

Further evidence for self-organization in English spelling. **Draft version**
Kristian Berg & Mark Aronoff
Kristian.berg@uni-oldenburg.de

Geoffrey Sampson makes two claims in his response to our article about self-organization in spelling (Berg & Aronoff, 2017): our explanation is redundant and the phenomena we address can be better captured in more traditional terms. In our article, we showed first that the relation between graphemic form and morphological function is isomorphic for some suffixes, e.g. the adjectival suffix *-ous*: In today's English, all and only adjectives with this morpheme are spelled with final <ous>, even though phonographically, more words could be spelled that way in American varieties (cf. *nervous/service*). We also demonstrated that this system evolved gradually in an unsupervised process of self-organization. Sampson does not contest our synchronic statement, but brings forth a variety of distinct accounts of our diachronic findings. Overall, he suggests that phonology and etymology are sufficient to explain why the spellings are the way they are: In British English, *nervous* and *service* have (and more importantly, *had*) distinct reduced vowels. The distinct spelling is thus explicable from phonology here. Other spellings, Sampson argues, follow a widespread convention preferring etymological spellings: *service* has an <i> in Latin, and <ous> stems from Latin *-os-*; that is why *<servous> is not the accepted form. No need for a distinct principle of self-organization, says Sampson, because all the data are accounted for already.

There are three misunderstandings in Sampson's response. We take responsibility for the first, though not for the second and third. Let us start with the first. Synchronically, we showed that graphemic form and morphological function are isomorphic. We then ask how this situation arose historically, but the fact is that we looked at only half the historical story. To show this, we need to step back a little and look at the relation between form and function. It is fine to talk about isomorphism as above – but it can be helpful to take the relation apart: We can distinguish between *uniformity* and *uniqueness*. Uniformity denotes the degree of consistency with which a function is represented in its form: Is there more than one spelling for the suffix? Today, *-ous* is spelled (almost) uniformly <ous>¹, but in earlier stages of English, there were variants (e.g. <us>, <ouse>), so we can say that the spelling of this adjective suffix has become more uniform over the centuries. Uniqueness flips the perspective; it denotes the degree of consistency with which a form represents a function. Is there more than one

¹ There is one systematic exception: In cases where *-ous* derivations are themselves derived with *-ity*, *-ous* is spelled <os> as in *generous* – *generosity*.

interpretation for this string of graphemes? While <ous> is unique in representing this one adjective suffix, other suffix spellings are not; <er> represents (among others) the agentive nominal suffix in <singer> and the comparative suffix in <longer>, but also a non-morphological word ending in <hammer>. Both uniformity and uniqueness are independent of phonology: they deal with the relation between spelling and morphology. When we take phonology into account, it is as circumstantial evidence: The fact that a number of words in today's American English could potentially be spelled with <ous> (e.g. *service*) is a further argument for the uniqueness.

Synchronically, we showed that *-ous*/*<ous>* is both unique and uniform – but diachronically, we investigated only the uniformity of the suffix. We asked whether Latinate loans like *status* were responsible for the decline of <us>-variant of *-ous* like <humerus> (short answer: probably not) – that is a question of uniqueness. But we did not investigate the uniqueness of each spelling variant in each time period, simply because the investigation was extensive enough already. That does not mean the question is uninteresting. A superficial search in the Helsinki corpus (a selection of English texts produced before 1710) indeed shows a number of words that end like variant spellings of the suffix *-ous* (<ous>, <ouse>), but which did not contain the suffix, e.g. <almous dedes> ('alms'), <felouse> ('fellows'), <alehous> ('alehouse') or <sparous> ('sparrows'). It remains an open question how unique the respective spellings were at any given time.

We now have data that show a striking case of the emergence of uniqueness over time in the spelling of the nominal and adjectival suffixes *-y*. Of the three suffixes that the letter <y> represents (adjectival *-y*, as in *windy*, nominal *-y* as in *harmony*, and diminutive *-y* as in *granny* (Bauer, Lieber & Plag 2013), one is leaving the set: the diminutive suffix is increasingly often spelled <ie>. To show this, we use a list of 431 OED entries with the diminutive suffix *-y* (suffix *-y*⁶ in OED terminology)² and check each variant spelling in the Corpus of Historical American English (CoHA, <http://corpus.byu.edu/coha/>). The corpus contains over 400 million words from 1810 to 2000. We searched 331 words³ in both spelling variants in the corpus and calculated the ratio of <y>-spellings for each word in each decade. Plotting the mean of these ratios over time results in the following graph:

² We would like to thank Oxford University Press for providing us with this list.

³ 100 words were excluded because a) they end with <ey>, and the <ie> variant is graphotactically impossible (e.g. <nursey>, *<nurseie>); b) the entry is a plural form (e.g. *panties*); or c) they are homographic to other forms in the list (e.g. *rummy* 'card game' vs. *rummy* 'odd person').

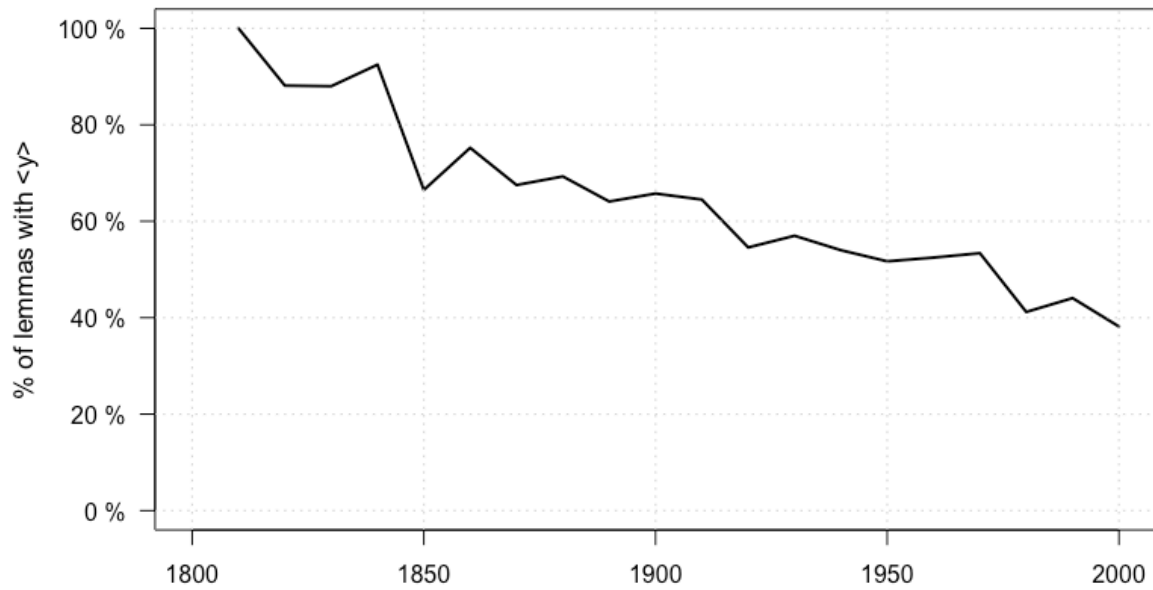


Figure 1: Percentage of diminutives with OED suffix -y⁶ that are spelled with <y> as opposed to <ie> in the Corpus of Historical American English.

The fraction of types that are spelled with <y> is steadily declining. More and more diminutives are spelled with <ie> instead; this is part of a trend that spans two centuries.

Not only does the general trend for the set of diminutives point towards <ie>, but newly coined diminutives also appear in the <ie>-spelling more and more often. We show this using the OED data on the first recorded usage. For each of the 183 words in the list of diminutives that first appear a) in singular form and b) 1780 or later (before that, there are too few new words to draw reliable conclusions), we noted the year of the first recording and whether this first recorded form ends with <y> or <ie>. In the following graph, the ratio of newly coined words with <y> (as a fraction of all newly appearing diminutives with <y> and <ie>) is plotted in 20-year bins:

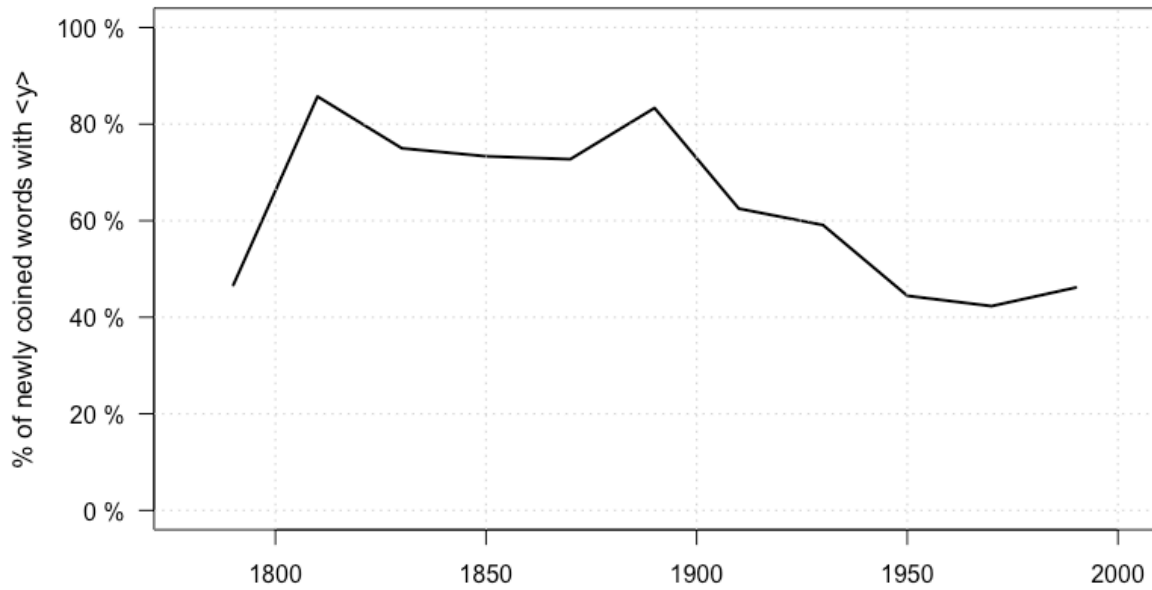


Figure 2: Percentage of newly appearing diminutives in the OED records with OED suffix $-y^6$ that are spelled with <y> as opposed to <ie>.

From 1800 on, the newly recorded diminutives also show a trend towards <ie>, from more than 80% with <y> in 1800-1820 to less than 50% in the second half of the 20th century. The lower level before 1800 can be explained by the fact that most new words at that time are Scots loans, as Sampson rightly points out; by 1800, however, the majority of words in each time span is formed in English, where the <y> spelling predominates throughout the 19th century. Both sets of data – the way diminutives are dominantly spelled and the way they are spelled when they first appear – support our hypothesis that the uniqueness of <y> is increasing by ‘outsourcing’ the diminutive suffix to the form <ie>.

But even if we limited the scope and investigate only uniformity diachronically, that is, without the arguments we just provided that the uniqueness is increasing for <y> – would that invalidate our point? Most certainly not. We take the fact that spelling was standardized without external influence as an argument for the self-organization of the writing system. This claim is of course controversial and far from self-evident. It taps into the age-old discussion of who is responsible for the standardization of English: printers or scholars? The long-established view was that it was the former: “the printer with his professional sense of the importance of the mechanical side of his art, always strives for complete consistency and regularity” (Krapp 1909: 172). But then Brengelman (1980) famously argued the case for 16th and 17th century scholars: “There is no evidence whatever of any printing house taking a serious interest in the regularization of English spelling. [...] There is abundant evidence, on

the other hand, that linguistic scholars knew each other's work well, that schoolmasters followed their recommendations, and that the spelling practice that emerged late in the [17th] century was the result of this collaboration" (Bregelman 1980: 333f).⁴ Unfortunately, though, the data are not as clear as Bregelman suggests. Take one of the major "improvements brought about by seventeenth-century scholars", namely "the rationalization of the use of final *e*" (Bregelman 1980: 347): silent final <e> had become a marker for vowel quality, distinguishing the phonological vowels in e.g. *mad* and *made*. That in turn meant that final <e> stopped being used after double consonants (as in 16th century spellings like <sunne>, <badde>). We can track this second transition in a diachronic corpus, and we do just that in a small study for seven monosyllabic words (*tell*, *cut*, *mad*, *skin*, *back*, *son*, *cross*) in the large corpus Early English Books Online (EEBO, ~25,000 texts in phase one, ~500 million words). Figure 3 shows the amount of <CCe>-spellings for each of these words over time⁵:

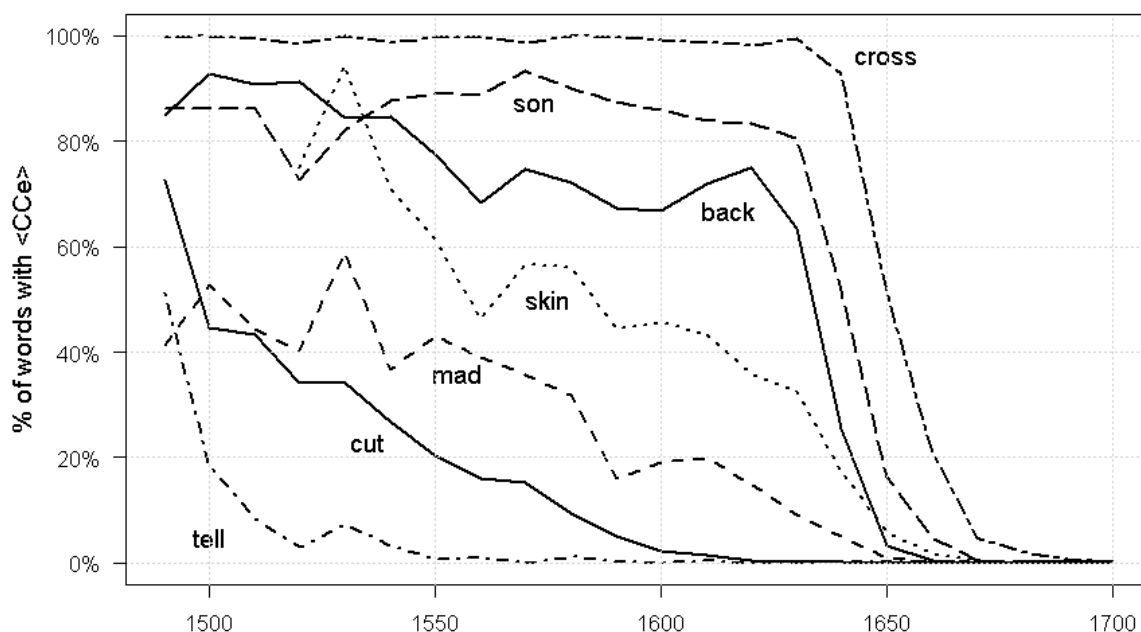


Figure 3: Percentage of occurrences of <telle>, <cutte>/<cvtte>, <madde>, <skinne>, <backe>, <sonne>/<fonne> and <crossse>/<croffe> as a fraction of all occurrences of these words. Database: EEBO corpus.

⁴ Interestingly, the arguments for the scholars often take on a teleological flavor. For example, Carney (1994: 467) notes that "standardization [...] was too well-designed to be a simple settling down of printing-house practices." Darwin long ago taught us to doubt evidence from alleged design. In this case, a lack of "serious interest" in spelling on the part of the printers is precisely what one would expect of a system that was self-organizing rather than deliberately planned.

⁵ For each of the ten words, we searched for potential forms with and without <CCe> (e.g. <sinne>, <finne> vs. <sin>, <fin>, <sinn>, <finn>). We then computed the average ratio of <CCe> tokens compared to all tokens in ten year spans.

The final stage of standardization happens very quickly around 1640-1660, but many of the words show a downward trend for <CCe> variants as early as 1550. That is earlier than the first scholars who, Brengelman (1980) suggests, brought about the change (Levins 1570, Mulcaster 1582, Coote 1596). The structure of the data in figure 3 also speaks for a self-organizing system: Each word has its own trajectory; this can be seen as a graphemic instance of lexical diffusion (cf. e.g. Labov 2005). If printers had actually followed the scholars' advice, the movement should have been a lot more uniform.

But if the standardization of English can be traced neither to printers nor to scholars, how did it come about? The alternative is simple: It was both the scholars and the printers, but neither in the way envisaged by Krapp (1909) or Brengelman (1980). Printers standardized English, but not intentionally. They experimented with different spellings, but only within bounds. Their interest was commercial; they tried to anticipate the way people would like to read and set type accordingly.⁶ This is the textbook definition of self-organization: The emergence of order on a global scale by actions on a local scale (cf. e.g. de Boer 2005)⁷. Where does that leave the scholars? Theirs is the function of a positive feedback loop, as Cummings (1988: 8) suggests: "They [the scholars, KBMA], figuratively speaking, were describing the mean toward which the orthographic system was regressing."

Sampson's second misunderstanding is conceptually related to the first one. Sampson argues that the concept of self-regulation is redundant because the data can be explained by the "convention" of keeping the spelling of Latin words intact. Variants like <publick> were abandoned in favor of other variants closer to the Latin original like <public>. Throughout his response, he seems to conceptualize this convention as a kind of constant in the history of English. But this convention, too, is clearly something that has evolved, and the <ick>-spellings of *-ic* are a particularly persistent reminder of that fact, as we show in our original article. It is the result of self-organization (Sampson himself acknowledges the possibility in one of the final sentences). The fact that many readers could read Latin and Greek certainly contributed to the appeal of the respective forms, but again, it took quite some time until

⁶ Cf. Scragg (1974: 74): "[S]uccessful printers from Caxton onwards have been primarily businessmen, and their only concern as regards spelling has been to provide their public with what is most acceptable". Cf. also Tyrkkö (2013: 157): "The principle aim of the printer was to produce an attractive book that customers would purchase rather than waste time doggedly following the spelling of the manuscript when few authors and fewer customers had any interest in spelling".

⁷ This is what Keller (1994: 61ff.) terms 'invisible hand phenomena'.

words like *public* were standardized. That history indicates that conventions are linguistic abstractions over patterns of usage – and it is these patterns that are immediately relevant to the language community.

Finally, Sampson’s third misunderstanding is a meta-theoretical one. He asserts that “data cannot constitute evidence for a novel theory, if they are already convincingly accounted for independent of that theory.” Following such advice would doom progress in any area of inquiry. A new account is judged by whether it provides us with a greater understanding of a wider range of phenomena than we possessed beforehand, not by whether other accounts of some of the phenomena already exist. In our article, we showed how morphological uniformity and uniqueness help us to make greater sense of English spelling, both synchronically and diachronically. We never denied the importance of other factors or the value of other explanations, including those that Sampson suggests.

One hallmark of written English over the last millennium has been the wholesale importation of words from other languages written in the Roman alphabet, spelling and all, words like *spaghetti* from Italian, *triage* from French, *schloss* from German, and *laddie* from Scots, among many thousands from many languages. This long-standing practice has wrought havoc with the systematicity of English spelling. In some cases, though, including those that we have discussed, it has led to change, for example the shift of diminutive <y>⁶ to <ie> under the influence of Scots that we have demonstrated here. But this change, like many others, makes sense only in the light of the polar organizational drives towards uniqueness and uniformity that we introduced as the centerpieces of our article.

References

- Berg, Kristian & Mark Aronoff. 2017. Self-organization in the spelling of English suffixes: the emergence of culture out of anarchy. *Language* 93. 37-64.
- Brengelman, Frederick. 1980. Orthoepists, printers and the rationalisation of English spelling. *Journal of English and Germanic Philology* 79. 332–54.
- Carney, Edward. 1994. *A survey of English spelling*. London: Routledge.
- Coote, Edward. 1596. *The English Schoole-maister*. London: Widow Orwin, for Ralph Jackson and Robert Dextar.
- Cummings, D. W. (1988). *American English Spelling*. Baltimore, MD: Johns Hopkins University Press.

- de Boer, Bart. 2005. Self-organisation in Language, In: C. Hemelrijk (ed.) Proceedings of the international workshop on Self-organization and Evolution of Social behavior. Cambridge: Cambridge University Press. 123–139.
- Keller, Rudi. 1994. On language change. The invisible hand in language. London: Routledge.
- Krapp, George Philip. 1909. Modern English: Its growth and present use. New York: Charles Scribner's Sons.
- Labov, William. 2005. Principles of Linguistic Change. Volume 1: Internal Factors. Oxford: Blackwell.
- Levins, Peter. 1570. Manipulus Vocabulorum. London: Henrie Bynneman, for John Waley.
- Mulcaster, Richard. 1582. The First Part of the Elementarie. London: Vautroullier.
- Scragg, Donald G. 1974. A history of English spelling. Manchester: University Press.
- Tyrkkö, Jukka. 2013. Printing houses as communities of practice. Orthography in early modern medical books. In: Kopaczyk, Joanna & Andreas H. Jucker (eds.), Communities of Practice in the History of English. Amsterdam: Benjamins. 151–176.