



Julia Schlüter*

Tracing the (re-)emergence of /h/ and /j/ onsets through 350 years of books: Mergers and merger reversals at the interface of phonetics and phonology

<https://doi.org/10.1515/flih-2019-0009>

Received April 23, 2018; revised December 04, 2018; accepted January 02, 2019

Abstract: This paper investigates the (re-)emergence of onset consonants in English loans from French, Latin and Greek, spelt with initial <u> (> /ju:/; e.g. *union*, *use*), initial <eu> (> /ju:/; e.g. *eulogy*, *euphemism*), or initial <h> (e.g. *habit*, *homogeneous*). It analyses Google Books data, exploiting the occurrence of the article allomorph *a* (rather than *an*) as a diagnostic of consonantal realisation. The analysis yields a fine-grained description of the (re-)emergence of consonantal onsets. It shows that their emergence has been a gradual process and has not reached completion yet. On a theoretical level, the paper discusses the interaction between categorical phonological processing and fine-grained phonetic distinctions in an exemplar-based framework. It also sheds light on the question of (near-)mergers and their potential reversibility.

Keywords: h-dropping, glide formation, filled-onset constraint, unmerging of (near-)mergers, categorical perception

1 Introduction

This study focuses on English nouns borrowed from French, Latin, or Greek, that start with the letters <h> as in *habit*, <u> as in *union*, or <eu> as in *eulogy*, and have re-introduced or developed consonantal onsets during Modern English. These changes are difficult to reconstruct, as they are not reflected in spelling: In words like *habit*, the initial <h> was once silent, while the initial /j/ in words like *union* or *eulogy* is not represented in the orthography. I approach these changes on the empirical basis of Google Books Ngrams. This huge database contains a large number of relevant tokens following the indefinite article, which is realised as *an* before vowels, and as *a* before consonants. Its allomorphy allows inferences as to whether a word is pronounced with an initial vowel or consonant

*Corresponding author: Julia Schlüter, Department of English and American Studies, University of Bamberg, Bamberg, Germany, E-mail: julia.schlueter@uni-bamberg.de

(despite its uninformative spelling). The large number of such cases in Google Books Ngrams makes it possible to reconstruct the emergence of onset consonants at a high temporal resolution, while revealing the impact of various contextual factors at a level of detail that is normally reached only by precise phonetic measurements.

For this study, I retrieved loanwords beginning with <h>, <u>, or <eu> that were sufficiently often attested after the indefinite article. Although this paper is about pronunciation, I will – for practical purposes – keep referring to the three word types by their initial letters. The retrieved items were further divided into subsets according to the following criteria:

- (1) the presence or absence of initial <h>
- (2) the representation of their vocalic nucleus by <u>, <eu>, or some other letter
- (3) the length of their vocalic nucleus
- (4) the stress level of their first syllable (primary, secondary stress, or no stress)

This classification yielded 16 different subsets, represented in Table 1.

For each item in each subset, I then determined how many of its occurrences were preceded by *a* or *an* at different points in time. As expected in a quantitative study sampling millions of exponents of the change from diverse sources, the procedure revealed gradual shifts in usage. However, a closer analysis also pointed to interesting differences between the rates of consonant emergence in the phonologically defined subclasses, and provided a fine-grained and differentiated picture of the spread of the (re-)emerging onset consonants.

At the same time, the study yielded implications for phonological theory, as the interpretation of the observed shift patterns depends on one's model of phonological and lexical representations. Here, I adopt an exemplar-based model, according to which a speaker's mental lexicon stores lexemes as clouds of multiple, phonetically detailed variants. I show that such a model can explain not only the fine-grained patterning of the change, but also sheds light on the question of how phonemic near-mergers can persist (such as the one between <u>- and <eu>-initial words) and even unmerge (as in <h>-initial words). Finally, the changes discussed represent further evidence of the universal preference for filled syllable onsets.

The following section surveys extant research on the (re-)emergence of consonantal onsets. Section 3 details the methods used for extracting data from Google Books Ngrams and preparing them for analysis. Section 4 reports results, and Section 5 attempts to explain patterns in the data with regard to (asymmetrical) relations between spellings, phonemes, and phones, relating them to issues revolving around the merging and unmerging of phonemes. Section 6 summarizes the findings and their theoretical implications in terms of an exemplar-based model.

Table 1: Characteristics of initial syllables, type numbers and example words.

		initial <h>			
		long nucleus	other vowels	short nucleus	
nucleus <ɪ>	primary stress	36 types <i>unified, unicorn, unit, urine, use, unison, ululating, usual, uvula ...</i>	5 types <i>huge, human, humorous, humour, humourist</i>	18 types <i>hierarchy, hero, harlot, hermit, host, hydrogen, hybrid, hormonal ...</i>	40 types <i>history, horrible, habit, heretic, heritage, hectic, heroine ...</i>
	secondary stress	27 types <i>unification, universal, usability, utilitarian, unilingual ...</i>	2 types <i>humanitarian, humiliation</i>	13 types <i>homosexual, hypothetical, hotel, hyperbolic, herculean, hydroxide ...</i>	18 types <i>hesitation, histrionic, horizontal, histological, hospitality ...</i>
	zero stress	20 types <i>unique, united, usurped, utility, utensil, utopia, ubiquity ...</i>	4 types <i>humane, humanity, humiliating, humility</i>	30 types <i>hyperbole, hermetic, harmonic, hierarchic, holistic, hygienic, hotelier ...</i>	39 types <i>historic, historical, habitual, homologous, horrific, hysteria ...</i>
nucleus <eu>	primary stress	14 types <i>eucalypt, eucharist, eulogy, eumuch, euphemism ...</i>			
	secondary stress	9 types <i>eucalyptus, eucharistic, eudiometer, eulogistic ...</i>			
	zero stress	22 types <i>eulogium, euphonic, euphoria, Eustachian ...</i>	1 type <i>heuristic</i>		

2 Previous research on emerging onset consonants

Historical evidence for the pronunciation of word onsets is scarce because spellings are less variable than pronunciations. Apart from orthoepic works (surveyed in Dobson 1968), most findings are derived from phonologically conditioned allomorphy in function words such as the indefinite article or possessive determiners. The following survey is limited to the three types of loanwords this paper focuses on.

The initial <u> in Romance loans like *use* or *union* first represented the loan phoneme /y:/ (< Old French /y/), a rounded high front vowel. By the end of the

fifteenth century this vowel had merged with native /iʊ/ or /eʊ/ (as in *chew*, *true*; *dew*, *hew*). The vocalic centre shifted to the second component, which lengthened to /u:/, while the /i/ was reanalysed as (part of) the onset. This yielded the sequence /ju:/, which satisfied the filled-onset constraint. In conservative, formal registers, /y:/ was retained longer, and /ju:/ was regarded as “barbarous” by orthoepists (cf. Dobson 1968: 700–713; Lass 1992: 50–55, Lass 1999: 99). This suggests that the representation of the change in written texts may have been delayed. The exact dating of /j/-emergence is controversial: Lass (1999: 100) dates it between 1650 and 1700, while Dobson (1968: 709) claims it started “at least as early as the last decade of the sixteenth century (and probably as early as the 1560’s)”. He does not date the completion of the change, but points out that it happened first in word-initial positions¹ and that, towards the end of the seventeenth century, /ju:/ was more likely than /iʊ/ (1968: 709, 712). Drawing on collocations with the indefinite article allomorphs in historical literature collections, Schlüter (2006: 53–54) reports first signs of the change before 1580, which confirm Dobson’s early dating.

The diphthong represented by <eu> in Greek loanwords like *eulogy* joined the native diphthong /eʊ/ when the words entered the English language. It passed through /eʊ/ and merged with the reflex of /iʊ/, later changing to /ju:/ (Dobson 1968: 789–799). According to Dobson, /eʊ/ and /iʊ/ began to merge at the end of the sixteenth century in northern and eastern dialects. The merger proceeded slowly south- and westwards and was completed in the standard by the last third of the seventeenth century. He argues that /ju:/ appeared earlier in <u>-words than in <eu>-words and that the difference was neutralized by the late seventeenth century. Lass (1999: 100), on the other hand, claims that the merger of /iʊ/ and /eʊ/ coincided with the appearance of the glide in the second half of the seventeenth century. Dobson’s evidence against an early neutralization is Owen Price’s *English Orthographie* of 1668, which still points to a distinction between /iʊ/ and /eʊ/ (Dobson 1968: 798). Schlüter (2006: 53–54) investigates Romance and Greek loanwords beginning with /iʊ/ and /eʊ/ in large literature collections, and shows that the establishment of word-initial /j/ in both groups only gained momentum in the nineteenth century, and even then the two diphthongs did not fully merge. This is at odds with both Lass’s and Dobson’s accounts, possibly due to the limitation of Schlüter’s study to word initial diphthongs in loanwords. The prominence of the initial position and the late arrival of some loans might explain why the /iʊ/-/eʊ/ merger occurred later in them than in native words and word-medially.

¹ For methodological reasons, the present study concentrates exclusively on word-initial <u>.

Loanwords with initial <h> (like *habit*) evolved in a similar way to other <h>-initial words. In early Middle English, /h/ was no longer obligatorily realized (cf. Milroy 1983, Milroy 1992: 198–199; Lass and Laing 2010: 348; Minkova 2014: 107). Its deletion was most advanced in Midland and Southern varieties, from which the English standard was to develop later. Loss of initial /h/ continued a home-grown weakening of the consonant. It had begun before – and independently of – French influence (cf. Lutz 1991: 62; Schlüter 2009: 175). Being co-articulated with and strongly coloured by the vowels that followed it, /h/ may have been perceived as no more than the voiceless beginning of a vowel rather than as a segment in its own right. In that respect, it can be compared to the glottal stop reconstructed for Old English and attested in Present-day German.²

The re-establishment of initial /h/ in Standard English is commonly regarded as driven by spelling pronunciation (cf. Graband 1965: 223; Dobson 1968: 992; Wells 1982: 255; Lutz 1991: 60, note 113; Gimson 1994: 175; Minkova 2014: 107). This also applies to words from French (Strang 1970: 81), Latin, and Greek (cf. Fox and Wood 1968: 48; Pope 1973 [1934]: 91; Harviainen 1976): As had also become the rule in the native vocabulary, initial <h> was not pronounced, not even in stressed syllables, and it reappeared later than in native words (cf. Schlüter 2009).

Much research on the re-emergence of /h/ has focussed on dialectal differences (cf. Graband 1965: 222; Wells 1982: 255–256; Milroy 1983: 39–49, Milroy 1992: 137; Ihalainen 1994: 217; Crisma 2009: 135),³ on the stigma attached to /h/-dropping since the late eighteenth century (cf. Dobson 1968: 991; Strang 1970: 81; Lutz 1991: 59; Milroy 1983: 49, Milroy 1992: 140; Lass 1992: 61; Muggleston 2003: 95–134; Minkova 2014: 107), and on emotional emphasis (cf. Wells 1982: 252; Milroy 1992: 138–142). Language-internal factors, on the other hand, have received less attention. Exceptions are Strang (1970: 81) and Minkova (2014: 106–107), who consider the degree of stress on word-initial syllables, and Schlüter (2009), who shows that the etymological provenance, the quantity of vowels in initial syllables, and lexeme-specific factors also played a role.

² For more on the special status of initial <h> in English and the controversy surrounding its phonemic status, see Section 3.1 as well as Wells (1982: 253–256), Ogden (2017: 133), Minkova (2014: 101–102), Schlüter (2009: 176–177) and the references therein.

³ The geographic distribution of /h/-dropping is well documented for early and late Middle English as well as Present-day rural dialects (see LAEME; eLALME; Orton et al. 1978: Ph220, Ph221; Ramisch 2010).

3 Methodology

This study interprets data from Google Books Ngrams as evidence of phonetic and phonological variation and change. The full dataset is accessible at <https://osf.io/ht8se/>. This section explains the methodological steps and the rationale underlying my proposed interpretation.

3.1 Diagnostics

To decide whether a word-initial consonant is present, this study applies a well-established criterion, namely the phonologically conditioned realization of the indefinite article as either *a* or *an*. Before consonants the allomorph *a* is chosen, while before vowels *an* is chosen to avoid hiatus. Google Books Ngrams contain many tokens of <h>-, <u>-, and <eu>-words preceded by the indefinite article, and the proportion of its realizations as *a* vs. *an* varies and changes over time. I assume that the proportion of *a*-realizations correlates with the probability that the following words are perceived as starting with a consonant, and that this probability in turn correlates with the consonantal strength of their onsets.

Both emerging consonants considered here vacillate between a vocalic and a consonantal categorization, but in different ways. Phonetically, [j] hardly differs from the vowel [i]. Ogden (2017: 81) describes [j] “as a very short, non-syllabic version of [i]”. The production of [j] involves movement into and out of palatal approximation, but no steady state in between (cf. also Wright 2004: 36; Padgett 2008: 1937–1940). Diachronically, however, the duration of the steady state distinguishing [i] from [j] can increase or (in the present case) decrease. Whether the resulting sound is perceived as a consonant or a vowel is then determined by the phonology.

The sound [h] is characterized by glottal friction, but since the supralaryngeal part of the articulatory tract anticipates the following vowel, it can also be conceived as “a period of voicelessness superimposed on a vowel”, i.e. as part of the vowel (Ogden 2017: 132–133). The amount and duration of glottal friction would then determine the perception of the sound as vocalic or consonantal. As Wright (2004: 36) suggests, articulation is continuous and involves overlaps, creating redundancy in the speech signal and providing multiple cues for phoneme recognition. In other words, [h] can be co-articulated even if glottal friction represents no distinct segment in the speech signal. By extension, if co-articulated glottal friction drops below a certain threshold in diachronic change, [h] may cease to be identified as a consonantal segment preceding a vowel. If

the friction continues to be co-articulated, however, it can lead to a renewed strengthening and re-establish the original consonant.

By hypothesis, the proportion of writers and the number of times they select *a* before a word reflects the probability that the word was perceived as consonant-initial due to the consonantal strength of its onset. This interpretation is based on the view that lexemes are phonologically represented as clouds of exemplars, i.e. as large numbers of memory traces in the minds of speaker-listeners following encounters with the lexemes in actual language use. Exemplars include highly detailed phonetic information that is much more fine-grained than necessary for merely encoding categorical phonemic contrasts (cf. Bybee 2001: 52; Pierrehumbert 2001: 139–143; Bermúdez-Otero 2007: 512–513; Nycz 2013: 330). New encounters leave new memory traces and older traces decay. Sound change is therefore modelled as a gradual shift in the exemplar cloud (cf. Bybee 2001: 57–60; Pierrehumbert 2001: 148; Nycz 2013: 333). It is “both phonetically gradual and lexically gradual – that is, if words change gradually, and if each word changes at its own rate, then each word will encompass its own range of variation” (Bybee 2001: 41). As will be seen below, my data show that individual lexemes within a group can differ from one another substantially.

3.2 Database, items studied, and data visualization

The data for the present study come from the Google Books Ngram Viewer (cf. Michel et al. 2010), which is freely accessible.⁴ Google Books Ngrams contain 468 billion words from over 4.5 million English books (cf. Lin et al. 2012: 170). The study thus inherits the limitations that come with big but messy data (cf. Hiltunen et al. 2017, Section 4, and sources quoted there). They lack representativeness (oversampling books stocked in academic libraries), and metadata (except publication years), and include multiple copies, editions and reprints of the same works. In addition, there is no simple way of viewing individual Ngrams in context.

For present purposes, however, the affordances of the database clearly outweigh its limitations: The Google Books Ngram database covers the whole period from 1500 to 2008 at a high diachronic resolution. Unsurprisingly, data density increases over time and becomes sufficient for the present study from 1650 onwards. Due to the large amount of text, even low-frequency words are

⁴ See <https://books.google.com/ngrams> and <http://storage.googleapis.com/books/ngrams/books/datasetsv2.html> (accessed 24 June 2019).

attested sufficiently often for quantitative analysis, and although context-free, the bigrams provide enough information on the factors relevant for this study (see also Schlüter and Sönning in preparation).

Like other databases (e.g. those in Schlüter 2009), the Google Books library represents mostly standard written English and pastes over dialectal diversity. There is little hope of ever accumulating a set of dialectally differentiated texts from the past 350 years sufficiently large for tracing middle- and low-frequency items.⁵ To achieve phonetic-phonological depth, the present approach has no choice but to sacrifice diatopic detail.

The specific items in the study were selected from the complete list of Google Books bigrams for English (version 20120701) with *a* and *an* as the first element. The 170 <h>-initial types were selected from the list of words preceded by *an*, and the 83 <u>- and 45 <eu>-initial types were chosen from the list of words preceded by *a*, arranged in descending order of frequency. This made it easier to exclude Germanic words manually. In loanwords, initial /h/ was weaker than in native ones throughout the period studied (thus producing more collocations with *an*; see Sections 2 and 4.2). Among words beginning with <u>, the /j/-glide emerged only in Romance loanwords, and Germanic <u>-initial words were never preceded by *a*.⁶

As tokens of each lexeme type, I counted the most common spelling variants (e.g. *heretic/heretick/heretike*, *eudaemon/eudaimon/eudemmon*, *uniformity/uniformitie/vniformity*) as well as obvious OCR mistakes (e.g. *hofpital*, *hoftile*, *ufage*). I also counted morphological variants that did not differ in terms of the relevant variables (e.g. *hyperbole* under *hyperbola*; *historical* and *historically* under *historic*; *hydraulics* and *hydraulically* under *hydraulic*, *eulogia* under *eulogium*). Loanwords with a markedly foreign ring were excluded from the data, such as *homme*, *hombre*, *habeas (corpus)*, *honnete (homme)*, *haute (couture)*, *uomo*, *upana*, *upadhi*, *upasaka*. I also discarded lexemes that were first attested with *a* or *an* after the year 1900 (e.g. *histogram*, *unionized*, *eucaryote*), as well as the highly exceptional roots *hour*, *honour*, *heir*, *herb* and *homage* which preserve a mute <h> to the present day (the latter two mostly in American English), and all their derivatives. Finally, I excluded a significant number of originally Germanic (Frankish) words that were borrowed into continental (Norman) French and re-borrowed into English after the Norman invasion. These had a pronounced initial /h/ at the time of borrowing (e.g. *hardy*, *heraldic*, *helmet*, *habergeon*,

⁵ A study of the national varieties of British and American English is however feasible based on Google Books Ngrams, and is presented in Schlüter and Sönning (in preparation).

⁶ Spellings beginning with <eu> are limited to loanwords from Greek, so manual categorization according to etymology was not required for this group.

hamlet, *harbinger*, *heinous* etc.; see Pope 1973 [1934]: 41, 94; Schlüter 2006: 44–45; Minkova 2014: 106). The numbers for the included bigram types are shown in Table 1; the total number of tokens analysed was around 80 million.

All items in the data are of Romance or Greek origin, and pronounced with initial /j/, /h/, or /hj/ in Present-Day Standard English. I manually categorized them with regard to two further properties of their initial syllables, namely their stress level and the quantity of their nuclear vowels.⁷ At this stage, I excluded words with variable stress or vowel quantities (e.g. *harassment*, *hostess*, *hydrated*, *hegemony*).⁸

Items with lexical stress on their initial syllable were classified as ‘primary stress’ (e.g. *únified*, *éucalypt*, *hístory*); items with a rhythmically prominent initial syllable at a distance of two or three syllables from the primary stress were classified as ‘secondary stress’ (e.g. *ùnificátion*, *èucalýptus*, *hèsitátion*); items beginning with an unstressed syllable, usually adjacent to a stressed syllable, were classified as ‘zero stress’ (e.g. *únique*, *eulógium*, *hístóric*). For <h>-initial words, a further distinction was made between vocalic nuclei of different quantities. Regardless of stress levels, long monophthongs and diphthongs (e.g. *héro*, *hýpothétical*, *hermétic*) make initial syllables more prominent than short monophthongs (e.g. *hórrible*, *hìstriónic*, *habítual*). Of course, <u>- and <eu>-words all have a long monophthong.

To visualize the data, an interactive application was developed using the Shiny package in R (cf. Chang et al. 2016). Below, only some visualizations are shown, but readers are invited to create more at <https://osf.io/ht8se/> and find the documentation in Schlüter and Vetter (to appear).

4 Results: Zooming into the data

4.1 New onsets: Initial <u> and <eu>

First, I compare lexemes with <u>- and <eu>-spellings. As pointed out, previous research agrees neither on the beginnings of the changes, nor their endpoints,

⁷ For words that would be assigned to different categories in Received Pronunciation (RP) and in General American (GA), I opted for the British version due to its greater historical depth, which is more in line with the long-term perspective of the present study. Thus, *hero*, *uronic*, *eurypterid* etc. were categorized as having long rather than short initial vowels, and vice versa for *horrible*, *hospital* etc.

⁸ Pronunciations were checked against the EPD, LPD, and OED, and etymologies against the OED.

nor the timing of the merger. Therefore, Ngram data are explored to establish the dates more precisely. The first analysis compares both onset types globally, by aggregating data across lexemes and per decade from 1650 onwards. It describes changes in the average proportion of instances of *a* among all instances of the indefinite article. The ‘average’ was calculated by giving the same weight to all type-specific proportions, which were first established separately for each lexeme type. This prevents a bias towards types with high token frequencies.

Figure 1 shows the results for the 83 <u>-initial and 45 <eu>-initial types. It shows data points for every decade beginning in 1650 (provided the frequency of a lexeme per decade is at least 20). The share of *a*, interpreted as a measure of onset strength, is indicated on the y-axis.⁹

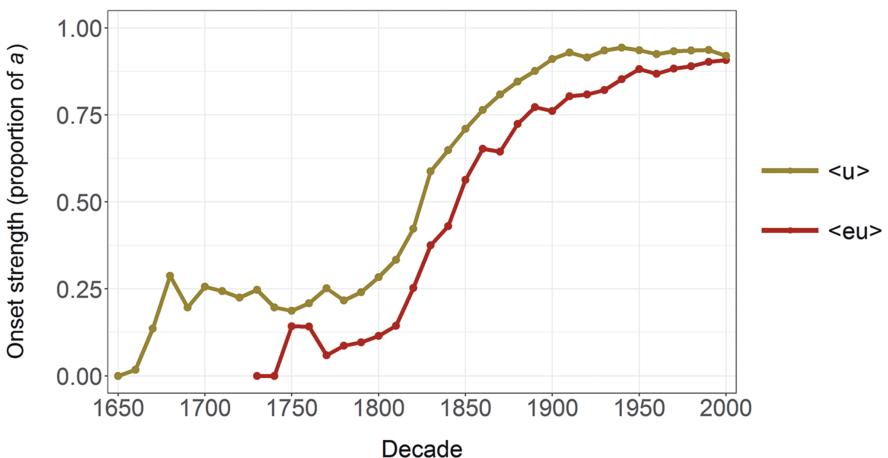


Figure 1: Mean proportion of *a* before <u>- and <eu>-initial words.

The two groups show a near-monotonous increase of the proportion of *a*, describing two prototypical S-curves. The development of the /j/-glide (i.e. its reflection in printed books) started in the second half of the seventeenth century, but made little progress for about a century. The curve for <u>-initial words indicates a beginning in line with Lass (1999: 100), but later than what both Dobson (1968: 709; second half sixteenth century) and Schlüter (2006: 53–54; mid-sixteenth century) find. The delay may reflect transitory stigmatization and orthoepic influence, which was certainly stronger in the academic library collections in Google Books than in the less formal text types of fiction and drama

⁹ Absolute numbers are not indicated, but can be accessed in the Shiny app.

(investigated in Schlüter 2006); or it may be due to the fact that Google Books includes reprints and re-editions of older books. It may also be an effect of all lexemes being more or less recent borrowings from their original languages, in which they were vowel-initial. Educated writers (who are certainly well represented among users of these loans) may have known this, and the spelling may additionally have inclined them to use *a* rather than *an* before <u>-initial lexemes. As to the further development of the change, the steep transition phase typical of S-curves is largely contained in the nineteenth century. The present status quo is approached around 1900.

In contrast, <eu>-words start to appear in sizeable numbers only in the second half of the eighteenth century, lagging behind <u>-words. The likelihood of writers treating words of both classes as consonant-initial seems to increase with time. Although the rapid transition from vowel-initial to consonant-initial treatment of <eu>-words also takes place in the nineteenth century, there is a small but constant difference between the two groups in every single decade. In the steep segments of the two S-curves it amounts to around 20 percentage points, i.e. <u>-words were 20 percentage points more likely to be considered consonant-initial than <eu>-words. In diachronic terms, the emergent glide manifests itself roughly 20 years earlier in <u>- than in <eu>-words. The switch from a majority of vowel-initial perceptions to a majority of consonant-initial perceptions happens between the 1820s and 1830s for <u>- and between the 1840s and 1850s for <eu>-words.

Neither Lass (1999: 100) nor Dobson (1968: 789–799) mention this difference, both dating the merger to the late seventeenth century. However, Figure 1 suggests that it approached completion only in the twentieth century. Of course, Lass and Dobson discuss not only loanwords, but Germanic words as well, and loans might be expected to lag behind in their integration into the native sound system. Still, Dobson (1968: 709) claims that the glide developed faster in initial position, which applies to all items investigated here. Generally, <u>-words appear to be less specific in meaning and application than the <eu>-words, many of which are highly specialized, known only to a small community and written more often than pronounced. Thus, the influence of spelling may make <eu>-words more liable to be treated as vowel-initial than <u>-words. In addition, a large portion of English words beginning with <u> behave as consonant-initial, but only a small portion of those beginning with <e> do. Therefore, the difference between the dating of Lass and Dobson and the one suggested here may be due to factors that the present study cannot control for.

On the other hand, the statistical patterns that can be detected in very large databases may reflect subtle phonetic differences that contemporary speakers would not be aware of. As mentioned in Section 2, the last piece of historical

evidence for a noticeable distinction between <eu>- and <u>-words is an orthoepic work published in 1668 (cf. Dobson 1968: 798). While later speakers may not have noticed such a difference anymore, the strongly quantitative perspective adopted here suggests a contrast between the two groups that disappeared only in the second half of the twentieth century. Thus, we seem to be dealing with a surprisingly persistent near-merger, or a perception-production asymmetry (cf. Labov 1994: 349–418): Averaged across many speakers and tokens, /j/-onsets of slightly different degrees of consonantal strength were produced, and individual speakers stored those they encountered in their mental lexicons. This resulted in a probabilistic difference between categorical perceptions of <u>- and <eu>-words by individuals as consonant- or vowel-initial.

More evidence in favour of this interpretation comes from a subdivision of <u>- and <eu>-words according to the stress levels of their initial syllables, as in Figures 2 and 3.

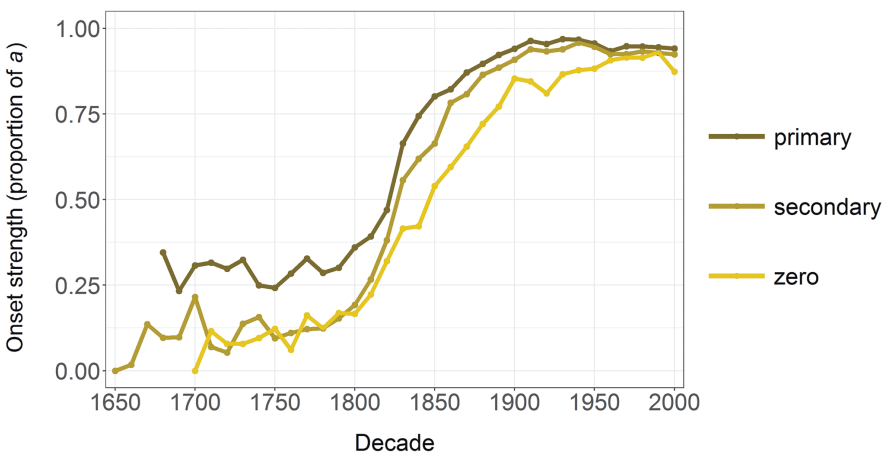


Figure 2: Mean proportion of *a* preceding <u>-initial words by stress level of initial syllable.

The two figures support the hypothesis that the perceptibility of onsets increases with the prominence of the initial syllable: A fully stressed first syllable lends articulatory strength and perceptual salience even to weak consonants, while empty onsets are associated with weakly stressed or unstressed syllables. Primary-stressed items were the first and the most likely to show evidence of a consonantal onset, closely followed by items with secondary stress; items with zero stress were last. The differences between groups are small but remarkably consistent, at least after numbers for <eu>-words with secondary stress have

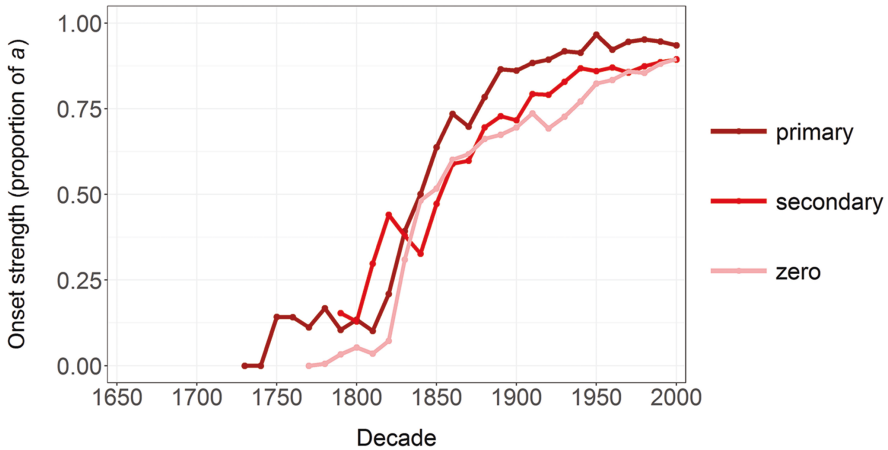


Figure 3: Mean proportion of *a* preceding <eu>-initial words by stress level of initial syllable.

consolidated. Thus, the cumulative effect on many tokens reveals that the emergence of /j/ correlated positively with the degree of stress.

It is worth noting that the change seems not to have reached completion: Words like *usurper*, *ubiquity*, *euphemistic*, *eustachian* etc. still occur with a residue of 6 to 12% *an*. Several explanations are possible: One may be a mere spelling effect. After all, the words start with an orthographic vowel, and individual writers may not have heard them pronounced.¹⁰ Another explanation may be the persistence of conservative or spelling-induced pronunciations, which are occasionally reported.¹¹

4.2 Onsets coming back: Initial <h>

As mentioned in Section 2, initial /h/ was slowly re-established after what looks like its nearly complete loss as a phoneme in relevant varieties of early Middle English.

10 A simple Google Books search suggests that many scientific and medical books (where technical terms in <u>- and especially <eu>- are abundant) in the late twentieth century are authored by non-native speakers, who may rely more on spelling.

11 On <https://english.stackexchange.com/questions/280921/a-or-an-ubiquitous>, “a question and answer site for linguists, etymologists, and serious English language enthusiasts”, the following short exchange can be found (accessed 24 June 2019):

1. I wonder how all of the people using “an ubiquitous” pronounce it? – sumelic Oct 18 '15 at 20:55
2. They probably pronounce it “oobiquitous”. I have heard several people say it that way. – terminex9 Oct 23 '15 at 21:29

Note that native Germanic words with initial <h> are largely disregarded in the present study. In them, initial /h/ may have survived from Old English after being reduced to non-phonemic glottal friction and submerged in early Middle English. Ancient Latin (and Greek) had also possessed an /h/, which only disappeared in the post-classical stage. The typical Romance loanwords investigated here can safely be assumed to have been /h/-less when they were borrowed. As a rule, loanwords tend to adapt to native sound patterns (cf. Uffmann 2015), and so Romance words followed the Germanic ones in their further development, with the exception of only three Romance roots (*hour*, *honour*, *heir*) and their derivatives, which have remained /h/-less and are excluded from this study.

Despite the weakness of /h/, orthographic conventions in both French and English preserved <h>-spellings quite consistently (see Minkova [1991: 159], Minkova [2006: 162], but cf. Lutz [1991: 60], Crisma [2009: 155] and Minkova [2014: 105], who find a greater insecurity in Romance loanwords in Middle English). As mentioned, spelling is often considered as the main reason why /h/-pronunciations re-emerged. However, as my data suggest, the re-emergence of /h/ seems to have proceeded slowly and in highly differentiated ways. It seems to have been a natural, organic change spanning many generations, which allows more subtle explanations than mere reference to spelling.

It is obvious that etymological factors have played a role. Germanic <h>-words have arguably never been 100 percent /h/-less, but preserved a subliminal presence of the onset, while Romance (and Greek) loanwords assumed this feature with a considerable delay. The overall evolutionary trends for native Germanic words and Romance loanwords are visualized in Figure 4, which does not

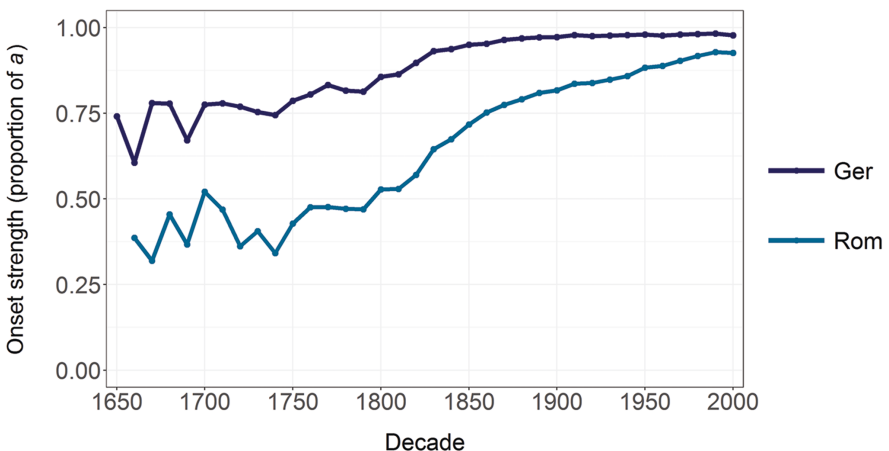


Figure 4: Mean proportion of a preceding typical <h>-initial words by etymological origin.

consider the less prototypical loanwords excluded from this study (see Section 3.2) nor the items investigated separately in Section 4.3.

Figure 4 shows a striking delay of /h/-emergence in Romance loanwords, which is unlikely to have been caused by the etymological difference per se, but probably reflects a difference in the relative prominence of initial /h/. In other words, Romance loans may have had weaker onsets to start with. They followed the native trend towards increasing consonantal strength, yet lagged behind native words by about one and a half centuries. Today, a residual average of 7% of Romance loanwords are still perceived as vowel-initial.

Once again, a subdivision of <h>-initial loanwords according to the stress level of their first syllables strongly suggests that the perceptibility of initial /h/ had a phonetic basis. Figure 5 shows that /h/ was re-established first in initially stressed words, followed by words with a secondary stress on the initial syllable, and then (with a substantial delay) by words with unstressed initial syllables.

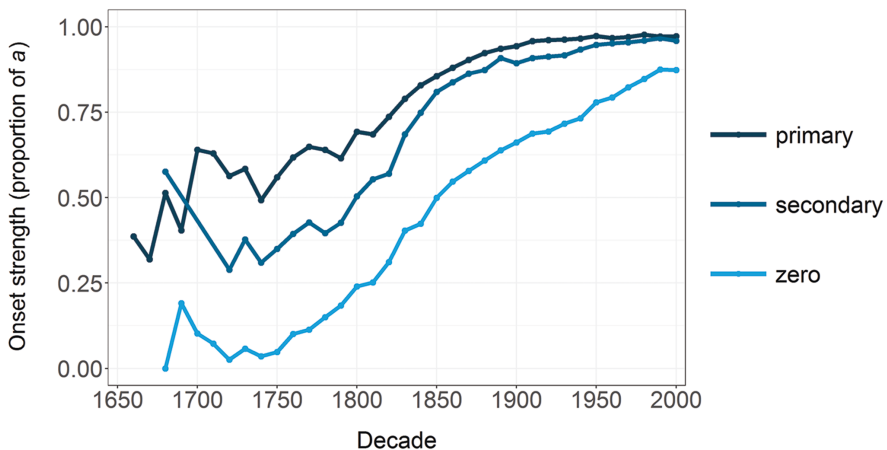


Figure 5: Mean proportion of a preceding typical <h>-initial loanwords of Romance (or Greek) origin by stress level of initial syllable.

An interpretation in terms of onset prominence receives further support if one considers words with long vowels in their initial syllables separately from words with short vowels (a distinction that does not apply to <u>- and <eu>-initial words, which invariably have long nuclei).

Most of the data points in Figure 6 support the hypothesis that syllables with long vowels are more prominent than syllables with short vowels, and that their prominence extends to their onsets and makes them more perceptible. Excepting

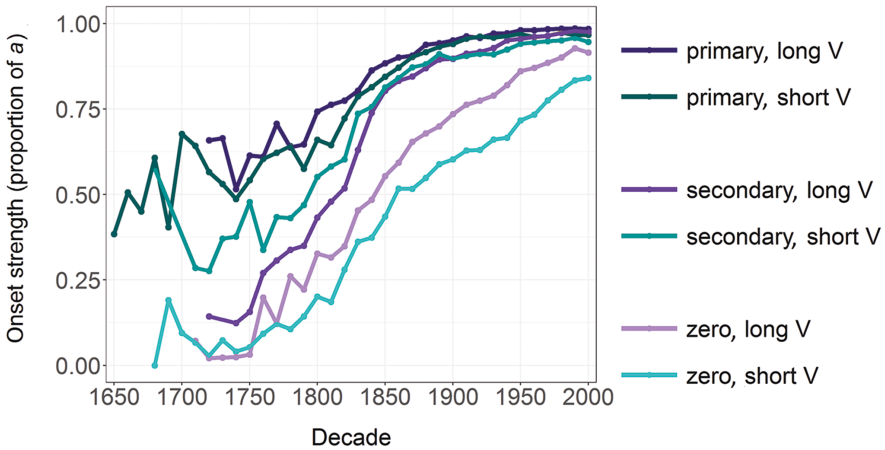


Figure 6: Mean proportion of *a* preceding typical <h>-initial loanwords by stress level and vowel quantity of initial syllable.

items with secondary stress prior to 1900, the /h/-onset manifests itself earlier before a long vowel than before a short vowel. Items with zero initial stress show the largest contrast. Combining the two independent prominence parameters of stress and vowel quantity, the delay in /h/-emergence between the most and the least prominent types amounts to around 150 years.¹²

Crucially, the differentiated pattern of /h/-emergence in Figure 6 correlates in a phonetically plausible way with the phonological prominence of the onsets. There is no way in which spelling pronunciation or orthoepic precepts could account for the pattern. Furthermore, the fact that the process spans more than five centuries and takes the shape of multiple superimposed S-curves suggests that it was a naturally unfolding change rather than a consciously imposed change in norms.

4.3 Onset clusters: Initial <hu> and <heu>

A final piece of evidence of the diagnostic power of article choice is provided by lexemes that combine (re-)emerging /h/ and /j/ in their onsets, thereby forming a novel consonant cluster. Since type numbers are comparatively small, Figure 7

¹² Notably, this applies within the homogeneous group of Romance loanwords. The range of variation is much larger if native words and re-borrowed loanwords are taken into consideration.

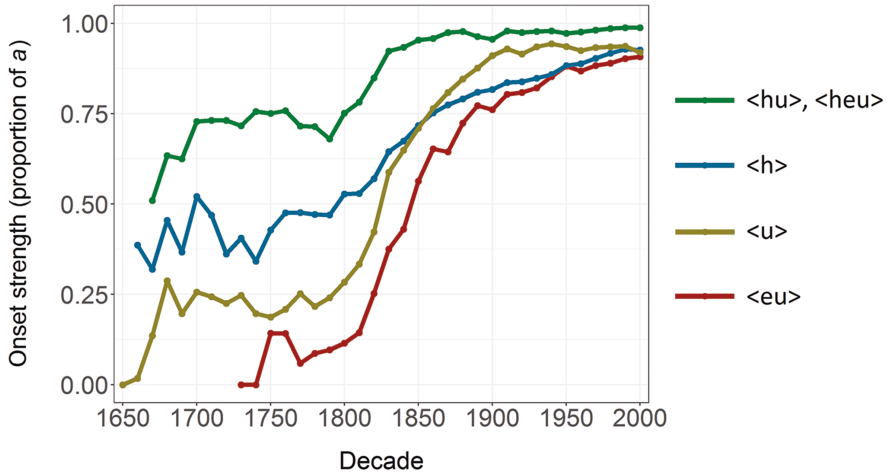


Figure 7: Mean proportion of *a* preceding loanwords of Romance (or Greek) origin by emergent onset type.

zooms back out of fine-grained prominence differences and surveys the four main types of onset on a general level. Items with <hu> and <heu> are grouped together, as only one <heu>-word, namely *heuristic*, is frequent enough for analysis. The hypothesis is that the cluster /hj/ developing in <hu>- or <heu>-words will be clearly perceived as consonantal ahead of the singletons /h/ or /j/. When /h/ and /j/ were not yet sufficiently strong on their own to be perceived as viable onsets, their combined strength would have satisfied the onset requirement.

The results in Figure 7 are perfectly in line with that prediction. In every single decade, the combination of /h/ and /j/ is more likely to be perceived as consonantal than single /j/- or /h/-onsets. The figure also shows that /hj/-initial items had already left behind simple /h/-initial ones in the second half of the seventeenth century. This supports Dobson's (1968: 709) and Schlüter's (2006: 53–54) early dating of the emergence of /j/, i.e. before 1650, which is less evident in Lass's (1999: 100) data and those presented in Section 4.1. Since <hu>- and <heu>-words started with consonant letters, the stigmatization of consonantal pronunciations of <u>- and <eu>-initial words apparently did not extend to them.

Figure 7 also allows us to compare relative rates of change. The re-establishment of /h/ follows a rather flat S-curve, while the emergence of /j/ in <u>- and <eu>-words proceeds considerably faster. The spurt in the increasing onset strength of /hj/-initial words between 1790 and 1830 appears to be the combined effect of the gradients for /j/ and /h/. Even at the turn of the twenty-first century the difference between <h>-words with and without /j/ has not disappeared. As a

result, Romance <hu>-initial lexemes have reached a similar level of onset strength as Germanic <h>-initial words (seen in Figure 4). Once again, this supports the view that co-occurrence rates with indefinite article allomorphs indicate the perceived consonantal strength of word-initial sounds, and can plausibly be interpreted in phonetic terms.

5 Discussion

Explanations for these empirical effects can be discussed at two levels: In synchrony, at the interface between phonetics and phonology; in diachrony, for the modelling of near-mergers and their unmerging.

5.1 Phonology, phonetics, orthography, and orthoepy

The results reported in the previous section suggest that the consonantal strength of the sounds surfacing as initial [j] and [h] today has been gradually increasing during the last 350 years. They also show that the rate of this increase depended, among other things, on the relative prominence of the syllables in which they occurred.

The results imply that the influence of orthography may have been smaller than previously suggested. <h>-spellings were remarkably stable even in early Middle English (cf. Minkova 1991: 159, Minkova 2006: 162), and if they had been the only cause of /h/-re-emergence, /h/ would not have become re-established at different rates in words with different degrees of initial stress or vowel quantities, or with different etymologies. Furthermore, /j/-emergence in <u>- and <eu>-initial words displays the very same dependence on prominence, and cannot have been motivated by spelling. Thus, the impact of orthography on the re-emergence of /h/ is likely to have been ancillary.

Explaining the changes in terms of orthoepic prescriptions is also problematic. First, the failure to pronounce /h/ began to be stigmatized only in the second half of the eighteenth century (cf. Mugglestone 2003: 98), but /h/ had begun to re-emerge much earlier. Second, prescriptions to pronounce /h/ when etymologically appropriate ought to have affected words across the board and more or less simultaneously, but the data show that this has not been the case. Finally, prescribed norms of the ‘correct’ use of the indefinite article in the eighteenth century were logically inconsistent and varied between authors (cf. Sundby et al. 1991: 177–178).¹³

¹³ Before *hospital*, *habitual*, *Herculean*, *humble*, and *humoursome*, some recommended *an*, while others recommended *a* before *habit*, *harlot*, *hermit*, *hero*, *historian*, *history*, *host*, *heroic*,

In contrast, the account I am proposing for the empirical observations is largely system-internal, reflecting the well-established preference for filled onsets (Vennemann 1988). As has been shown, /h/ and /j/ (re-)emerged earlier in prominent syllables than in less prominent ones, which is expected because prominent ones tolerate more complex consonantal onsets or – in the absence of a consonant – tend to attract one, even a prothetic one (cf. Mailhammer et al. 2015: 454, 462). In both cases, the preference for filled onsets seized an opportunity afforded by historical circumstances. In the case of /h/, this was its latent presence in native Germanic words along with the <h>-spellings in Romance and Greek loans. In the case of /j/, this was the fact that the initial part of /iʊ/ (from /y/ or /εʊ/) could be easily reinterpreted as a glide.

The huge set of bigrams analysed also suggests that the (re-)emergence was gradual. I have interpreted this to mean that the consonantal strength of initial /h/ and /j/ increased over time. This deserves some discussion since the evidence that was considered reflects around 80 million case-by-case decisions made by diverse authors on the use of *a* or *an* before words beginning with <h>, <u> or <eu>. What is the link between large sets of categorical decisions between *a* and *an* on the one hand, and gradient phonetic strength on the other?

It is commonly acknowledged that phonetic (allophonic) realizations are infinitely variable, while phonological perception is categorical and determined by boundaries established in a perceiver's phonological system. To justify a phonetic interpretation of the data, three additional assumptions are made.

- (1) Phonetic continua exist not only between neighbouring phonemes in the sound system, but the phonetic expression of (at least some) phonemes can also vary on a continuum ranging from full realization to zero.
- (2) There is clear evidence that diachronic change can shift the phonetic strength of a variable consonant along this continuum.
- (3) Despite inherent phonetic variability, speakers perceive their own phonological output in the same categorical way as hearers perceive the output of others. Writers do the same, based on their internal speech.

I assume that over the last 350 years, the realizations of emerging /j/ and /h/ shifted on a gradient of consonantal strength ranging from being practically imperceptible to being fully consonantal. The strength of a specific word-initial realization varied with time, etymological provenance, the presence of a reinforcing second consonant, and the prominence of the syllable. If consonantal

hideous, horrid, humorous, and – again – humble and humoursome. Before union, universe, unanimous, universal, and useful some authors prescribed an; others prescribed a before unicorn, uniformity, unison, unit, usurer, uniform, universal, and – again – union and useful.

strength was below a critical threshold, the word would have been perceived as vowel-initial, triggering *an*. If it was above the threshold, the word would have been perceived as consonant-initial, triggering *a*. As the variable phonetic strength of word-initial /j/ and /h/ realizations increased over time, so did the probability that they would be perceived as onset consonants (rather than parts of initial vowels), and this would in turn have increased the proportion of *a*-allomorphs before words beginning with <u>, <eu> or <h>. Thus, it is plausible to interpret the proportions of *a*-uses as a measure of phonetic onset strength, particularly when they correlate with factors such as syllabic prominence, whose impact on phonetic onset strength is undeniable.

5.2 (Re-)emergence, merging, and (un)merging

As I have argued, the phonetic realizations of <h>-, <u>-, and <eu>-initial words could have had consonantal properties (such as glottal friction, or palatal approximation without a steady state), without (yet) being perceived as actually consonant-initial. This possibility, made plausible by the patterns in my data, has the potential of shedding light on a question that has been controversially debated in historical linguistics for some time: Are there such things as phonological near-mergers, and can near-mergers be reversed?

Hickey (2004: 130) defines near-mergers in the Labovian sense as follows: “a speaker consistently makes a small articulatory difference between items of two lexical sets but cannot distinguish these auditively, specifically when the pronunciations are offered to the speaker for evaluation” (see Labov et al. 1972), adding that “speakers cannot hear the phonetic distinction which linguists tease out in a spectrographical analysis”. This view has met with considerable resistance, but is receiving an increasing amount of support (see Labov et al. 1991; Labov 1994: 349–418; Maguire et al. 2013: 233–236; Gordon 2015: 188–190). As I would like to argue, the developments discussed here can also be interpreted as evidence of near-mergers.

First, it has been shown that <u>- and <eu>-onsets did not fully merge in the seventeenth century, as previously proposed. Although the last historical source displaying awareness of a distinction between these two classes dates back to 1668, my data show that a difference regarding the selection of *a* or *an* has persisted for three and a half centuries after that date. A probabilistic difference of 20 percentage points, or a diachronic lag of around 20 years (see Figure 1), will not have been consciously perceived by contemporary speakers, but it may well be represented in the distributed mental lexicons of a speech community (as modelled in exemplar-based theories).

Second, the Middle English loss of /h/ in words beginning with the letter <h> can also be conceptualized as a merger of /h/ and zero. The fact that /h/ re-emerged where it was etymologically justified suggests that this merger may in fact have been a near-merger. The proposed line of argument is as follows: Realizations of Old English /h/ ceased to be perceived as phonological segments in early Middle English. They were not completely lost, however, but kept being transmitted as co-articulation features of the following vowels (probably as short periods of weak and possibly voiceless glottal friction) across many generations. Thereby, the phonetic difference between truly empty onsets and weakly aspirated ones could increase and become distinctive again, leading to the unmerging of a near (or apparent) merger (cf. Labov 1987, Labov 1994: 349–418). As one would expect from this perspective, and as has been shown, /h/ did indeed re-emerge first where the residual articulation was best preserved, namely in native Germanic words with initial <h>.

Hickey (2004: 131) remains agnostic about such a possibility: “There are no reported cases of [the unmerging of a near-merger] happening”. In particular, he gives little credit to the idea that speakers would maintain subliminal distinctions “for future ‘unmerging’ of mergers”. He suggests that if cases of merger reversal should be reported, contact with another (e.g. social or regional) variety lacking the merger would be the obvious explanation. Alternatively, he adds, a merger may not have fully diffused in the lexicon and its reversal may start from items that were not affected by it (Hickey 2004: 134–135).

As far as /h/-loss is concerned, however, my own evidence from the dialectally and socially stratified Helsinki Corpus attests to the large-scale absence of a (perceivable) /h/ in early Middle English, most certainly in the norm-providing Southern and Midland dialects, and even in the onsets of Germanic words (Schlüter 2009: 178).¹⁴ At the same time, the pattern of /h/-re-emergence reported here reveals a clear impact of language-internal factors such as stress on the initial syllable and the quantity of its nuclear vowel. In describing a prototypical S-curve, the change carries the signature of a language-internal development. Thus, it clearly seems to have happened “by linguistic means”, to borrow Labov’s phrasing (1994: 311, rendering Garde).

In sum, both the long persistence of a subliminal difference between <eu>- and <u>-words as well as the pattern of /h/-re-emergence suggest that near-mergers can exist, and that subliminal phonetic distinctions between the sounds involved may survive for long periods, representing plausible causes of merger reversals.

14 67 out of 67 instances of native Germanic <h>-initial words in the earliest subperiod (1150–1250) combine with *an* rather than *a*, and until 1420 very little happens to change that situation (cf. Schlüter 2009: 177–180).

6 Conclusion

This paper has presented results collected from a newly available set of big data (c. 80 million hits in Google Books Ngrams) that shed new light on long-standing empirical and theoretical questions. On the empirical side, it has been shown that the re-establishment of initial /h/ in the large class of Romance loanwords was gradual, stretched across a long period, exhibited an intricate patterning, and remains incomplete to this day. Two steeper trajectories of change have been observed for loanwords beginning with orthographic <u> and <eu>, in which a /j/-glide emerged to fill the onset position. These changes probably began in the later seventeenth century, but as far as the onset of <u>-initial loanwords in a corpus of written Standard English is concerned, the steep part of the S-curve describing the transition from vowel-initial to consonant-initial interpretations falls within the nineteenth century. It is paralleled by another S-curve with the same steep slope some 20 years later describing the development of <eu>-initial words.

On the theoretical side, these data have been used to shed light on the way phonetic realizations interface with phonological perception, and to sketch possible implications for phonological (near-)mergers and their potential for reversal.

The interpretations proposed here may not be the only ones compatible with the data, but they are plausible and account for a variety of aspects, including differences between individual lexemes (viewable under <https://osf.io/ht8se/>), synchronic variation and gradual shifts in usage across time. In particular, the interpretations integrate well with exemplar-based models, which assume variation in the tokens that speakers store in their exemplar clouds. In this view, subtle co-articulation effects (such as a voiceless onset of a vowel or some accompanying glottal friction) may be enough to ensure the survival of traces of a former /h/-onset, possibly even across several generations of speakers. By the same rationale, the exemplar clouds representing the /j/-parts in <u>-initial words may be differently composed than clouds representing the /j/-parts in <eu>-initial words, differing with regard to the average initial consonant strength of the exemplars in them. Moreover, each lexeme is represented by its own set of exemplars, occupying its own area on the gradient of strength. The addition of increasingly strong realizations to the cloud can lead to a switch in the perception from vowel-initial to consonant-initial. Little by little, the frequency with which speakers categorize the initial phoneme as consonantal, and/or the number of lexemes to which this happens, and/or the proportion of speakers that analyse instances as consonantal, may increase. Thus, the re-emergence of /h/ (seen as the unmerging of /h/ and zero) and the incomplete merger of <u>- and <eu>-words (through maintenance of an imperceptible contrast) become explicable.

This study has investigated populations of attested uses, and sampled vast numbers of discrete data points. It has revealed the gradual frequency shifts that are expected in diachronic change, with systematic variation on every level (between speakers, groups and sub-groups of lexemes and individual instances of use). It has shown that the distribution of variants is skewed by the pull exerted by the filled-onset constraint towards the (re-)emergence of consonantal articulations. Methodologically, it has demonstrated that big and messy data cropped from centuries-old prints can be exploited to gain indirect information about phonetic distinctions too fine for speakers to perceive, pointing to ways in which historical phonology can benefit from affordances that have become available in the digital age.

Acknowledgements: I am greatly indebted to Fabian Vetter for extracting relevant bigrams from the Google Books Ngrams raw data and for implementing the application with which the figures in this article were generated (accessible at <https://osf.io/ht8se/>). Further thanks are due to the participants of the dpt17 workshop at the University of Vienna in September 2017, to the editors of this volume, in particular Niki Ritt for his exceptional editorial support, and to two anonymous reviewers.

References

- Bermúdez-Otero, Ricardo. 2007. Diachronic phonology. In Paul de Lacy (ed.), *The Cambridge Handbook of Phonology*, 497–518. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511486371.022> (accessed 24 June 2019).
- Bybee, Joan. 2001. *Phonology and language use*. (Cambridge Studies in Linguistics 94.) Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511612886> (accessed 24 June 2019).
- Chang, Winston, Joe Cheng, J. J. Allaire, Yihui Xie & Jonathan McPherson. 2016. *Shiny: Web application framework for R*. <https://shiny.rstudio.com/> (accessed 24 June 2019).
- Crisma, Paola. 2009. Word-initial *h*- in Middle and Early Modern English. In Donka Minkova (ed.), *Phonological weakness in English: From Old to Present-Day English*, 130–167. Houndmills, Basingstoke, Hampshire & New York, NY: Palgrave Macmillan.
- Dobson, Eric J. 1968. *English pronunciation 1500–1700, vol. II: Phonology*, 2nd edn. Oxford: Clarendon.
- eLALME. Benskin, Michael, Margaret Laing, Vasilis Karaiskos, Keith Williamson & The University of Edinburgh. 2013–. *An electronic version of A Linguistic Atlas of Late Mediaeval English*. <http://www.lel.ed.ac.uk/ihd/elalme/elalme.html> (accessed 24 June 2019).
- EPD. Jones, Daniel. 2011. *English pronouncing dictionary*. Peter Roach, Jane Setter & John Esling (eds.). 18th edn. Cambridge: Cambridge University Press. CD-ROM edition.
- Fox, John & Robin Wood. 1968. *A concise history of the French language*. Oxford: Blackwell.

- Gimson, Alfred Charles. 1994. *Gimson's pronunciation of English*, 5th edn. revised by Alan Cruttenden. London: Arnold.
- Google Books Ngram Viewer. <https://books.google.com/ngrams> and <http://storage.googleapis.com/books/ngrams/books/datasetsv2.html> (accessed 24 June 2019).
- Gordon, Matthew J. 2015. Exploring chain shifts, mergers, and near mergers as changes in progress. In Patrick Honeybone & Joseph Salmons (eds.), *The Oxford handbook of historical phonology*, 173–190. Oxford: Oxford University Press.
- Graband, Gerhard. 1965. *Die Entwicklung der frühneuenglischen Nominalflexion: Dargestellt vornehmlich auf Grund von Grammatikerzeugnissen des 17. Jahrhunderts*. Tübingen: Narr.
- Harviainen, Tapani. 1976. On the loss of the Greek /h/ and the so-called aspirated rhō. *Studia Orientalia Electronica* 45. 1–88. <https://journal.fi/store/issue/view/3272> (accessed 24 June 2019).
- Hickey, Raymond. 2004. Mergers, near-mergers and phonological interpretation. In Christian Kay, Carole Hough & Irené Wotherspoon (eds.), *New perspectives on English historical linguistics, vol. II: Lexis and transmission*, 125–137. Amsterdam: Benjamins.
- Hiltunen, Turo, Joe McVeigh & Tanja Säily. 2017. How to turn linguistic data into evidence? In Turo Hiltunen, Joe McVeigh & Tanja Säily (eds.), *Big and rich data in English corpus linguistics: Methods and explorations*. (Studies in Variation, Contacts and Change in English 19). Helsinki: University of Helsinki. <http://www.helsinki.fi/varieng/series/volumes/19/introduction.html> (accessed 24 June 2019).
- Ihalainen, Ossi. 1994. The dialects of England since 1776. In Robert Burchfield (ed.), *The Cambridge history of the English language, vol. V: English in Britain and overseas: Origins and development*, 197–274. Cambridge: Cambridge University Press.
- Labov, William. 1987. The overestimation of functionalism. In René Dirven & Vilém Fried (eds.), *Functionalism in linguistics*, 311–332. Amsterdam: Benjamins.
- Labov, William. 1994. *Principles of linguistic change, vol. I: Internal factors*. Oxford: Blackwell.
- Labov, William, Mark Karen & Corey Miller. 1991. Near-mergers and the suspension of phonemic contrast. *Language Variation and Change* 3. 33–74. doi:10.1017/S0954394500000442 (accessed 24 June 2019).
- Labov, William, Malcah Yaeger & Richard Steiner. 1972. *A quantitative study of sound change in progress*. Philadelphia: U.S. Regional Survey.
- LAEME. Laing, Margaret. 2013–. *A Linguistic Atlas of Early Middle English, 1150–1325*. Version 3.2. The University of Edinburgh. <http://www.lel.ed.ac.uk/ihd/laeme2/laeme2.html> (accessed 24 June 2019).
- Lass, Roger. 1992. Phonology and morphology. In Norman Blake (ed.), *The Cambridge history of the English language, vol. II: 1066–1476*, 23–155. Cambridge: Cambridge University Press.
- Lass, Roger. 1999. Phonology and morphology. In Richard Hogg (ed.), *The Cambridge history of the English language, vol. III: 1476–1776*, 56–186. Cambridge: Cambridge University Press.
- Lass, Roger & Margaret Laing. 2010. In celebration of early Middle English 'H'. *Neuphilologische Mitteilungen* 111(3). 345–354.
- Lin, Yuri, Jean-Baptiste Michel, Erez Lieberman Aiden, Jon Orwant, Will Brockman & Slav Petrov. 2012. Syntactic annotations for the Google Books Ngram Corpus. Association for Computational Linguistics. In *Proceedings of the 50th annual meeting of the Association for Computational Linguistics, 8–14 July 2012*, 169–174. Jeju, Republic of Korea. <http://aclweb.org/anthology/P12-3029> (accessed 24 June 2019).
- LPD. Wells, John. 2008. *Longman pronunciation dictionary*, 3rd edn. Harlow: Pearson Longman. CD-ROM edition: *Longman pronunciation coach*.

- Lutz, Angelika. 1991. *Phonotaktisch gesteuerte Konsonantenveränderungen in der Geschichte des Englischen*. Tübingen: Niemeyer.
- Maguire, Warren, Lynn Clark & Kevin Watson. 2013. Introduction: What are mergers and can they be reversed? *English Language and Linguistics* 17(2). 229–239. doi:10.1017/S1360674313000014 (accessed 24 June 2019).
- Mailhammer, Robert, David Restle & Theo Vennemann. 2015. Preference laws in phonological change. In Patrick Honeybone & Joseph Salmons (eds.), *The Oxford handbook of historical phonology*, 450–466. Oxford: Oxford University Press.
- Michel, Jean-Baptiste, Yuan Kui Shen, Aviva P. Aiden, Adrian Veres, Matthew K. Gray, The Google Books Team, Joseph P. Pickett, Dale Hoiberg, Dan Clancy, Peter Norvig, Jon Orwant, Steven Pinker, Martin A. Nowak & Erez Lieberman Aiden. 2010. Quantitative analysis of culture using millions of digitized books. *Science* 331(6014). 176–182. doi:10.1126/science.1199644 (accessed 24 June 2019).
- Milroy, James. 1983. On the sociolinguistic history of /h/-dropping in English. In Michael Davenport, Erik Hansen & Hans Frede Nielsen (eds.), *Current topics in English historical linguistics*, 37–53. Odense: Odense University Press.
- Milroy, James. 1992. *Linguistic variation and change: On the historical sociolinguistics of English*. Oxford: Blackwell.
- Minkova, Donka. 1991. *The history of final vowels in English: The sound of muting*. Berlin/New York: Mouton de Gruyter.
- Minkova, Donka. 2006. Phonetic naturalness vs. orthography in the formation of a standardized consonantal inventory of English. In Ursula Schäfer (ed.), *The beginnings of standardization: Language and culture in fourteenth-century England*, 147–175. Frankfurt a. M. & New York: Lang.
- Minkova, Donka. 2014. *A historical phonology of English*. Edinburgh: Edinburgh University Press.
- Mugglestone, Lynda. 2003. *‘Talking proper’: The rise of accent as social symbol*, 2nd edn. Oxford: Clarendon.
- Nycz, Jennifer. 2013. New contrast acquisition: Methodological issues and theoretical implications. *English Language and Linguistics* 17(2). 325–357. doi:10.1017/S1360674313000051 (accessed 24 June 2019).
- OED. *Oxford English dictionary online*. 2000–. Oxford University Press. <http://dictionary.oed.com/> (accessed 24 June 2019).
- Ogden, Richard. 2017. *An introduction to English phonetics*, 2nd edn. Edinburgh: Edinburgh University Press.
- Orton, Harold, Stewart Sanderson & John David Allison Widdowson. 1978. *The linguistic atlas of England*. London: Croom Helm & Munich: Hueber.
- Padgett, Jaye. 2008. Glides, vowels, and features. *Lingua* 118(12). 1937–1955. doi:10.1016/j.lingua.2007.10.002 (accessed 24 June 2019).
- Pierrehumbert, Janet B. 2001. Exemplar dynamics: Word frequency, lenition and contrast. In Joan Bybee & Paul J. Hopper (eds.), *Frequency and the emergence of linguistic structure*, 137–157. Amsterdam/New York: Benjamins.
- Pope, Mildred K. 1973 [1934]. *From Latin to Modern French with especial consideration of Anglo-Norman: Phonology and morphology*. Manchester: Manchester University Press.
- Ramisch, Heinrich. 2010. Analysing linguistic atlas data: The (socio-)linguistic context of H-dropping. *Dialectologia* (special issue I). 175–184. <https://www.raco.cat/index.php/Dialectologia/article/viewFile/242110/324722> (accessed 24 June 2019).

- Schlüter, Julia. 2006. A small word of great interest: The allomorphy of the indefinite article as a diagnostic of sound change from the sixteenth to nineteenth centuries. In Nikolaus Ritt, Herbert Schendl, Christiane Dalton-Puffer & Dieter Kastovsky (eds.), *Medieval English and its heritage: Structure, meaning and mechanisms of change*, 37–59. Frankfurt: Lang.
- Schlüter, Julia. 2009. Consonant or 'vowel'? A diachronic study of the status of initial <h> from early Middle English to nineteenth-century English. In Donka Minkova (ed.), *Phonological weakness in English: From Old to Present-Day English*, 168–196. Houndmills, Basingstoke, Hampshire & New York, NY: Palgrave Macmillan.
- Schlüter, Julia & Lukas Sönning. In preparation. Comparing strengths and limitations of linguistic corpora and big data resources: Evidence for the pronunciation of initial <h> in the BNC, COCA and Google Ngrams. In Ole Schützler & Julia Schlüter (eds.), *Data and methods in corpus linguistics: Comparative approaches*. Cambridge: Cambridge University Press.
- Schlüter, Julia & Fabian Vetter. To appear. An interactive visualization of Google Books Ngrams with R and Shiny: Exploring *a(n)* historical increase in onset strength in *a(n)* huge database. In *Journal of Data Mining and Digital Humanities* (special issue). Bettelou Los, Benjamin Molineaux & Martti Mäkinen (eds.), *Visualizations in Historical Linguistics* 21. Atelier Digit_Hum. <https://hal.archives-ouvertes.fr/hal-02149498> (accessed 24 June 2019).
- Strang, Barbara M. H. 1970. *A history of English*. London: Methuen.
- Sundby, Bertil, Anne Kari Bjørge & Kari E. Haughland. 1991. *A dictionary of English normative grammar 1700–1800*. Amsterdam: Benjamins.
- Uffmann, Christian. 2015. Loanword adaptation. In Patrick Honeybone & Joseph Salmons (eds.), *The Oxford handbook of historical phonology*, 644–665. Oxford: Oxford University Press.
- Vennemann, Theo. 1988. *Preference laws for syllable structure and the explanation of sound change: With special reference to German, Germanic, Italian, and Latin*. Berlin: Mouton de Gruyter.
- Wells, John C. 1982. *Accents of English I: An introduction*. Cambridge: Cambridge University Press.
- Wright, Richard. 2004. A review of perceptual cues and cue robustness. In Bruce Hayes, Robert Kirchner & Donca Steriade (eds.), *Phonetically based phonology*, 34–57. Cambridge: Cambridge University Press.