Julia Schlüter

A small word of great interest: the allomorphy of the indefinite article as a diagnostic of sound change from the sixteenth to nineteenth centuries*

1. Introduction

It is the aim of this study to shed more light on a set of largely unexplored sound changes that were under way between the sixteenth and nineteenth centuries. For this purpose, the indefinite article with its two allomorphs will be exploited as an indicator of the phonetic quality of the following sound. The analysis rests on the premise that the short form of the article *a* is regularly placed before lexemes that are considered to begin with a consonantal segment, whereas the long form *an* precedes lexemes considered as vowel-initial. Since early Middle English, when the loss of final /n/ before consonants was completed, this distributional rule has remained essentially unchanged, with very little variation as far as Standard English is concerned (cf. Schlüter forthc. a).

The occurrence in English of V-initial content words with first-syllable stress is an innovation of the ME period, which appears to have been fostered (though not initiated) by the influx of French loanwords (cf. Minkova 2000b:506-507; 2003:149-160; and the critical review in Schlüter forthc. a). Even so, the universal preference for syllables to begin with consonantal onsets (cf. Vennemann

---

* I take this opportunity to foreground a person who is very much present in the background of the work that has gone into this paper as well as into all my past (and, for that matter, future) ventures into English linguistics. I dedicate this paper to my supervisor Günter Rohdenburg on the happy occasion of his 65th birthday on 28th July 2005, with congratulations and best wishes. Working under his guidance has always been an immensely inspiring and gratifying experience, for which I owe him my heartfelt thanks. The present research is part of a larger research project under Günter Rohdenburg's direction. I acknowledge the financial support received from the German Research Foundation (DFG; grant number RO 2271/1-3) and the Lise Meitner post-doctoral fellowship awarded by the North-Rhine Westfalian Ministry of Science and Research.

1988:13, 21) has remained in vigour – the allomorphy of the indefinite article being the most visible of its manifestations. It is assumed that the final /n/ of the article transfers (resyllabifies) to the onset of a following V-initial lexeme, thereby optimizing syllable structure (cf. Nespor and Vogel 1986:65; Lutz 1991:61, note 115).

The present contribution contends that a number of sound changes have taken place in the course of the ModEn era that can be subsumed under a tendency to (re-)introduce or strengthen consonantal onsets in formerly V-initial lexemes. Focussing on written Standard English, the analysis will deal with the following phenomena:

(1) Lexemes beginning etymologically with /h/ had apparently lost this initial consonant in the ME era. It has been shown (cf. e.g. Lutz 1991:62, or Schlüter forthc. a) that the loss, frequently attributed to French influence, was more probably a natural continuation of home-grown developments in English and that it was not completed, but left some phonetic traces, which provided the germs of a renewed strengthening in the Modern period (cf. Schlüter forthc. a). It is this restitution of /h/-onsets that will come under closer scrutiny in section 2.1 of this contribution. The influence of French phonology will be discussed as a factor distinguishing the group of Romance loanwords from the native Germanic word stock.

(2) Section 2.2 relativizes the explanatory force of purely etymological arguments in that it points out the sharp contrast between the mainstream development of Romance loanwords and the strikingly exceptional items *hour*, *honour*, *honest* and *heir*, which fail to reintroduce an initial /h/. It also hints at possible explanations for the deviant behaviour of these stems.

(3) Among Germanic as well as Romance <h>-initial words, some evolved an additional glide that came to reinforce the onset, yielding the initial consonant cluster /hj/. Section 2.3 sets out to measure the effect of this privilege with regard to the filling of the onset position.

(4) While (1) to (3) all relate to the prominent class of <h>-initial lexemes, the allomorphy of the indefinite article also tells a story about the evolution of new initial consonants in formerly V-initial lexemes. Due to the methodologically necessary restriction of this study to the written standard language, the cases investigated in

section 2.4 are limited to the development of the /j/-onset in loanwords such as *eunuch, eulogy, Europe* as well as *unit, urine, use*.

Further relevant examples of newly created consonantal onsets known from dialectal usage since the EModE period are the formation of a prothetic /j/ in *yer* for *ere*, *yerst* for *erst*, *yerth* for *earth*, and of a prothetic /w/ in *uonli* for *only*, *wuts* for *oats*.[1] The forms *whole* and *whore*, whose modern spellings may reflect traces of a phonological /w/-glide, also belong here, but are more properly treated together with <h>-words in a study of Standard English (see section 2.1). The only case in point that happens to have been adopted into the standard is the prothetic /w/ in *one* and *once*. However, since *once* is rare after the indefinite article and *one* occurs mostly in fixed collocations like *such a(n) one*, *many a(n) one* and *so* Adj *a(n) one*, where special conditions apply, both have been excluded from this study.

Investigating sound change in written texts is usually a challenging enterprise. In the present case, however, the availability of phonotactically distinct variants of the indefinite article provides a relatively direct clue to the realization of the following sound or sounds. In addition, the extremely high frequency of the article and the availability of electronically readable texts allow us to describe synchronic contrasts and diachronic trends with a precision that surpasses that of earlier approaches, including Dobson's (1968) monumental compilation.

2. Analyses

The data on which the following analyses are based include four collections of dramatic and non-dramatic prose spanning the sixteenth to nineteenth centuries: one drama collection (*English Prose Drama*, *EPD*), and three collections of non-dramatic literature (mostly novels), namely the *Early English Prose Fiction* (*EEPF*), *Eighteenth-Century Fiction* (*ECF*) and *Nineteenth-Century Fiction* (*NCF*) collections. The four centuries under consideration were divided into segments of about 40 years each, with the exception of the longer first subsection,

[1] For more comments and further evidence, see Dobson (1968:995-1001), which is the source of the examples given. Cf. furthermore Ihalainen (1994:202, 213).

in which the *EEPF* coverage of the early years is still relatively thin, while the *EPD* collection sets in only in 1540. The years referred to throughout this discussion invariably represent the publication dates of the works or, in the case of plays, the dates of first performances, if they antedate the first printed publications. Table 1 charts the resulting division and the number of texts and word counts for each subsection.

Table 1. Composition of the database

Subperiod	**1518-1580**	**1581-1620**	**1621-1660**	**1661-1700**	**1701-1740**	**1741-1780**	**1781-1820**	**1821-1860**	**1861-1903**
Prose	*EEPF*	*EEPF*	*EEPF*	*EEPF*	*ECF*	*ECF*	*NCF*	*NCF*	*NCF*
No. of works	23	71	52	65	41	52	56	91	103
Million words	1.42	2.67	2.83	3.12	2.30	8.03	7.57	16.38	15.74
Drama	*EPD*	*EPD*	*EPD*	*EPD*	*EPD*	*EPD*	*EPD*	*EPD*	*EPD*
No. of works	6	62	69	172	226	197	388	315	173
Million words	0.19	1.15	1.44	4.38	3.91	2.80	6.28	4.22	2.63
Total									
Million words	**1.62**	**3.82**	**4.27**	**7.50**	**6.21**	**10.83**	**13.85**	**20.60**	**18.36**

These subcorpora were subjected to a set of full-text searches that were directed at both variants of the indefinite article followed by the typical spellings for the onsets under discussion.[2] This procedure excluded all unwanted instances of lexemes with invariant, categorical C- or V-onsets.[3] In a subsequent manual classification of all matches, sporadically occurring words with an unetymological <h>-spelling, e.g. *hidea*, *horacle*, *hointment* etc., were discarded. This non-standard phenomenon is beyond the scope of the present study, but receives due attention in Häcker (1998, 2004) and Crisma (forthc.).

[2] Besides <h>-initial words, all words beginning with the digraph <wh> (occasionally representing /h/ as in *whole*, *whore*, etc.) and the graphemes <eu>, <u> and their allographs <ev> and <v> were searched and monitored.

[3] The *EEPF* corpus contains a total of 34 instances of *a* preceding V-initial words, which is a carry-over effect of the evolution of the indefinite article in many non-standard varieties of English. This extremely low number, which compares with over 20,000 occurrences of *an* before V in this corpus, testifies to the close adherence of the corpus texts to the standard norms.

2.1 A salad bowl of etymologies

Descriptions of the status of initial /h/ in Early and Late Modern English typically concentrate on factors such as the presence of stress on the syllable under consideration (cf. Strang 1970:81), the emotional emphasis on a word (cf. Milroy 1992:138-142), dialectal differences (cf. Graband 1965:222; Milroy 1983:39-40; 1992:137; Ihalainen 1994:217) and the increasing social stigma attached to /h/-dropping (cf. Dobson 1968:991; Strang 1970:81; Lutz 1991:59; Milroy 1983:49; 1992:140; Lass 1992:61; Mugglestone 1995:107-150). However, independently of prosodic, stylistic, diatopic and sociolinguistic considerations, the class of English <h>-initial lexemes represents a mixture of items of very diverse origins. Teasing apart the etymological intricacies involved in this heterogeneous class of lexemes is a worthwhile task that has never been addressed in any systematic way.

To start with, a substantial section of the English lexicon goes back to French sources: beginning in the thirteenth, peaking in the fourteenth century and persisting for many centuries, English has been tapping massively into the French vocabulary (cf. Pope 1934:424; Dekeyser 1986:256; Baugh and Cable 1993:164). Native French words are known to have lost the pronunciation of initial <h> since the Late Latin stage; so it follows that they were borrowed into English without a consonantal onset. The question of whether native English lexemes had preserved traces of a consonantal realization in the period of the most substantial borrowing from French is still debated (see in particular Schlüter forthc. a, where an affirmative answer is suggested). For the present study, it will be sufficient to bear in mind that the pronunciation of initial <h> was regaining ground in the course of the EModE and LModE eras.[4]

This suggests the hypothesis that Germanic and Romance <h>-words in the ModE period may have behaved differently in the course of the evolution reinstating their /h/-onsets. For the following

[4] In Schlüter (forthc. a), it is argued that this trend cannot be exclusively ascribed to the influence of a normative enforcement of a spelling pronunciation (as is almost unanimously done by linguists at a loss for a better explanation; cf. e.g. Graband 1965:223; Dobson 1968:992; Strang 1970:81; Lutz 1991:60, note 113; Gimson 1994:175).

analyses, it will be proposed that the relative collocation frequencies of these words with *a* and *an* can be taken as a direct clue to the consonantal strength of their onsets: the higher the percentage of *an*, the higher the number of occasions on which an <h>-initial word was judged to be V-initial, and inversely, the lower the percentage of *an*, the higher the number of occasions on which it was judged to be C-initial. For the consistency of this argument, it does not matter whether we assume a homogeneous speech community in which every speaker reacts spontaneously to every instance of an <h>-initial word by variably selecting *a* or *an*, or whether we assume a heterogeneous community in which an individual consistently uses either *a* or *an* in connection with a certain lexical entry. In either case, the statistical picture afforded by a corpus study will be the same. Crucially, however, my argument presupposes a gradient phonetic realization with a categorical phonological interpretation that switches between V and C as a certain threshold of consonantal strength is crossed. Evidence in favour of this claim will accumulate as the discussion proceeds (for further discussion, see Schlüter forthc. a).

Figure 1. The distribution of *a* and *an* before <h>-initial words of Romance (or Greek) and Germanic (or other) origin

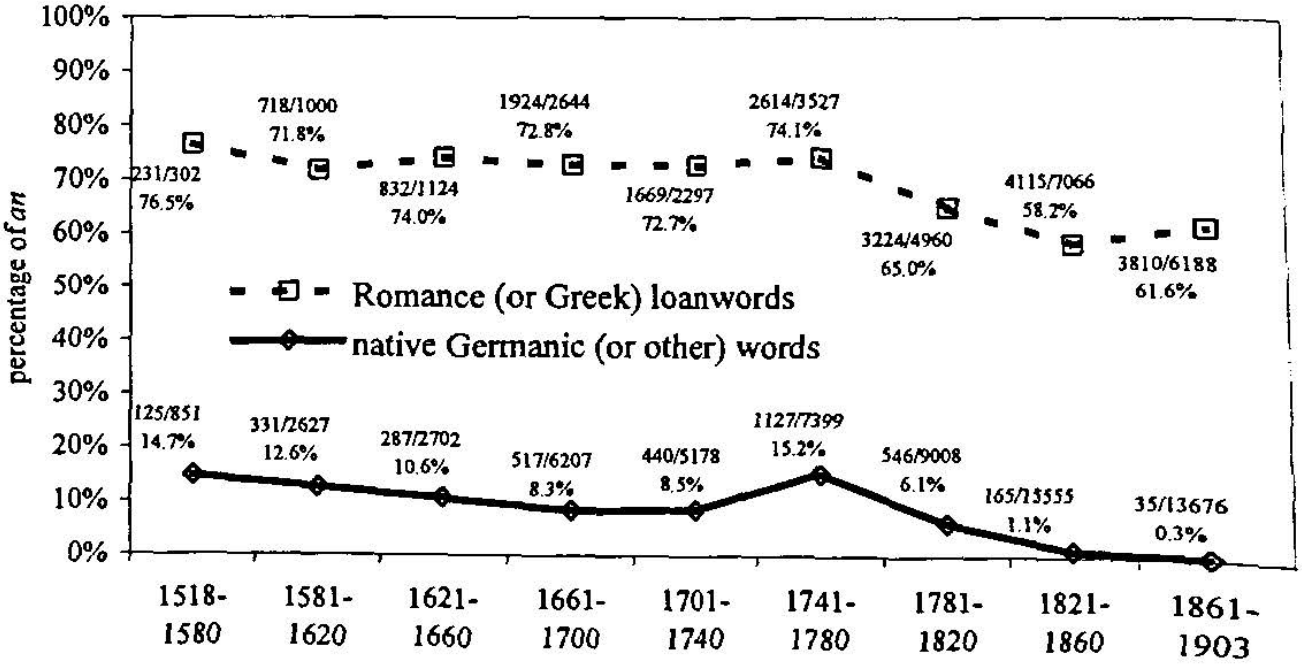


Figure 1 provides a first rough division of all corpus examples of <h>-words preceded by *a/an* according to their etymological sources.

The clear and remarkably constant difference of around 60 percent between the two groups[5] demonstrates that, on the above premise, Romance loanwords typically show a far weaker onset realization than Germanic words. In addition, the figure displays a slight downward trend in the use of *an*, which translates into a reinforcement of initial /h/, both Germanic and Romance. However, this presentation of the data hides some important distinctions, which make the contrast look less categorical and the diachronic trend more pronounced.

For a start, the category of Germanic <h>-words subsumes the stems *whole* and *whore* as well as their derivatives. Their <wh>-spelling, nowadays no more than a variant of <h>, gained currency in the fifteenth and sixteenth centuries and corresponded to a labialized pronunciation that extended also to other <(h)o>-initial words (cf. *OED*: s.v. *wh*; Dobson 1968:997-998). However, the glide-like realization of the onset was only short-lived, while the spelling is perpetuated to the present day. Besides, it should be noted that the category labelled 'native Germanic (or other) words' also comprises all loanwords that are not of Romance or Greek origin. However, members of this subcategory appear only in the eighteenth- and nineteenth-century data and are negligible in number. Persian/Urdu *hindoo* and *hookah*, Romanian *hospodar* and Hungarian *hussar* are cases in point. None of these words however exhibits any quantifiable deviations from the norm of Germanic <h>, so that there is no need for a separate treatment.

In contrast to the Germanic category, which can stand as it is, the only thing that unifies the <h>-words in the Romance (or Greek) category is the fact that they were usually adopted into English (more or

[5] The results are highly significant, which is indicated by the fact that the statistical error probability *p* as calculated by the chi-square test, i.e. the likelihood with which the observed contrast is only a chance distribution, is negligible for each of the chronological corpus sections: 1518-1580: $p = 9.60 \cdot 10^{-89}$ (***); 1581-1620: $p = 1.67 \cdot 10^{-278}$ (***); 1621-1660: $p \approx 0$ (***); 1661-1700: $p \approx 0$ (***); 1701-1740: $p \approx 0$ (***); 1741-1780: $p \approx 0$ (***); 1781-1820: $p \approx 0$ (***); 1821-1860: $p \approx 0$ (***); 1861-1903: $p \approx 0$ (***).

less directly) from French. Yet, this conglomerate is so heterogeneous that the corresponding curve in Figure 1 is no more than an artefact – none of the subgroups actually behaves in the way suggested by the curve. For one thing, the data subsume words of different Romance provenances, including above all Old French, but also Classical, Late and Mediaeval Latin, as well as many learned words of Greek origin, which typically arrived in English by the mediation of French.

More importantly, there is another subgroup that contrasts sharply with all others, namely those words that French had previously adopted from the Germanic languages spoken on the continent. Not surprisingly, the number of these lexemes is far from negligible, considering that Norman French in particular had undergone a strong influence from Frankish plus a slight one from Old Norse (cf. Pope 1934:13-14). Following the etymologies given in the *OED*, examples include *habergeon*, *halberd*, *halt*, *hamlet*, *harbinger*, *hardy*, *harlequin*, *harquebus*, *haste*, *hasty*, *hatchet*, *haunch*, *heinous*, *helmet*, *herald*, *heron* and *hobby*. Pope (1934:15, 41, 56, 94) claims that even in French, these lexemes preserved a pronounced initial /h/ as late as the sixteenth or seventeenth century. This fact can be expected to be mirrored in a deviant behaviour after their adoption into English.

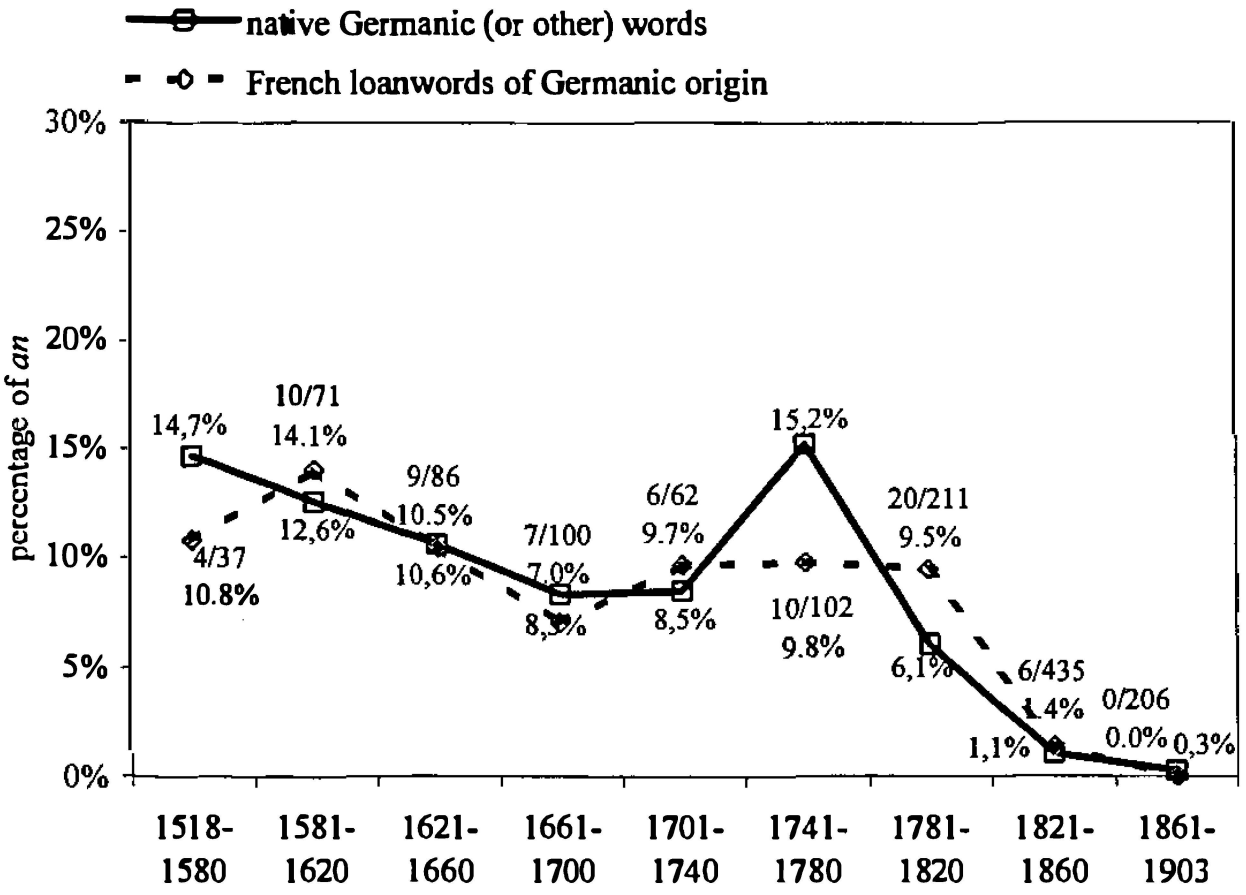
Figure 2 below extracts the Romance loanwords of Germanic origin and juxtaposes them with the category of straightforward Germanic words, which is reproduced from Figure 1 as a standard of comparison.

The data indicate that the detour taken by the doubly borrowed words remains without a noticeable effect on the realization of their onsets: the initial /h/ turns out to be practically identical in strength with that of their less travelled congeners.[6] This seems to corroborate Pope's view that the words concerned preserved the consonantal onset inherited from Germanic during their stopover in the French system.

[6] The results of the chi-square test show no noteworthy differences between the two groups, with the exception of the subperiod 1781-1820, in which straightforward Germanic words are slightly ahead in the strengthening of /h/: 1518-1580: $p = 0.51$; 1581-1620: $p = 0.71$; 1621-1660: $p = 0.96$; 1661-1700: $p = 0.63$; 1701-1740: $p = 0.74$; 1741-1780: $p = 0.12$; 1781-1820: $p = 0.04$ (*); 1821-1860: $p = 0.52$; 1861-1903: $p = 0.47$.

Incidentally, the etymological complication has left intact not only the peculiar onset but also the substance of the lexemes: thus, native *helm* and re-borrowed *helmet*, native *halter* and re-borrowed *halt*, native *hard* and re-borrowed *hardy* share so many phonological, morphological and semantic features that a limited period of spatial separation is unlikely to have obliterated the connection holding between them. On the diachronic dimension, Figure 2 furthermore shows that throughout the three centuries from 1518 to 1820, there is very little change in the pronunciation of initial <h> in Germanic words: there continues to be a residue of cases with a subliminal realization that only disappears in the course of the nineteenth century.

Figure 2. The distribution of *a* and *an* before originally Germanic <h>-initial words re-borrowed from French compared to straightforward native Germanic <h>-words



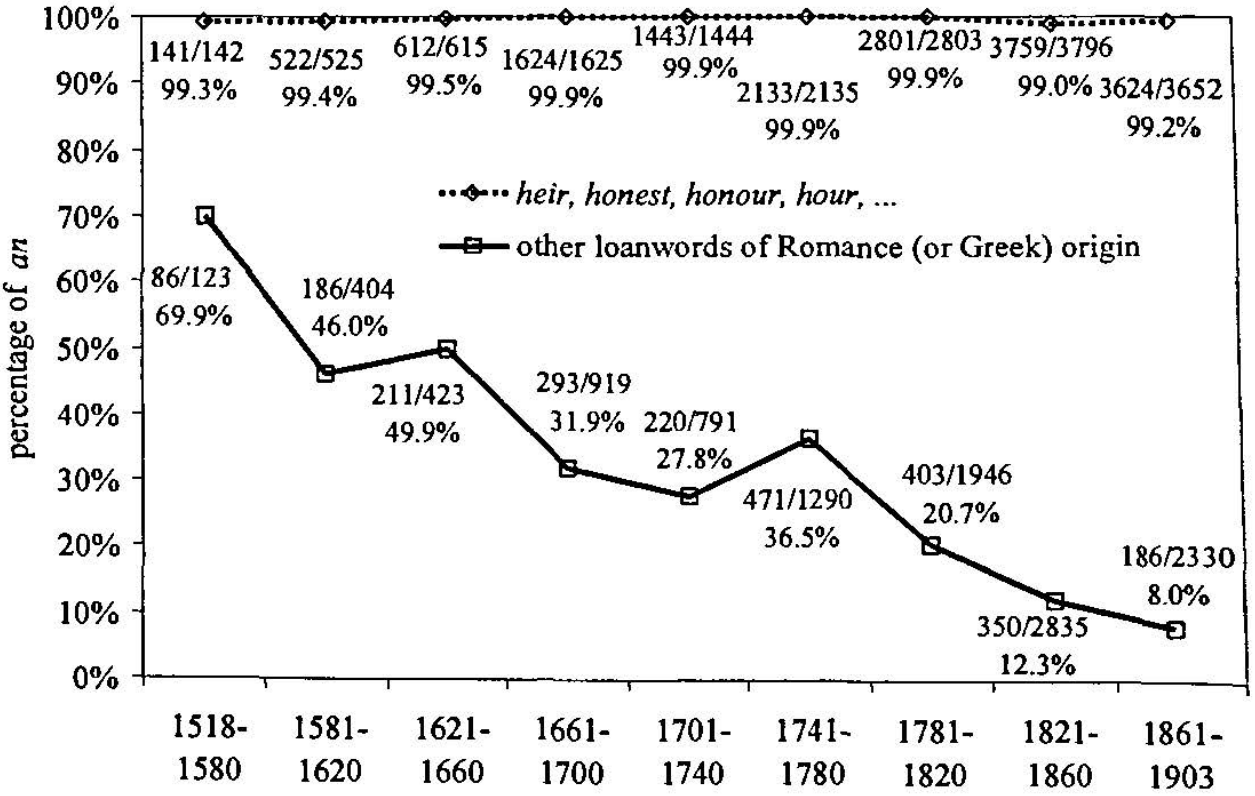
This concludes the analysis of the historical origins of English <h>, albeit in a preliminary way. As we have seen, the English lexicon

might be more adequately compared to a salad bowl of etymologies rather than a melting pot: the ingredients do not assimilate to an undifferentiated norm, but preserve the marks of their provenance. Thus, a major contrast between Germanic and Romance etymologies has been observed that cuts across the group of loanwords adopted from French.

2.2 They never come back: mute initial <h>

Even after the exclusion of words of Germanic origin from the group of Romance (plus Greek) loanwords in Figure 1, we are left with a varied set.

Figure 3. The distribution of *a* and *an* before Romance (or Greek) loanwords with a 'pronounced' or 'mute' initial <h>



Down to the present day, there is a split between Romance and Greek loanwords that have re-adopted a pronounced <h> and such that have been remarkably resistant to it, including first and foremost the

group *heir*, *honest*, *honour*, *hour* and their derivatives. Figure 3 differentiates between the two groups on the basis of the present-day situation in Standard British English (in American English, *herb* and possibly *hom(m)age* would be further candidates for the second category).

The results show a stagnant situation among the four stems that separates them from the remaining lexemes of Romance or Greek origin. While in the former group, there are never any signs of even an incipient comeback of pronounced initial /h/, in the latter, this change already has a firm foothold in the earlier part of the sixteenth century and catches on slowly but continuously.[7] As a result, in the late nineteenth century, an almost categorical distinction can be observed between two classes of Romance loanwords, which cannot be explained as an effect of etymological differences.

A closer look at the data for individual lexemes shows that the reintroduction of initial /h/ spreads by lexical diffusion: not all lexemes are affected at the same time and at an equal rate. Thus, *hospital*, *host(ess)*, *hotel* and *humble* as well as some non-initially stressed adjectives (*habitual*, *hereditary*, *heroic*, *historic(al)*, *hysteric(al)*) are among the last to consistently introduce a pronounced consonantal onset. This effect may in large part be due to the increasing influence of prescriptive norms in the eighteenth and nineteenth centuries.[8] However, what is still at stake is an account of the striking divergence between the groups depicted in Figure 3.

---

[7] The statistical difference between the two groups is very highly significant throughout: 1518-1580: $p = 1.01 \cdot 10^{-11}$ (***); 1581-1620: $p = 4.77 \cdot 10^{-80}$ (***); 1621-1660: $p = 9.73 \cdot 10^{-84}$ (***); 1661-1700: $p \approx 0$ (***); 1701-1740: $p = 1.73 \cdot 10^{-305}$ (***); 1741-1780: $p \approx 0$ (***); 1781-1820: $p \approx 0$ (***); 1821-1860: $p \approx 0$ (***); 1861-1903: $p \approx 0$ (***).

[8] In their detailed survey of eighteenth-century English grammars, Sundby, Bjørge and Haughland (1991:177-178) find partly conflicting orthoepic pronouncements on whether to use *a* or *an* before different Romance and Greek <h>-words: before *habitual, Herculean, honest, humble* and *humoursome* some recommend the use of *an*; before *heroic, hideous* and *horrid,* but again also *humble, humorous* and *humoursome* others recommend the use of *a*, which illustrates the somewhat arbitrary nature of normative precepts.

A suggestive remark comes from Graband (1965:224), who relates the exceptionality of the stems *heir*, *honest*, *honour* and *hour* to their extreme frequency of occurrence. A glimpse at the absolute numbers included in Figure 3 gives an impression of this striking imbalance between the two groups: tokens representing only these four stems are on average 1.5 times as frequent in the collocation with the article as all other tokens taken together, even though the latter category contains an incommensurately larger number of stems and derivatives (e.g. *haberdasher*, *habit*, *hackney*, *halo*, *harass*, *harlot*, *harmony*, *harpoon*, *haughty*, *haunt*, *hazard*, to give just a few examples from the beginning of the list). The effect of frequency is investigated in more detail in Schlüter (forthc. a). Suffice it to say at this point that a moderate correlation can be detected between a high frequency of the string *a/an* + <h>-lexeme and an /h/-less pronunciation. This effect can be explained by the high degree of contextual predictability of the items and the economy of articulatory effort in frequent collocations. In addition, items that are current in ordinary spoken discourse can be expected to be particularly immune to the influence of spelling on their pronunciation. Thus, *honest* and *honour* and the extraordinarily frequent *hour* have a particular predisposition to remain /h/-less (the same is less true of *heir*). The explanatory potential of frequency becomes manifest also among other loanwords of Romance or Greek origin: the more frequent they are, the slower they are to reintroduce a pronounced initial /h/ (cf. Schlüter forthc. a).

2.3 More equal than others: /hj/-clusters

In early Middle English, the three /h/-containing stem-initial consonant clusters /hn/, /hl/ and /hr/ inherited from Old English (e.g. *hnutu* 'nut', *hlud* 'loud', *hræfn* 'raven') had been reduced to /n/, /l/ and /r/ as a first step in the weakening process of initial /h/ (cf. Lutz 1991:29-37; Lass 1992:62). Since /h/ had from then on no longer clustered with any other consonants, assessing the consonantal force of lexeme onsets has so far only taken into account /h/ alone. However, in the course of the EModE period, the new cluster /hj/ evolved as a consequence of a phonetic change in a ME vowel: by the end of the fifteenth century, the native diphthongs /iu/ and /eu/ and the

loan phoneme /y:/ (from Old French /y/), which had in all but the most cultivated registers merged with the former, had only one common realization /iu/. Due to a lengthening of the second element, the phoneme developed into the glide plus vowel combination /ju:/ (cf. Dobson 1968:700-713; Lass 1992:50-55; 1999:99). According to Lass (1999:100), this shift took place between 1650 and 1700; in contrast, Dobson's (1968:709) evidence indicates that it began "at least as early as the last decade of the sixteenth century (and probably as early as the 1560's)" and first affected vowels in initial position. The increased rate of change at the beginnings of words can be explained with reference to the preference for filled onsets: English words typically carry stress on the first (stem-initial) syllable, which is therefore particularly prominent, and loanwords tended to assimilate to this pattern relatively fast (cf. Minkova 2000a:440, 451). Thus, an empty onset under the primary stress can be expected to constitute a particularly serious infraction of the preference for filled onsets, and the development of a glide came in handy to amend this. Dobson (1968:705) furthermore informs us that the /ju:/ pronunciation was considered "barbarous" by orthoepists of the EModE period, so that we have to reckon with a delayed penetration of the change into the standard language that was at the basis of written texts. As to the date of the completion of the change, Dobson remains silent, but towards the end of the seventeenth century, he considers /ju:/ more likely than /iu/ (1968:712).

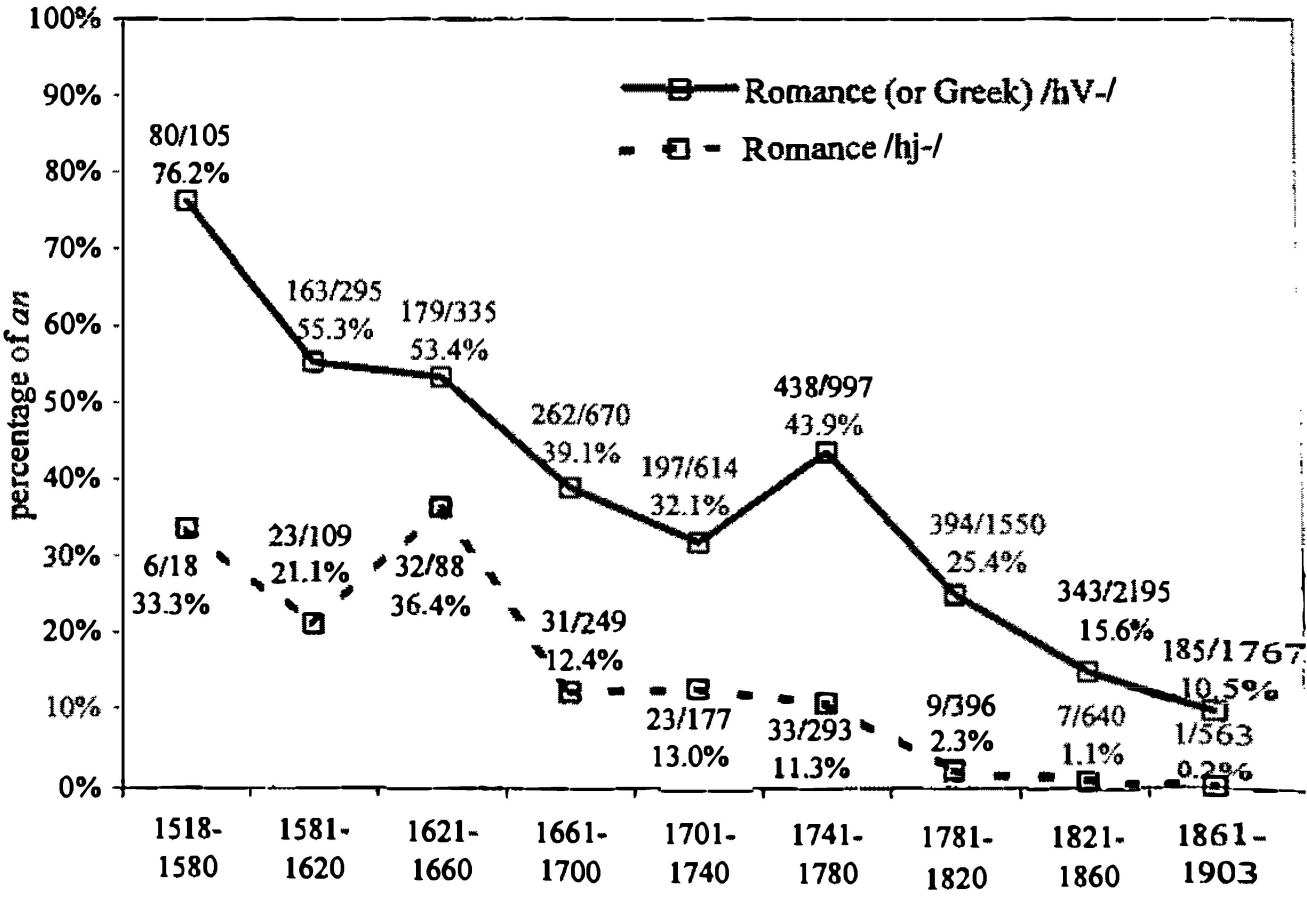
The concordances of <h>-words studied in the previous sections contain a considerable number of lexemes, of both native and foreign origin, whose first vowel underwent the evolution just described, so that they standardly begin in /hju:/ in the present day.[9] Examples of Germanic origin include mainly *hewer* and *hue*, and the most frequent Romance loanwords concerned are *huge*, *human/-e/-ity*, *humility/ -iation/-iating* and *humour/-ist/-ous*. Compared to other <h>-initial lexemes with a pure /h/-onset followed by a vowel (/hV/), these items

[9] On the variant realization of the cluster /hj/ as [ç], see Gimson (1994:192) and Lutz (1991:57, notes 103 and 104).

by hypothesis boast a stronger onset, combining the consonantal force of an increasingly articulated /h/ and the newly forming glide /j/.

Focusing on Romance (or Greek) words,[10] the dataset represented by the black curve in Figure 3 (already excluding the group *heir, honest, honour, hour*) can be split up further into two subcategories: the /hj/-items can be predicted to be ahead of /hV/-items of the same etymological provenance in the abandonment of the long variant of the indefinite article. Figure 4 illustrates the results of this subdivision.

Figure 4. The distribution of *a* and *an* before Romance (or Greek) loanwords with /hV/- and /hj/-onsets



The figure shows a clear internal differentiation within the class of loanwords in the expected direction: in every diachronic subsection of the corpus, we find a statistically highly significant contrast between

[10] Parallel findings have emerged from a study of native lexemes, which is not described here since the amount of data for Germanic /hj/-onsets is less satisfactory.

the respective proportions of *an*-selection.[11] This indicates that the development of the /j/-glide out of the diphthongal nucleus in the first syllable increases the adequacy of the onset to function as fully consonantal. After the exclusion of the /hj/-items, the remaining Romance or Greek <h>-words appear even more conservative as far as the reinforcement of initial /h/ and the concomitant abandonment of *an* in favour of *a* is concerned.

Diachronically, Figure 4 suggests that the impact of the glide formation was clearly noticeable even in the earliest subcorpus, which confirms Dobson's early dating of the change. The implementation of the change, however, just like the resurrection of initial /h/, seems to have spanned an extended period of far more than three centuries. Astonishingly, while /hj/-items virtually behave like full consonant-initial words by the turn of the nineteenth century, items with a pure /h/-onset still preserve a considerable margin of variability, indicating that the reintroduction of initial /h/ across the board is still incomplete at the end of the century.[12]

In sum, closer analysis of Romance <h>-initial lexemes shows that the formation of a glide on the basis of the LME diphthong /iu/ – which proceeds simultaneously with but independently of the re-emergence of pronounced /h/ – gives rise to an inequality among members of these two etymological categories: the combination of two consonantally weak, but stabilizing sounds significantly boosts the consonantal strength of the onset cluster /hj/, leaving single /h/ far behind. This concludes the discussion of the richly varied category of <h>-initial words and leads over to a study of two related sets of vowel-initial lexemes.

---

[11] The chi-square test yields the following error probabilities: 1518-1580: $p = 0.00022$ (***); 1581-1620: $p = 9.77 \cdot 10^{-10}$ (***); 1621-1660: $p = 0.0044$ (***); 1661-1700: $p = 1.29 \cdot 10^{-14}$ (***); 1701-1740: $p = 5.92 \cdot 10^{-17}$ (***); 1741-1780: $p = 1.77 \cdot 10^{-24}$ (***); 1781-1820: $p = 3.50 \cdot 10^{-24}$ (***); 1821-1860: $p = 8.03 \cdot 10^{-23}$ (***); 1861-1903: $p = 4.27 \cdot 10^{-15}$ (***).

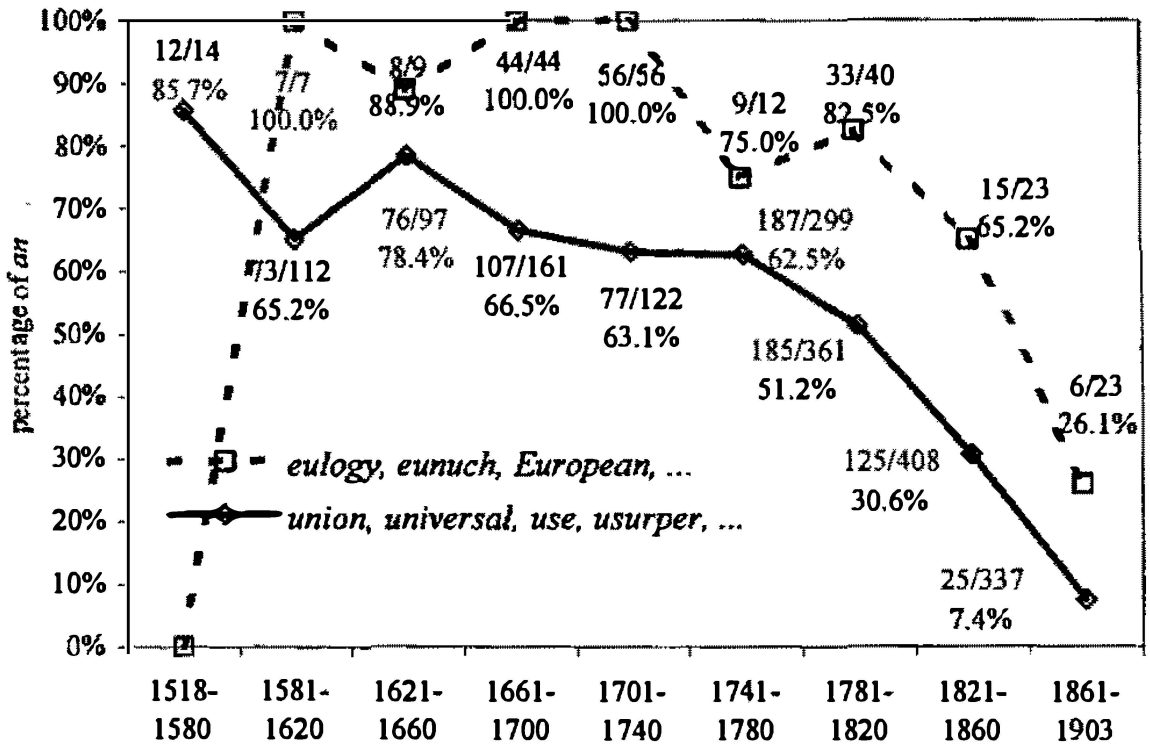
[12] The non-negligible remains of weak initial <h> in Present-Day English are analyzed in detail in Schlüter (2003:83-88; forthc. b). Suffice it to say that the absence of stress on the initial syllable is the main factor raising the percentage of *an*-selections, in particular in British English.

2.4 Upstart onsets: the development of /ju:/

The emergence of the glide plus vowel combination /ju:/ can also be traced in the evolution of lexemes without an initial <h>. There are two classes of lexemes that are at issue here. One class goes back to French loanwords etymologically beginning in /y/, which had commonly been replaced by ME /iu/, as has already been pointed out. To this class belong the lexemes *union/-que/-form*, *unit/-y/-ed*, *universe/-ity/-al/-ality*, *use/-ful/-less*, *usual/-rer*, *usurper/-ation* and *unanimous/-ity*. The second, smaller, class goes back to Greek roots, which also arrived in English via Latin and French and had an initial diphthong /ɛu/ in Middle English (cf. Dobson 1968:799; Lass 1992:50). The principal representatives of this class that can be found following the indefinite article are *eulogy/-ium*, *eunuch* and *European*. According to Dobson (1968:798-799), the first component of the ME diphthong /ɛu/ was raised, passing through /eu/ and eventually merging with the reflex of /iu/. As shown above, this latter sound must at that time have been developing the initial glide /j/, and former /ɛu/ followed suit. Dobson suggests that the loss of the distinction between the two phonemes began possibly as early as the end of the fourteenth century in Northern and Eastern dialects, spread to the South in the sixteenth century, and affected the more conservative careful standard pronunciation around the mid-seventeenth century. According to this chronology, the realization /ju:/ must have appeared earlier for the <u>-words than for the <eu>-words, but the difference should have been neutralized by the late seventeenth century. In Lass' (1999:100) account of the chronology, the merger and the development of /ju:/ coincided in the second half of the seventeenth century. The <u>-initial words would correspondingly be expected to develop the initial glide no earlier than the <eu>-words.

The corpus study outlined in Figure 5 measures the choice of the allomorphs of the indefinite article as a diagnostic of the transition of the two word classes from V-initial to C-initial realizations.

Figure 5. The distribution of *a* and *an* before two categories of originally V-initial lexemes forming new /j/-onsets



The graph reveals the expected continuous establishment of a consonantal onset, making the full form of the article increasingly superfluous. As already stated in section 2.3, the effects of the change are visible long before Lass' (1999:99) dating. It is also obvious that the <u>-words were more advanced than the <eu>-words, as a consequence of the fact that former /ɛu/ joined into the evolution of /ju:/ only in the mid-seventeenth century.[13] A comparison of the <u>-initial words to the <hu>-initial words in Figure 4 (the grey curve) moreover exhibits an obvious difference between these categories, which follows from the summation of the consonantal properties

[13] The results of the chi-square test are statistically highly significant in all subcorpora in which the dataset is sufficiently large. In some subsections, the rarity of <eu>-words makes the test not applicable (n.a.). 1518-1580: –; 1581-1620: n.a.; 1621-1660: n.a.; 1661-1700: $p = 7.60 \cdot 10^{-06}$ (***); 1701-1740: $p = 1.46 \cdot 10^{-07}$ (***); 1741-1780: n.a.; 1781-1820: $p = 0.00017$ (***); 1821-1860: $p = 0.00057$ (***); 1861-1903: $p = 0.0020$ (**).

contributed by both the /h/- and /j/-sounds: the developing glide on its own takes considerably longer to acquire a full consonantal status than in combination with the glottal fricative.[14]

While Dobson's dating of the incipient change is largely confirmed by the corpus data portrayed in Figure 5, the further time course disagrees with his findings in two respects. For one thing, the establishment of /ju:/ in <u>- as well as <eu>-words gains momentum mainly as late as the nineteenth century, and its endpoint is still pending at the turn of the twentieth century. For another, even in the last subperiod investigated, we witness a robust and statistically reliable difference between the reflexes of former /iu/ and /ɛu/. This necessitates a substantial revision of the chronology proposed by Dobson (1968:798), who dates the last convincing piece of evidence of a distinction very precisely to a description by the orthoepist Owen Price in the year 1668.

While this is not the place to discuss the problem of phonemic mergers in any detail, suffice it to point to a way in which the discrepancy between empirical measures and the judgement of members of a speech community can be reconciled. Labov (1994:349-418) adduces six examples of 'near mergers', i.e. cases in which speakers of various contemporary dialects consistently produce a difference between two sounds that they are at the same time unable to distinguish perceptually. If Dobson's information drawn from orthoepists' observations holds true, and if the indefinite article is a valid measure of the realization of onsets as well, this suggests the conclusion that we are dealing with a production-perception asymmetry of the kind described as a 'near merger', which may have been carried to completion only in the twentieth century.[15]

---

[14] The difference between <u>- and <hu>-words is statistically highly significant throughout: 1518-1580: $p = 0.0030$ (**); 1581-1620: $p = 3.87 \cdot 10^{-11}$ (***); 1621-1660: $p = 7.21 \cdot 10^{-09}$ (***); 1661-1700: $p = 1.28 \cdot 10^{-29}$ (***); 1701-1740: $p = 1.75 \cdot 10^{-19}$ (***); 1741-1780: $p = 4.01 \cdot 10^{-38}$ (***); 1781-1820: $p = 1.29 \cdot 10^{-53}$ (***); 1821-1860: $p = 7.19 \cdot 10^{-45}$ (***); 1861-1903: $p = 3.46 \cdot 10^{-10}$ (***).

[15] The status of /h/ in Middle English is argued to be another instance of a near merger in Schlüter (forthc. a).

3. Conclusion

The discussion in this contribution has focussed on a set of sound changes stretching all across the EModE and LModE eras and sharing the characteristic of adding a consonantal onset to formerly V-initial lexemes. This leads to a gradual reduction of the range of application of the long form of the indefinite article. The functional motivations promoting these changes can be assumed to include a tendency to optimize the syllable structure of the lexemes concerned as well as a minimization of the necessity to resyllabify the final /n/ of the article (which negatively affects the phonological integrity of both the article and the following word).

The two principal changes considered here are the reintroduction or reinforcement of initial /h/ in lexemes containing this sound in their etymologies (sections 2.1–2.3) and the new formation of the glide /j/ in reflexes of the ME vowels /iu/, /eu/, /y/ and /ɛu/ (sections 2.3–2.4). The analysis in section 2.3 has shown that the effects of the two changes add up in environments where both overlap. As additional factors, the etymological origin and the frequency of <h>-initial words have been brought into play. Section 2.1 has provided evidence in favour of a strong effect of the phonology of the language of origin, which remains noticeable in loanwords for a long time after their adoption. Thus, Romance loanwords preserve a particularly weak /h/ for several centuries in English, and Germanic loanwords re-adopted from French also preserve the marks of their provenance in the form of an undiminished consonantal value of their onset, which is indicative of a continued realization of /h/ in French. Section 2.2 has demonstrated the conservative effect of high frequency, which opens up an avenue to explaining the continued /h/-lessness of the stems *hour*, *honour*, *honest* and *heir*.

The interpretation of the statistical data presented in this contribution is open as to the source of the variation observed: it can be put down to variation in the individual language user, who responds spontaneously to the conditions obtaining in a particular instance of the lexical items discussed, or it can be ascribed to interindividual variation with every speaker (or writer) adhering to an

invariable lexical representation of each item that includes or excludes an onset consonant. The former view suggests a homogeneous speech community, the latter implies a heterogeneous one, but the most likely hypothesis is that both types of variability combine to produce the highly complex picture painted by the above analyses. Though there is no direct evidence, we have reason to assume that the actual phonetic strength of a consonantal onset plays a major role in judging a lexeme to be either C- or V-initial. This binary distinction (which is mirrored in the selection of *a* or *an*) depends on whether a certain threshold of consonantal strength is crossed.[16] Consonantal strength, however, is a gradual phenomenon dependent on systematic parameters that are beyond the conscious control of language users. Thus, even educated speakers of Early Modern English and Late Modern English can hardly be expected to be aware of the etymological distinction between native, borrowed and re-borrowed lexemes. In addition, lexeme frequency is a gradual phenomenon that translates into stochastic probabilities rather than into unique lexical representations. Finally, the additive treatment of developing /h/- and /j/-onsets speaks clearly in favour of an account in terms of phonetic strength that comes to bear in the assignment of *a* or *an* to the lexeme in question.

To summarize the combined effects of this set of factors, a relative chronology of the changes studied can be established: when the relevant curves from Figures 2 to 5 are superimposed on each other, there is practically no crossover between two lines (with the exception of straightforward and re-borrowed Germanic <h>-words, which are virtually indistinct). On the basis of these data, a hierarchy of consonanthood can be established, which is visualized in Table 2. This hierarchy is consistent on the synchronic as well as diachronic dimensions, i.e. a higher hierarchical position translates into both a

[16] The competing interpretation according to which the variability is an artefact of a categorical change in lexical representations distributing along a social and historical dimension cannot be completely excluded. However, the phonetic parameters underlying the change and the systematic character of the variation pattern are accommodated more easily in a model in which the interplay of several functional factors enables language users to come to phonetically grounded decisions in every single instance of the lexemes considered.

relatively higher level of consonanthood at a given point in time and a chronologically earlier transition to a C-initial status.

Table 2: Hierarchy of consonantal force for the onsets studied

← least consonantal					→ most consonantal
Romance loanwords with 'mute' <h>	**Romance or Greek loanwords with /ɛu/ > /ju:/**	**Romance loanwords with /iu/ > /ju:/**	**Romance loanwords with /hV/**	**Romance loanwords with /hiu/ > /hju:/**	**Native Germanic or other words with /h/**
hour, honour, honest, heir, ...	*Eulogy, eunuch, European, ...*	*union, universal, use, usurper, ...*	*habit, hackney, heroic, historical, hospital, ...*	*huge, human, humility, humour, ...*	*hair, hand, house, husband, whole, whore, ... hindoo, hookah, ... halberd, halt, hardy, haste, heinous, ...*

What remains to be said in conclusion is that the allomorphy of the indefinite article has proved to be a particularly instructive diagnostic of sound changes in progress. Its frequency and accessibility to electronic full-text searches have allowed us to refine the categorization of the lexemes at the centre of the analysis step by step and to establish and quantify a phonetically plausible synchronic and diachronic gradation of their onsets. The suitability of the indefinite article as a tool for measuring historical sound change is what makes this small word so deserving of linguistic interest.

References

Electronic text databases

Early English Prose Fiction (*EEPF*). 1997. Electronic Book Technologies Inc. and Chadwyck-Healey. Cambridge. In association with the Salzburg Centre for Research on the English Novel SCREEN.

Eighteenth-Century Fiction (*ECF*). 1996. Electronic Book Technologies Inc. and Chadwyck-Healey. Cambridge.

English Prose Drama (*EPD*). 1996/1997. Electronic Book Technologies Inc. and Chadwyck-Healey. Cambridge.

Nineteenth-Century Fiction (*NCF*). 1999/2000. Electronic Book Technologies Inc. and Chadwyck-Healey. Cambridge.

Secondary sources

Baugh, Albert C. and Thomas Cable. 1993. *A history of the English language*. 4th ed. London: Routledge.

Crisma, Paola. (forthc.). "Were they 'dropping their aitches'? A quantitative study of *h*-loss in Middle English". *English Language and Linguistics*.

Dekeyser, Xavier. 1986. "Romance loans in Middle English: a re-assessment". In: Kastovsky, Dieter and Aleksander Szwedek (eds.). *Linguistics across historical and geographical boundaries: in honour of Jacek Fisiak on the occasion of his fiftieth birthday. Volume I: Linguistic theory and historical linguistics*. Berlin: de Gruyter: 253-265.

Dobson, Eric J. 1968. *English pronunciation 1500-1700. Volume II: Phonology*. 2nd ed. Oxford: Clarendon.

Gimson, Alfred Charles. 1994. *Gimson's pronunciation of English*. 5th ed., revised by Alan Cruttenden. London: Arnold.

Graband, Gerhard. 1965. *Die Entwicklung der frühneuenglischen Nominalflexion: Dargestellt vornehmlich auf Grund von Grammatikerzeugnissen des 17. Jahrhunderts*. Tübingen: Narr.

Häcker, Martina. 1998. "Why is there no /h/-dropping in Scots? Loss and insertion of /h/ as a contact phenomenon in British English dialects". In: Strässler, Jürg (ed.). *Tendenzen europäischer Linguistik: Akten des 31. Linguistischen Kolloquiums, Bern 1996*. Tübingen: Niemeyer: 71-76.

Häcker, Martina. 2004. "Intrusive [h] in Present-Day English accents and ‹h›-insertion in medieval manuscripts: hypercorrection or functionally-motivated language use?" In: Kay, Christian (ed.). *New perspectives on English historical linguistics*, vol. II. Amsterdam: Benjamins: 109-123.

Ihalainen, Ossi. 1994. "The dialects of England since 1776". In: Burchfield, Robert (ed.). *The Cambridge history of the English language. Volume V: English in Britain and overseas: origins and development*. Cambridge: CUP: 197-274.

Labov, William. 1994. *Principles of linguistic change. Volume I: Internal factors*. Oxford: Blackwell.

Lass, Roger. 1992. "Phonology and morphology". In: Blake, Norman (ed.). *The Cambridge history of the English language. Volume II: 1066-1476*. Cambridge: CUP: 23-155.

Lass, Roger. 1999. "Phonology and morphology". In: Hogg, Richard (ed.). *The Cambridge history of the English language. Volume III: 1476-1776*. Cambridge: CUP: 56-186.

Lutz, Angelika. 1991. *Phonotaktisch gesteuerte Konsonantenveränderungen in der Geschichte des Englischen.* Tübingen: Niemeyer.

Milroy, James. 1983. "On the sociolinguistic history of /h/-dropping in English". In: Davenport, Michael, Erik Hansen and Hans Frede Nielsen (eds.). *Current topics in English historical linguistics.* Odense: Odense University Press: 37-53.

Milroy, James. 1992. *Linguistic variation and change: on the historical sociolinguistics of English.* Oxford: Blackwell.

Minkova, Donka. 2000a. "Middle English prosodic innovations and their testability in verse". In: Taavitsainen, Irma, Terttu Nevalainen, Päivi Pahta and Matti Rissanen (eds.). *Placing Middle English in context.* Berlin: de Gruyter: 431-459.

-----. 2000b. "Syllable ONSET in the history of English". In: Bermúdez-Otero, Ricardo, David Denison, Richard M. Hogg and Chris B. McCully (eds.). *Generative theory and corpus studies: a dialogue from 10 ICEHL.* Berlin: de Gruyter: 499-540.

-----. 2003. *Alliteration and sound change in Early English.* Cambridge: CUP.

Mugglestone, Lynda. 1995. *'Talking proper': the rise of accent as social symbol.* Oxford: Clarendon.

Nespor, Marina and Irene Vogel. 1986. *Prosodic phonology.* Dordrecht: Foris.

OED 2 on CD-ROM. 1995. Version 1.13. Oxford: OUP and Rotterdam: AND Software B.V.

Pope, Mildred K. 1934. *From Latin to Modern French with especial consideration of Anglo-Norman: phonology and morphology.* Manchester: Manchester University Press.

Schlüter, Julia. 2003. "Phonological determinants of grammatical variation in English: Chomsky's worst possible case". In: Rohdenburg, Günther and Britta Mondorf (eds.). *Determinants of grammatical variation in English.* Berlin: de Gruyter: 69-118.

-----. (forthc. a). "Consonant or 'vowel'? A diachronic study of initial ‹h› from Early Middle English to nineteenth-century English". In: Minkova, Donka (ed.). *Phonetic and phonological weakness in the history of English* (working title).

-----. (forthc. b). "Rhythm and grammar". In: Rohdenburg, Günter and Julia Schlüter (eds.). *One language, two grammars? Differences between British and American English.* (SEL.) Cambridge: CUP.

Strang, Barbara M.H. 1970. *A history of English.* London: Methuen.

Sundby, Bertil, Anne Kari Bjørge and Kari E. Haughland. 1991. *A dictionary of English normative grammar 1700-1800.* Amsterdam: Benjamins.

Vennemann, Theo. 1988. *Preference laws for syllable structure and the explanation of sound change: with special reference to German, Germanic, Italian, and Latin.* Berlin: de Gruyter.