Tracing the (re-)emergence of onset consonants through 500 years of books: Big data on a detail of historical English phonetics and phonology

Julia Schlüter

In this talk I focus on the evolution of three types of onset consonants in word-initial position, all of which share a trend towards reinforcement, thus implementing the universal preference for filled syllable onsets.

- One group includes words with an initial <u> borrowed into English from French and Latin (e.g. *union*, *universe*, *use*, *usurper*), in which the Middle English diphthong /iu/ developed into the glide-plus-vowel combination /ju:/. This change has been controversially dated to the second half of the 16th century (Dobson 1968: 709) or a century later (Lass 1999: 100).

- The second group comprises loanwords of Greek origin spelled with an initial <eu> (e.g. *eulogy*, *eunuch*, *European*), whose original diphthong /ɛu/ merged with the reflex of /iu/ (Dobson 1968: 789-799) and joined into the evolution to /ju:/ – at a period of time that still needs to be determined.

- The third and most diversified group is provided by words spelled with an initial <h>, which was gradually re-introduced in the pronunciation. This large group is constituted by native (*hand*, *help*, *harbour*, *half*, *happy*, *hard*) as well as borrowed words. The latter fall into several categories according to the stress level of their initial syllables (primary: *habit*, *hero*; secondary: *homogeneous*, *hieroglyphic*; zero: *historical*, *habitual*) and the length of their initial vowels (short: *horrible*, *homologous*; long: *hierarchy*, *hyperbole*). They also contain special subgroups, such as items of Germanic origin re-borrowed from French (e.g. *hardy*, *heraldic*, *helmet*) and items preserving a mute <h> to the present day (*hour*, *honest*, *heir*). Various cross-category-memberships are also possible, e.g. items with <h> followed by <u> or <eu> (*humility*, *heuristic*).

These changes can be viewed as a partial recovery from a change that had affected the onsets of stressed syllables in Middle English (cf. Minkova 2003: 149-60): The former requirement of a filled onset position had been downgraded to a mere preference.

Concerning the group of <h>-initial words, I have already shown (Schlüter 2009) that initial /h/ – though well-established in the Old English period – had become virtually mute in Middle English and during the Modern English period underwent a gradual process of re-establishment, which is slightly but demonstrably different between the several subgroups portrayed above. In this paper, the exact timing of these developments and interactions between the abovementioned factors will be scrutinized at a level of granularity unattained up until now.

As a diagnostic of the changes under investigation, this study will exploit the variability of the indefinite article *a/an*, which since early Middle English (when the loss of its final /n/ before consonant-initial words was complete) has displayed a highly regular and consistent distribution. The phonotactic constraints underlying this mechanism can be taken to be the preference for filled onsets (even at the expense of resyllabification of the /n/ across a word boundary) and the prevention of hiatuses and consonant clusters within tightly bound prosodic units.

The paper will provide evidence from a textbase at a scale unavailable until recently: It will draw on the raw data of the Google Books Ngram Viewer, which provides access to the full text of over 4.5 million books in English, totalling over 468 billion words, selected from the Google books collection (cf. Lin et al. 2012). The sheer size of this mega-corpus and its coverage of roughly five centuries afford us the possibility to attain a maximal diachronic resolution, to distinguish highly specific groups of lexical items and even to trace the diffusion of the observed consonantal changes across lexical units while spotting deviant items. To this end, the raw data extracted from Google Books will be visualized in

various ways through an interactive web application based on the RStudio environment (Chang et al. 2016). This allows us to graph the data individually and as groups. Changes take the shape of S-curves and can be viewed as a combination of many superimposed individual curves. This prototypical outcome and the fact that parallel findings have been collected from scholarly editions of sizeable literature databases (though at a much smaller scale; cf. Schlüter 2006) lend further plausibility to the results.

An explorative analysis of the data confirms the relevance of the lexical characteristics investigated on a synchronic as well as diachronic level. Besides establishing the empirical facts of the (re-)emergence of consonantal onsets in English, the present study allows for some more theoretical conclusions regarding the mental processing of word-initial sounds. The data suggest that syllable stress, vowel quantity, consonantal reinforcement by a /j/-glide and the etymological dimension in the case of <h>-initial words conjointly determine the consonantal force of the onset, which varies on a scale from fully consonantal to practically imperceptible. This gradual phonetic feature maps onto a binary phonological distinction with a cut-off point: Below a certain threshold, the lexical item is perceived as vowel-initial (and thus combined with *an*); above, it is perceived as consonant-initial (and preceded by *a*). The multiple factors influencing this perception appear to be weighed by speakers in a case-by-case decision on whether a word is treated as consonant- or vowel-initial.

By extension, this argument can be applied to the diachronic plane, where it provides an avenue to understanding how weakened etymological /h/ has remained in place without however being perceived as a consonantal onset during much of the Middle English era: Weak consonants can retain a subliminal presence below speakers' level of awareness. A diachronic increase in onset strength can lead to a switch from vowel-initial to consonant-initial and thereby from determiner *an* to *a* (by individual speakers/by tokens of occurrence). In line with the unmerging of phonemic mergers for which Labov (1994: 349-418) provides an intriguing explanation, this assumption helps to explain the astonishing re-emergence of a seemingly lost phoneme.

References

Chang, Winston; Cheng, Joe; Allaire, J.J.; Xie, Yihui & McPherson, Jonathan. 2016. Shiny: Web Application Framework for R. <https://shiny.rstudio.com/>

Dobson, Eric J. 1968. English Pronunciation 1500-1700. Volume II: Phonology. Second edition. Oxford: Clarendon.

Labov, William. 1994. Principles of Linguistic Change. Volume I: Internal Factors. Oxford, UK/Cambridge, Mass.: Blackwell.

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

Lin, Yuri; Michel, Jean-Baptiste; Lieberman Aiden, Erez; Orwant, Jon; Brockman, Will & Petrov, Slav. 2012. "Syntactic Annotations for the Google Books Ngram Corpus." Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics. Jeju, Republic of Korea, 8-14 July 2012. 169-174.

Minkova, Donka. 2003. Alliteration and Sound Change in Early English. Cambridge: Cambridge University Press.

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: Ritt, Nikolaus; Schendl, Herbert; Dalton-Puffer, Christiane & Kastovsky, Dieter (eds.) Medieval English and its Heritage: Structure, Meaning and Mechanisms of Change. Frankfurt: Lang. 37-59.

Schlüter, Julia. 2009. "Consonant or 'vowel'? A diachronic study of the status of initial ‹h› from early Middle English to nineteenth-century English." In: Minkova, Donka (ed.) Phonological Weakness in English: From Old to Present-Day English. Houndmills, Basingstoke, Hampshire and New York, NY: Palgrave Macmillan. 168-196.