *Early* ($prop_{most} = 0.00$, 9 pmw) and *likely* (0.98, 10 pmw) have the highest frequency scores in this subgroup and exhibit widely divergent strategy preferences. The near-categorical analyticity value of the latter adjective can be ascribed to the observation that *likely* comes with an infinitival complement in the vast majority of cases; this contextual constraint has been shown to significantly increase the odds of the analytic comparative (cf.

Mondorf 2009b: 57-78, Cheung & Zhang 2016: 573-574) and superlative (cf. the questionnaire-based findings presented in Section 3.3). The relatively high analyticity value of *costly* (0.49, 0.48 pmw) can be taken to reflect the urge to avoid segmental identity in ***costliest***; judging by the general superlative frequency trend for /-li/-disyllables, we would have expected this adjective to be strongly inclined towards *-est*. Among /-i/-disyllables, the exceptionally high analyticity of *worthy* (0.76, 0.19 pmw) can likewise be linked to the observation that especially when used predicatively or postnominally, this adjective customarily comes with a prepositional or infinitival complement, thus creating an environment which may prompt *more-*/*most*-support (cf. Mondorf 2009b: 77-78).

Several studies of the comparative alternation have provided empirical support for the claim that while /-i/-disyllables exhibit a preference for the synthetic variant, /-li/-disyllables favor the analytic variant (Hilpert 2008: 407, 409; Cheung & Zhang 2016: 573; Chua 2018: 466, *inter alia*). Leech & Culpeper (1997: 359-360), by contrast, report that in their data, /-li/-disyllables are more analytic than /-i/-disyllables but still solidly biased towards the synthetic comparative in absolute terms. They further note that this observation may be due to the influence of a single very frequent type in their data, namely *likely*, which shows a near-categorical preference for *most*. The general trends that can be identified for /-i/- and /-li/-disyllables in the current data (cf. Figure 21) converge by and large with those presented by Leech & Culpeper (1997: 359-360) for the comparative. Especially in the low and intermediate ranges of superlative frequency, /-li/-disyllables exhibit considerable variation; unexpectedly, however, the trend line for this subgroup does not approach categorical *-est* in the high-frequency band (i.e. above approximately 1 pmw). Similar to the disproportionate influence of *likely* reported in Leech & Culpeper (1997: 359-360), this can be attributed to the observation that the category of /-li/-disyllables comprises only two adjective types in this frequency range, namely *early* (0.00, 9.34 pmw) and *likely* (0.98, 9.73 pmw), as I have noted in the previous paragraph. Since the estimates in the high-frequency range of this subgroup are greatly influenced by the idiosyncratic behavior of these two types (notably with diametrically opposed strategy preferences), caution must be exercised when interpreting the predicted trend in this frequency region, as indicated by the wide uncertainty intervals in

Figure 21. In Section 4.7, I will revisit the key question raised at the beginning of this paragraph: How can the slightly diverging gradation strategy preferences of /-i/- and /-li/-disyllables be accounted for? In addressing this question, I will pursue a usage-based line of argument, zooming in on differences in the productivity of the phonological subgroups under scrutiny. The current dataset includes seven times as many /-i/-disyllables (n = 574) as /-li/-disyllables (n = 82).

Disyllables ending in /-ə(r)/ (n = 43) exhibit a (near-)categorical preference for *most* in the ranges of low and very low superlative frequency (between 0.001 and 0.1 pmw). This general bias towards the analytic variant ties in with the findings of previous studies of the comparative alternation (Mondorf 2003: 281, 2009b: 27-28; Sönning & Hartmann 2019, *inter alia*). As noted earlier, we can interpret this superlative trend as a parallel to the comparative, in which the strong preference for *more* is, by hypothesis, due to segmental *horror aequi* effects, i.e. a reluctance to render two identical sound sequences in immediate adjacency (e.g. bitt**erer** /-ərə(r)/). It will be interesting to scrutinize in Section 4.6.5, however, whether the impact of this analogy is as sizeable as we would assume. In the band of intermediate frequency (between 0.1 and 1 pmw), we can identify five types with alternating strategy preferences: *severe* (0.79), *sincere* (0.32), *tender* (0.54), *clever* (0.15) and *bitter* (0.45). These types are rather heterogeneous with regard to two somewhat interdependent determinants that have not been quantitatively operationalized in the present study, namely etymology and lexical stress; these factors therefore appear of little use for explaining the special behavior of the five adjectives in focus, although the data lend support to the assumption that final stress – the predominant stress pattern among disyllabic adjectives of Romance origin – generally increases the odds of the analytic variant (cf. Hilpert 2008: 409), e.g. *obscure* (0.94), *mature* (0.98), *secure* (0.96), *austere* (1.00). The disproportionally high share of *-est*-tokens of the five outliers under scrutiny may at least in part be due to the observation by Leech & Culpeper (1997: 368-369) that the synthetic superlative "can have a particular affective force". Reconsidering the collocations in which these adjectives occur in our data, we indeed find moderate support for this hypothesized pragmatic aspect: *Sincerest*, for instance, frequently collocates with nouns denoting emotion, e.g. *condolences*, *apologies*, *thanks* and *wishes*. Further, this form can be thought of as invariant in the idiom "Imitation is the sincerest

form of flattery." Strikingly, this idiom accounts for one in five superlative tokens of *sincere* in the current dataset. The only disyllable in /-ə(r)/ with a categorical preference for *-est* is the adjectival ad hoc formation *glamour*, for which the data contain a single token:

(62) Roger Moore was the most elegant, suave, self confidence [sic], relaxed, finest, **glamourest** and best Bond ever. [GloWbE US 349909]

Among disyllables ending in /-l/ (n = 198), the vast majority of adjectives that participate in the (near-)categorical preference for *most* contain one of three suffixes, namely *-al* (e.g. *vital*, *crucial*, *vocal*), *-ile* (e.g. *servile*, *hostile*, *docile*) and *-ful* (*fearful*, *thankful*, *thoughtful*). It is evident from Bauer, Lieber & Plag (2013: 293) that *-al* represents a non-native adjective-forming affix attaching exclusively to bound bases of Latin origin. These uses of *-al* have never been productive in the history of English and must be distinguished from cases such as *behavioral* or *traditional*, in which the suffix is productively used to derive an adjective from a free lexeme in present-day English ({*tradition*} + {*al*} but {*crucial*}; see Bauer, Lieber & Plag 2013: 291, 293). A similar caveat applies to *-ile*: This suffix reflects the Latin third declension adjectival suffix *-ilis* (*docilis*, *hostilis*), which marks deverbal and denominal adjective formations in Latin; in English, however, adjectives ending in *-ile* are strictly monomorphemic and were borrowed as such from Latin or French. By contrast, the native, i.e. Germanic suffix *-ful* is productive in present-day English but limited to "bases that denote psychological states or processes" (Bauer, Lieber & Plag 2013: 305). Schmid (2005: 170) rightly points out that from a cognitive perspective, denominal adjectives derived by *-ful* can be regarded as shading into compounds involving the free lexical morpheme *full*. The observation that vowel reduction (/fʊl/ → /-fəl/) occurs when *-ful* combines with a free morpheme, however, provides a convincing argument in favor of the view that the element under scrutiny has completed the transition to a derivational suffix. Mondorf's (2009b: 137) diachronic data suggest that although the synthetic comparative of disyllables in *-ful* did not fall out of use until approximately 1900, analytic gradation has always been the default for this morphological subgroup. The observation that these types

exhibit a categorical preference for *most* in our data further provides support for the view that *-ful* has the status of a derivational suffix in present-day English: If language users (still) treated adjectives such as *fearful* as compounds, we should expect a greater proportion of synthetic superlatives for these types by analogy to *full* (cf. Mondorf 2009b: 36), which is categorically synthetic in the current data. Interestingly, the present dataset contains a single synthetic superlative token of *mournful* ($prop_{most} = 0.5$; $n_{sup} = 2$), which may at least partly be driven by the urge to maintain structural consistency across a sequence of coordinated superlatives:

(63) We call those ages in which he gets so low the **mournfulest**, sickest and meanest of all ages. [GloWbE US 612009]

15 types deviate from the general analytic trend we have identified for this phonological subgroup: The morphophonological schemas instantiated by adjectives such as *uncool* ($prop_{most} = 0.00$; $n_{sup} = 3$) and *unkind* (0.50; 16), which does not participate in the solidly analytic tendency of disyllables ending in a consonant cluster, will be analyzed separately in Section 4.6.3. The alternating behavior of *cruel* ($prop_{most} = 0.09$; $n_{sup} = 1{,}189$) has received ample attention in the literature: Quirk et al. (1985: 462) list this type among those for which the use of analytic gradation appears to have increased in the recent history of English. In Peters (2000: 308-309), *cruel* is one of seven "crossover" adjectives, i.e. types that prefer *more* in the comparative but *-est* in the superlative. In Sönning & Hartmann's (2019) comparative data, this adjective is indeed more analytic than in my superlative data ($prop_{more} = 0.4$; $n_{comp} = 20$) but still predominantly synthetic. A phonological detail may have explanatory power for the observed alternation of *cruel*: The OED provides both /kruːl/ and /ˈkruːəl/ as variants (notably irrespective of regional variety), thus suggesting that this type can be mono- and disyllabic. The exemplar network proposed in Section 4.3 warrants the hypothesis that morphophonologically similar adjectives exhibit the same gradation strategy preferences. In the language users' experience-based exemplar networks, *cruel* can thus be expected to have formed connections of varying strengths to clusters of mono- and disyllabic adjectives in /-l/. Given that these two phonological subgroups appear to differ in terms of their general strategy preferences

(cf. Figure 21), we can reasonably assume that the variable cluster membership of this adjective results in alternating superlative strategy choices.

The remaining 16 outliers in the group of disyllables ending in /-l/ have a common phonological structure:

(consonant +) front/central vowel/diphthong (+ nasal) + plosive + /ə/ + /l/

In the data at hand, this schema is instantiated by adjectives such as *humble* ($prop_{most}$ = 0.48; n_{sup} = 250), *gentle* (0.32; 187), *noble* (0.31; 545) and *subtle* (0.58; 189). Most of these types score in the ranges of low and intermediate frequency (between 0.01 and 1 pmw); these are the frequency ranges in which we would expect morphophonological clusters to be formed on the basis of lexical connections in the network of exemplars (cf. Sections 4.2 and 4.3). Our data also include types that instantiate the schema specified above but exclusively use the *most*-superlative (cf. Figure 23), thereby conforming with the overall trend observed for disyllables in /-l/. Notably, however, these types score in the range of very low frequency, namely *brittle* (1.00; 3) and *supple* (1.00; 2).[91] These adjectives have gained only low lexical strength and have established few and weak lexical connections in our usage-based exemplar network; their categorical bias towards *most* can be taken either as a reflex to default to the globally more productive strategy or as a tendency to comply with the general gradation strategy preferences of disyllables in /-l/.

[91] Further, note the categorical behavior of *stable* (0.99; 287), which scores in the range of intermediate frequency (0.37 pmw). Strikingly, this adjective shows no signs of sensitivity to phonological *horror aequi* in the analytic superlative (*most stable*).

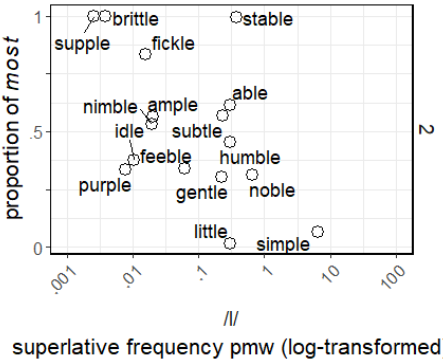

Figure 23. Proportions of *most* for disyllables ending in plosive + /ə/ + /l/ (n = 16)

It is still unclear, however, why types pertaining to this schema appear to be more amenable to the synthetic superlative than most other types in the subgroup under scrutiny. As with *cruel*, my explanation focusses on the phonology of these adjectives, more specifically on the length and syllabic structure of the synthetic superlative. While the types in question are disyllabic in the positive form, the situation is less clear when we attach *-est*: The CEPD (Jones et al. 2006) is remarkably inconsistent in transcribing the synthetic superlative of these adjectives, suggesting that their *-est*-variants can be considered di- or trisyllabic: For eight of the 16 types in focus (*gentle*, *subtle*, *noble*, *nimble*, *feeble*, *purple*, *able* and *fickle*), both variants are listed (e.g. /ˈdʒen.tᵊl.ɪst/ and /ˈdʒent.lɪst/); for *humble*, *simple*, *ample* and *idle*, the OED commits to the disyllabic option. *Little* is the only type in this cluster for which the synthetic superlative is exclusively trisyllabic according to the OED (/ˈlɪt.ᵊl.ɪst/). It is apparent that the syllabification of the forms in focus depends on the extent to which /ə/ is elided between the plosive and /l/. Ryu & Hong (2013) examine the effect of three linguistic factors on schwa deletion in conversational speech, namely stress environment, sonority distance and lexical frequency. Their findings for these variables indicate that

(i) the odds of schwa deletion are higher in post-stress than in pre-stress position,

(ii) the greater the difference in sonority between the two consonants that appear before and after the schwa, the higher the odds of schwa deletion, and

(iii) schwa is more likely deleted in words of high as compared to low lexical frequency.

Within all adjectives instantiating the schema under scrutiny, schwa occurs in post-stress position. In Ryu & Hong (2013: 323), sonority distance is measured on a five-point scale from +2 (greatest) to -2 (smallest), with the distance between the consonants in focus, i.e. between a plosive and the lateral /l/, assigned the highest possible value on this scale. As for lexical frequency, we have observed that most of the alternating types pertaining to this group are considerably more frequent than the very rare and categorically analytic types *brittle* ($prop_{most}$ = 1.00; n_{sup} = 3) and *supple* (1.00; 2). We can reasonably assume that speakers tend to retain schwa in the synthetic superlative of these very infrequent types, thus treating their *-est*-forms as trisyllabic; since schwa deletion can be taken to be more common in the synthetic superlative of the alternating disyllables ending in plosive (+ /ə/) + /l/, forms such as *subtlest*, *noblest* and *gentlest* are presumably considered disyllabic by speakers. This hypothesized feature puts the types in focus on a par with monosyllables in /-l/, a group for which my data suggest a strong preference for *-est*.

For disyllabic types ending in a consonant cluster (n = 96), the data at hand suggest a solidly categorical preference for *most* in all frequency ranges. This overall trend ties in with Mondorf's (2009b: 32) findings for the comparative. It is worthy of note that the great majority of types in this phonological subgroup is of Latinate origin. This group includes 16 types which, due to their ending in /-st/, can be assumed to prefer *most* for reasons of segmental *horror aequi*; these types can be divided into two cohorts, namely monomorphemic disyllables of Latinate origin (e.g. *modest*, *robust*, *honest*) and types such as *sexist*, *racist* and *leftist*, which should be considered denominal noun formations converted to adjectives. While we find a sizeable number of types ending in a voiceless plosive + /t/ (e.g. *abstract*, *compact*, *intact*, *inept*), the most frequent final consonant cluster among disyllabic adjectives is /-nt/. Orthographically, the ending of these types is realized either as <-ent> (urgent, fervent, silent) or <-ant> (*flagrant*, *poignant*, *vibrant*).

In the intermediate frequency range of this phonological subgroup, we find two types with a strong but not categorical bias towards *most*, namely *pleasant* ($prop_{most}$ = 0.80; n_{sup} = 190) and *profound* (0.90; 592), both of which were examined in the questionnaire study presented in Chapter 3. The special behavior of *pleasant* is also apparent in Mondorf's (2009b: 152-153) historical data on the comparative: In her data from the period between 1700 and 1900, three out of four comparative forms of *pleasant* are analytic while this adjective exclusively takes *more* in her 20th-century corpus material. A close qualitative look at the superlative tokens for *pleasant* and *profound* in my GloWbE dataset suggests that the synthetic variant of these types predominantly occurs in contexts that warrant an absolute interpretation of the superlative (cf. Section 2.2), i.e. a superlative without a qualifying scope of reference. Example (64) represents one such case involving *profoundest*:

(64) [...] I found, in the words of this great man, whom I, for one, must always listen to with the **profoundest** respect, a sentence passed on the critic's business, which seems to justify every possible disparagement of it. [GloWbE GB 320526]

In these uses, *profoundest* and *pleasantest* frequently collocate with nouns denoting attitude or emotion such as *profoundest respect*, *contempt* and *sympathy*, and *pleasantest surprise*, *sensations* and *feelings*. These collocational preferences lend additional support to Leech & Culpeper's (1997: 368-369) hypothesis that the synthetic superlative is semantically associated with an affective and emotive force. Further, the *-est*-superlative of the two types in focus can be observed to instantiate the Superlative-*of*-N construction, in which the superlative occurs in its absolute sense (Section 2.2) and likewise has an affective charge to it, as in the following example:

(65) He sang a little song or two, of the sort he used to sing at his dinner-parties, curled himself up in the straw, and had an excellent night's rest and **the pleasantest of dreams**.
[GloWbE GB 69955]

In analyzing the large category of disyllables whose final segments were classed as "other" (n = 449), it will again be worthwhile to zoom in

on even smaller morphophonological subgroups. We will focus on types of non-native, i.e. Latinate origin before turning to types of native, i.e. Germanic origin.

Disyllables ending in *-ic* (n = 41, e.g. *poetic*, *frantic*, *hectic*) and *-ous* (n = 33, e.g. *anxious*, *callous*, *vicious*) are categorically analytic regardless of superlative frequency; these types are Latinate in origin, with *-ic* representing the French suffix *-ique* (and thus ultimately the Latin etymon *-ic*, as in *poetic-us*) and *-ous* reflecting the Latin derivational suffix *-ōsus* ('abounding in, full of, characterized by, of the nature of', as the OED notes). As for the latter group, the looming clash of two sibilants in the synthetic form (as in *famou**s**e**s**t* /ˈfeɪmə**s**ɪ**s**t/) is possibly an additional factor promoting the use of *most* (cf. Kytö & Romaine 1997: 342).

A similarly categorical preference for the analytic superlative can be observed for disyllables in *-id* (n = 35, e.g. *sordid*, *splendid*, *valid*), which mirrors the Latin derivational suffix *-idus*. While the data at hand contain rare and idiosyncratic examples of the synthetic variant for selected types such as *frigid* ($prop_{most}$ = 0.67; n_{sup} = 3), *vivid* (0.99; 207) and *horrid* (0.94; 36), the only type that shows a systematic bias towards *-est* is *stupid* (0.24; 943). Strikingly, with a superlative frequency of 1.22 pmw, this adjective is significantly more frequent in the superlative than the other adjectives in this subgroup and can thus be assumed to have gained lexical autonomy from the cluster of disyllables in *-id* in the network model of adjective exemplars proposed in Section 4.3. The preponderance of *stupidest* in my data can at least in part be ascribed to phonological *horror aequi* effects, i.e. the urge to avoid the repetition of /st/ in immediate adjacency (*mo**st** **st**upid*).

A moderate degree of variation is observed among the subgroup of oxytonic (i.e. finally-stressed) disyllables ending in <-te> (n = 16, e.g. *irate*, *remote*; cf. Figure 24), which reflects the Latin derivational and inflectional suffix *-tus* (e.g. *iratus*, *remotus*). In the range of very low superlative frequency, these types are categorically analytic, e.g. *dilute* (1.00; 1), *effete* (1.00; 1), *irate* (1.00; 5) and *ornate* (1.00; 13). While this trend still appears relatively solid for types of low and intermediate frequency, as indicated by *concrete* (1.00; 52), *astute* (0.98; 64) and *acute* (0.89; 121), we identify three types with a sizeable degree of alternation, namely *polite* (0.66; 89), *remote* (0.46; 604) and *minute* (0.46; 146). Interestingly, the former two adjectives are considered crossover adjectives in Peters (2000: 308-309),

i.e. there is evidence to suggest that they prefer *more* in the comparative but *-est* in the superlative. Although *polite* is still predominantly analytic in the current superlative data, the comparative data presented by Sönning & Hartmann (2019) tentatively support the crossover hypothesis for the two adjectives in question, yielding analyticity values of $prop_{more} = 0.85$ for *polite* ($n_{comp} = 27$) and $prop_{more} = 0.68$ for *remote* ($n_{comp} = 243$). Strikingly, an even greater discrepancy between the analyticity scores in the comparative and superlative can be observed for *minute* ($prop_{most} = 0.46$; $n_{sup} = 146$; $prop_{more} = 1.00$; $n_{comp} = 3$) despite the low comparative token count in the BNC. In her attempt to account for crossover types, Peters' (2000: 309-310) draws on a variable that was not controlled in the current analysis, namely *position*. If we accept that (i) attributive use increases the probability of the synthetic form (Mondorf 2009b: 80-90) and (ii) attribution is the primary function of the superlative (Claridge 2007a), it appears legitimate to assume that positional differences may underlie the diverging strategy trends that have been identified for the comparative and superlative of the three adjectives in focus. By hypothesis, collocational preferences play a key role, too: 43 of the 83 tokens in the present corpus render *minutest* in the collocation *minutest detail*(*s*).

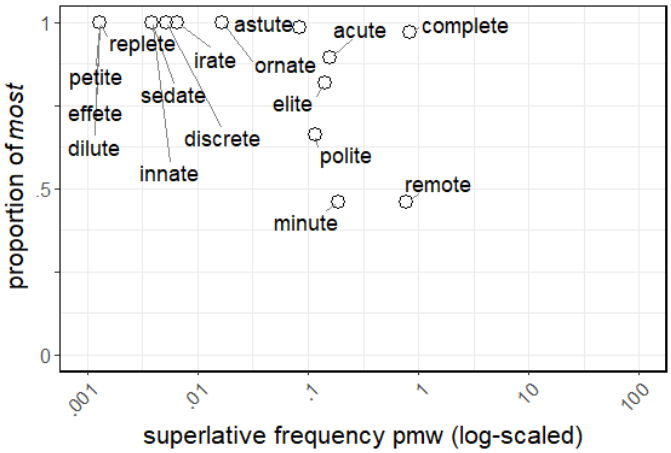

Figure 24. Proportions of *most* for oxytonic disyllables in <-te> (n = 16)

In line with Mondorf's (2009b: 149-150, 153-154) comparative data, disyllables containing the native English suffixes *-less* (n = 63, e.g. *reckless*, *pointless*, *needless*) and *-ish* (n = 49, e.g. *fiendish*, *hawkish*, *stylish*) exclusively take *most* irrespective of superlative frequency. It is only for *foolish* ($prop_{most}$ = 0.98; n_{sup} = 54) that we record a single *-est*-token.

Disyllables ending in <-some> (n = 16) are solidly analytic in the current dataset (cf. the top panel in Figure 25). In the range of very low frequency, we identify a single type with deviant behavior, namely *lonesome* (0.00; 1). In the ranges of low and intermediate frequency, some types such as *irksome* (0.91; 11), *wholesome* (0.95; 22) and *awesome* (0.92; 315) occasionally allow the *-est*-superlative but take *most* in the vast majority of cases. The only adjective in this subgroup to display alternating behavior is *handsome* (0.49; 136). The historical data presented by Mondorf (2009b: 145-146) suggest that this adjective predominantly formed the synthetic comparative in the period from 1500 to 1900 but exhibited a stable preference for *more* in the 20th century. This tendency is also borne out by Sönning & Hartmann's (2019) BNC data, in which three out of four comparative tokens of *handsome* involve *more* ($prop_{more}$ = 0.76; n_{comp} = 29). Varietal differences seem to account for some of the variation we observe for this adjective: A comparison of the varietal subsections of the current dataset implies that *handsome* has a greater proportion of *most* in GloWbE GB ($prop_{most}$ = 0.56; n_{sup} = 62) as compared to GloWbE US (0.45; 74). The analyticity values obtained from the questionnaire data presented in Section 3.3.2 exhibit a similar varietal discrepancy, although both American ($prop_{most}$ = 0.66) and British respondents (0.83) generally appear to be biased towards *most handsome*.

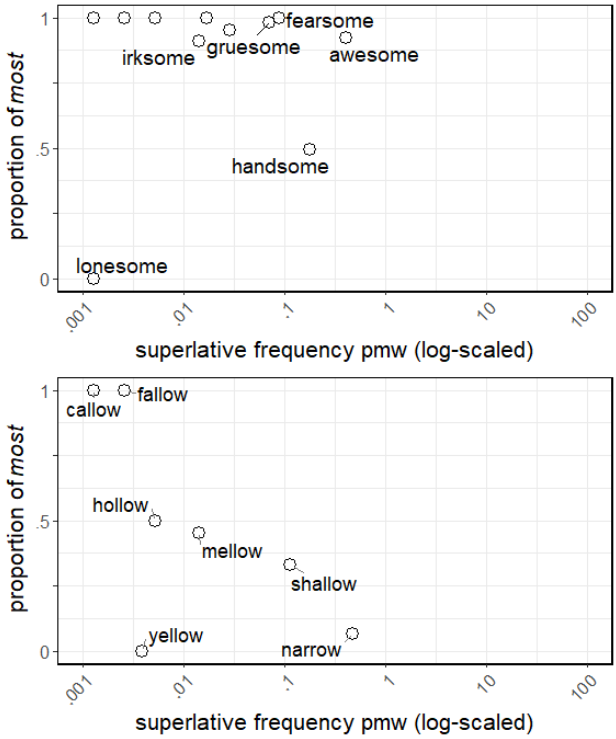

Figure 25. Proportions of *most* for disyllables in <-some> (top, n = 16) and <-ow> (bottom, n = 7)

Peters (2000) is, to my knowledge, the only study to have treated disyllables ending in <-ow> (/əʊ/) as a separate subcategory. The bottom panel in Figure 25 hints at a sizeable degree of variation among this group of n = 7 types. In the range of very low frequency, we identify types which are either categorically analytic, such as *callow* (1.00; 1) and *fallow* (1.00; 2), or synthetic, such as *yellow* (0.00; 3), with *hollow* (0.5; 4) alternating between the two superlative variants. With increasing superlative frequency, this group appears to show a gradual trend towards *-est*. Since my data contain only three disyllables in <-ow> that score in the bands of low and intermediate frequency, namely *mellow* (0.45; 11), *shallow* (0.33; 88) and *narrow* (0.07; 367), this observation is extremely tentative. By and large, the analyticity values of these three types converge with those pre-

sented by Peters (2000: 304) for the comparative, suggesting that the vocalic final segment /-əʊ/ readily takes the degree-marking suffixes *-er* and *-est*.

The synthetic bias observed for *eerie* (0.33; 21) and *bonnie* (0.00; 4) is possibly due to the fact that these two types share a common final segment with disyllables ending in /-i/, a subgroup that has been shown to strongly favor *-est* (cf. Figure 21).[92]

Polysyllabic (i.e. tri(+)syllabic) types

Adjectives with three or more syllables virtually exclusively take *most* (cf. Figure 21). Systematic deviations from this trend can merely be found among *un*-prefixed trisyllables ending in /-i/ and /-li/, two subgroups that will be dealt with in more detail in Section 4.6.3. In the ranges of very low frequency, selected polysyllabic types ending in /-i/ are attested with *-est* but are solidly analytic on the whole, e.g. *shadowy* (0.80; 5) and *touristy* (0.93; 14). Among polysyllables in /-i/ (n = 144) and /-li/ (n = 45), we further identify *hapax legomena* whose single superlative token is synthetic, e.g. *cartoony* (0.00; 1), *chocolaty* (0.00; 1), *squirrely* (0.00; 1) and *tinselly* (0.00; 1). If we ignore for a moment the special case of *un*-prefixed polysyllables in /-i/ and /-li/, the only type that shows alternating behavior in the range of low superlative frequency is *slippery* (0.33; 24); as discussed in the context of disyllabic adjectives such as *gentle* and *subtle*, we can assume that the degree of schwa elision in the synthetic superlative is related to the syllabification of this form, to the effect that speakers may consider *slipperiest* trisyllabic (/ˈslɪp.ᵊri.ɪst/).

In the remaining final segment categories, the synthetic superlative is extremely rare and idiosyncratic. Among polysyllabic types with final /-l/ (n = 840), only twelve are attested with *-est*; interestingly, all these types are suffixed by *-ful* or *-able*, e.g. *reputable* (0.97; 69), *pitiful* (0.96; 23) and *indisputable* (0.75; 4). In all instances of this kind, the superlative has an emphatic component and unambiguously warrants an absolute interpretation, as in the following example:

[92] The decision to assign these two adjectives to the final segment category "other" was motivated in Section 4.5.2.

(66) I find in Johnson's Books the **indisputablest** traces of a great intellect and great heart; -- ever welcome, under what obstructions and perversions soever. [GloWbE US 612009]

My data do not contain a single synthetic superlative token of a polysyllabic type in /-ə(r)/ (n = 46). Only two types in the large category of polysyllables ending in a consonant cluster (n = 250) occur with *-est* at all, namely *innocent* (0.98; 121) and *important* (1.00; 33,696); the superlative tallies for these types include two synthetic tokens each. Finally, among polysyllabic adjectives with "other" final segments (n = 1,102), notably the largest phonological subcategory in the current dataset, five types are attested in the synthetic superlative: Four of them end in *-ous* and comprise a single *-est*-token, namely *virtuous* (0.98; 45), *fabulous* (0.99; 90), *ridiculous* (1.00; 543) and *serious* (1.00; 1,596). Although predominantly analytic, *favorite* (0.95; 348) stands out due to its disproportionately large number of synthetic superlatives (n_{est} = 16). Again, syllabification may be central to our understanding of this observation: Due to a variable degree of schwa elision, *favorite* can be trisyllabic /ˈfeɪ.vər.ɪt/ or disyllabic /ˈfeɪv.rɪt/ according to the CEPD (Jones et al. 2006). In the latter case, this adjective bears phonological resemblance to disyllables ending in /-t/, many of which readily form the *-est*-superlative (e.g. *quiet*).

4.6.3 Prefixation by *un-*

In this section, we will be concerned with a research question that was raised in the previous section and reserved for a separate analysis: How do the superlative strategy preferences of an adjective change if we attach the *un*-prefix to it?

Figure 26 illustrates the proportions of *most* for the n = 196 adjectives that are attested both with and without *un-* in my data, with the analyticity score of the positive,[93] i.e. non-prefixed form shown on the horizontal axis and the corresponding value of the negated, i.e. prefixed form on the ver-

[93] It is important to note that in the current section, the term *positive* does not refer to the ungraded base form of an adjective (i.e. as opposed to the *comparative* and *superlative* degrees) but denotes a non-negated adjective type (e.g. *tidy*) to which there exists a negated counterpart in the data (*un-tidy*).

tical axis. Thus, datapoints that lie above the diagonal line at x = y represent adjective pairs for which the prefixed variant is more analytic than the non-prefixed variant. The reverse applies to datapoints below this line. In Figure 26, the n = 196 datapoints are arranged by the final segment and the number of syllables of the non-negated positive adjective. In the description of results, the analyticity scores and superlative token counts for individual adjective pairs will be provided in the following format: *positive adjective* (pos $prop_{most}$; pos n_{sup} / neg $prop_{most}$; neg n_{sup}).

The current dataset contains n = 15 monosyllabic adjectives that have a negated counterpart formed by the *un*-prefix. Two of these types, namely *cool* (0.00; 2,161 / 0.00; 3) and *fine* (0.00; 9,645 / 0.00; 1), exclusively take *-est* both in the prefixed and non-prefixed form. As for the remaining 13 monosyllables, prefixation by *un-* either constitutes a knockout criterion for *-est*, e.g. *true* (0.10; 717 / 1.00; 6) and *fair* (0.08; 548 / 1.00; 59), or results in alternating strategy choices, as in the cases of *kind* (0.07; 570 / 0.41; 22) and *clean* (0.05; 489 / 0.60; 5).

Two different trends can be observed for disyllabic adjectives (n = 49): Non-prefixed adjectives in the final segment categories *cluster* (n = 1, *pleasant*), /-ə(r)/ (n = 1, *aware*), /-l/ (n = 12) and *other* (n = 12) turn categorically analytic if prefixed by *un-*. Notably, most of these types already have a great proportion of *most* in the positive. As for disyllables in /-i/ (n = 15) and /-li/ (n = 8), by contrast, the data at hand suggest that prefixation by *un-* results in sizeable variation, especially if the respective positives are strongly biased towards *-est*. In these cases, the negated counterparts can be found at both extremes of the analyticity spectrum, e.g. *pretty* (0.01; 699 / 0.00; 1), *lazy* (0.07; 227 / 0.00; 1), *canny* (0.10; 20 / 1.00; 7) and *comfy* (0.21; 53 / 1.00; 1). Strikingly, this categorical behavior is limited to instances in which the negated form only has a single superlative token. Alternating strategy preferences in the negated form are limited to predominantly synthetic positives that score in the ranges of intermediate or high superlative frequency such as *lucky* (0.01; 401 / 0.14; 69), *happy* (0.05; 2,335 / 0.48; 134) and *friendly* (0.21; 418 / 0.74; 19). If a positive type exhibits a preference for *most*, this preferred strategy appears to be the default for its negated counterpart, e.g. *godly* (0.64; 14 / 1.00; 5). At odds with this trend, one in five superlative tokens of *unworthy* (0.80; 10) and *unlikely* (0.82; 615) is synthetic although *worthy* (0.76; 148) and *likely* (0.98; 7,537)

are solidly analytic in my data. We further identify two types that are exclusively synthetic in the negated but not in the positive form, namely *lovely* (0.20; 407 / 0.00; 1) and *crafty* (0.43; 28 / 0.00; 1); however, since the negated proportions of *most* are based on a single superlative token, this should not be taken as indicative of a systematic trend.

Polysyllabic types (n = 131) are analytic across the board if negated by *un-*. This observation is hardly surprising if we consider that except for *slippery*, whose special behavior was discussed in the previous section, all these types also categorically take *most* in the non-negated form.

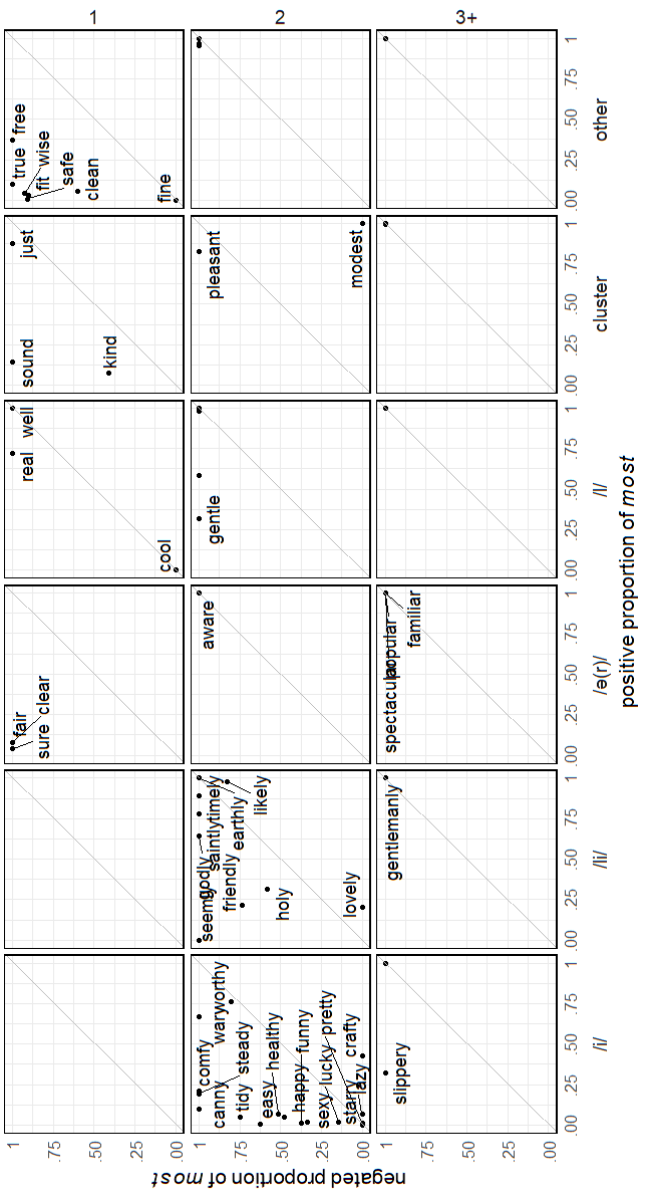

Figure 26. Proportions of *most* for n = 196 adjectives and their *un*-prefixed antonyms in the combined dataset. Diagonal line: negated proportion of *most* = positive proportion of *most*

To account for the patterns observed in *un*-prefixed adjectives, we first turn to Mondorf's (2009b: 37-38) proposal, which emphasizes the role of right-branching structure. Subsequently, the phenomena are examined from a usage-based perspective, highlighting the interplay of (morpho)phonological similarity and superlative frequency.

In cognitive linguistic terms, the morphological structure of adjectives can influence their processing and the preferred direction of affixation. Prefixed adjectives such as *unhappy* are initially left-branching: the prefix *un-* attaches to the base adjective *happy*, requiring the negation to be processed before the lexical meaning of the stem can be fully integrated. However, when these adjectives undergo comparative or superlative formation by suffixation (*unhappier*, *unhappiest*), the derivational operation occurs on the right. Following Mondorf (2009b: 37-38), the presence of a right-attached suffix effectively renders the overall word structure right-branching, which aligns with the incremental, left-to-right processing preferences typical of English. This structural property provides a cognitive explanation for why *un*-prefixed adjectives can more readily form synthetic comparatives and superlatives, partially mitigating the general tendency for longer adjectives to be graded analytically with *more/most*.

From a usage-based perspective, a consideration of two key determinants may add to our understanding of the diverging superlative strategy preferences of antonymic pairs such as *happy/unhappy*, namely (morpho)phonological similarity and superlative frequency.

Phonology and morphology: The morphophonological determinants that are standardly invoked in studies of gradation strategy choice are the final segment and the length of the adjective (measured in terms of the number of syllables). It is apparent that while prefixation by *un-* increases the length of the positive type, the final segment remains unaffected. With this in mind, let us briefly consider three scenarios for the exemplary adjective pair *happy/unhappy*: First, if all trisyllabic types defaulted to *most*, as a didactical rule of thumb dictates (i.e. the "trisyllabic-plus rule"; cf. Bauer, Lieber & Plag 2013: 114), our data should not contain a single token of *unhappiest*. Second, if a speaker's superlative choice for *unhappy* were entirely insensitive to the *un*-prefix and exclusively motivated by formal similarity to *happy*, we should find that the gradation strategy preferences of *unhappy* and *happy* are identical. Neither scenario is fully supported by

the data: *happy* (0.05; 2,335) is near-categorically synthetic, *unhappy* (0.48; 134) exhibits alternating behavior. The third scenario that shall be discussed can be considered a hybrid of the first two. A usage-based view of gradation strategy choice proposes that the degree to which the strategy preferences of adjectives converge is correlated with their morphophonological similarity. It is obvious, therefore, that in a network of phonological and semantic connections as put forward by Bybee (1985), *unhappy* has strong lexical connections to *happy* and slightly weaker ones to other disyllabic adjectives in /-i/. Since this formal subgroup is largely biased towards *-est*, we would expect *unhappy* to form the synthetic superlative in the vast majority of cases. In addition, however, *unhappy* is connected to other adjectives that are prefixed by *un-*. According to the current data, this prefix is considerably more productive with polysyllabic types (n = 131) than with mono- (n = 15) and disyllabic ones (n = 49). With polysyllabic types being the exclusive domain of *most* (cf. Section 4.6.2), it appears reasonable to assume that the alternating behavior of *unhappy* results from its formal similarity to a fairly large group of disyllables in *-y* on the one hand and an even larger group of polysyllabic types containing the *un*-prefix on the other, although these all have strong similarity with their shorter non-negated counterparts. It is further important to reiterate that morphophonological similarity is a matter of degree and that the formal criteria of length and final segment should be analyzed in combination with a determinant that constitutes the second cornerstone of a usage-based view of gradation strategy choice, namely gradation frequency.

Superlative frequency: In Section 4.3, we discussed the central role of superlative frequency in a usage-based perspective of gradation strategy choice. Along these lines, differences in superlative frequency can be interpreted as reflecting different degrees of lexical strength; this metric is assumed to influence the connectivity of morphophonologically similar types in the proposed network of adjectival exemplars. Types of low superlative frequency have few and weak connections to more frequent types with a similar morphophonological structure. Thus, by hypothesis, for strategy alignment to occur between positive adjectives and their negated counterparts, both types must be sufficiently frequent in the superlative. In terms of superlative frequency, two key observations can be made for the n = 196 types that are attested both with and without *un-* in the current

data. First, for only 15 types (13 polysyllabic and two disyllabic, namely *earthly* and *equal*), the prefixed form has a higher superlative frequency than the respective non-prefixed form. In the vast majority of cases, therefore, it is the positive, i.e. non-negated type that has greater lexical strength as well as more and stronger lexical connections to formally similar adjectives. Second, just under 50 % of negated types have only one (n = 64) or two tokens (n = 31) in the current data. The datapoints for these extremely rare negated types do not hint at a systematic trend; while strategy alignment with the non-negated counterpart can be identified for some types such as *lazy* (0.07; 227 / 0.00; 1) and *lovely* (0.20; 407 / 0.00; 1), divergent preferences are attested for others, e.g. *clear* (0.07; 1,349 / 1.00; 2) and *gentle* (0.32; 187 / 0.00; 1). Although the relatively low superlative token counts of types containing the *un*-prefix in the current dataset do not permit a precise conclusion about the role of frequency in the strategy choice of adjectives prefixed by *un-*, the superlative frequencies of both the negated and the corresponding non-negated type should be taken into account in future studies of this comparison.

Based on a joint investigation of morphophonological factors and superlative frequency, the present section has offered usage-based insights into the superlative strategy trends of adjectives negated by *un-* and their non-negated counterparts. More work is needed to elucidate how exactly the gradation frequencies of the antonyms in focus, or rather the interplay of these metrics, can be related to superlative strategy choice. A study that intends to address this question should also consider that adjective pairs such as *likely* and *unlikely* may also differ on other levels of linguistic analysis such as semantics, syntax and phraseology.

4.6.4 Superlative strategy choice in the GB and US subsections

Our next step is to identify similarities and differences between the main reference varieties BrE and AmE with regard to superlative strategy choice. This comparison relies on a separate analysis of the GB and US subsections of the current GloWbE data, which means that two beta-binomial models were fit, one for each main reference variety under scrutiny. Figure 27 illustrates the superlative frequency trends predicted by the varietal models for GloWbE GB (blue) and GloWbE US (red). The error bands visualize the respective 50 % and 90 % uncertainty estimates. The

predictions are arranged according to number of syllables and final segment.

Across the board, the superlative frequency trends estimated for BrE and AmE are remarkably homogeneous. Slight divergences can merely be observed in the ranges of very low and low frequency for monosyllabic adjectives ending in /-l/ and a consonant cluster; in both subcategories, rarely graded types have marginally higher proportions of *most* in GloWbE US. Since the results come with considerable uncertainty in these frequency ranges, however, this observation should not be taken as indicative of a generally greater bias towards the analytic superlative in AmE. It is hardly surprising that the patterns for the varietal subsections largely converge with those that were estimated on the basis of the combined dataset (cf. Section 4.6.2). These patterns can be described as follows:

- A sizeable degree of variation in superlative strategy choice can be observed for monosyllabic types in the ranges of very low and low superlative frequency. Monosyllables of intermediate and high superlative frequency are solidly synthetic regardless of their final segment.
- Disyllabic types in /-i/ and /-li/ are predominantly synthetic but show alternating behavior in all frequency ranges. Irrespective of frequency, disyllables in the remaining categories of final segment prefer *most*, with selected types ending in /-ə(r)/ and /-l/ exhibiting moderate variation in the ranges of intermediate and high superlative frequency.
- Polysyllabic types are categorically analytic in all frequency ranges and irrespective of final segment. Notable exceptions to this overall trend are polysyllables in /-i/ and /-li/: In these final segment categories, many *un*-prefixed disyllables readily take *-est*, as was shown in Section 4.6.3.

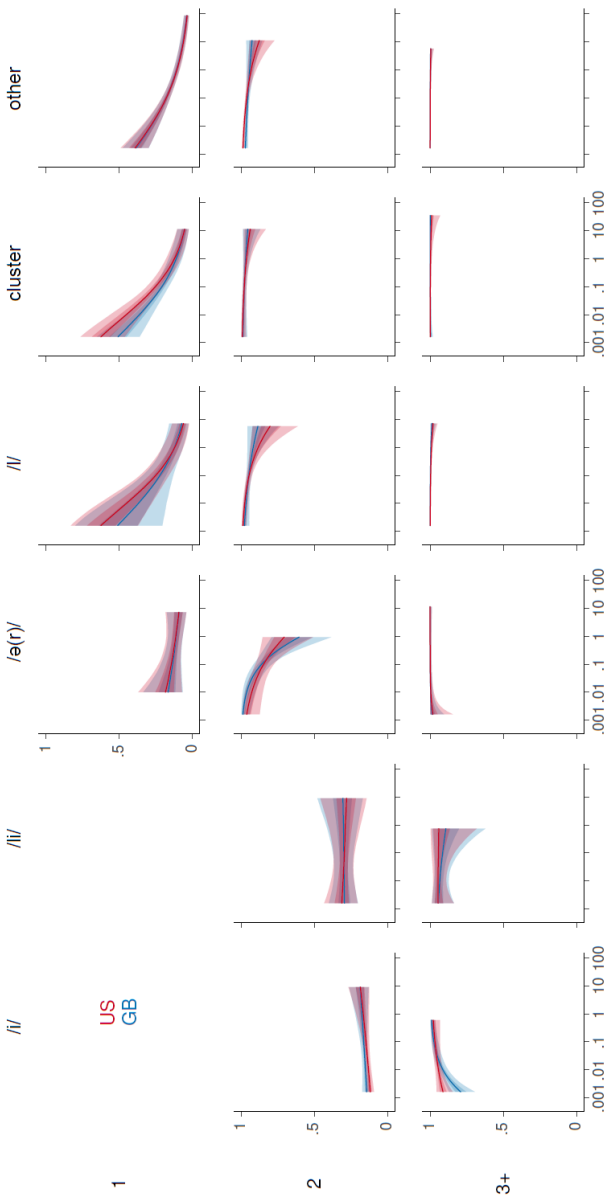

Figure 27. Superlative frequency trends predicted for the US (red) and GB sections of GloWbE (blue). The predictions are arranged according to number of syllables and final segment(s). Error bands visualize the 50 % and 90 % uncertainty estimates.

4.6.5 Comparatives and superlatives

Figure 28 contrasts the superlative frequency trends that emerge from the current analysis (top) with the comparative frequency trends reported by Sönning & Hartmann (2019) (bottom).[94]

The fact that the present study was closely modelled on Sönning & Hartmann (2019), both in terms of the general methodological procedure and the statistical analysis, ensures the comparability of the estimated trends. Nevertheless, interpretational difficulties may arise from the fact that I rely on external evidence as reference points. The BNC material used by Sönning & Hartmann (2019) differs from my GloWbE material in quantitative and qualitative terms. While the present corpus exclusively consists of texts gleaned from the Internet and can be taken to represent both BrE and AmE, the BNC is made up of 90 % written and 10 % spoken material from British sources.[95] It has further been noted that since the current corpus is approximately 7.75 times as large as the BNC, my analysis of the superlative extends further to the left on the gradation frequency spectrum, thereby encompassing very rarely graded types. Importantly, therefore, types with a single token have a comparative frequency of 0.01 pmw in Sönning & Hartmann's (2019) dataset but a superlative frequency of 0.001 pmw in the current one. The overall share of hapax legomena is slightly higher in the comparative (31.5 %) as compared to the superlative data (26.8 %).

[94] Note that in this figure, the frequency trends based on the comparative data slightly deviate from the trends presented in Sönning & Hartmann (2019, slide 22). In contrast to Sönning & Hartmann (2019), who report *subject-specific* estimates, the present analysis presents *population-averaged* estimates. See Sönning (2020: 45) for a summary of the difference between these types of summary measures.

[95] It seems legitimate to object that varietal differences may skew the present analysis since my superlative dataset comprises data from GloWbE GB and US. The results presented in Section 4.6.4, however, suggest only marginal differences between the US and GB sections of GloWbE in terms of the estimated superlative frequency trends. It thus appeared most expedient not to confine the current analysis of the superlative to the trends that can be predicted on the basis of my GloWbE GB data.

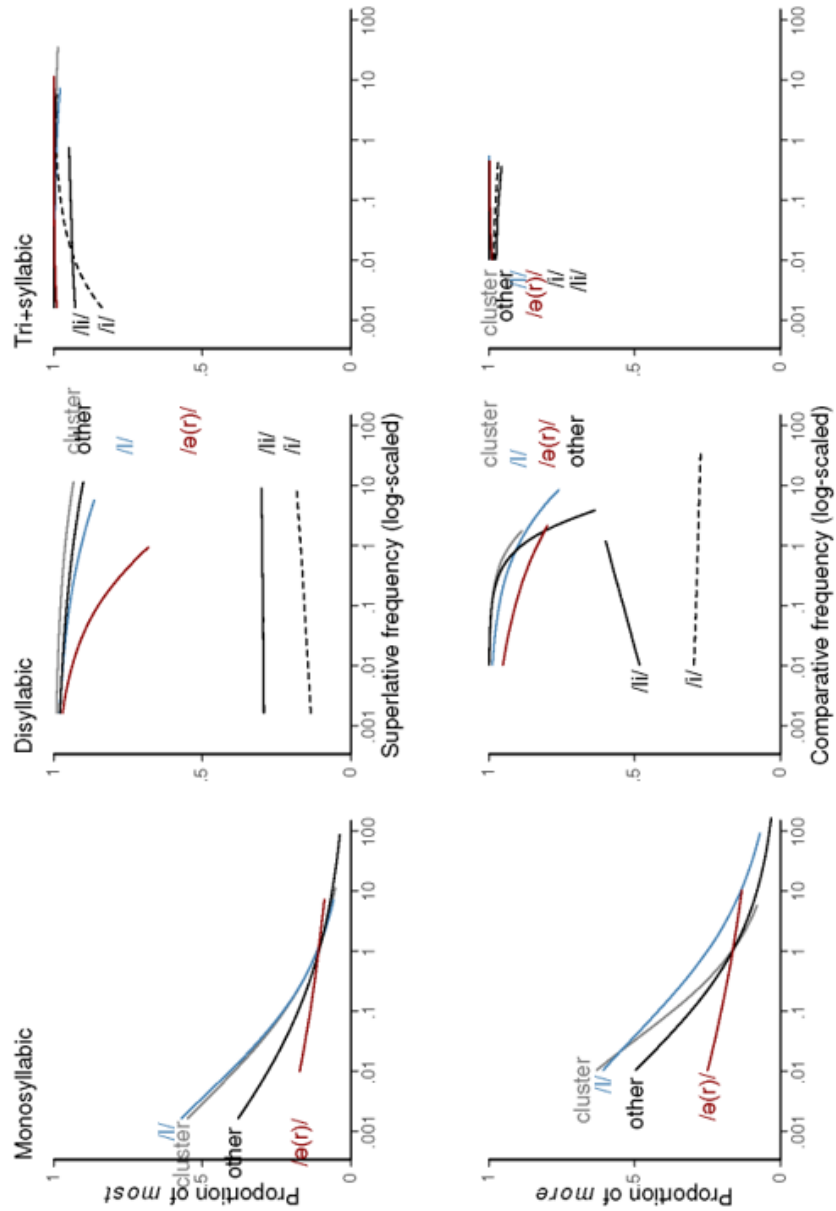

Figure 28. Summary of superlative (top) and comparative frequency trends (bottom) by phonological subgroup (BNC data on the comparative from Sönning & Hartmann 2019)

As for monosyllabic types, similar gradation frequency trends can be identified in the comparative and superlative. Across all final segments under scrutiny, sizeable alternation is attested among rarely and very rarely graded types. In the ranges of intermediate and high gradation frequency, the trend lines exhibit a marked decline in the proportion of analytic forms and eventually converge to categorical *-er* and *-est* for (very) frequently graded types (above a relative frequency of approximately 1 pmw). Despite the overall uniformity of these trends, slight differences between the comparative and superlative are apparent from Figure 28. Comparing the trajectories for individual final segments, we observe that in the range of low gradation frequency (i.e. between .01 and .1 pmw), monosyllables are marginally more analytic in the comparative than in the superlative. However, this observation should not be taken as indicative of a systematic trend: The wide error bands in Figure 21 (Section 4.6.2) reveal the imprecision of the estimates in this range of superlative frequency; this is due to (i) the small number of data points in these frequency ranges, and (ii) the sizeable variation among monosyllabic adjectives in each subcategory of final segment.

It is among disyllables that we find the most substantial divergences between the comparative and superlative frequency trends. While the comparative of disyllables in /-ə(r)/ is solidly analytic irrespective of comparative frequency, the superlative trajectory hints at sizeable variation especially in the range of intermediate superlative frequency. In the right-hand plot of Figure 29, we can identify six disyllables in /-ə(r)/, /-ɪə(r)/ and /-ɪr/ that stand out due to their comparatively low proportion of *most*, namely *slender*, *severe*, *tender*, *bitter*, *sincere* and *clever*. Figure 29 further reveals that apart from *slender*, all these types are more analytic in the comparative than the superlative (compare the data points labelled in red).

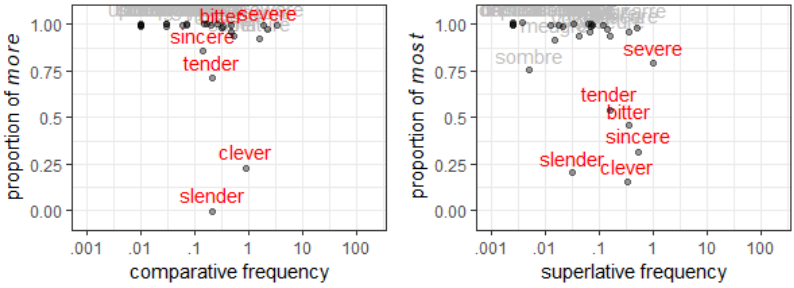

Figure 29. Gradation frequency scores and proportions of *more*/*most* for the n = 28 disyllabic adjectives ending in /-ə(r)/ that are attested both in the current superlative data (right) and Sönning & Hartmann's (2019) comparative data (left)

In line with Mondorf (2009b: 24-30), this observation can be taken to suggest that comparative strategy choice is indeed constrained by effects of segmental *horror aequi*: By hypothesis, language users tend to prefer the analytic comparative of adjectives in /-ə(r)/ due to the adjacency of identical segments in the synthetic competitor (e.g. ***bitterer*** /-ərə(r)/), although this effect is assumedly stronger in AmE than BrE due to a higher degree of rhoticity.

The special behavior of disyllables in /-i/ and /-li/ relative to all other final segment categories is apparent both from the comparative and superlative frequency trends in Figure 28. These types exhibit a moderate bias towards the synthetic variant; both in the comparative and superlative, and irrespective of gradation frequency, this tendency is slightly more pronounced in the /-i/-category. The key question that emerges from this observation is this: Why are disyllables in /-i/ slightly more synthetic than disyllables in /-li/ both in the comparative and superlative? We will return to this issue in Section 4.7.

The comparative and superlative frequency trends of disyllables in the final segment categories /-l/, *cluster* and *other* are largely identical. We observe a virtually categorical preference for the analytic variant in all frequency ranges, however with a notable degree of variation particularly among types in the range of intermediate frequency such as *humble*, *gentle*, *hollow* and *shallow* (cf. Section 4.6.2).

For polysyllabic adjectives ending in /-ə(r)/, /-l/ and final segments that were classed as *other*, the analytic form is the default both in the comparative and superlative. The data appear to suggest, however, that while

polysyllables in /-li/ and /-i/ are near-categorically drawn to *more* irrespective of comparative frequency, some of these types occasionally take *-est* in the range of very low superlative frequency, albeit to a limited degree. The results I have presented in Sections 4.6.2 and 4.6.3 imply that virtually all these types are disyllables prefixed by *un-* (e.g. *unlucky, unhappy, unfriendly*). The observation that this slight trend is borne out by the current superlative data but not by Sönning & Hartmann's (2019) comparative data may be owing to the low number of hits in these low-frequency bands and might thus be a methodological artefact. Future work may be able to account for this perceived inconsistency by investigating the comparative and superlative frequency trends on the basis of identical corpus material.

4.7 Discussion

On the basis of the empirical insights gleaned in Section 4.6, this section discusses the research questions that have previously been formulated. First I will shed light on the gradation strategy trends of disyllables in /-i/ and /-li/, zooming in on differences in type and token frequency between these morphophonological subcategories. Next I will provide a critical assessment of the explanatory power of *more-*/*most*-support with regard to gradation frequency. Adopting a crossvarietal perspective, I will then assess the superlative trends in the data from the US and GB sections of GloWbE. On a similar note, I will address the trends that have become apparent in the comparison of the present superlative data Sönning & Hartmann's (2019) comparative data (cf. Section 4.6.5). Finally, I will explore the implications of the observed patterns for the theoretical assumptions on morphological processing and storage of adjectives that were advanced in Sections 4.1 to 4.3.

Let us first consider a central trend that emerged both from the present superlative data and the comparative data provided by Sönning & Hartmann (2019): Disyllabic types in /-i/ and /-li/ deviate from all other formally defined subcategories of disyllables in that they have a relatively great proportion of synthetic comparative and superlative forms; this preference is slightly more marked in the superlative and among disyllables in /-li/ (as compared to those in /-i/) and remains remarkably stable across the frequency spectrum. Within-group variability is more sizeable in the /-li/-subgroup, notably in all frequency ranges.

The line of argument I will pursue in the following to account for these findings focusses on differences in *type* and *token* frequency. It is apparent from Figure 30 that the current dataset includes approximately seven times as many /-i/-disyllables (n = 570) as /-li/-disyllables (n = 80). In both phonological subgroups under scrutiny, hapax legomena (i.e. $n_{sup} = 1$) make up just under a third of all types (2 /i/: $prop_{hapax}$ = 32.5 %; 2 /li/: $prop_{hapax}$ = 30.0 %). Further, approximately two thirds of types in each subcategory have a gradation frequency of < 0.1 pmw (2 /li/: 67.5 %, 2 /i/: 66.1 %).

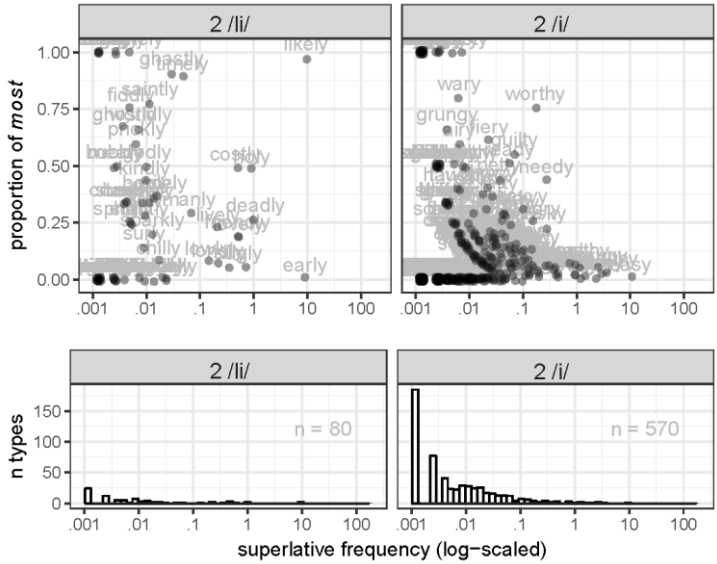

Figure 30. Superlative frequency trends (top) and type frequency distributions of disyllabic adjectives in *-ly* and *-y* (bottom)

Importantly, as I have proposed above, the subgroups in focus are homogeneous not only in phonological but also in *morphological* terms. The derivational suffix *-y* is significantly more productive in present-day English than the derivational suffix *-ly*: Bauer, Lieber & Plag (2013: 304-305) find that *-y* and *-ish* are the most productive derivational suffixes in English, "attach[ing] to any kind of available base, including compounds and phrases". Inspecting the types in the present dataset, we can in fact find *-y* attached to nouns (e.g. *dirty*, *cheesy*, *cloudy*), verbs (e.g. *bouncy*, *grabby*,

picky) and adjectives (e.g. *crispy*, *yellowy*), while *-ly* is restricted to nominal bases (e.g. *friendly*, *beastly*; cf. Bauer, Lieber & Plag 2013: 290, 304).[96]

Two assumptions are integral to a usage-based perspective of gradation strategy choice, namely (i) the strength and productivity of an adjectival schema is influenced by the type frequency of the schema itself (or rather, the adjectival cluster from which the schema can be abstracted) and the token frequency of the adjective exemplars comprised by the underlying cluster, and (ii) types that instantiate the same schema have similar gradation strategy preferences.

Both subcategories in focus comprise types which, by means of their high superlative token frequency, can be assumed to have gained lexical autonomy from clusters of morphophonologically similar types on the basis of which adjectival schemas emerge, e.g. *early* ($prop_{most}$ = 0.00, $freq_{sup}$ = 9.34 pmw), *likely* (0.98, 9.73 pmw) and *easy* (0.00, 10.25 pmw). By hypothesis, the superlative forms of these types have a whole-word representation and play no role in the superlative strategy choice of very rarely graded types, which is determined by analogy with morphophonologically similar types. The size of a cluster, and hence the strength of the schema that emerges from it, largely depends on the number and strength of the lexical relations between its members. Types that are tightly integrated in a cluster due to their many and strong lexical connections are those that, in terms of the frequency spectrum applied in the current study, score in the ranges of low and intermediate superlative frequency (i.e. 0.01 < $freq_{sup}$ > 1 pmw). It is apparent from Figure 30 that for disyllables in /-i/ (n = 184), the number of data points in these frequency bands is eight times as high as for disyllables in /-li/ (n = 23). This can be taken to suggest that the exemplar cluster comprised by disyllables in /-i/ is considerably larger as compared to the cluster of disyllables in /-li/, and the schema that emerges from this cluster is significantly more robust yet still quite variable. In other words, disyllables in /-i/ that very rarely form the superlative have a larger pool of exemplars with which to align their strategy choice. The observation that both subgroups contain hapax legomena whose single superlative token is analytic can be hypothesized to reflect a

[96] Recall that types in which *-y* attach to compounds (e.g. *hot-tubby*) and phrases (e.g. *secret-agenty*) were weeded out in the first stage of pre-processing (cf. Section 4.5.2). For the same reason, the data do not contain any instances of *-ly* combining with a noun phrase (e.g. *innerworldly*), which is another possibility mentioned in Bauer, Lieber & Plag (2013: 304).

tendency to resort to the globally more productive strategy if strategy alignment by analogy with morphophonologically similar types fails. The data support the assumption that this tendency is stronger among the subgroup comprising fewer types (i.e. disyllables in /-li/), although, judging by Sönning & Hartmann's (2019) findings for the comparative, we would have expected a higher proportion of *most* among disyllables in /-li/ in the range of very low superlative frequency (compare the comparative and superlative frequency trends in Figure 28). However, this apparent non-convergence of the gradation frequency trends should not be ascribed to systematic linguistic differences between the comparative and superlative: It has been noted that the corpora used in the current study and in Sönning & Hartmann (2019) differ with regard to several aspects, most notably in terms of size (cf. Section 4.5).

In the light of the patterns identified in Section 4.6, a critical assessment of the explanatory power of *more-*/*most*-support with regard to gradation frequency appears mandatory. Mondorf (2009b: 41-42) observes a positive relation between comparative frequency and the proportion of *more*, concluding that "[t]he less entrenched an adjective is in any of the two comparative forms, the more likely it is to require *more*-support". The results of the current study provide support for this trend: Especially in the phonological subcategories of disyllables in /-i/ and /-li/ as well as monosyllables irrespective of final segment, my data hint at a sizeable degree of alternation for types that rarely form the superlative. According to the *more*-support principle (Mondorf 2009b: 7), one of the three advantages of the analytic variant is that it consists of two lexical components with separate functions while in the case of the synthetic competitor, two functions are fused into a single complex lexeme. Incidentally, practical evidence for this hypothesized advantage in cognitive processing can be derived from the fact (observed in Section 4.5.2 on data collection and coding) that in the pre-processing phase of this study, I found analytic tokens significantly harder to disambiguate than synthetic tokens. Further, the comparison of scaling factors in Figure 17 indicated that the set of analytic tokens contained more false positives than the set of synthetic tokens under analysis. The one-to-one correspondence of form and function that is afforded by the analytic variant comes at the expense of reduced cue reliability: The inflectional superlative suffix *-est* is a highly reliable cue; perceptual competition is limited to idiosyncratic cases such as

real, whose synthetic superlative form is phonologically identical with the deadjectival noun *realist* (/ˈrɪəlɪst/) formed by derivation. For the analytic superlative, by contrast, quantifier uses of *most* and elatives complicate the situation; it emerged that in many cases, a thorough analysis of the broader context is required to resolve the ambiguity (cf. example (32) in Section 4.5.2). However, this practical observation is by no means sufficient to serve as a foundation for challenging the *more*-support hypothesis. As noted earlier, Lohmann's (2010) main point of criticism of *more*-support is that this notion is occasionally extended to predictors for which the assumed relation between formal explicitness and processing economy is less clear. This appears to be the case for the predictor in focus. The usage-based perspective adopted in this chapter suggests an alternative explanation for the high proportions of *most* that have been observed for (very) rarely graded types in the phonological subcategories in focus: Due to their lack of lexical connections in the exemplar network, these adjectives have not yet been able to align their strategy choice with formally similar ones; by reflex, these types opt for the strategy that is more productive in absolute terms, i.e. analytic gradation.

In Section 4.6.4, we adopted a crossvarietal perspective, comparing the superlative frequency trends in the data gleaned from the US and GB sections of GloWbE. The results that have been presented do not hint at any systematic differences between the main reference varieties in focus. These insights are consistent with the questionnaire-based findings reported in Chapter 3, which suggest a high degree of crossvarietal homogeneity with regard to the effects of contextual constraints on the superlative alternation. As such, the current study adds to the body of evidence indicating that grammatical alternations are remarkably stable across lects (Szmrecsanyi et al. 2017; Heller, Szmrecsanyi & Grafmiller 2017, *inter alia*).

A similar conclusion can be drawn for my comparison of the superlative and comparative frequency trends (Section 4.6.5), for which I relied on the comparative data provided by Sönning & Hartmann (2019) as reference points: On the whole, my analysis of gradation frequency trends according to phonological subgroup has revealed that despite the slightly diverging tendencies in the low-frequency range, where the estimates come with sizeable uncertainty, the systematic patterns underlying com-

parative and superlative choice do not differ appreciably. However, it appears imperative to reiterate the caveat that the validity of this comparison is compromised by the fact that the GloWbE material of the current study differs in many ways from the BNC material used by Sönning & Hartmann (2019). Importantly, the largely similar comparative and superlative trends observed in Section 4.6.5 cannot be taken to call into question the crossover hypothesis (Peters 2000), which accounts for the observation that selected types with alternating behavior seem to prefer the analytic comparative but the synthetic superlative. While the current analysis only occasionally zoomed in on the analyticity scores of individual types in the comparative and superlative, and thus does not warrant a verdict on the applicability of this concept, it appears plausible that the crossover tendencies that have been identified in Peters (2000) can at least in part be attributed to collocational effects; this scenario, which may constitute a promising avenue for future research, will be explored in greater detail in Chapter 5.

Motivated by the literature on morphological processing, my considerations in Section 4.2.1 focused on the question of whether the processing and storage of superlative forms can be accounted for in terms of a dual-route mechanism; in this empirical context, root frequency is standardly considered an indicator of morphological decomposition, and surface frequency (of the morphologically complex form) is treated as a diagnostic of whole-word processing (Baayen, Wurm & Aycock 2007: 448). While the current study did not measure root frequency (i.e. the positive frequency of an adjective), my operationalization of superlative frequency is firmly grounded in usage-based theory and can be taken to correspond with the metric of surface frequency.

The experimental evidence presented by Alegre & Gordon (1999) indicates that morphologically complex words (i.e. synthetic superlatives, in the current research context) above a threshold surface frequency of 6 pmw can be assumed to be processed and stored as a whole while morphologically complex words below this threshold are, by hypothesis, decomposed into their constituent morphemes and reassembled on demand. Psycholinguistic studies of inflectional and derivational forms in Dutch, on the other hand, suggest that whole-word processing occurs below the proposed threshold of 6 pmw (Baayen, Dijkstra & Schreuder 1997;

Baayen, Wurm & Aycock 2007). Since previous research on morphological processing has chiefly relied on reaction time experiments, it appears legitimate to ask how the current corpus-based investigation can contribute to the dual-route debate. The usage-based view of gradation strategy choice predicts that very frequently graded adjectives have gained lexical autonomy from morphophonologically similar adjectives in the similarity network (cf. Section 4.3). These types categorically form either the synthetic or analytic superlative, and we can reasonably assume that the preferred variant has a rote representation in the mental lexicon. Thus, is there a superlative frequency value above which we can exclusively find types with a categorical preference for one of the two variants? We can infer from Figure 20 (Section 4.6.1) that variability in strategy choice is virtually non-existent above a superlative and comparative frequency of approximately 3 pmw. Of the 38 types that score above this threshold in the current superlative data, only *common* ($prop_{most}$ = 0.96; $freq_{sup}$ = 12 pmw) and *simple* (0.06; 6 pmw) exhibit perceivable but marginal deviations from the respectively categorical variant. Based on these insights, it is far from unlikely that holistic representations are available for superlatives below the supposed threshold of 6 pmw, possibly already for types with a superlative frequency higher than 3 pmw. Although previous empirical studies on dual-route processing have predominantly been concerned with morphologically complex words (inflected or derived), i.e. words that are realized as a single lexical unit on the linguistic surface, the hypothesis that the analytic superlatives of high-frequency types such as *important* (44 pmw), *popular* (13 pmw) and *famous* (8 pmw) are also stored holistically is compatible with usage-based construction grammar. With several studies indicating that highly frequent multi-word sequences such as *all over the place* are likely to be stored as a whole (Arnon & Snider 2010; Tremblay & Baayen 2010; McCauley & Christiansen 2014, *inter alia*), we should also expect such effects for analytic superlatives (i.e. *most* + Adj bigrams). Future studies should harness the potential of experimental designs to determine whether rote representations are available for both synthetic and analytic superlatives, and further, above which threshold whole-word frequency effects are likely to arise.

It is evident from the corpus-based insights offered in this chapter that the interplay of two central predictors of the superlative alternation, namely frequency of use and morphophonological type, can adequately

be explained on the basis of usage-based theory. Drawing on Bybee's (1985, 1988) dynamic network model of lexical representation and connectionist mechanisms of categorization (cf. Sections 4.2.2 and 4.2.3), I argued that, by hypothesis, the mental network of adjective exemplars is arranged according to two conceptual metrics, i.e. lexical connections and lexical strength. This entailed an empirical focus on *superlative frequency* and the customary formal determinants of the adjective, namely the number of syllables and the final segment. Morphophonologically similar adjectives of low or intermediate lexical strength form exemplar clusters from which superlative schemas are abstracted; thus, adjectives that are strongly interconnected within the same cluster should be expected to exhibit the same strategy preferences. Types of very low lexical strength, by contrast, only have few and rather loose connections in the exemplar network; their strategy choice often deviates from the preference we would predict on the basis of morphophonological criteria. This scenario can be adduced to account for the observation made by Hilpert (2008: 395) that *easy* categorically takes *-er* while *queasy*, which only rarely forms the comparative, prefers *more*. Since the structure of the mental network of adjective exemplars persistently changes on the basis of new input, superlative category assignment should be considered a highly dynamic process.

According to Bybee (1988: 125-126), the network of lexical representations is structured not only by phonological but also by *semantic* (dis)similarity. Along these lines, we should expect the largely identical strategy preferences of synonyms such as *broad* and *wide* to be due to the strong lexical connections of these types in the mental network. This theoretical scenario is highly problematic in empirical terms, however, since it raises the question of how semantic (dis)similarity can reliably be quantified, especially considering that many semantically similar adjectives also share phonological characteristics. Pending further evidence, there is little reason to believe that the similar gradation strategy preferences of adjectives such as *broad* and *wide* can be attributed to their semantic relation.

5 Conclusion and outlook

The study at hand explored four sources of variation in the choice of superlative strategy: The questionnaire-based investigation presented in Chapter 3 shed light on seven contextual and four social (i.e. speaker-related) predictors while the corpus analyses in Chapter 4 examined the systematic relationship between the morphophonological characteristics of the adjective and superlative frequency, i.e. the rate at which an adjective occurs in the highest degree.

My results for the four speaker-related predictors under analysis, namely *age*, *gender*, *education* and *variety*, imply that on the whole, superlative strategy choice is relatively stable across social categories. Nevertheless, we have observed interesting trends for *age* and *gender*: A marginal increase in the use of the analytic superlative has been identified in apparent time, with the difference in the proportion of *most* between 20- and 80-year-olds amounting to 5 percentage points. This trend is consistent with previous corpus-based investigations of the comparative alternation indicating that *most* has indeed gained ground in the latter half of the 20th (Leech & Culpeper 1997: 370-371; Leech et al. 2009: 264-267) as well as the early 21st century (Säily, González-Díaz & Suomela 2018: 179). As for *gender*, my questionnaire data hint at a mildly greater proportion of *most* among the elicited ratings of female as compared to male informants.

In Section 3.4.3, these findings were related to established sociolinguistic theory. It was argued that the observed increase in analytic superlatives in the period under scrutiny mirrors a change from below that was driven by two societal developments: The decline of institutionalized prescriptivism as of the 1970s (Crystal 2004: 523-525, 2012: 507-510) and the innovations of the digital age may have promoted a trend towards increased communicative pragmatism and efficiency. With two (near-)equivalent linguistic structures at their avail, speakers therefore tend to opt for the more explicit and easier-to-process variant at an increasing rate, in order to facilitate comprehension. Following Mondorf (2003, 2009b), we should expect this processing advantage to be afforded by the analytic superlative. This assumed trend towards *most* is possibly spearheaded by female speakers, who, according to Tagliamonte (2012: 63), act as vanguards of linguistic change in the great majority of cases. Pending further

empirical insight, however, the scenario of a female-led apparent-time increase in analytic superlatives should be treated with caution, particularly because it premises different degrees of sensitivity to contextual complexity among younger and older speakers as well as women and men. Thus, future studies that intend to shed light on the speaker-related predictors of gradation strategy choice should investigate whether differences between social categories are in fact conditional on or apply irrespective of differences in contextual complexity.

My analysis of *education* was largely exploratory in nature. The results revealed no appreciable differences in the overall proportion of *most* between speakers who have obtained a university degree and those who have not. A similar conclusion was drawn for the predictor *variety*: The data at hand suggest that native speakers of BrE, AmE, AustrE and CanE differ only marginally in their use of gradation strategy. These findings are consistent with those of previous studies suggesting that grammatical alternations are relatively stable across varieties (Szmrecsanyi et al. 2017; Heller, Szmrecsanyi & Grafmiller 2017: 21, *inter alia*).

The questionnaire data presented in Chapter 3 further enabled us to assess the effect of seven contextual constraints on the choice of superlative variant, all of which were previously investigated for their role in the comparative alternation. For four of these predictors, namely *segmental* horror aequi, *syntactic position*, *infinitival complement* and *end weight*, I predicted that the test conditions should result in greater proportions of *most* than the respective control conditions. To motivate these predictions for the syntactic and pragmatic predictors under scrutiny, I argued that by analogy with Mondorf's (2003, 2009b: 6) *more*-support hypothesis, the analytic superlative should be preferred in environments of increased complexity since (i) the initial degree marker serves as a signal to the recipient that the following adjective phrase is complex and urges them to devote increased processing effort to this phrase, (ii) it assigns each function a separate lexeme, thus obviating the need to mentally decompose a dimorphemic unit, and (iii) it keeps constituent recognition domains shorter and more continuous. Along these lines, then, speakers can be expected to constantly perform a tradeoff between formal explicitness and processing complexity, using *most* in contexts of high complexity and *-est* in environments of low complexity. The data I have presented lend support to this assumed cognitive compensatory mechanism in the domains of

syntax and pragmatics (i.e. for *syntactic position*, *infinitival complement* and *end weight*). Our analysis of the two phonological constraints that were operationalized in the questionnaire study (*rhythmic harmony* and *segmental* horror aequi) exposed structural deficits of elicitation methods in the investigation of morphosyntactic alternations

As argued by Lohmann (2010: 308), the link between formal explicitness and cognitive complexity is not immediately apparent in the context of phonological identity avoidance; instead, a nodal-activation account as proposed by Schlüter (2005) appears to have greater explanatory power for effects of this kind (cf. Section 3.4.1). Thus, although we can legitimately assume that *most*-support conditions a speaker's choice of superlative form at least in selected domains of linguistic analysis, the scope of this cognitive compensatory mechanism should not be stretched too far: The corpus analyses presented in Chapter 4 imply that contextual complexity effects as elicited through my questionnaire (Chapter 3) are clearly subordinate to the interplay of gradation frequency and the morphophonological characteristics of the adjective.

Although the body of literature attesting to the role of contextual constraints in gradation strategy choice is constantly growing, several gaps in our knowledge of this set of predictors remain to be addressed in future studies. As for constraints that have been linked to *more*- and *most*-support, it will be worthwhile to investigate, for instance, whether learners of English exhibit the same degree of sensitivity to contextual complexity as native speakers. For all we know, it appears reasonable to hypothesize that non-native speakers of English tend to adhere more closely to prescriptive rules on gradation strategy choice than native speakers. These rules typically focus on the morphophonological characteristics of the adjective, i.e. the number of syllables and final segment(s), and do not factor in contextual constraints. Along these lines, then, we should expect to find that differences between contexts of low and high complexity result in more limited differences in the probability of observing the analytic variant in material produced by learners as compared to native speakers of English. Further, as pointed out in Section 3.4.2, future research should compare the effects of the assumed cognitive compensatory mechanism in comparative and superlative strategy choice on the basis of identical linguistic material to reevaluate Cheung & Zhang's (2016: 579) conclusion that "'most-support' does not seem to fare as well as more-support". However,

researchers that intend to address this question should consider that the two degree forms in focus differ on multiple levels of linguistic analysis (cf. Chapter 2). Thus, we need to be cautious when comparing the effects of contextual complexity on the comparative and superlative alternations.

The corpus-based analysis of superlative frequency and the formal characteristics of the adjective in Chapter 4 originated from a usage-based perspective of gradation strategy choice. The model that was proposed to account for the interplay of superlative frequency and the formal characteristics of the adjective combines two design features, namely (i) connectionist principles of category assignment and (ii) Bybee's (1985, 1988, 1995) network architecture of lexical representations.

In Bybee's (1985, 1988, 1995) schema-based network model, lexical representations are constantly derived and adapted upon exposure to linguistic input. The arrangement of these lexical representations, or exemplars, hinges on the interplay of two conceptual characteristics, namely lexical strength and lexical connections. Adjective exemplars that have strong lexical connections to each other form a cluster due to their morphophonological similarity. It is upon these exemplar clusters that language users abstract schemas for superlative formation. Adjective exemplars with a high superlative token frequency have a high degree of lexical strength and only few and weak lexical connections to morphophonologically similar exemplars in the network. Due to their lexical autonomy, these types are largely independent of exemplar clusters and thus, by hypothesis, do not contribute to the formation of superlative schemas.

The analyses in Chapter 4 were guided by two assumptions about superlative strategy choice that could be derived from Bybee's (1985, 1988, 1995) schema-based network model, namely that (i) the strength and productivity of an adjectival schema is influenced by the type frequency of the schema itself (or rather, the adjectival cluster from which the schema can be abstracted) and the token frequency of the adjective exemplars comprised by the underlying cluster, and (ii) types that instantiate the same morphophonological schema have similar gradation strategy preferences.

Based on these assumptions, my empirical analysis zoomed in on the superlative frequency trends among subgroups that were defined by the combination of two morphophonological characteristics: the number of syllables (mono-, di-, polysyllabic) and the final segment(s) of the adjective

(/-i/, /-li/, /-ə(r)/, /-l/, cluster, other). We observed that among monosyllabic types regardless of final segment(s) variation in superlative strategy choice is relatively sizeable in the ranges of very low and low superlative frequency while types of intermediate and high frequency exhibit a robust preference for the synthetic variant. For disyllables in /-i/ and /-li/, the (limited) preference for the synthetic superlative seems independent of frequency. Among the remaining subgroups, i.e. disyllables in /-ə(r)/, /-l/, cluster and other final segments as well as polysyllables, on the contrary, the (near-)categorical trend towards *most* is largely irrespective of superlative frequency. Only among polysyllables in /-i/ did we identify moderate variation in the ranges of very low and low frequency: Individual *un*-prefixed trisyllabic types such as *unhappy* and *unhealthy* occasionally take *-est* und thereby instantiate the superlative schema of their non-negated counterparts to which they have strong lexical connections in the hypothesized network of adjective exemplars.

In sum, a noticeable share of the variation observed among adjectives in the same morphophonological subgroup can be explained by differences in superlative token frequency. Further, our comparison of the subgroups of disyllables in /-i/ and /-li/ revealed that the strength of a superlative schema depends on the size of the exemplar cluster upon which the schema is abstracted (i.e. its type frequency). The observation that the interplay of these frequency metrics has explanatory power for variation in superlative strategy choice lends support to the bedrock of usage-based theory: Experience with linguistic input shapes our mental representation of language.

While the connectionist exemplar model of superlative strategy choice proposed in Chapter 4 accounts for two influential sources of variation, namely frequency and the morphophonological properties of the adjective, it makes no provisions for the two remaining types of predictors that have been addressed in this study. Thus, it remains desirable of future research to consider how the effects of contextual constraints and speaker variables can conceptually be integrated into a connectionist exemplar model of superlative strategy choice.

Both empirical approaches that were combined in this thesis adopted a crossvarietal perspective on superlative strategy choice, examining systematic differences between the main reference varieties British and American English. My analysis of the material gleaned from the GB and

US sections of GloWbE (Section 4.6.4) revealed only negligible differences between the superlative frequency trends in the varietal datasets under scrutiny. This verdict ties in with the questionnaire-based findings that on the whole, British and American English are relatively homogeneous with regard to the effects of contextual constraints on the alternation in focus (Sections 3.3.2 and 3.3.4). This can be taken to suggest that neither variety is systematically more inclined than the other to compensate contextual complexity by resorting to the analytic variant, which, by analogy with Mondorf (2003, 2009b), is assumed to be cognitively advantageous. However, differences between the varieties became more sizeable as soon as we zoomed in on individual constraints and adjectives. Based on the 23 adjectives covered by my questionnaire, the findings can be taken to indicate that the more categorical an adjective is by virtue of its morphophonological properties, the lower the degree of cross-lectal variation. It remains open for future research to extend the scope of analysis to other varieties of English, also encompassing varieties located in the Outer and Expanding circles of Kachru's (1992) model of World Englishes.

The analyses in Section 4.6.5 further enabled a comparison of the superlative frequency trends that emerged from the current GloWbE data and the comparative frequency trends reported by Sönning & Hartmann (2019) on the basis of their BNC data. We observed that these frequency trends were largely uniform across all morphophonological subgroups; similar to the findings obtained for the crossvarietal comparison, however, we identified considerable discrepancies between the obtained proportions of more and most for individual adjectives. A close inspection of disyllabic adjectives in /-ə(r)/ suggested that especially in the range of intermediate gradation frequency, most of these types have a higher share of analytic forms in the comparative as compared to the superlative. This observation was attributed to segmental *horror aequi*, i.e. the tendency to avoid the adjacency of identical segments in the synthetic comparative (/ərə(r)/). It will be worthwhile for future work to explore in greater detail the systematic differences between the comparative and superlative alternations on the basis of identical data sources. An investigation of this kind should also enable a critical assessment of the concept of crossover adjectives (Peters 2000). Although the findings I have presented do not allow a definite conclusion on this matter, we may assume that if adjec-

tives indeed showed crossover behaviour, syntactic and collocational differences between comparatives and superlatives could be central to accounting for this phenomenon.

Different collocational preferences – both between the comparative and superlative of the same adjective and between the synthetic and analytic variants of the same degree – have hitherto largely been neglected as a potential source of variation in gradation strategy choice. The case study that will be presented in the following may constitute a useful vantage point for future work on the role of collocational preferences in the alternations under investigation.

In Section 3.3.2, I compared the proportions of *most* for the 23 adjectives in the BrE questionnaire data to those reported by Cheung & Zhang (2016) on the basis of their BNC data. It is for *remote* (BrE = .61, BNC = .35) that I obtained the most sizeable differences in analyticity between the questionnaire and corpus data. Let us now take a closer look at the superlative + noun (N) bigrams involving *remotest* and *most remote* in the BNC: 90 % of synthetic superlative tokens (64/71) occur in attributive position, as opposed to only 62 % of analytic superlative tokens (24/39). The observation that the analytic variant is more likely to be found in predicative and nominal position than its synthetic competitor is in keeping with my findings for the predictor syntactic position as well as those presented by Mondorf (2009b: 80-90) for the comparative. Remarkably, however, the nouns that respectively combine with *remotest* and *most remote* in attributive position pertain to markedly different semantic domains: While *most remote* virtually exclusively occurs in collocations with nouns denoting physical location, e.g. *parts* (3 tokens), *regions* (3), *areas* (3) and *corners* (2), to name the four most frequent types, *remotest* appears to favor abstract nouns denoting cognitive concepts such as *idea* (12), *chance* (5), *intention* (3) and *notion* (2); further, these collocations of *remotest* are always negated in the current BNC data, that is *not the remotest* + N. In attributive use, therefore, the superlative variants of this adjective seem to have divergent collocational preferences. In all eight questionnaire items that were designed to elicit the superlative of *remote*, this adjective was combined – either in attributive or nominal use – with nouns denoting physical location, namely *village*, *place*, *stretch* and *part*. These insights may at least in part account for the substantial discrepancy between the analyticity values obtained in the questionnaire study and those gleaned from

the BNC. The studies by Mondorf (2007; 2009b: 91-97) can serve as a valuable starting point for future research on the role of semantic and collocational differences in gradation strategy choice. In this context, it also appears promising to analyze idiosyncratic constructions that seem to come with a very restricted degree of variation: Bouso's (2024) corpus study, for instance, investigates the Superlative Object Construction (e.g. *She worked her hardest.*) from a usage-based perspective.

In examining the role of collocations in degree strategy choice, it will further be interesting to ask, from the perspective of dual-route morphological processing, whether n-grams such as *not the remotest* + N are processed and stored as a whole or decomposed and reassembled on demand. We can legitimately assume that frequency of use plays a crucial part in this research context.

According to Cohen, Manion & Morrison (2007: 141), triangulation can be defined as an "attempt to map out, or explain more fully, the richness and complexity of human behavior by studying it from more than one standpoint". Although the concept of triangulation was first established in the social sciences, the linguistic community has increasingly recognized the need for methodological triangulation in the research of language variation and change (see, for instance, the volumes edited by Baker & Egbert 2016 and Egbert & Baker 2020). The thesis at hand triangulated two default methods in linguistic research to explore four sources of variation in superlative strategy choice. Since the researcher employing methodological triangulation intends to compensate the biases and weaknesses inherent in any single method, it appears worthwhile to reconsider some of the key drawbacks and assets of the methods that were used in the present work.

The questionnaire study presented in Chapter 3 aimed to quantify the effects of seven contextual constraints and four speaker-related predictors on the superlative alternation. The pool of questionnaire items consisted of 120 sentence prompts. These were organized into 60 item pairs, each of which included a test and a control item. The control items were constructed to serve as a neutral baseline, a condition in which contextual complexity is reduced to a minimum. Test items were derived by manipulating the control sentence with regard to a single contextual predictor. Thus, my general rationale underlying the questionnaire design was to approximate what can essentially be considered a monofactorial design:

If we obtained different proportions of *most* for the test and control item of the same item pair, this difference can be attributed to the effect of the manipulated variable. This opportunity to isolate the effect of a single contextual constraint constitutes a major advantage over a corpus-based analysis of naturally produced linguistic material.

Inspecting the cluster-level predictors on the basis of the null model (Section 3.3.4), however, I cautioned that the process of manipulating an item along a single contextual dimension is error-prone and requires great diligence on behalf of the researcher. This inspection further highlighted the need for multilevel statistical analysis to systematically explore the sources of variation in the data. In the current context, it appeared particularly urgent to quantify the variation that could be attributed to the 23 adjectives used for item construction, which were not the focus of attention in the present investigation but merely a necessary component of the instrument. Our quality checks in Section 3.3.4 suggested that on the whole, the adjectives that were selected for item construction allowed for a sufficiently large degree of variation by speaker-related factors and contextual constraints. Since the questionnaire was only based on a small number of adjectives, most of which showed alternating behavior, the results we obtained clearly have little predictive power for the effects of frequency variables and the morphophonological properties of the adjective.

Another challenge posed by the current questionnaire-based elicitation method pertains to the high degree of artificiality of the instrument as well as the process of data collection. The results obtained for the two phonological constraints under scrutiny (*rhythmic harmony* and *segmental* horror aequi) and the two persistence predictors (*synthetic* and *analytic persistence*) were largely inconsistent with those of previous studies. This perceived inconsistency was taken to imply that my questionnaire was rather ill-suited to elicit the assumed effects of phonological identity avoidance and structural priming, both of which can be accounted for along the lines of nodal activation. For studies that intend to quantify the effects of these contextual constraints, naturally produced linguistic material may provide a more reliable data basis.

In addition, we should not rule out the possibility that the current questionnaire data are influenced by effects of social desirability (Krug & Sell 2013: 75): Although explicitly instructed to go by their initial feeling,

respondents may have tried to comply with prescriptive norms on gradation strategy choice while filling in the questionnaire, thereby deviating from spontaneous language production. Some of my respondents indeed pointed out to me that they were aware of such rules of thumb that exclusively centered on the morphophonological properties of the adjective, and that they occasionally felt tempted to resort to these rules. Further, respondents may have made efforts to be internally consistent in their choice of superlative form, thus opting for the same superlative form in the test and control conditions of the same item pair regardless of the linguistic context. Even though the items of a pair were presented within the largest possible distance of each other, it is very likely that my respondents were able to recall earlier choices. This may have had the effect of shrinking the differences between test and control item, thereby making the findings more conservative, overall. Put differently, the differences we did observe then seem even more notable, since they managed to survive this levelling effect.

The corpus study presented in Chapter 4 relied on the GB and US sections of the GloWbE corpus to investigate the interplay of superlative frequency and the morphophonological properties of the adjective. Since the analyses were based on (written) online material, it is legitimate to assume that large parts of the text under scrutiny – particularly formal texts such as (online) newspaper articles – may have been pre-written in a word-processing program. The grammar checkers that are customarily implemented in these programs may have induced authors to revise their choices of superlative strategy, thereby potentially reducing variation in the data.

Further, it has been noted that GloWbE also comprises text from informal blogs, which can be regarded as representing conversational language in a written format, sometimes with little or no revision. While the influence of spell-checking software appears less problematic for this text type, material from this source often contains misspellings, typos and missing spaces between words. It appears reasonable to assume that these errors might have affected the accuracy of the automatic tagger. This might have led to a small number of superlative forms not being classified as such and, as a result, might have produced inaccurate type and token counts. Closely related to this issue, it has been noted in Section 4.5.2 that the automatic tagger occasionally confuses instances of quantifier *most*

and uses of *most* as a superlative degree marker. Although I tried to disambiguate these cases to the best of my abilities, it cannot be ruled out that the proportions of analytic tokens in the final dataset are slightly perturbated.

The analytical focus of my corpus-based approach to the superlative alternation was to examine how differences in type and token frequency affect the choice of gradation strategy. It appeared expedient, therefore, to include a large number of both alternating and categorical types in the analysis and disregard the variation that can be attributed to contextual predictors. The questionnaire study presented in Chapter 3, by contrast, which explicitly zoomed in on the effects of contextual constraints, required a careful selection of a small number of adjectives that were non-categorical by virtue of their morphophonological characteristics.

The present study hopes to have illustrated the added value of combining different methodological approaches to the study of the superlative alternation. Even though it has been stated in the literature that adjective gradation is among the best-researched alternation phenomena in the English language (cf. Szmrecsanyi 2006: 63), this is arguably a corpus-centered view. With its questionnaire-based judgement tasks, the current study adds to the increasingly diverse body of empirical approaches to adjective gradation (e.g. Boyd 2007; LaFave 2015; Kunter 2017). The fact that the choice between synthetic and analytic superlatives is subject to a multifactorial set of influences puts a prime on the ability to exert control over contextual factors when assessing the relative weight of different constraints that have been put forward in corpus-based work. Methodological pluralism and the triangulation of different types of data will undoubtedly enhance our knowledge of this showcase of grammatical variation.

Bibliography

Primary sources/databases

The British National Corpus (BNC): BNC Consortium/Oxford University Computing Services. 2007. Version 3.0 (BNC XML edition). URL http://www.natcorp.ox.ac.uk/.

The Corpus of Contemporary American English (COCA): Davies, Mark. 2008–). The Corpus of Contemporary American English (COCA): 1 billion words, 1990-2019. URL https://www.english-corpora.org/coca/.

The Corpus of Global Web-Based English (GloWbE): Davies, Mark. 2013. Corpus of Global Web-Based English: 1.9 billion words from speakers in 20 countries. URL https://www.english-corpora.org/glowbe/.

The Oxford English Dictionary Online (OED): *Oxford English Dictionary*: The definitive record of the English language. 2nd edition. Oxford: Oxford University Press. URL https://www.oed.com/.

Secondary sources

Abbot-Smith, Kirsten & Michael Tomasello. 2006. Exemplar-learning and schematization in a usage-based account of syntactic acquisition. *The Linguistic Review* 23 (3). 275-290.

Abel, Beate. 2003. *Sprecherurteile zur Dekomponierbarkeit englischer Idiome: Entwicklung eines Modells der lexikalischen und konzeptuellen Repräsentation von Idiomen bei Muttersprachlern und Nichtmuttersprachlern*. Berlin: Mouton de Gruyter.

Ackema, Peter & Ad Neeleman. 2005. Word-formation in optimality theory. In: Pavol Štekauer & Rochelle Lieber (eds.), *Handbook of word-formation*. Dordrecht: Springer. 285-313.

Al Wer, Enam. 2002. Education as a speaker variable. In: Aleya Rouchdy (ed.), *Language contact and language conflict in Arabic: Variations on a sociolinguistic theme*. London: Routledge. 41-53.

Alegre, Maria & Peter Gordon. 1999. Frequency effects and the representational status of regular inflections. *Journal of Memory and Language* 40. 41-46.

Altenberg, Bengt. 1982. *The genitive v. the of-construction: A study of syntactic variation in the 17th century English*. Lund: Gleerup.

Ambridge, Ben. 2018. Against stored abstractions: A radical exemplar model of language acquisition. *First Language* 40 (5-6). 509-559.
Andersen, Paul Kent. 1983. *Word order typology and comparative constructions*. Amsterdam: Benjamins.
Arnold, Jennifer E., Anthony Losongco, Thomas Wasow & Ryan Ginstrom. 2000. Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering. *Language* 76 (1). 28-55.
Arnon, Inbal & Neal Snider. 2010. More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language* 62 (1). 67-82.
Ash, Sharon. 2002. Social Class. In: Jack K. Chambers, Peter Trudgill & Natalie Schilling-Estes (eds.), *The handbook of language variation and change*. Malden: Blackwell. 402-422.
Aslin, Richard N. & Elissa L. Newport. 2012. Statistical learning: From acquiring specific items to forming general rules. *Current Directions in Psychological Science* 21 (3). 170-176.
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.
Baayen, R. Harald, Ton Dijkstra & Robert Schreuder. 1997. Singulars and plurals in Dutch: Evidence for a parallel dual-route model. *Journal of Memory and Language* 37 (1). 94-117.
Baayen, R. Harald & Michael Ramscar. 2015. Abstraction, storage and naive discriminative learning. In: Ewa Dabrowska & Dagmar Divjak (eds.), *Handbook of cognitive linguistics*. Berlin: Mouton de Gruyter. 100-120.
Baayen, R. Harald, Lee H. Wurm & Joanna Aycock. 2007. Lexical dynamics for low-frequency complex words: A regression study across tasks and modalities. *The Mental Lexicon* 2 (3). 419-463.
Bailey, Guy. [2]2013. Real and apparent time. In: Jack K. Chambers & Natalie Schilling-Estes (eds.), *The handbook of language variation and change*. Chichester: Wiley-Blackwell. 312-332.
Bailey, Guy, Tom Wikle, Jan Tillery, & Lori Sand. 1991. The apparent time construct. *Language Variation and Change* 3 (3). 241-264.
Baker, Paul & Jesse Egbert (eds.). 2016. *Triangulating methodological approaches in corpus linguistic research*. New York: Routledge.
Bauer, Laurie. 1994. *Watching English change: An introduction to the study of linguistic change in standard Englishes in the twentieth century*. London: Longman.

Bauer, Laurie, Rochelle Lieber & Ingo Plag. 2013: *The Oxford reference guide to English morphology*. Oxford: Oxford University Press.
Baugh, Albert Croll & Thomas Cable. [4]1993. *A history of the English language*. London: Routledge.
Bayley, Robert. [2]2013. The quantitative paradigm. In: Jack K. Chambers & Natalie Schilling-Estes (eds.), *The handbook of language variation and change*. Chichester: Wiley-Blackwell. 85-107.
Behaghel, Otto. 1909. Beziehungen zwischen Umfang und Reihenfolge von Satzgliedern. *Indogermanische Forschungen* 25. 110-142.
Beland, Nikolai. 2021. The superlative alternation in present-day English: Questionnaire data. DataverseNO, V1, UNF: 6:RKi85c6Q4ObGsRs1yg7J/g== [fileUNF]. doi:10.18710/ NL8UQC.
Beland, Nikolai. 2022. The superlative alternation in British and American English: Questionnaire-based insights. In: Manfred Krug, Ole Schützler, Fabian Vetter & Valentin Werner (eds.), *Perspectives on contemporary English: Structure, variation, cognition*. Frankfurt: Lang. 59-90.
Benor, Sarah & Roger Levy. 2006. The chicken or the egg? A probabilistic analysis of English binomials. *Language* 82. 233-278.
Berg, Thomas. 2000. The position of adjectives on the noun-verb continuum. *English Language* 4 (2). 269-293.
Berlage, Eva. 2014. *Noun phrase complexity in English*. Cambridge: Cambridge University Press.
Bethlehem, Jelke G. & Silvia Biffignandi. 2012. *Handbook of web surveys*. Hoboken: Wiley.
Biber, Douglas. 1988. *Variation across speech and writing*. Cambridge: Cambridge University Press.
Biber, Douglas. 1995. *Dimensions of register variation: A cross-linguistic comparison*. Cambridge: Cambridge University Press.
Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. 1999. *Longman grammar of spoken and written English*. London: Longman.
Blumenthal-Dramé, Alice. 2012. *Entrenchment in usage-based theories: What corpus data do and do not reveal about the mind*. Berlin: Mouton de Gruyter.
Boberg, Charles. 2012. Standard Canadian English. In: Raymond Hickey (ed.), *Standards of English: Codified varieties around the world*. Cambridge: Cambridge University Press. 159-178.
Bock, Kathryn J. 1986. Syntactic persistence in language production. *Cognitive Psychology* 18 (3). 355-387.

Bock, Kathryn J., Gary S. Dell, Franklin Chang & Kristine H. Onishi. 2007. Persistent structural priming from language comprehension to language production. *Cognition* 104 (3). 437-458.

Bock, Kathryn J. & Zenzi M. Griffin. 2000. The persistence of structural priming: Transient activation or implicit learning? *Journal of Experimental Psychology: General* 129 (2). 177-192.

Bock, Kathryn J. & Anthony Kroch. 1989. The isolability of syntactic processing. In: Greg N. Carlson & Michael K. Tanenhaus (eds.), *Linguistic structure in language processing*. Dordrecht: Springer. 157-196.

Bolinger, Dwight. 1967. Adjectives in English: Attribution and predication. *Lingua* 18. 1-34.

Bolinger, Dwight. 1977. *Meaning and form*. London: Longman.

Booij, Geert E. 2013. Morphology in Construction Grammar. In: Thomas Hoffmann & Graeme Trousdale (eds.), *The Oxford Handbook of Construction Grammar*. Oxford: Oxford University Press. 255–273.

Bouso, Tamara. 2024. Towards a usage-based characterisation of the English Superlative Object Construction. *Constructions and Frames* 16 (1). 100-129.

Boyd, Jeremy K. 2007. *Comparatively speaking: A psycholinguistic study of optionality in grammar*. Unpublished PhD dissertation. San Diego: University of California PhD thesis. URL https://escholarship.org/uc/item/1wq3j7bk.

Branigan, Holly P., Martin J. Pickering & Alexandra A. Cleland. 1999. Syntactic priming in written production: Evidence for rapid decay. *Psychonomic Bulletin & Review* 6 (4). 635-640.

Branigan, Holly P., Martin J. Pickering, Andrew J. Stewart & Janet F. McLean. 2000. Syntactic priming in spoken production: Linguistic and temporal interference. *Memory & Cognition* 28 (8). 1297-1302.

Braun, Albert. 1982. *Studien zu Syntax und Morphologie der Steigerungsformen im Englischen*. Bern: Francke.

Brems, Lieselotte. 2007. The grammaticalization of small size nouns. *Journal of English Linguistics* 35 (4). 293-324.

Bresnan, Joan. 2011. Acquiring syntactic variation in English: A cross-constructional study. Plenary talk presented at the 6th International Conference on Language Variation in Europe (ICLAVE 6), Freiburg.

Bresnan, Joan, Anna Cueni, Tatiana Nikitina & R. Harald Baayen. 2007. Predicting the dative alternation. In: Gerlof Bouma, Irene Krämer & Joost Zwarts (eds.), *Cognitive foundations of interpretation*. Amsterdam: Royal Netherlands Academy of Arts and Sciences. 69-94.

Bresnan, Joan & Marilyn Ford. 2010. Predicting syntax: Processing dative constructions in American and Australian varieties of English. *Language* 86 (1). 168-213.

Bresnan, Joan & Jennifer Hay. 2008. Gradient grammar: An effect of animacy on the syntax of give in New Zealand and American English. *Lingua* 118 (2). 245-259.

Brezina, Vaclav, Robbie Love & Karin Aijmer (eds.). 2018. *Corpus approaches to contemporary British speech: Sociolinguistic studies of the Spoken BNC2014*. New York: Routledge.

Brugmann, Karl. 1909. *Das Wesen der lautlichen Dissimilationen*. Leipzig: Teubner.

Bürkner, Paul-Christian. 2017. brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software* 80 (1). 1-28.

Busemeyer, Jerome R. & Adele Diederich. 2010. *Cognitive modeling*. Los Angeles: Sage.

Butterworth, Brian. 1983. Lexical representation. In: Brian Butterworth (ed.), *Language production*. Volume II: Development, writing and other language processes. London: Academic Press. 257-294.

Bybee, Joan L. 1985. *Morphology: A study of the relation between meaning and form*. Amsterdam: Benjamins.

Bybee, Joan L. 1988. Morphology and lexical representation. In: Michael T. Hammond & Michael Noonan (eds.), *Theoretical morphology: Approaches in modern linguistics*. San Diego: Academic Press. 119-141.

Bybee, Joan L. 1995. Regular morphology and the lexicon. *Language and Cognitive Processes* 10 (5). 425-455.

Bybee, Joan L. 2006. From usage to grammar: The mind's response to repetition. *Language* 82 (4). 711-733.

Bybee, Joan L. 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press.

Bybee, Joan L. & Clay Beckner. 2009. Usage-based theory. In: Bernd Heine & Heiko Narrog (eds.), *The Oxford handbook of linguistic analysis*. Oxford: Oxford University Press. 828-856.

Bybee, Joan L. & Mary A. Brewer. 1980. Explanations in morphophonemics: Changes in Provençal and Spanish preterite forms. *Lingua* 52. 201-242.

Bybee, Joan L. & Carol L. Moder. 1983. Morphological classes as natural categories. *Language* 59 (2). 251-270.
Bybee, Joan L. & Dan I. Slobin. 1982. Rules and schemas in the development and use of the English past tense. *Language* 58 (2). 265-289.
Bybee, Joan L. & Sandra Thompson. 1997. Three frequency effects in syntax. *The Annual Proceedings of the Berkeley Linguistics Society* 23. 378-388.
Cameron, Deborah. 1985. *Feminism and linguistic theory*. London: Macmillan.
Caramazza, Alfonso, Alessandro Laudanna & Cristina Romani. 1988. Lexical access and inflectional morphology. *Cognition* 28 (3). 297-332.
Carlier, Anne, Walter de Mulder & Béatrice Lamiroy. 2012. Introduction: The pace of grammaticalization in a typological perspective. *Folia Linguistica* 46 (2). 287-301.
Chafe, Wallace. 1982. Integration and involvement in speaking, writing, and oral literature. In: Deborah Tannen (ed.), *Spoken and written language: Exploring orality and literacy*. Norwood: Ablex. 35-54.
Cheshire, Jenny. 1987. Syntactic variation, the linguistic variable, and sociolinguistic theory. *Linguistics* 25 (2). 257-282.
Cheshire, Jenny. 2002. Sex and gender in variationist research. In: Jack K. Chambers, Peter Trudgill & Natalie Schilling-Estes (eds.), *The handbook of language variation and change*. Malden: Blackwell. 423-443.
Cheung, Lawrence & Longtu Zhang. 2014. Inflectional/periphrastic (I/P) alternation of English superlatives in coordination. Presentation given at the Second Asia Pacific Corpus Linguistics Conference (APCLC 2014), Hongkong.
Cheung, Lawrence & Longtu Zhang. 2016. Determinants of the synthetic–analytic variation across English comparatives and superlatives. *English Language and Linguistics* 20 (3). 559-583.
Chomsky, Noam. 1964. *Current issues in linguistic theory*. Berlin: Mouton de Gruyter.
Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.
Chua, Deborah. 2018. Understanding comparative alternation in *y*-adjectives: What else might we need? *Journal of Linguistics* 54 (3). 459-491.
Clahsen, Harald & Claudia Felser. 2006. Grammatical processing in language learners. *Applied Psycholinguistics* 27 (1). 3-42.

Claridge, Claudia. 2007a. The superlative in spoken English. In: Roberta Fachinetti (ed.), *Corpus linguistics 25 years on*. Amsterdam: Rodopi. 121-148.

Claridge, Claudia. 2007b. Constructing a corpus from the web: Message boards. In: Marianne Hundt, Nadja Nesselhauf & Carolin Biewer (eds.), *Corpus linguistics and the web*. Amsterdam: Rodopi. 87-108.

Clark, Herbert H. & Susan E. Haviland. 1977. Comprehension and the given-new contract. In: Roy O. Freedle (ed.), *Discourse production and comprehension*. Norwood: Ablex. 1-40.

Cleland, Alexandra & Martin Pickering. 2003. The use of lexical and syntactic information in language production: Evidence from the priming of noun-phrase structure. *Journal of Memory and Language* 49 (2). 214-230.

Cohen, Louis, Lawrence Manion & Keith Morrison. ^{6}2007. *Research methods in education*. New York: Routledge.

Cook, Paul & Graeme Hirst. 2012. Do web corpora from top-level domains represent national varieties of English? In: *Proceedings of the 11th International Conference on Textual Data Statistical Analysis, Liège, Belgium (JADT 2012)*. 281-293. URL http://ftp.cs.toronto.edu/pub/gh/Cook+Hirst-2012.pdf.

Cordes, Anne-Kristin. 2017. The roles of analogy, categorization, and generalization in entrenchment. In: Hans-Jörg Schmid (ed.), *Entrenchment and the psychology of language learning: How we reorganize and adapt linguistic knowledge*. Berlin: Mouton de Gruyter. 269-288.

Couper-Kuhlen, Elizabeth. 1986. *An introduction to English prosody*. Tübingen: Niemeyer.

Croft, William. 1991. *Syntactic categories and grammatical relations*. Chicago: University of Chicago Press.

Crystal, David. 2004. *The stories of English*. London: Allen Lane.

Crystal, David. 2011. *Internet linguistics: A student guide*. London: Routledge.

Crystal, David. 2012. Into the twenty-first century. In: Lynda Mugglestone (ed.), *The Oxford history of English*. Oxford: Oxford University Press. 488-513.

Culicover, Peter W. & Ray S. Jackendoff. 1999. The view from the periphery: The English comparative correlative. *Linguistic Inquiry* 30. 543-571.

Curme, George O. 1931. *A grammar of the English language*. Volume II: Syntax. Boston: Heath.

D'Arcy, Alexandra. 2014. Functional partitioning and possible limits on variability: A view of adjective comparison from the vernacular. *Journal of English Linguistics* 42 (3). 218-244.

Dąbrowska, Ewa. 2008. The effects of frequency and neighbourhood density on adult speakers' productivity with Polish case inflections: An empirical test of usage-based approaches to morphology. *Journal of Memory and Language* 58 (4). 931-951.

Davies, Mark & Robert Fuchs. 2015. Expanding horizons in the study of World Englishes with the 1.9 billion word Global Web-based English Corpus (GloWbE). *English World-Wide* 36 (1). 1-28.

Décieux, Jean Philippe, Alexandra Mergener, Kristina Marliese Neufang & Philipp Sischka. 2015. Implementation of the forced answering option within online surveys: Do higher item response rates come at the expense of participation and answer quality? *Psihologija* 45 (4). 311-326.

Derouvray, Cristel & Mick P. Couper. 2002. Designing a strategy for reducing "no opinion" responses in web-based surveys. *Social Science Computer Review* 20 (1). 3-9.

Dewaele, Jean-Marc. 2018. Online questionnaires. In: Aek Phakiti, Peter DeCosta, Luke Plonsky & Sue Starfield (eds.), *The Palgrave handbook of applied linguistics research methodology*. London: Palgrave Macmillan. 269-286.

Diessel, Holger. 2016. Frequency and lexical specificity in grammar: A critical review. In: Heike Behrens & Stefan Pfänder (eds.), *Experience counts: Frequency effects in language*. Berlin: Mouton de Gruyter. 209-237.

Diessel, Holger. 2017. Usage-based linguistics. In: Mark Aronoff (ed.), *Oxford research encyclopedia of linguistics*. Oxford: Oxford University Press. URL https://doi.org/10.1093/acrefore/9780199384655.013.363.

Diessel, Holger. 2019. *The grammar network: How linguistic structure is shaped by language use*. Cambridge: Cambridge University Press.

Diessel, Holger & Martin Hilpert. 2016. Frequency effects in grammar. In Mark Aronoff (ed.), *Oxford research encyclopedia of linguistics*. New York: Oxford University Press. URL https://doi.org/10.1093/acrefore/9780199384655.013.120.

Divjak, Dagmar. 2019. *Frequency in language: Memory, attention and learning*. Cambridge: Cambridge University Press.

Dollinger, Stefan. 2015. *The written questionnaire in social dialectology: History, theory, practice*. Amsterdam: Benjamins.

Dörnyei, Zoltán & Tatsuya Taguchi. [2]2010. *Questionnaires in second language research: Construction, administration, and processing*. New York: Routledge.
Egbert, Jesse & Paul Baker (eds.). 2020. *Using corpus methods to triangulate linguistic analysis*. New York: Routledge.
Ehret, Katharina, Christoph Wolk & Benedikt Szmrecsanyi. 2014. Quirky quadratures: On rhythm and weight as constraints on genitive variation in an unconventional data set. *English Language* 18 (2). 263-303.
Eitelmann, Matthias. 2016. Support for end-weight as a determinant of linguistic variation and change. *English Language and Linguistics* 20 (3). 395-420.
Ellis, Nick C. 1996 Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition* 18 (1). 91-126.
Ellis, Nick C. 2002. Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition* 24 (2). 143-188.
Ellis, Nick C. & Fernando Ferreira-Junior. 2009. Construction learning as a function of frequency, frequency distribution, and function. *The Modern Language Journal* 93 (3). 370-385.
Erdmann, Peter. 1988. On the principle of 'weight' in English. In: Caroline Duncan-Rose & Theo Vennemann (eds.), *On language: Rhetorica, phonologica, syntactica. A festschrift for Robert P. Stockwell from his friends and colleagues*. London: Routledge. 325-339.
Fábregas, Antonio. 2014. Adjectival and adverbial derivation. In: Rochelle Lieber & Pavol Štekauer (eds.), *The Oxford handbook of derivational morphology*. Oxford: Oxford University Press. 276-295.
Faiß, Klaus. 1977. *Aspekte der englischen Sprachgeschichte*. Tübingen: Narr.
Farkas, Donka F. & Katalin É. Kiss. 2000. On the comparative and absolute readings of superlatives. *Natural Language and Linguistic Theory* 18 (3). 417-455.
Fauconnier, Gilles. 1975. Pragmatic scales and logical structure. *Linguistic Inquiry* 6 (3). 353-375.
Fillmore, Charles. 1988. The mechanisms of "construction grammar". *The Annual Proceedings of the Berkeley Linguistics Society* 14. 35-55.
Fillmore, Charles, Paul Kay & Mary O'Connor. 1988. Regularity and idiomaticity in grammatical constructions: The case of *let alone*. *Language* 64 (3). 501-538.

Fodor, Jerry A. 1983. *The modularity of mind: An essay on faculty psychology*. Cambridge, MA: MIT Press.
Frauenfelder, Uli H. & Robert Schreuder. 1992. Constraining psycholinguistic models of morphological processing and representation: The role of productivity. In: Geert Booij & Jaap van Marle (eds.), *Yearbook of morphology 1991*. Dordrecht: Springer. 165-183.
Frazier, Lyn, Alan Munn & Charles Clifton. 2000. Processing coordinate structures. *Journal of Psycholinguistic Research* 29 (4). 343-370.
Fries, Udo. 1993. The comparison of monosyllabic adjectives. In: Andreas H. Jucker (ed.), *The noun phrase in English: Its structure and variability*. Heidelberg: Winter. 25-44.
Fuhrhop, Nanna & Petra Vogel. 2010. Analytisches und Synthetisches im deutschen Superlativ. In: Dagmar Bittner & Livio Gaeta (eds.), *Kodierungstechniken im Wandel: Das Zusammenspiel von Analytik und Synthese im Gegenwartsdeutschen*. Berlin: Mouton de Gruyter. 83-97.
Gergel, Remus. 2015. *Most* historically. In: Chiara Gianollo, Agnes Jäger & Doris Penka (eds.), *Language change at the syntax-semantics interface*. Berlin: Mouton de Gruyter. 101-123.
Gerwin, Johanna. 2014. *Ditransitives in British English dialects*. Berlin: Mouton de Gruyter.
Gibson, Edward. 2000. The dependency locality theory: A distance-based theory of linguistic complexity. In: Alec Marantz, Yasushi Miyashita & Wayne O'Neil (eds.), *Image, language, brain: Papers from the First Mind Articulation Project Symposium*. Cambridge, MA: MIT Press. 95-126.
Givón, Talmy. 1991. Isomorphism in the grammatical code: Cognitive and biological considerations. *Studies in Language* 15 (1). 85-114.
Goldberg, Adele E. 1995. *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.
Goldberg, Adele E. 1996. Construction grammar. In: Edward Keith Brown & Jim Miller (eds.), *Concise encyclopedia of syntactic theories*. Oxford: Pergamon. 68-71.
Goldberg, Adele E. 2003. Constructions: A new theoretical approach to language. *Trends in Cognitive Sciences* 7 (5). 219-224.
Goldblum, Naomi. 2001. *The brain-shaped mind: What the brain can tell us about the mind*. Cambridge: Cambridge University Press.
Gomez, Rebecca L. & LouAnn Gerken. 1999. Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition* 70 (2). 109-135.

González-Díaz, Victorina. 2004. Review article on Günter Rohdenburg & Britta Mondorf (eds.). 2003. Determinants of grammatical variation in English. *English Language and Linguistics* 8 (2). 360-373.
González-Díaz, Victorina. 2006a. On the nature and distribution of English double periphrastic comparison. *The Review of English Studies, New Series* 57 (232). 623-664.
González-Díaz, Victorina. 2006b. The origin of English periphrastic comparatives. *English Studies* 87 (6). 707-739.
González-Díaz, Victorina. 2008. *English adjective comparison: A historical perspective*. Amsterdam: Benjamins.
Gosling, Samuel, Simine Vazire, Sanjay Srivastava & Oliver John. 2004. Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires. *American Psychologist* 59 (2). 93-104.
Grafmiller, Jason. 2014. Variation in English genitives across modality and genres. *English Language* 18 (3). 471-496.
Grafmiller, Jason & Benedikt Szmrecsanyi. 2018. Mapping out particle placement in Englishes around the world: A study in comparative sociolinguistic analysis. *Language Variation and Change* 30 (3). 385-412.
Graziano-King, Janine & Helen Smith Cairns. 2005. Acquisition of English comparative adjectives. *Journal of Child Language* 32 (2). 345-373.
Gregory, Michelle L. & Laura A. Michaelis. 2001. Topicalization and left-dislocation: A functional opposition revisited. *Journal of Pragmatics* 33 (11). 1665-1706.
Grice, Paul H. 1975. Logic and Conversation. In: Peter Cole & Jerry L. Morgan (eds.), *Syntax and semantics*. Volume 3: Speech Acts. New York: Academic Press. 41-58.
Gries, Stefan Th. 2005. Syntactic priming: A corpus-based approach. *Journal of Psycholinguistic Research* 34 (4). 365-399.
Gries, Stefan Th. 2015. The most under-used statistical method in corpus linguistics: Multi-level (and mixed-effects) models. *Corpora* 10 (1). 95-125.
Gries, Stefan Th. 2017. Syntactic alternation research: Taking stock and some suggestions for the future. *Belgian Journal of Linguistics* 31 (2). 8-29.

Grieve, Jack, Douglas Biber, Eric Friginal & Tatiana Nekrasova. 2011. Variation among blogs: A multi-dimensional analysis. In: Alexander Mehler, Serge Sharoff & Marina Santini (eds.), *Genres on the web: Computational models and empirical studies*. Dordrecht: Springer. 303-322.

Grofulović, Jelena & Vladimir Ž. Jovanović. 2016. A study of gender-conditioned use of inflectional and periphrastic forms in English adjective comparison. *Facta Universitatis* 14 (2). 189-208.

Günther, Christine. 2018. *The rich, the poor, the obvious*: Arguing for an ellipsis analysis of "adjectives used as nouns". In: Alex Ho-Cheong Leung & Wim van der Wurff (eds.), *The noun phrase in English: Past and present*. Amsterdam: Benjamins. 77-112.

Halliday, Michael A. K. 1967. Notes on transitivity and theme in English: Part 2. *Journal of Linguistics* 3 (2). 199-244.

Hawkins, John A. 1992. Syntactic weight versus information structure in word order variation. In: Joachim Jacobs (ed.), *Informationsstruktur und Grammatik*. Wiesbaden: Verlag für Sozialwissenschaften. 196-219.

Hawkins, John A. 1995. *A performance theory of order and constituency*. Cambridge: Cambridge University Press.

Hawkins, John A. 1999. Processing complexity and filler-gap dependencies across grammars. *Language* 75 (2). 244-285.

Hawkins, John A. 2001. Why are categories adjacent? *Journal of Linguistics* 37 (1). 1-34.

Hawkins, John A. 2004. *Efficiency and complexity in grammars*. Oxford: Oxford University Press.

Hay, Jennifer. 2001. Lexical frequency in morphology: Is everything relative? *Linguistics* 39 (6). 1041-1070.

Hay, Jennifer. 2003. *Causes and consequences of word structure*. New York: Routledge.

Heckathorn, Douglas D. 2002. Respondent-driven sampling II: Deriving valid population estimates from chain-referral samples of hidden populations. *Social Problems* 49 (1). 11-34.

Heerwegh, Dirk & Geert Loosveldt. 2002. An evaluation of the effect of response formats on data quality in web surveys. *Social Science Computer Review* 20 (4). 471-484.

Heerwegh, Dirk & Geert Loosveldt. 2008. Face-to-face versus web surveying in a high-internet-coverage population: Differences in response quality. *Public Opinion Quarterly* 72 (5). 836-846.

Heine, Bernd. 1997. *Cognitive foundations of grammar*. Oxford: Oxford University Press.
Heller, Benedikt, Benedikt Szmrecsanyi & Jason Grafmiller. 2017. Stability and fluidity in syntactic variation world-wide: The genitive alternation across varieties of English. *Journal of English Linguistics* 45 (1). 3-27.
Helm, Gerhard. 1991. *Symbolische und konnektionistische Modelle der menschlichen Informationsverarbeitung: Eine kritische Gegenüberstellung*. Berlin: Springer.
Hilpert, Martin. 2008. The English comparative – language structure and language use. *English Language and Linguistics* 12 (3). 395-417.
Hilpert, Martin. 2010. Comparing comparatives: A corpus-based study of comparative constructions in English and Swedish. In: Hans Christian Boas (ed.), *Contrastive studies in construction grammar*. Amsterdam: Benjamins. 21-41.
Hilpert, Martin. 2014. *Construction grammar and its application to English*. Edinburgh: Edinburgh University Press.
Hinrichs, Lars & Benedikt Szmrecsanyi. 2007. Recent changes in the function and frequency of standard English genitive constructions: A multivariate analysis of tagged corpora. *English Language and Linguistics* 11 (3). 437-474.
Hinton, Geoffrey E., James L. McClelland & David E. Rumelhart. 1986. Distributed representations. In: James L. McClelland, David E. Rumelhart & the PDP Research Group (eds.), *Parallel distributed processing: Explorations in the microstructure of cognition*. Vol. 1: Foundations. Cambridge, MA: MIT Press. 77-109.
Hirotani, Masako, Lyn Frazier & Keith Rayner. 2006. Punctuation and intonation effects on clause and sentence wrap-up: Evidence from eye movements. *Journal of Memory and Language* 54 (3). 425-443.
Hoffmann, Sebastian. 2004. Are low-frequency complex prepositions grammaticalized? On the limits of corpus data – and the importance of intuition. In: Hans Lindquist & Christian Mair (eds.), *Corpus approaches to grammaticalization in English*. Amsterdam: Benjamins. 171-210.
Hoffmann, Thomas. 2019. *English comparative correlatives: Diachronic and synchronic variation at the lexicon-syntax interface*. Cambridge: Cambridge University Press.
Hoffmann, Thomas, Jakob Horsch & Thomas Brunner. 2019. The more data, the better: A usage-based account of the English comparative correlative construction. *Cognitive Linguistics* 30 (1). 1-36.

Hohaus, Vera, Sonja Tiemann & Sigrid Beck. 2014. Acquisition of comparison constructions. *Language Acquisition* 21 (3). 215-249.
Huddleston, Rodney D. & Geoffrey K. Pullum. 2002. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.
Humboldt, Wilhelm von. 1836. *Über die Verschiedenheit des menschlichen Sprachbaues und ihren Einfluss auf die geistige Entwicklung des Menschengeschlechts*. Berlin: Dümmler.
Humboldt, Wilhelm von. 1999 [1836]. *On language: On the diversity of human language construction and its influence on the mental development of the human species* [Über die Verschiedenheit des menschlichen Sprachbaues und ihren Einfluss auf die geistige Entwicklung des Menschengeschlechts]. Ed. Michael Losonsky. Translated by Peter Heath. Cambridge: Cambridge University Press.
Jackendoff, Ray. 1972. *Semantic interpretation in generative grammar*. Cambridge, MA: MIT Press.
Jäschke, Katja & Ingo Plag. 2016. The dative alternation in German-English interlanguage. *Studies in Second Language Acquisition* 38 (3). 485-521.
Jenset, Gard B., Barbara McGillivray & Michael Rundell. 2018. The dative alternation revisited: Fresh insights from contemporary British spoken data. In: Vaclav Brezina, Robbie Love & Karin Aijmer (eds.), *Corpus approaches to contemporary British speech: Sociolinguistic studies of the Spoken BNC2014*. New York: Routledge. 185-208.
Jespersen, Otto. 1949. *A Modern English grammar on historical principles*. Part VII: Syntax. Copenhagen: Munksgaard.
Jones, Daniel, Peter Roach, Jane Setter & John Esling. ^{15}1997. *Cambridge English pronouncing dictionary*. Cambridge: Cambridge University Press.
Jones, Daniel, Peter Roach, Jane Setter & John Esling. ^{17}2006. *Cambridge English pronouncing dictionary*. Cambridge: Cambridge University Press.
Jusczyk, Peter W. 1997. *The discovery of spoken language*. Cambridge, MA: MIT Press.
Just, Marcel A. & Patricia A. Carpenter. 1980. A theory of reading: From eye fixations to comprehension. *Psychological Review* 87 (4). 329-354.
Kachru, Braj. B. ^{2}1992. Teaching World Englishes. In: Braj B. Kachru (ed.), *The other tongue: English across cultures*. Urbana: University of Illinois Press. 355-365.

Kager, René. 1989. *A metrical theory of stress and destressing in English and Dutch*. Utrecht: University of Utrecht PhD dissertation.
Kalton, G., Julie Roberts & D. Holt. 1980. The effects of offering a middle response option with opinion questions. *The Statistician* 29 (1). 65-78.
Keizer, Evelien. 2007. *The English noun phrase: The nature of linguistic categorization*. Cambridge: Cambridge University Press.
Kelly, Michael H. & Kathryn J. Bock. 1988. Stress in time. *Journal of Experimental Psychology: Human Perception and Performance* 14 (3). 389-403.
Kırkıcı, Bilal. 2012. The L2 production of English comparative structures: More support for L2 developmental changes. *Çankaya University Journal of Humanities and Social Sciences* 9 (1). 1-15.
Klavan, Jane & Dagmar Divjak. 2016. The cognitive plausibility of statistical classification models: Comparing textual and behavioral evidence. *Folia Linguistica* 50 (2). 355-384.
Koller, Erwin. 2007. *Zur Grammatik von Elativ und Superlativ*. Mannheim: Leibniz-Institut für Deutsche Sprache.
König, Ekkehard. 1994. English. In: Ekkehard König & Johan van der Auwera (eds.), *The Germanic languages*. London: Routledge. 532-565.
Köpcke, Klaus-Michael. 1998. The acquisition of plural marking in English and German revisited: Schemata versus rules. *Journal of Child Language* 25 (2). 293-319.
Kortmann, Bernd & Benedikt Szmrecsanyi. 2012. Introduction. In: Bernd Kortmann & Benedikt Szmrecsanyi (eds.), *Linguistic complexity: Second language acquisition, indigenization, contact*. Berlin: Mouton de Gruyter. 6-34.
Kövecses, Zóltan. 2000. *American English: An introduction*. Peterborough, Ontario: Broadview Press.
Kreuter, Frauke (ed.). 2013. *Improving surveys with paradata: Analytic uses of process information*. Hoboken: Wiley.
Krosnick, Jon A. 1991. Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology* 5 (3). 213-236.
Krug, Manfred & Katrin Sell. 2013. Designing and conducting interviews and questionnaires. In: Manfred Krug & Julia Schlüter (eds.), *Research methods in language variation and change*. Cambridge: Cambridge University Press. 69-98.
Kruschke, John K. 1992. ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review* 99 (1). 22-44.

Kunter, Gero. 2015. Effects of processing complexity in perception and production: The case of English comparative alternation. In: Vito Pirrelli, Claudia Marzi & Marcello Ferro (eds.), *Word structure and word usage: Proceedings of the NetWordS Final Conference, Pisa, March 30-April 1, 2015*. 32-36.

Kunter, Gero. 2017. *Processing complexity and the alternation between analytic and synthetic forms in English*. Düsseldorf: University of Düsseldorf postdoctoral thesis.

Kytö, Merja & Suzanne Romaine. 1997. Competing forms of adjective comparison in Modern English: What could be more quicker and easier and more effective? In: Terttu Nevalainen & Leena Kahlas Tarkka (eds.), *To explain the present: Studies in the changing English language in honour of Matti Rissanen*. Helsinki: Société Néophilologique. 329-352.

Kytö, Merja & Suzanne Romaine. 2000. Adjective comparison and standardization processes in American and British English from 1620 to the present. In: Laura Wright (ed.), *The development of standard English 1300-1800: Theories, descriptions, conflicts*. Cambridge: Cambridge University Press. 171-194.

Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.

Labov, William. 1994. *Principles of linguistic change*. Volume 1: Internal factors. Oxford: Wiley-Blackwell.

Labov, William. 2001. *Principles of linguistic change*. Volume 2: Social factors. Oxford: Wiley-Blackwell.

LaFave, Nathan. 2015. The most apt experimental investigation of English comparative and superlative formation. *University of Pennsylvania Working Papers in Linguistics* 21 (1). Art. 16.

LaFave, Nathan. 2016. Social factors and lexical frequency influencing English adjective gradation in speech and CMC. In: Lauren Squires (ed.), *English in computer-mediated communication: Variation, representation, and change*. Berlin: Mouton de Gruyter. 301-326.

Langacker, Ronald W. 1987. *Foundations of cognitive grammar*. Volume I: Theoretical prerequisites. Stanford: Stanford University Press.

Langacker, Ronald W. 2008. *Cognitive grammar: A basic introduction*. Oxford: Oxford University Press.

Larson, Richard K. 1988. On the double object construction. *Linguistic Inquiry* 19 (3). 335-391.

Lass, Roger. 1994. *Old English: A historical linguistic companion*. Cambridge: Cambridge University Press.

Laudanna, Alessandro, William Badecker & Alfonso Caramazza. 1992. Processing inflectional and derivational morphology. *Journal of Memory and Language* 31 (3). 333-348.

Leben, William R. 1973. *Suprasegmental phonology*. Cambridge, MA: Massachusetts Institute of Technology PhD dissertation.

Lee, Ming-Wei & Julie Gibbons. 2007. Rhythmic alternation and the optional complementiser in English: New evidence of phonological influence on grammatical encoding. *Cognition* 105 (2). 446-456.

Leech, Geoffrey & Jonathan Culpeper. 1997. The comparison of adjectives in recent British English. In: Terttu Nevalainen & Leena Kahlas Tarkka (eds.), *To explain the present: Studies in the changing English language in honour of Matti Rissanen*. Helsinki: Société Néophilologique. 353-374.

Leech, Geoffrey, Marianne Hundt, Christian Mair & Nicholas Smith. 2009. *Change in contemporary English: A grammatical study*. Cambridge: Cambridge University Press.

Levelt, Willem J. M. 1989. *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Levelt, Willem J. M. & Stephanie Kelter. 1982. Surface form and memory in question answering. *Cognitive Psychology* 14 (1). 78-106.

Levshina, Natalia. 2018. Probabilistic grammar and constructional predictability: Bayesian generalized additive models of *help* + (*to*) infinitive in varieties of web-based English. *Glossa* 3 (1): 55. 1-22.

LimeSurvey. 2018. *LimeSurvey: An open source survey tool*. Version 3.5.1. LimeSurvey GmbH Hamburg, Germany. URL http://www.limesurvey.org.

Lindquist, Hans. 1998. The comparison of English disyllabic adjectives in *-y* and *-ly* in present-day British and American English. In: Hans Lindquist, Staffan Klintborg, Magnus Levin & Maria Estling (eds.), *The major varieties of English. Papers from MAVEN 97, Växjö 20-22 November 1997*. Växjö: Växjö University. 205-212.

Lindquist, Hans. 2000. *Livelier* or *more lively*? Syntactic and contextual factors influencing the comparison of disyllabic adjectives. In: John M. Kirk (ed.), *Corpora galore: Analyses and techniques in describing English*. Papers from the 19th International Conference on English Language Research on Computerised Corpora (ICAME 1998). Amsterdam: Rodopi. 125-132.

Litcofsky, Kaitlyn A. & Janet G. van Hell. 2019. Bi-directional evidence linking sentence production and comprehension: A cross-modality structural priming study. *Frontiers in Psychology* 10: 1095.

Lohmann, Arne. 2010. Review article on Mondorf, Britta. 2009. More support for *more*-support: The role of processing constraints on the choice between synthetic and analytic forms. *Corpus Linguistics and Linguistic Theory* 6 (2).301-311.

Lohmann, Arne. 2014. *English coordinate constructions. A processing perspective on constituent order*. Cambridge: Cambridge University Press.

Loureiro-Porto, Lucía. 2017. ICE vs GloWbE: Big data and corpus compilation. *World Englishes* 36 (3). 448-470. DOI: 10.1111/weng.12281.

Love, Robbie, Claire Dembry, Andrew Hardie, Vaclav Brezina & Tony McEnery. 2017. The Spoken BNC2014: Designing and building a spoken corpus of everyday conversations. *International Journal of Corpus Linguistics* 22 (3). 319-344.

MacKay, Donald G. 1987. *The organization of perception and action: A theory for language and other cognitive skills*. New York: Springer.

MacWhinney, Brian. 1978. The acquisition of morphophonology. *Monographs of the Society for Research in Child Development* 43 (1-2).

Mair, Christian. 1990. *Infinitival complement clauses in English: A study of syntax in discourse*. Cambridge: Cambridge University Press.

Mair, Christian. 2003. Gerundial complements after *begin* and *start*: Grammatical and sociolinguistic factors, and how they work against each other. In: Günter Rohdenburg & Britta Mondorf (eds.), *Determinants of grammatical variation in English*. Berlin: Mouton de Gruyter. 329-345.

Mair, Christian. 2006. Tracking ongoing grammatical change and recent diversification in present-day standard English: The complementary role of small and large corpora. In: Antoinette Renouf & Andrew Kehoe (eds.), *The changing face of corpus linguistics*. Amsterdam: Rodopi. 355-376.

Mair, Christian. 2015. Response to Davies and Fuchs. *English World-Wide* 36 (1). 29-33.

Maniaci, Michael R. & Ronald D. Rogge. [2]2014. Conducting research on the Internet. In: Harry T. Reis & Charles M. Judd (eds.), *Handbook of research methods in social and personality psychology*. Cambridge: Cambridge University Press. 443-470.

Marcus, Gary. F., S. Vijayan, Shoba Bandi Rao & Peter M. Vishton. 1999. Rule learning by seven-month-old infants. *Science* 283 (5398). 77-80.

Matsui, Chie. 2010. *–er* type or *more* type adjectives of comparison in English. *Studia Neophilologica* 82 (2). 188-202.

Matthews, Peter H. 2014. *The positions of adjectives in English*. Oxford: Oxford University Press.
McCarthy, John. 1986. OCP effects: Gemination and antigemination. *Linguistic Inquiry* 17 (2). 207-263.
McCauley, Stewart & Morton Christiansen. 2014. Acquiring formulaic language: A computational model. *The Mental Lexicon* 9. 419-436.
McClelland, James L. & David E. Rumelhart. 1981. An interactive activation model of context effects in letter perception. Part I: An account of basic findings. *Psychological Review* 48. 375-407.
McClelland, James L. & Karalyn Patterson. 2002. Rules or connections in past-tense inflections: What does the evidence rule out? *Trends in Cognitive Sciences* 6 (11). 465-472.
McDonald, Janet L, Kathryn J. Bock & Michael H. Kelly. 1993. Word and world order: semantic, phonological, and metrical determinants of serial position. *Cognitive Psychology* 25 (2). 188-230.
McElreath, Richard. 2016. *rethinking: Statistical Rethinking book package*. R package version 1.59.
McElreath, Richard. 2020. *Statistical rethinking: A Bayesian course with examples in R and Stan*. Boca Raton: CRC Press.
Menn, Lise & Brian MacWhinney. 1984. The repeated morph constraint: Toward an explanation. *Language* 60 (3). 519-541.
Michaelis, Laura A. 2004. Type shifting in construction grammar: An integrated approach to aspectual coercion. *Cognitive Linguistics* 15 (1). 1-67.
Michaelis, Susanne & Martin Haspelmath. 2017. Analytic and synthetic: Typological change in varieties of European languages. In: Isabelle Buchstaller & Beat Siebenhaar (eds.), *Language variation – European perspectives VI: Selected papers from the 8th International Conference on Language Variation in Europe (ICLaVE 8), Leipzig 2015*. Amsterdam: Benjamins. 1-17.
Mitchell, Bruce & Fred C. Robinson. [7]2007. *A guide to Old English*. Malden: Blackwell.
Mondorf, Britta. 2002. The effect of prepositional complements on the choice of synthetic or analytic comparatives. In: Hubert Cuyckens & Günter Radden (eds.), *Perspectives on prepositions*. Tübingen: Niemeyer. 65-78.
Mondorf, Britta. 2003. Support for *more*-support. In: Günter Rohdenburg & Britta Mondorf (eds.), *Determinants of grammatical variation in English*. Berlin: Mouton de Gruyter. 251-304.

Mondorf, Britta. 2009a. Synthetic and analytic comparatives. In: Günter Rohdenburg & Julia Schlüter (eds.), *One language, two grammars? Differences between British and American English*. Cambridge: Cambridge University Press. 86-107.

Mondorf, Britta. 2009b. *More support for* more-*support: The role of processing constraints on the choice between synthetic and analytic comparative forms*. Amsterdam: Benjamins.

Mondorf, Britta. forthcoming. Grammatical Differences Between British and American English. In: Carol A. Chapelle (ed.), *The Encyclopedia of Applied Linguistics*.

Monsell, Stephen. 1991. The nature and locus of word frequency effects in reading. In: Derek Besner & Glyn W. Humphreys (eds.), *Basic processes in reading: Visual word recognition*. Hoboken: Taylor & Francis.148-197.

Mukherjee, Joybrato. 2015. Response to Davies and Fuchs. *English World-Wide* 36 (1). 34-37.

Mustanoja, Tauno F. 1960. *A Middle English syntax*. Part I: Parts of speech. Helsinki: Société Néophilologique.

Nelson, Gerald. 2015. Response to Davies and Fuchs. *English World-Wide* 36 (1). 38-40.

Nespor, Marina & Irene Vogel. 1989. On clashes and lapses. *Phonology* 6 (1). 69-116.

Nosofsky, Robert M. 1988. Similarity, frequency, and category representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 14 (1). 54-65.

Nosofsky, Robert M. 1992. Exemplars, prototypes, and similarity rules. In: Alice F. Healy, Stephen Michael Kosslyn & Richard M. Shiffrin (eds.), *From learning theory to connectionist theory: Essays in honor of William K. Estes*. Volume II. Hillsdale: Erlbaum. 149-167.

Nosofsky, Robert M., John K. Kruschke & Stephen C. McKinley. 1992. Combining exemplar-based category representations and connectionist learning rules. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 18 (2). 211-233.

Olsen, Susan. 1988. Das „substantivierte" Adjektiv im Deutschen und Englischen: Attribuierung vs. syntaktische „Substantivierung". *Folia Linguistica* 22. 337-372.

Peschl, Markus. 1996. Repräsentation in natürlichen und künstlichen (konnektionistischen) neuronalen Systemen. In: Gerhard Strube (ed.), *Wörterbuch der Kognitionswissenschaft*. Stuttgart: Klett-Cotta. 579-580.

Peters, Pam. 2000. Paradigm split. In: Christian Mair & Marianne Hundt (eds.), *Corpus linguistics and linguistic theory: Papers from the twentieth international conference on English language research using computerized corpora (ICAME 20) Freiburg im Breisgau 1999*. Amsterdam: Rodopi. 301-312.

Peters, Pam. 2015. Response to Davies and Fuchs. *English World-Wide* 36 (1). 41-44.

Pfänder, Stefan & Heike Behrens. 2016. Experience counts: An introduction to frequency effects in language. In: Heike Behrens & Stefan Pfänder (eds.), *Experience counts: Frequency effects in language*. Berlin: Mouton de Gruyter. 1-20.

Pickering, Martin J. & Holly P. Branigan. 1998. The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language* 39 (4). 633-651.

Pickering, Martin J. & Victor S. Ferreira. 2008. Structural priming: A critical review. *Psychological Bulletin* 134 (3). 427-459.

Pierrehumbert, Janet B. 2001. Exemplar dynamics: Word frequency, lenition and contrast. In: Joan L. Bybee & Paul J. Hopper (eds.), *Frequency and the emergence of linguistic structure*. Amsterdam: Benjamins. 138-157.

Pinker, Steven. 1999. *Words and rules: The ingredients of language*. New York: Perennial.

Pinker, Steven & Alan Prince. 1988. On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition* 28 (1-2). 73-193.

Pinker, Steven & Michael T. Ullman. 2002. The past and future of the past tense. *Trends in Cognitive Sciences* 6 (11). 456-463.

Plag, Ingo. 1998. Morphological haplology in a constraint-based morpho-phonology. In: Wolfgang Kehrein & Richard Wiese (eds.), *Phonology and morphology of the Germanic languages*. Berlin: Mouton de Gruyter. 199-216.

Pospeschill, Markus. 2004. *Konnektionismus und Kognition: Eine Einführung*. Stuttgart: Kohlhammer.

Potter, Mary C. & Linda Lombardi. 1998. Syntactic priming in immediate recall of sentences. *Journal of Memory and Language* 38 (3). 265-282.

Potter, Simeon. 1969. *Changing English*. London: André Deutsch.

Pound, Louise. 1901. *The comparison of adjectives in English in the XV and the XVI century*. Heidelberg: Carl Winter.

Poutsma, Hendrik. 1914. *A grammar of late modern English*. Part II: The parts of speech. Section I A: Nouns, adjectives and articles. Groningen: Noordhoff.
Prasada, Sandeep & Steven Pinker. 1993. Generalisation of regular and irregular morphological patterns. *Language and Cognitive Processes* 8 (1). 1-56.
Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1985. *A comprehensive grammar of the English language*. Harlow: Longman.
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/.
Radatz, Hans-Ingo. 2001. *Die Semantik der Adjektivstellung: Eine kognitive Studie zur Konstruktion "Adjektiv + Substantiv" im Spanischen, Französischen und Italienischen*. Tübingen: Niemeyer.
Rauh, Gisa. 2010. *Syntactic categories: Their identification and description in linguistic theories*. Oxford: Oxford University Press.
Rivero, María-Luisa. 1980. On left-dislocation and topicalization in Spanish. *Linguistic Inquiry* 11 (2). 363-393.
Rohdenburg, Günter & Julia Schlüter. 2000. Determinanten grammatischer Variation im Früh- und Spätneuenglischen. *Sprachwissenschaft* 25. 443-496.
Rohdenburg, Günter. 1995. On the replacement of finite complement clauses by infinitives in English. *English Studies* 76 (4). 367-388.
Rohdenburg, Günter. 1996. Cognitive complexity and increased grammatical explicitness in English. *Cognitive Linguistics* 7 (2). 149-182.
Rohdenburg, Günter. 2003. *Horror aequi* and cognitive complexity as factors determining the use of interrogative clause linkers. In: Günter Rohdenburg & Britta Mondorf (eds.), *Determinants of grammatical variation in English*. Berlin: Mouton de Gruyter. 205-250.
Rohdenburg, Günter. 2009. Reflexive structures. In: Günter Rohdenburg & Julia Schlüter (eds.), *One language, two grammars? Differences between British and American English*. Cambridge: Cambridge University Press. 166-181.
Rohr, Anny. 1929. *Die Steigerung des neuenglischen Eigenschaftswortes im 17. und 18. Jahrhundert mit Ausblicken auf den Sprachgebrauch der Gegenwart*. Giessen: University of Giessen PhD dissertation.
Rosch, Eleanor. 1978. Principles of categorization. In: Eleanor Rosch & Barbara Lloyd (eds.), *Cognition and categorization*. Hillsdale: Erlbaum. 27-48.

Rosch, Eleanor & Carolyn B. Mervis. 1975. Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology* 7 (4). 573-605.
Rosenbach, Anette. 2002. *Genitive variation in English: Conceptual factors in synchronic and diachronic studies*. Berlin: Mouton de Gruyter.
Rosenbach, Anette. 2005. Animacy versus weight as determinants of grammatical variation in English. *Language* 81 (3). 613-644.
Roßmann, Joss. 2017. *Satisficing in Befragungen: Theorie, Messung und Erklärung*. Wiesbaden: Springer.
Röthlisberger, Melanie. 2018. *Regional variation in probabilistic grammars: A multifactorial study of the English dative alternation*. Leuven: KU Leuven unpublished doctoral dissertation.
Röthlisberger, Melanie, Jason Grafmiller & Benedikt Szmrecsanyi. 2017. Cognitive indigenization effects in the English dative alternation. *Cognitive Linguistics* 28 (4). 673-710.
Rumelhart, David E. & James L. McClelland (eds.). 1986. *Parallel distributed processing: Explorations in the microstructure of cognition*. Cambridge, MA: MIT Press.
Rumelhart, David E. & James L. McClelland. 1986. On learning the past tenses of English verbs. In: David E. Rumelhart & James L. McClelland (eds.), *Parallel distributed processing: Explorations in the microstructure of cognition*. Volume 2: Psychological and Biological Models. Cambridge, MA: MIT Press. 216-271.
Rusiecki, Jan. 1985. *Adjectives and comparison in English: A semantic study*. London: Longman.
Ryu, Na-Young & Sung-Hoon Hong. 2013. Schwa deletion in the conversational speech of English: The role of linguistic factors. *Linguistic Research* 30. 313-333.
Saffran, Jenny R., Richard N. Aslin & Elissa L. Newport. 1996. Statistical learning by 8-month-old infants. *Science* 274 (5294). 1926-1928.
Säily, Tanja, Victorina González-Díaz & Jukka Suomela. 2018. Variation in the productivity of adjective comparison in present-day English. In: Vaclav Brezina, Robbie Love & Karin Aijmer (eds.), *Corpus approaches to contemporary British speech: Sociolinguistic studies of the Spoken BNC2014*. London: Routledge. 159-184.
Salganik, Matthew J. & Douglas D. Heckathorn. 2004. Sampling and estimation in hidden populations using respondent-driven sampling. *Sociological Methodology* 34 (1). 193-240.
Sanchez-Stockhammer, Christina. 2018. *English compounds and their spelling*. Cambridge: Cambridge University Press.

Sarkar, Deepayan. 2008. *Lattice: Multivariate data visualization with R*. New York: Springer.
Schade, Ulrich. 1999. *Konnektionistische Sprachproduktion*. Wiesbaden: Deutscher Universitätsverlag.
Schibsbye, Knud. [2]1970. *A modern English grammar*. London: Oxford University Press.
Schlegel, August Wilhelm von. 1818. *Observations sur la langue et la littérature provençales*. Paris: Librairie grecque-latine-allemande.
Schlüter, Julia. 2001. Why *worser* is better: The double comparative in 16th- to 17th-century English. *Language Variation and Change* 13. 193-208.
Schlüter, Julia. 2005. *Rhythmic grammar: The influence of rhythm on grammatical variation and change in English*. Berlin: Mouton de Gruyter.
Schlüter, Julia. 2009. Phonology and grammar. In: Günter Rohdenburg und Julia Schlüter (eds.), *One language, two grammars? Differences between British and American English*. Cambridge: Cambridge University Press. 108-129.
Schlüter, Julia. 2015. Rhythmic influence on grammar: Scope and limitations. In: Ralf Vogel & Ruben van de Vijver (eds.), *Rhythm in cognition and grammar: A Germanic perspective*. Berlin: Mouton de Gruyter. 179-205.
Schmid, Hans-Jörg. 2005. *Englische Morphologie und Wortbildung: Eine Einführung*. Berlin: Erich Schmidt.
Schmidt, Richard W. & Carol F. McCreary. 1977. Standard and superstandard English: Recognition and use of prescriptive rules by native and non-native speakers. *TESOL Quarterly* 11 (4). 415-429.
Schonlau, Matthias & Vera Toepoel. 2015. Straightlining in web survey panels over time. *Survey Research Methods* 9 (2). 125-137.
Schütze, Carson T. 2016. *The empirical base of linguistics: Grammaticality judgements and linguistic methodology*. Berlin: Language Science Press.
Selkirk, Elisabeth. 1977. Some remarks on noun phrase structure. In: Peter W. Culicover, Thomas Wasow & Adrian Akmajian (eds.), *Formal syntax: Papers from the MSSB-UC Irvine Conference on the Formal Syntax of Natural Language, June 9 - 11, 1976, Newport Beach, California*. New York: Academic Press. 285-316.
Sereno, Joan A. & Allard Jongman. 1997. Processing of English inflectional morphology. *Memory & Cognition* 25 (4). 425-437.
Sharvit, Yael & Penka Stateva. 2002. Superlative expressions, context and focus. *Linguistics and Philosophy* 25 (4). 453-504.

Shih, Stephanie, Jason Grafmiller, Richard Futrell & Joan Bresnan. 2015. Rhythm's role in genitive construction choice in spoken English. In: Ralf Vogel & Ruben van de Vijver (eds.), *Rhythm in cognition and grammar: A Germanic perspective*. Berlin: Mouton de Gruyter. 207-233.

Silva, Renita & Harald Clahsen. 2008. Morphologically complex words in L1 and L2 processing: Evidence from masked priming experiments in English. *Bilingualism: Language and Cognition* 11 (2). 245-260.

Sinclair, John. 1991. *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Sonnenstuhl-Henning, Ingrid. 2003. *Deutsche Plurale im mentalen Lexikon: Experimentelle Untersuchungen zum Verhältnis von Speicherung und Dekomposition*. Berlin: Mouton de Gruyter.

Sönning, Lukas. 2018. Multilevel models for corpus linguistic research: A conceptual introduction. Unpublished manuscript. URL https://osf.io/hpvwz/.

Sönning, Lukas. 2020. *Phonological variation in German learner English*. Bamberg: Bamberg University Press.

Sönning, Lukas. 2022. Basic statistical methods for language research. Unpublished manuscript. URL https://osf.io/v37pw/.

Sönning, Lukas & Stefan Hartmann. 2019. The English comparative alternation revisited: A fresh look at theory and data. Presentation held at ICAME40, Neuchâtel, Switzerland, June 2019. URL https://osf.io/bgrp8/.

Spencer, Andrew & Gergana Popova. 2015. Periphrasis and inflection. In: Matthew Baerman (ed.), *The Oxford handbook of inflection*. Oxford: Oxford University Press. 197-230.

Sprouse, Jon. 2011. A validation of Amazon Mechanical Turk for the collection of acceptability judgments in linguistic theory. *Behavior research methods* 43 (1). 155-167.

Stan Development Team. 2020. *RStan: The R interface to Stan*. R package version 2.19.3. URL http://mc-stan.org/.

Stefanowitsch, Anatol. 2013. Collostructional analysis. In: Thomas Hoffmann & Graeme Trousdale (eds.), *The Oxford handbook of construction grammar*. Oxford: Oxford University Press. 290-306.

Stefanowitsch, Anatol. 2020. *Corpus linguistics: A guide to the methodology*. Berlin: Language Science Press.

Stefanowitsch, Anatol & Stefan Th. Gries. 2003. Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8 (2). 209-243.

Stemberger, Joseph P. & Brian MacWhinney. 1986. Frequency and the lexical storage of regularly inflected forms. *Memory & Cognition* 14 (1). 17-26.
Strang, Barbara. 1970. *A history of English*. London: Methuen.
Sudman, Seymour & Norman M. Bradburn. 1982. *Asking questions: A practical guide to questionnaire design*. San Francisco: Jossey-Bass.
Sweet, Henry. Reprint. 2014. *A new English grammar: Logical and historical*. Volume 1: Introduction, phonology, and accidence. Cambridge: Cambridge University Press. Original edition, 1891.
Szmrecsanyi, Benedikt. 2006. *Morphosyntactic persistence in spoken English: A corpus study at the intersection of variationist sociolinguistics, psycholinguistics, and discourse analysis*. Berlin: Mouton de Gruyter.
Szmrecsanyi, Benedikt. 2009. Typological parameters of intralingual variability: Grammatical analyticity versus syntheticity in varieties of English. *Language Variation and Change* 21 (3). 319-353.
Szmrecsanyi, Benedikt. 2012. Analyticity and syntheticity in the history of English. In: Terttu Nevalainen & Elizabeth Closs Traugott (eds.), *The Oxford handbook of the history of English*. Oxford: Oxford University Press. 654-665.
Szmrecsanyi, Benedikt. 2016. An analytic-synthetic spiral in the history of English. In: Elly van Gelderen (ed.), *Cyclical change continued*. Amsterdam: Benjamins. 93-112.
Szmrecsanyi, Benedikt. 2019. Exploring probabilistic grammar(s) in varieties of English. Plenary presentation held at the 8th Biennial International Conference on the Linguistics of Contemporary English (BICLCE8), Bamberg, Germany, 26-29 September 2019.
Szmrecsanyi, Benedikt, Jason Grafmiller, Joan Bresnan, Anette Rosenbach, Sali Tagliamonte & Simon Todd. 2017. Spoken syntax in a comparative perspective: The dative and genitive alternation in varieties of English. *Glossa* 2 (1). 1-27.
Taft, Marcus & Kenneth I. Forster. 1975. Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior* 14 (6). 638-647.
Tagliamonte, Sali A. 2012. *Variationist sociolinguistics: Change, observation, interpretation*. Malden: Wiley-Blackwell.
Tagliamonte, Sali A. 2013. Comparative Sociolinguistics. In: Jack K. Chambers & Natalie Schilling-Estes (eds.), *The handbook of language variation and change*. Chichester: Wiley-Blackwell. 729-762.

Tagliamonte, Sali A. & Harald R. Baayen. 2012. Models, forests, and trees of York English: *Was/were* variation as a case study for statistical practice. *Language Variation and Change* 24 (2). 135-178.
Taylor, John R. 2002. *Cognitive grammar*. Oxford: Oxford University Press.
Taylor, John R. 2012. *The mental corpus: How language is represented in the mind*. Oxford: Oxford University Press.
Thomas, Michael S. C. & James L. McClelland. 2012. Connectionist models of cognition. In: Ron Sun (ed.), *The Cambridge handbook of computational psychology*. Cambridge: Cambridge University Press. 23-58.
Tiffin-Richards, Simon P. & Sascha Schroeder. 2018. The development of wrap-up processes in text reading: A study of children's eye movements. *Journal of Experimental Psychology: Learning, memory, and cognition* 44 (7). 1051-1063.
Tomasello, Michael. 1992. *First verbs: A case study of early grammatical development*. Cambridge: Cambridge University Press.
Tomasello, Michael. 2003. *Constructing a language. A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
Tomasello, Michael. 2008. Konstruktionsgrammatik und früher Erstspracherwerb. In: Kerstin Fischer & Anatol Stefanowitsch (eds.), *Konstruktionsgrammatik: Von der Anwendung zur Theorie*. Tübingen: Stauffenburg. 19-36.
Tottie, Gunnel. 2002. *An introduction to American English*. Malden: Blackwell.
Tourangeau, Roger, Mick P. Couper & Fred Conrad. 2004. Spacing, position, and order: Interpretive heuristics for visual features of survey questions. *Public Opinion Quarterly* 68 (3). 368-393.
Tremblay, Antoine & Harald Baayen. 2010. Holistic processing of regular four-word sequences: A behavioural and ERP study of the effects of structure, frequency, and probability on immediate free recall. In: David Wood (ed.), *Perspectives on formulaic language: Acquisition and communication*. London: Continuum. 151-173.
Uhrig, Peter. 2015. Why the principle of no synonymy is overrated. *Zeitschrift für Anglistik und Amerikanistik* 63 (3). 323-337.
Ullman, Michael T. 2001. The neural basis of lexicon and grammar in first and second language: the declarative/procedural model. *Bilingualism: Language and Cognition* 4 (2). 105-122.

Ullman, Michael T. 2005. A cognitive neuroscience perspective on second language acquisition: The declarative/procedural model. In: Cristina Sanz (ed.), *Mind and context in adult second language acquisition: Methods, theory, and practice*. Washington, DC: Georgetown University Press. 141-178.

Van Gelderen, Elly. 2006. *A history of the English language*. Amsterdam: Benjamins.

Vogel, Irene & István Kenesei. 1990. Syntax & semantics in phonology. In: Sharon Inkelas & Draga Zec (eds.), *The Phonology-syntax connection*. Chicago: University of Chicago Press. 339-363.

Vosberg, Uwe. 2003. The role of extractions and *horror aequi* in the evolution of *-ing*-complements in Modern English. In: Günter Rohdenburg & Britta Mondorf (eds.), *Determinants of grammatical variation in English*. Berlin: Mouton de Gruyter. 305-327.

Wasow, Thomas. 1997. End-weight from the speaker's perspective. *Journal of Psycholinguistic Research* 26 (3). 347-361.

Weiner, E. Judith & William Labov. 1983. Constraints on the agentless passive. *Journal of Linguistics* 19 (1). 29-58.

Wickham, Hadley. 2016. *ggplot2: Elegant graphics for data analysis*. New York: Springer.

Wickham, Hadley. 2017. *tidyverse: Easily install and load the 'Tidyverse'*. R package version 1.2.1. URL https://CRAN.R-project.org/package=tidyverse.

Wilson, Rosemary & Jean-Marc Dewaele. 2010. The use of web questionnaires in second language acquisition and bilingualism research. *Second Language Research* 26 (1). 103-123.

Woods, Nicola J. 1997. The formation and development of New Zealand English: Interaction of gender-related variation and linguistic change. *Journal of Sociolinguistics* 1 (1). 95-125.

Yip, Moira. 1988. The obligatory contour principle and phonological rules: A loss of identity. *Linguistic Inquiry* 19 (1). 65-100.

Zec, Draga & Sharon Inkelas. 1990. Prosodically constrained syntax. In: Sharon Inkelas & Draga Zec (eds.), *The phonology-syntax connection*. Chicago: University of Chicago Press. 365-378.

Zwicky, Arnold M. 1969. Phonological constraints in syntactic descriptions. *Papers in Linguistics* 1 (3). 411-463.

Zwicky, Arnold M. & Geoffrey K. Pullum. 1986. The principle of phonology-free syntax: Introductory remarks. *Ohio State Working Papers in Linguistics* 32. 63-91.

University
of Bamberg
Press

This book investigates the alternation between synthetic (*-est*) and analytic (*most*) superlative forms in English through a two-part empirical study. The first part is based on a large-scale questionnaire examining the influence of social and speaker-related variables such as variety, age, gender, and education, alongside linguistic and complexity-based factors at phonological, syntactic, and pragmatic levels. The study also explores morphosyntactic persistence, assessing how preceding patterns of gradation influence subsequent variant choice. Multifactorial analyses reveal systematic interactions between social and linguistic predictors, supporting the hypothesis that the cognitive facilitation strategy known as *more*-support (Mondorf 2009) extends to the superlative, while highlighting the limits of a purely complexity-based account.

The second part complements these findings with a corpus-based study grounded in usage-based construction grammar and exemplar-theoretic assumptions. Drawing on the GloWbE corpus (US and GB sections), it examines superlative alternation with focus on the morphophonological properties of the adjective, frequency effects, *un*-prefixed adjectives (e.g. *happy* vs. *unhappy*), and the processing and representation of complex forms (holistic vs. decomposed). Cross-varietal and comparative analyses reveal usage patterns that contextualize and extend the questionnaire findings.

From a methodological perspective, the dissertation's central contribution lies in the triangulation of elicitation and usage data, demonstrating how a multimethod approach enriches our understanding of grammatical variation and morphosyntactic alternations. By integrating controlled elicitation with authentic corpus evidence, the study illuminates the cognitive, linguistic, and social mechanisms underlying superlative strategy choice and provides both theoretical and methodological insights for future research on grammatical variation.

ISBN: 978-3-98989-079-4

www.uni-bamberg.de/ubp