

John Benjamins Publishing Company



This is a contribution from *Crossing Boundaries through Corpora. Innovative corpus approaches within and beyond linguistics*.

Edited by Sarah Buschfeld, Patricia Ronan, Theresa Neumaier, Andreas Weilinghoff and Lisa Westermayer.

© 2024. John Benjamins Publishing Company

This electronic file may not be altered in any way.

The author(s) of this article is/are permitted to use this PDF file to generate printed copies to be used by way of offprints, for their personal use only.

Permission is granted by the publishers to post this file on a closed server which is accessible to members (students and staff) only of the author's/s' institute, it is not permitted to post this PDF on the open internet.

For any other use of this material prior written permission should be obtained from the publishers or through the Copyright Clearance Center (for USA: www.copyright.com).

Please contact rights@benjamins.nl or consult our website: www.benjamins.com

Tables of Contents, abstracts and guidelines are available at www.benjamins.com

New approaches to investigating change in derivational productivity

Gender and internal factors in the development of *-ity* and *-ness*, 1600–1800

Tanja Säily, Martin Hilpert & Jukka Suomela

University of Helsinki | Université de Neuchâtel | Aalto University

We study the productivity of the suffixes *-ness* and *-ity* in seventeenth- and eighteenth-century letters in the *Corpora of Early English Correspondence*. We analyze the role of gender and five internal factors: etymology, the word class of the base, branching structure, semantics, and occurrence in possessive constructions. We develop statistical and visual methods that facilitate diachronic comparisons within factors and between competing suffixes; our basic measure is the proportion of types of interest out of all relevant types, and we utilize permutation testing to assess the statistical significance of our findings. Our results support and refine the earlier finding of a male-led increase in the productivity of *-ity* and provide new information on the interplay of gender and internal factors.

Keywords: morphological productivity, nominal suffixes, historical sociolinguistics, Construction Grammar, methodology

1. Introduction

This paper addresses historical changes in the productivity of English derivational affixes. We aim to shed new light on this topic by adopting a perspective that includes constructional and sociolinguistic aspects. Our focus is on the nominal suffixes *-ness* and *-ity* in seventeenth- and eighteenth-century letters in the *Corpora of Early English Correspondence* (CEEC). Previous work has indicated that the productivity of the borrowed, more learned and prestigious suffix *-ity* increased during this time and that women were lagging behind in this development, whereas the productivity of the native *-ness* remained stable and nearly free from sociolinguistic variation (Säily 2014; cf. Rodríguez-Puente 2020). This

research, however, was based solely on type frequencies, and it did not account for several intralinguistic factors of interest. In order to better understand variation and change in the productivity of *-ness* and *-ity*, in addition to the social factor of gender, we analyze the role of five language-internal factors, namely etymological source (borrowed/derived/inherited), the word class of the base, branching structure (binary/left/right; Hilpert 2013), semantic type (state/thing/person/collectivity; Romaine 1985), and occurrence in possessive constructions (cf. Säily 2014: 253).

To this end, we develop new statistical and visual methods that facilitate diachronic comparisons within factors and between competing suffixes. Similar to Rodríguez-Puente et al. (2022), our basic measure is the proportion of types of interest out of all relevant types, for example the proportion of right-branching *-ity* types out of all *-ity* types, and we make use of permutation testing to assess the statistical significance of our findings. However, while Rodríguez-Puente et al. (2022) used this approach to compare differences between registers within each time period, we extend the methodology so that we can analyze the significance of changes over time, as well as language-internal factors in terms of both type-based and token-based measures.

The rest of this paper is organized as follows. Section 2 discusses previous research on the suffixes, measuring variation and change in morphological productivity, and analyzing productivity change in Construction Grammar. Section 3 introduces our material and methods, Section 4 presents the results, and Section 5 concludes the paper with a discussion of the results and possibilities for future research.

2. Background

2.1 The suffixes *-ity* and *-ness*

The nominal suffixes *-ity* and *-ness* are typically used to derive abstract nouns from adjectives, as in (1a)–(b).

- (1) a. *productive* + *-ness* → *productiveness*
- b. *productive* + *-ity* → *productivity*

While *-ness* is native and cognate with, for example, the German suffix *-nis*, *-ity* emerged as an English suffix through lexical borrowing from French in the Middle English period and was later reinforced through Latin loanwords (Marchand 1969: 312–313). The suffixes can be regarded as part of the “diglossic” vocabulary of English (Adamson 1989), *-ness* being the “default” suffix (Bauer et al. 2013: 246)

and *-ity* the more prestigious and learned alternative. It is therefore perhaps to be expected that the use of *-ity* is more restricted and that more sociolinguistic variation has been found in its productivity than in that of *-ness* (Säily 2014).

The competition between *-ness* and *-ity* heightened during the large-scale expansion of vocabulary in the Early Modern English period, when the English language began to be used more widely, including in registers previously dominated by French or Latin (Nevalainen 1999: 332). Recent studies by Rodríguez-Puente (2020) and Rodríguez-Puente et al. (2022) show that *-ity* gained ground on *-ness* in all registers, starting from written registers and spreading towards speech-related ones.

In personal letters of the seventeenth and eighteenth centuries, too, the productivity of *-ity* increased over time, whereas that of *-ness* remained stable (Säily 2014). Analyzing the *CEEC* from a sociolinguistic perspective, Säily and Suomela (2009) and Säily (2016, 2018) found that women lagged behind in the use of *-ity* in the seventeenth century but that this difference largely disappeared in the eighteenth century, with the exception of letters written to close friends. The convergence was attributed by Säily (2016) to the possible development of a shared style of letter-writing among the eighteenth-century literati overrepresented in the corpus. This style utilized nominal suffixes in a more personally involved and interactive way than before, reflecting the “feminization” of eighteenth-century culture (Säily 2018: 214; McIntosh 2008: 231). In the context of writing to close friends, friendship being less stable and more open to negotiation than either closer or more distant relationships (Wolfson 1990), men seem to have employed the learned suffix *-ity* more frequently than women to amuse and impress their correspondents (Säily 2018: 204–205).

2.2 Measuring variation and change in morphological productivity

Productivity was defined by Bolinger (1948: 18) as “the statistically determinable readiness with which an element enters into new combinations”. A number of quantitative measures have been proposed for assessing the productivity of morphological processes. The major measures include the number of different words containing the morpheme in a corpus, or *types*; the number of types occurring only once in the corpus, or *hapax legomena*;¹ and diachronically, the number

1. More specifically, Baayen (1993) proposes two hapax-based measures: the *category-conditioned degree of productivity*, or the ratio between the number of hapax legomena representing a given process and the total number of instances (tokens) of that process in the corpus, and the *hapax-conditioned degree of productivity*, or the ratio between the number of hapax legomena representing a given process and the total number of all hapax legomena in the corpus. These, along with type frequency, are argued to capture different aspects of productivity (see also Baayen 2009).

of types not occurring in previous periods in the corpus, or *new types* (Baayen 1993; Cowie & Dalton-Puffer 2002). While these type-based measures can be used to compare the productivity of different processes within the same corpus, their application becomes more complex when the aim is to analyze variation across (sub)corpora of unequal size, such as subcorpora representing different social groups or time periods. This is because type frequencies grow nonlinearly with corpus size, which means that normalizing them is unjustifiable and could in fact yield misleading results (Baayen 1992: 113; Säily & Suomela 2009: 96–97; Säily 2011: 127).

The problem of comparing type counts has been addressed in a few different ways in previous research. One solution has been to take a sample of each of the larger subcorpora to bring them down to equal size with the smallest subcorpus (Gaeta & Ricca 2006). However, this means that a great deal of the data is discarded, which is not ideal because the amount of data is typically very limited to start with. Furthermore, there is no obvious way of estimating the statistical significance of the differences observed between the samples. As an improvement, Berg (2021) suggests taking multiple samples of the same size and averaging the results, measuring the proportion of new types out of all types. While this approach utilizes more of the data, it is unusable in smaller corpora, where the number of new types (like that of hapax legomena; see Säily 2011) is likely to vary wildly and is hence not a reliable measure. Moreover, Berg's method of generating confidence intervals is problematic in that the intervals become narrower rather than wider when there is less data, making it difficult to estimate statistical significance. Another solution has been to use inter- or extrapolation to estimate how the number of types increases with corpus size in each subcorpus (Plag et al. 1999). As Plag et al. (1999: 217) acknowledge, however, this involves making assumptions about the data, such as “words appear randomly and independently in texts”.

Säily and Suomela (2009, 2017) present a partial solution to the problem. Rather than attempting to compare the subcorpora with each other, each subcorpus is compared with randomly composed subcorpora of the same size. These random subcorpora are sampled from the corpus as a whole and thus represent what is typical at the level of the entire corpus. To avoid making invalid assumptions about the data and to preserve discourse structure, the sampling is conducted at the level of texts or, in corpora designed for sociolinguistic research, informants. The sampling is done at each possible corpus size and can be plotted as type accumulation curves, with corpus size on the *x*-axis and the number of types on the *y*-axis; the actual subcorpora can then be plotted as points on these curves (Figure 1). In addition to making the most of the data, the method automatically provides a measure of statistical significance: if, for instance, only 0.1% of the random subcorpora have an equally low number of types as the actual

subcorpus, the number of types in the actual subcorpus is significantly low at $p < 0.001$ (Figure 1). Nevertheless, some issues remain. For instance, comparisons over time are difficult because the x -axis displays corpus size rather than time period. Furthermore, the method can only be used to analyze variation within a morpheme rather than between morphemes.

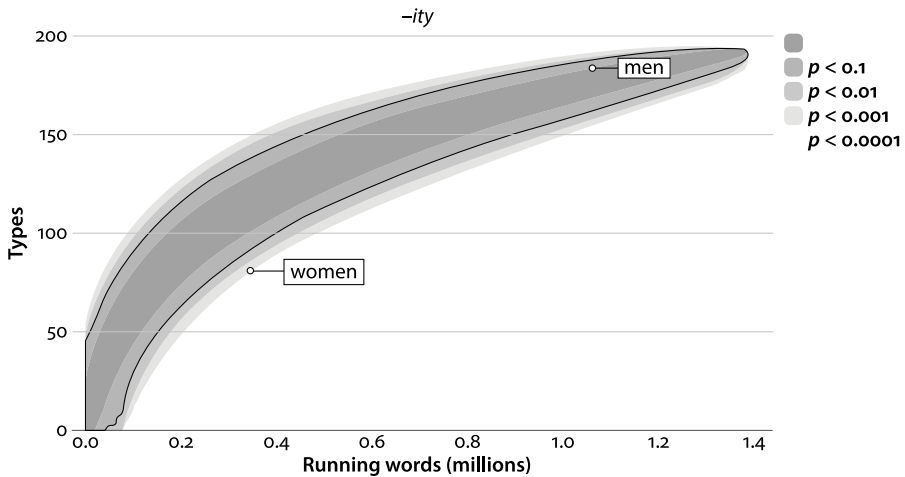


Figure 1. Type accumulation curves for *-ity* in the *CEEC*, 1600–1681. Women have significantly few *-ity* types. Based on Säily & Suomela (2009: Figure 5)

Rodríguez-Puente et al. (2022) introduce a method that puts time on the x -axis and enables the analysis of suffix competition. In brief, their idea is to treat competing suffixes as a linguistic variable and to measure the proportion of types of the incoming variant out of the variable as a whole over time (see Section 3.2 below for details). More generally, this method can be used to reliably estimate the proportion of any types of interest out of all relevant types; in this work, we develop it further to study both suffix competition and language-internal factors in the productivity of individual suffixes over time.

2.3 Productivity change in Construction Grammar

In this paper, we approach the development of *-ness* and *-ity* from the theoretical perspective of Construction Grammar (Hilpert 2019) and more specifically construction morphology (Booij 2010). Accordingly, we view the two suffixes as parts of partially schematic patterns with an open slot that speakers can use to create new lexical items, which are then connected to the schema and to each other in a network of constructions. Since morphological constructions of this kind are seen

as pairings of form and meaning, constructional research has been investigating their degrees of productivity in terms of their morpho-syntactic and semantic characteristics. Applied to diachrony, developments in productivity have been linked to changes that pertain to the form and meaning of morphological constructions, often in the form of hand-selected variables, but increasingly also to corpus-based measures that capture distributional characteristics of constructions.

Barðdal (2008: 27) offers a view on productivity that links the productivity of a construction to its type frequency and its coherence. The notion of coherence refers to the degrees of similarity that the instances of a construction exhibit with regard to their formal and functional characteristics. For example, the English ditransitive construction occurs with verbs such as *give*, *hand*, *send*, *offer*, or *bring*, which explicitly encode an intentional transfer of an object between someone who initiates the transfer and someone who receives the object. Examples of the ditransitive construction with these verbs are thus highly coherent. Type frequency and coherence are assumed to exhibit an inverse relation (Barðdal 2008: 35), so that with an increasing number of different types, the coherence of a construction will typically diminish. The relation between type frequency and coherence is taken up by Suttle and Goldberg (2011: 1254), who use it to develop the notion of “coverage”. The different types of a construction exhibit variability, so that some of the types are highly similar to one another, whereas others differ in one or more respects. The types of a construction can thus be viewed as being distributed across a variation space, in which some areas are densely filled, whereas other areas are only sparsely populated. Through a series of experiments, Suttle and Goldberg (2011: 1258) investigate the readiness with which participants accept new types of an existing schema, and their results indicate that higher coverage positively influences the acceptability of new types.

The variability found in the types of a construction can change over time. In a study that addresses the development of the derivational suffix *-ment*, Hilpert (2013) collects data from the *Oxford English Dictionary* (OED) and annotates all types that are recorded between 1250 and the 2000s in terms of several variables, which include the etymological source, the word class of the host element, the morphological branching structure, as well as meaning-related distinctions. A quantitative analysis of the types and their characteristics indicates that the diachronic productivity decrease of *-ment* is not a unified process but has to be understood as a development involving several subschemas that were productive at different moments in time.

Perek (2016) makes the case that diachronic changes in productivity can be fruitfully analyzed on the basis of distributional semantic methods, which capture degrees of similarity between the types of a construction on the basis of col-

locational data that is gathered from large corpora. He analyses the developing productivity of the English syntactic construction *V the hell out of NP*, exemplified by sentences such as *Snakes scare the hell out of him* or *It surprised the hell out of me*, focusing on the kinds of verbs that can appear as the main predicate, and how these verbs relate semantically to each other. Using distributional techniques, Perek is able to visualize the semantic space that is occupied by these verbs, and how the semantic coverage of that space develops over time. Corroborating the conclusions of Suttle and Goldberg (2011), he observes a rich-get-richer effect in which new verbs are more likely to appear in parts of the semantic space that are already densely populated. To illustrate, a tight semantic cluster that includes verbs such as *shock*, *scare*, *frighten*, *stun*, *surprise*, and other verbs with similar meanings is likely to attract new members. In a study that takes a similar approach, Hilpert (2018) investigates English noun-participle compounds such as *chocolate-covered* and *doctor-recommended* and notes diachronic changes in the productivity that is exhibited by different semantic clusters formed by the participles that appear in the construction. In line with the predictions of Barðdal (2008:35), certain clusters do not include many types, but are nonetheless relatively productive as a consequence of their internal coherence. For example, whereas forms such as *ginger-steamed* or *chipotle-spiced* contain participles with low type frequencies, their semantic coherence has the consequence that new noun-participle compounds with culinary meanings appear regularly (Hilpert 2018:103).

The present study adopts the constructional view of morphological schemas and changes in their productivity, but it tries to expand the perspective that has been taken in existing studies in order to bring the frameworks of Construction Grammar and Historical Sociolinguistics closer together (Hilpert 2017). Accordingly, our analysis includes some of the language-internal factors that were used in Hilpert's (2013) study of *-ment*, and it adds gender as a language-external variable.

3. Material and methods

3.1 Material: *Corpora of Early English Correspondence*

Personal letters are a promising source for historical sociolinguistics: they have identifiable writers and recipients representing all walks of life, and correspondence is an interactive, "speech-like" genre (Culpeper & Kytö 2010:17) that can be used to compensate for the lack of spoken data in the past. This was the rationale behind the compilation of the *CEEC* by a team led by Terttu Nevalainen and Helena Raumolin-Brunberg at the University of Helsinki from the early 1990s

onwards. The corpus family currently covers nearly four hundred years of the history of English, from the 1400s to 1800. It comprises about five million running words in 12,000 letters written by 1,200 informants. To reach a size large enough for quantitative investigation, the letters in the corpus were not painstakingly transcribed from original manuscripts scattered across archives and libraries, but rather sampled and digitized from previously published, original-spelling editions. The sampling was done per writer and 20-year period, and the aim was for each period to represent both men and women, all sections of the social hierarchy and a selection of regions of England. These and other social categories are included in the metadata associated with the corpus.²

Our focus is on the period 1600–1800 in the corpus for two reasons. Firstly, this time span covers a hundred years of both Early and Late Modern English, providing a window into the early modern expansion of vocabulary as well as the stylistically more elevated eighteenth century (Biber & Finegan 1997; McIntosh 1998: 23–24). Secondly, the amount of data in the corpus in this period (3.5 million words) is large enough to permit the analysis of relatively infrequent linguistic features like *-ness* and *-ity* over time and across some internal and external factors. We consider the external factor of gender because the factor has proved significant in previous research (Säily 2014) and because the data is relatively abundant: about a quarter of the running words are by women.

3.2 Methods

3.2.1 *Type-based measures*

One of the basic questions we will address in this study is how the productivity of *-ity* in comparison with the productivity of *-ness* has evolved over time. A key metric we will use is comparing how many *-ity* types (i.e., distinct words with the suffix *-ity*) there are in comparison with the total number of both *-ity* and *-ness* types. For example, in our data for the time period 1600–1679, there were in total 505 distinct *-ity* and *-ness* types, and out of those circa 38% were *-ity* types. Later, in the time period 1720–1800 we had in total 588 *-ity* and *-ness* types, and out of those circa 49% were *-ity* types. This suggests that the productivity of *-ity* is increasing in comparison with *-ness*, as the fraction of *-ity* types is increasing over time.

However, care is needed here for (at least) two reasons: First, there was more data in the later periods, with more *-ity* and *-ness* types overall; it is possible that

2. <<https://varieng.helsinki.fi/CoRD/corpora/CEEC/>>; see also Raumolin-Brunberg & Nevalainen (2007); Kaislaniemi (2018).

we accumulate new *-ity* and *-ness* types at different rates and hence perhaps the larger percentage for *-ity* in the later period is merely a consequence of that. Second, it might be the case that many of the new *-ity* types are, for instance, coming from letters written by a small number of individuals, and hence are not representative of the corpus as a whole.

To address both of these issues, we make use of the following three-stage approach (see Section 3.2.4 below for a comparison with prior work):

1. Build a model that represents the proportion of *-ity* types in the entire corpus
2. Compare each time period against this model
3. Visualize the results over all time periods

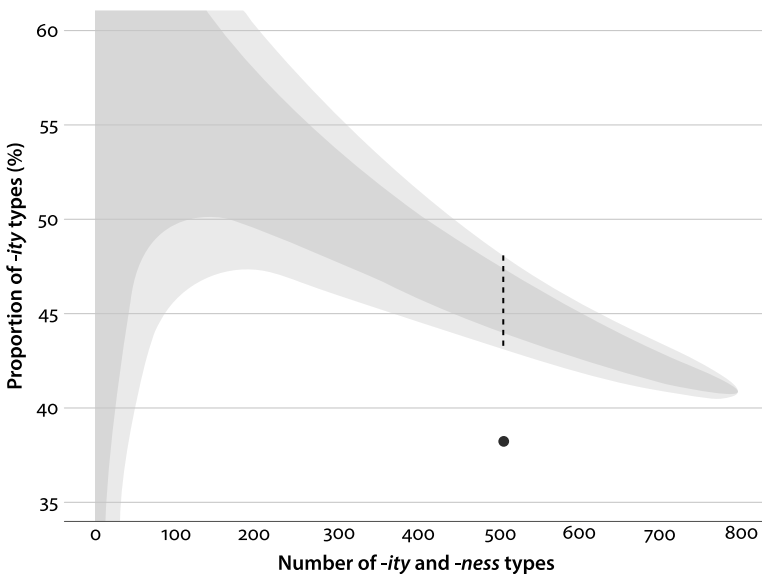


Figure 2. In the time period 1600–1679, there are in total 505 distinct *-ity* and *-ness* types, and only 193 of them, or circa 38%, are *-ity* types (shown with a dot). This is a significantly low number ($p < 0.0001$) in comparison with random subcorpora with the same total number of *-ity* and *-ness* types (dotted line across the shaded area)

In the first stage, we split the corpus into logical units, with each unit representing all texts written by one individual during a 20-year period. Then we construct 10 million random permutations (reorderings) of these units. Our computer program goes through each such reordering from start to end and keeps track of how the fraction of *-ity* types grows or shrinks as a function of the total number of *-ity* and *-ness* types. The results are shown with shading in Figures 2 and 3: The dark

area represents 80% of all reorderings, and the lighter area represents 95% of all reorderings.³

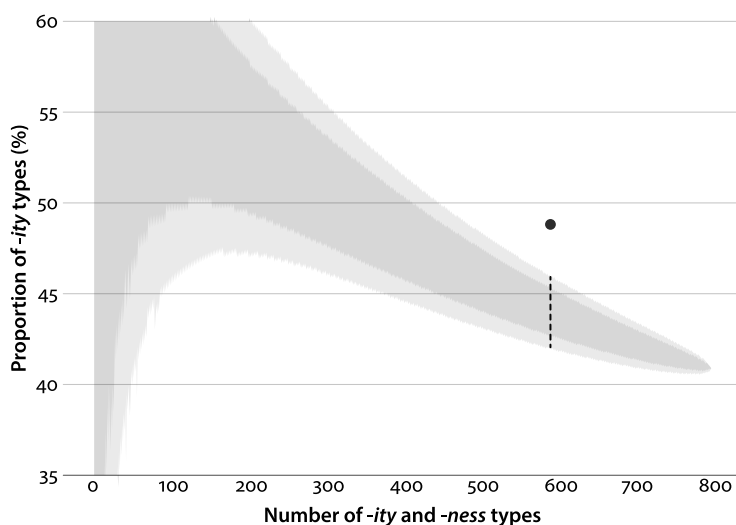


Figure 3. The proportion of *-ity* types is significantly high ($p < 0.0001$) in the time period 1720–1800 (dot) in comparison with random subcorpora with the same total number of *-ity* and *-ness* types (dotted line across the shaded area)

In the second stage, we compare each time period against the shaded areas. If the abundance of *-ity* does not depend on the time period, we would expect the dots marked in Figures 2 and 3 also to fall in the shaded areas. However, this is clearly not the case. For example, in Figure 2 we see that 38% is far below the typical proportions that we see when we look at a randomly reordered corpus with a similar number of *-ity* and *-ness* types (typical proportions would be closer to 43–48%). We can conclude that the proportion of *-ity* types is significantly low (in this case with $p < 0.0001$) in the early periods. Similarly, Figure 3 shows that the number is significantly high ($p < 0.0001$) in the later periods.

In the third stage, we summarize the findings over all time periods (see Figure 4). We have used here a sliding window of 80 years, with 20-year increments – the window is relatively long to ensure that there is a sufficient amount of data for each window position, but we use small increments so that we can more accurately pinpoint the time periods during which significant changes in language use occur. For each time period the dot shows the fraction of *-ity* types, while the

3. Note that to focus on the relevant parts of the figures, the y-axis sometimes does not start from zero.

shading shows where it would fall if the productivity of *-ity* was independent of the time period (the shading takes into account the number of *-ity* and *-ness* types that we have for each period and hence it is not constant). We see at a glance a clear increasing trend that is statistically significant. We will present and discuss our findings in more detail in Section 4.

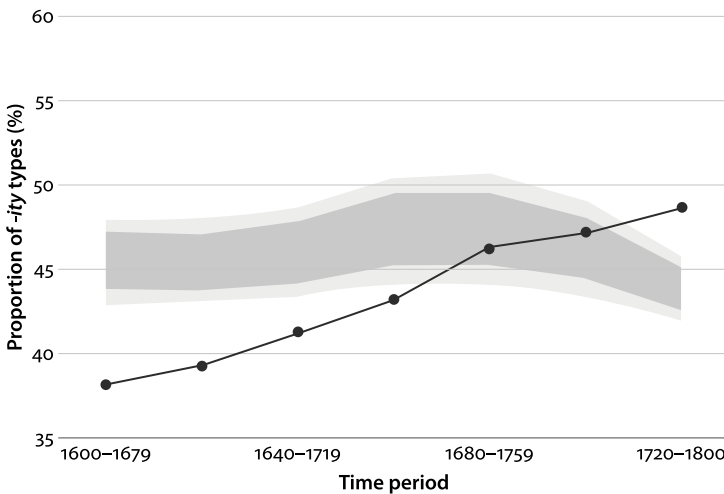


Figure 4. The line indicates the proportion of *-ity* types in each time period in our corpus. The shaded area indicates the expected range, under the null hypothesis that the use of *-ity* does not depend on the time period. The first dot corresponds to Figure 2, and the final dot corresponds to Figure 3

3.2.2 External and internal factors

While in Figure 4 we have only looked at *-ity* and *-ness* as a whole over different time periods, in this work we will also zoom in on more fine-grained questions related to both social and internal factors. First, we will consider the impact of the gender of the writer. To do that, we split the corpus into two parts: letters written by men and letters written by women. By constructing random reorderings for different subcorpora we can make two kinds of comparisons: are women in the early periods using *-ity* as diversely as all writers in the same time period (see Figure 7 in Section 4.1 for an example), and are they using *-ity* as diversely as women across all time periods (see Figure 8 in Section 4.1 for an example).

Inspired by the study of *-ment* by Hilpert (2013), we will consider the type-based internal factors described in Table 1.

Table 1. Type-based internal factors

	Alternatives	Examples
Etymological source	borrowed	<i>ability</i> (borrowed from French/Latin)
	derived	<i>oddity</i> (derived within English)
	inherited	<i>hardness</i> (inherited from Germanic)
Word class of the base	adjective	<i>ability</i> : <i>able</i>
	noun	<i>authorshipness</i> : <i>authorship</i>
	...	
Branching structure	binary	<i>[odd-ity]</i>
	left	<i>[[un-couth]-ness]</i>
	right	<i>[non-[conform-ity]]</i>

The classification of etymological source and the word class of the base relies on the OED, whereas that of branching structure is based on Hilpert (2013: 138–141). To analyze the internal factors, we simplify the classification into just two classes, for example, *-ity* types with adjectival vs. other bases. Then we can calculate and visualize the fraction of types with adjectival bases over time, and also assess the significance of possible trends that we observe. Note that the proportion of derived types indicates productivity of the suffix within English, whereas the proportion of right-branching types indicates productivity of the prefix (e.g., *non-* in *nonconformity*). The word class of the base is less clear-cut in terms of indicating productivity.

The majority of the letters in the *CEEC* were written by men, and hence when we zoom in on the subcorpora that represent men vs. women, we will have subcorpora of very different sizes. Visualizations similar to Figure 4 can be misleading in those cases, as the proportion of *-ity* types also depends on corpus size (recall Figure 2, which demonstrates that the proportion of *-ity* tends to decrease as the corpus size increases). To overcome this issue, for the purposes of visualization, we will show trend lines that represent the average values for a randomly created subcorpus with a fixed total number of types (for example, a subcorpus with exactly 50 *-ity* types); see Figures 5 and 6 in Section 4.1 for examples. We emphasize that this is primarily a visualization aid – to assess the statistical significance of each finding, we can use the method outlined in Section 3.2.1 directly, without such downsampling.

3.2.3 *Token-based measures*

While the branching structure is (usually) the property of a particular type, and it makes sense to study the diversity of right-branching types among all types,

we are also interested in internal factors that are related to individual tokens (Table 2).⁴

Table 2. Token-based internal factors

	Alternatives	Examples
Semantic type	state	<i>punctuality</i>
	thing	<i>electricity</i>
	office	<i>officiality</i>
	collectivity	<i>laity</i>
Occurrence in possessive constructions	<i>poss_pron Xness/ity</i>	<i>your kindness</i>
	<i>Xness/ity of Y</i>	<i>the wetness of the weather</i>
	none	

The classification of semantic types is based on Romaine (1985: 455–456), according to whom they represent the semantic route for *-ity* and *-ness* formations from more abstract to more concrete. Romaine argues that ‘state’ (comprising states, conditions, and qualities) is the original sense of most formations, which implies that the proportion of the ‘state’ sense indicates productivity of the suffix. The same word can belong to different semantic types depending on the context; for instance, *curiosity* can be classified as either ‘state’ or ‘thing’, and *highness* as either ‘state’ or ‘office’ (comprising offices, titles, persons, and occupations).

The possessive constructions that *-ness* and *-ity* often occur in were noted by Säily (2014: 253), who connects the use of possessive pronouns with a more personally involved style. To study the possessive constructions, we have annotated each occurrence of *-ity* and *-ness* by inspecting the surrounding context. However, determining the semantic type of a word requires slow manual work, and here we have only annotated a small sample of *-ity* instances (we focus on *-ity* because the vast majority of the *-ness* instances seem to represent the ‘state’ sense, so we do not expect to find much variation or change there). This sample of 188 tokens comprises instances of *-ity* types that were derived within English and that have adjectival bases in the sense that at least one of the word classes provided by the OED for the base is ‘adjective’.

Once we have the individual tokens annotated, we can easily calculate the proportion of *-ity* tokens that occur with a possessive pronoun out of all *-ity*

4. As *-ness* and *-ity* do not occur in the *s*-genitive construction frequently enough for analysis in this corpus, the handful of instances of the *s*-genitive have been classified in the ‘none’ category.

tokens. However, we also need to determine whether any observed trends are statistically significant. To do this, we use the same basic idea as what was described in Section 3.2.1, but we keep track of the number of tokens instead of the number of types. To implement this with the help of the computer program developed by Rodríguez-Puente et al. (2022), we simply make each token unique by adding a distinct counter.

One challenge with token-based measures is the impact of the common, lexicalized types *business*, *highness*, *holiness*, *society*, and *university*. These were discovered by inspecting a list of the most frequent types and their tokens in context. To ensure that the findings are not related to the use of such words, we also repeated our experiments with these five words excluded, which however did not have a major impact on the results.

3.2.4 *Prior work and key new ideas*

For type-based measures, Rodríguez-Puente et al. (2022) used a permutation testing method similar to what we described in Section 3.2.1 to compare the productivity of *-ity* across registers. They also developed a computer program that we used as a starting point in this work. However, in Rodríguez-Puente et al. (2022), the key idea was to compare, for example, speech-related registers in one time period with all registers in the *same* time period. In this work we also take an orthogonal approach: we compare one time period with the entire corpus across *all* time periods. This results in new visualizations similar to those of Figure 4 that enable us to directly assess the statistical significance of changes over time. Rodríguez-Puente et al. (2022) only considered differences between registers; while, for example, Säily and Suomela (2009, 2017) used methods based on permutation testing to study differences between genders, we are not aware of prior work that has used similar techniques to study language-internal factors in this context. The use of the tool by Rodríguez-Puente et al. (2022) in the study of token-based measures is also new.

To speed up the computationally intensive parts of this work, we extended the tool by Rodríguez-Puente et al. (2022) so that it can be used in a massively parallel fashion in a typical supercomputing environment (together with the Slurm Workload Manager).⁵ However, the tool can also be used with a normal desktop or laptop computer.

5. The supercomputing extension is now integrated into the freely available TypeRatio tool (Suomela 2023), with sample scripts available at <<https://github.com/suomela/type-ratio-csc>>.

4. Results

4.1 Suffix competition

As already seen in Figure 4, the proportion of *-ity* types out of all *-ity* and *-ness* types significantly increases over time. In other words, *-ity* gains ground on *-ness* in the CEEC during the seventeenth and eighteenth centuries. Figure 5 shows this phenomenon as trend lines in randomly built subcorpora of the same, fixed size. We investigate the trends at three corpus sizes (a total of 50/100/150 *-ity* and *-ness* types); all of them display the same trend, which supports the hypothesis that the increase is a real phenomenon within the corpus. However, the exact proportions of *-ity* types are slightly different, with the smallest subcorpora exhibiting the largest proportions. This is consistent with our observation in Section 3.2 that the proportion of *-ity* types tends to decrease as the corpus size increases (which suggests that the productivity of *-ity* is lower than that of *-ness*).

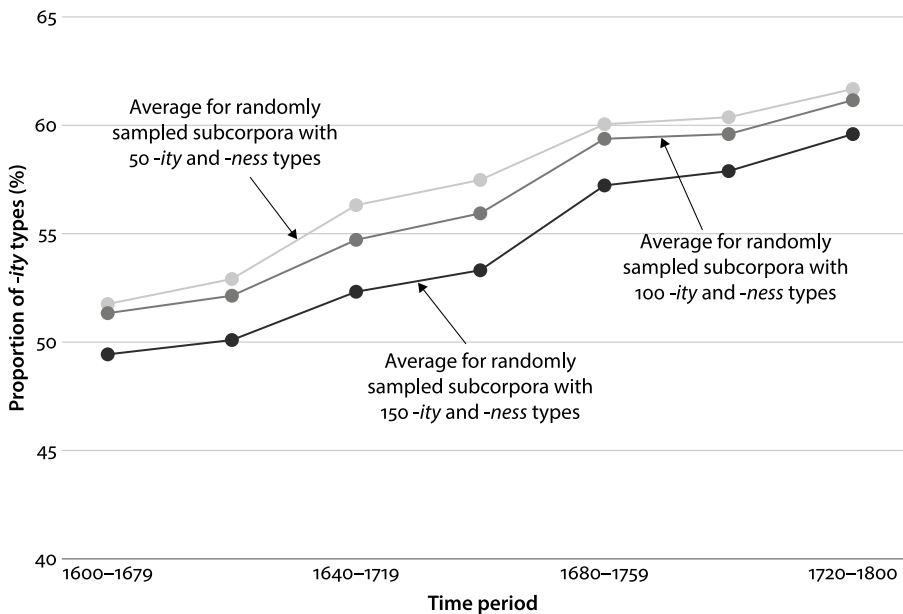


Figure 5. The proportion of *-ity* types out of all *-ity* and *-ness* types over time. Trend lines (from lightest to darkest): average proportion in randomly sampled subcorpora with a total of 50/100/150 *-ity* and *-ness* types. See Figure 4 for the statistical significance of the change

Adding the social factor of gender, Figure 6 indicates a steady growth in the proportion of *-ity* types for men, whereas women seem to lag behind for the first two periods (covering the seventeenth century). After that, they quickly catch up (by the period 1660–1739). Figure 7 shows that the gender difference in the first two periods is statistically significant. Focusing on the significance of change over time, Figure 8 shows that the increase over time within the female subcorpus is statistically significant; the same holds for the male subcorpus (not shown).

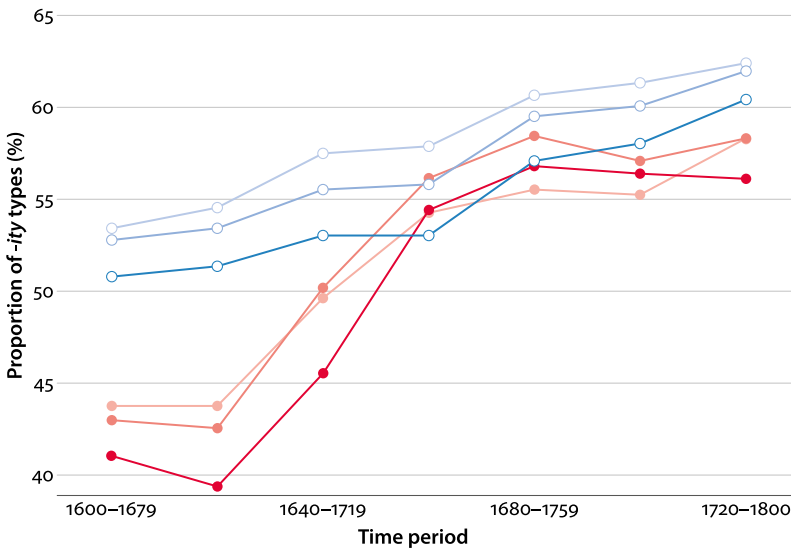


Figure 6. Change in the proportion of *-ity* over time by gender (*blue circles* = men, *red dots* = women). Trend lines (from lightest to darkest): average proportion in randomly sampled subcorpora with a total of 50/100/150 *-ity* and *-ness* types

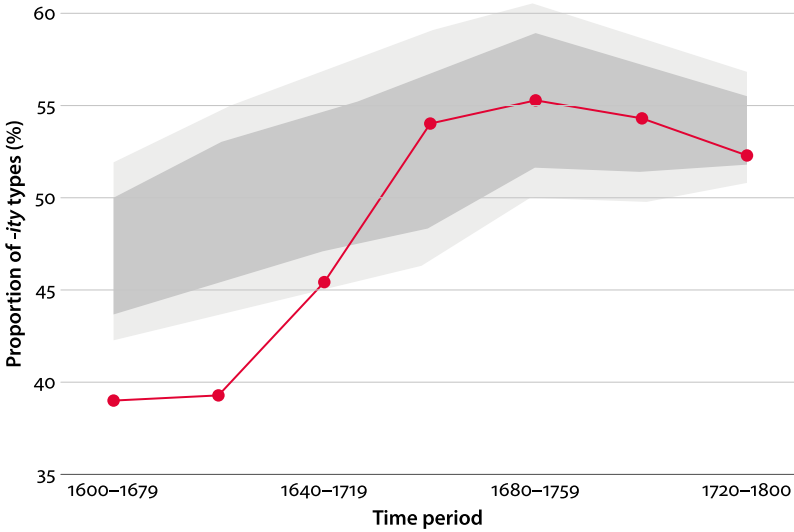


Figure 7. Comparison of women in each period (red line) with all writers in each period (shaded area). Women’s proportion of *-ity* is statistically significantly low in the first two periods ($p < 0.0008$ in 1600–1679, $p < 0.0007$ in 1620–1699)

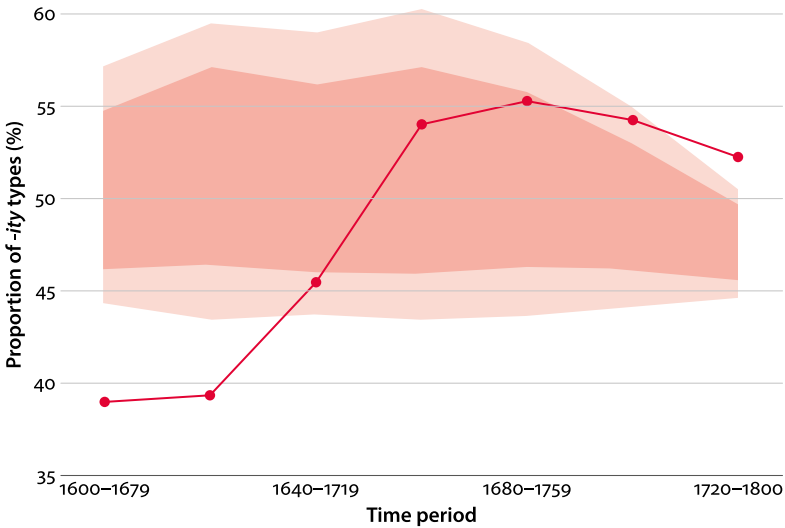


Figure 8. Comparison of women in each period with women in all periods. The proportion of *-ity* is statistically significantly low in the first two periods ($p < 0.0002$ in 1600–1679, $p < 0.0011$ in 1620–1699) and significantly high in the last period ($p < 0.0011$ in 1720–1800)

4.2 Internal factors: Type-based measures

Zooming in on internal factors within each suffix, we first consider etymological source. Virtually all of the *-ness* types are either derived or (in a few cases) inherited, and this does not vary or change over time. When we investigate the proportion of *-ity* types derived within English out of all *-ity* types, however, there is a significant change over time, which indicates increasing productivity of the suffix (Figure 9).

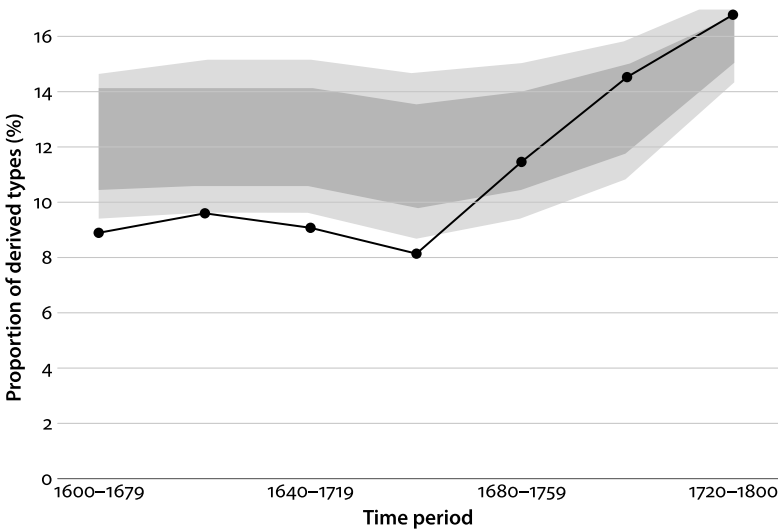


Figure 9. Proportion of *-ity* types derived within English out of all *-ity* types. Comparison of writers in each period with writers in all periods. The proportion is significantly low in the first four periods ($p < 0.020$, $p < 0.031$, $p < 0.012$, $p < 0.011$, respectively)

Figure 10 shows the increasing trend by gender. Again, women lag behind in the seventeenth century – in fact, they exclusively use borrowed types like *opportunity* and *necessity* – but then they quickly catch up with men, and the proportion of derived *-ity* types only really starts to grow when women join men in using them. The gender difference is statistically significant in the first period (Figure 11).

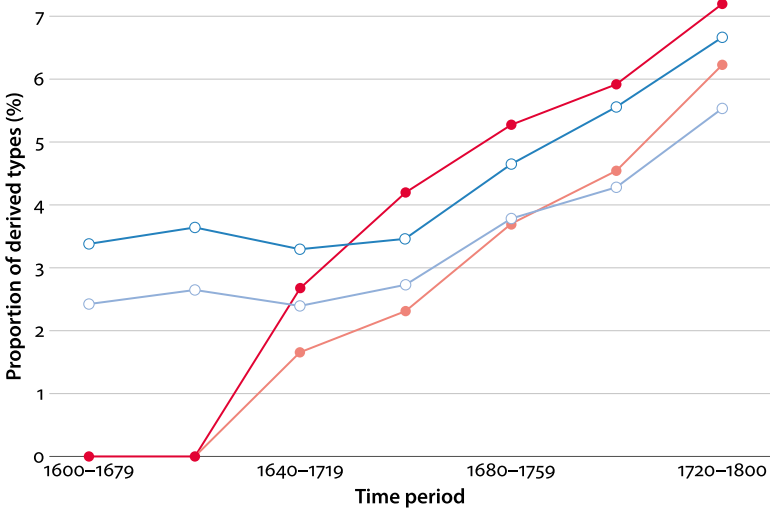


Figure 10. Change in the proportion of *-ity* types derived within English over time by gender (blue circles = men, red dots = women). Trend lines (from lighter to darker): average proportion in randomly sampled subcorpora with a total of 25/50 *-ity* types

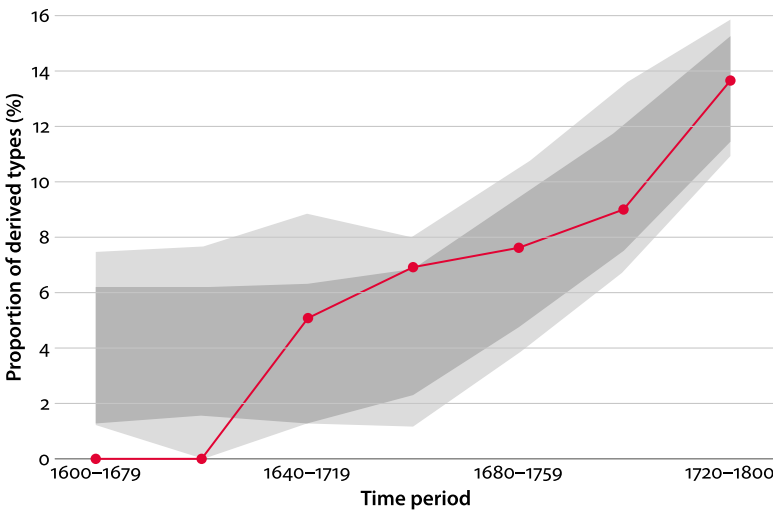


Figure 11. Comparison of women in each period with all writers in each period. Women's proportion of *-ity* types derived within English is significantly low in the first period ($p < 0.0007$)

In terms of the word class of the base, there are no clear trends over time or by gender for either suffix. For *-ity*, however, a weak but consistent increase is observable in the proportion of adjectival bases (Figure 12). Moreover, a qualitative examination of the *-ity* types with non-adjectival bases shows that with the exception of *mobility* ‘the mob’, they tend to be either borrowed or right-branching, indicating that the productive use of *-ity* is concentrated on adjectival bases.

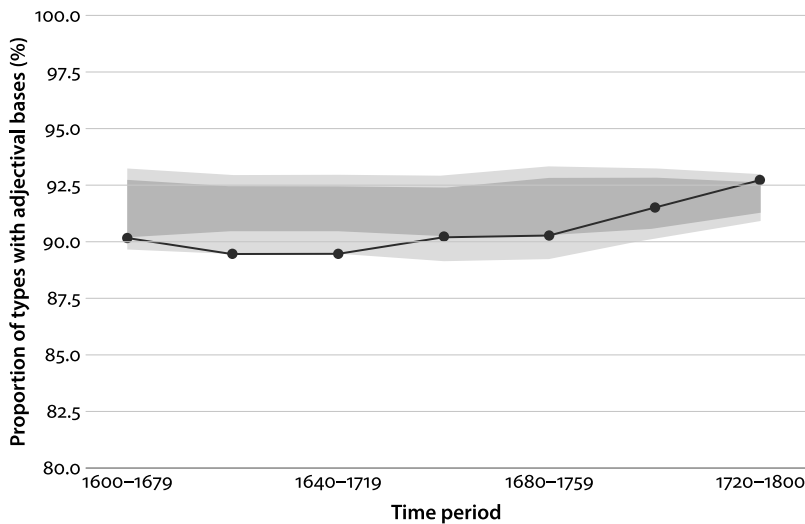


Figure 12. Proportion of *-ity* types with adjectival bases out of all *-ity* types. Comparison of writers in each period with writers in all periods. The proportions in the second and third periods are somewhat low ($p < 0.036$)

Moving on to branching structure, the proportion of right-branching *-ity* types out of all *-ity* types seems to decrease over time somewhat consistently and significantly (Figure 13). This again indicates increasing productivity of *-ity*. For *-ness*, there seems to be an opposite tendency, but it is completely non-significant (Figure 14). There are no statistically confirmable trends by gender for either suffix.

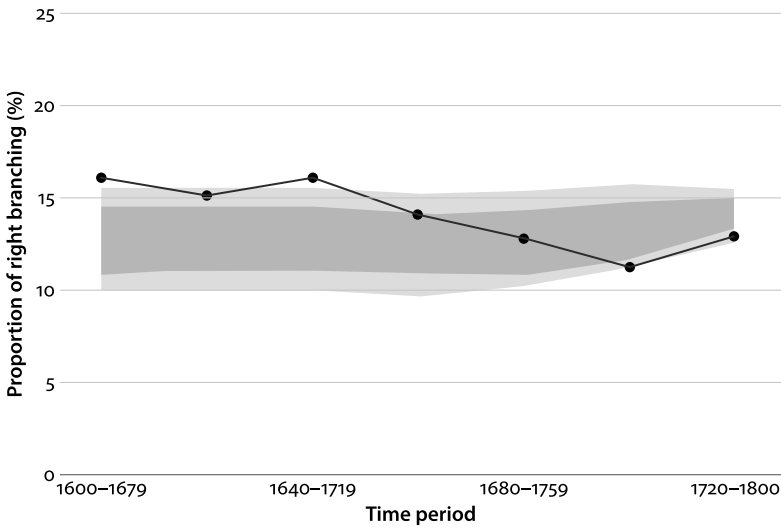


Figure 13. Proportion of right-branching *-ity* types out of all *-ity* types. Comparison of writers in each period with writers in all periods. The proportion is significantly high in the first and third periods ($p < 0.011$ and $p < 0.012$, respectively)

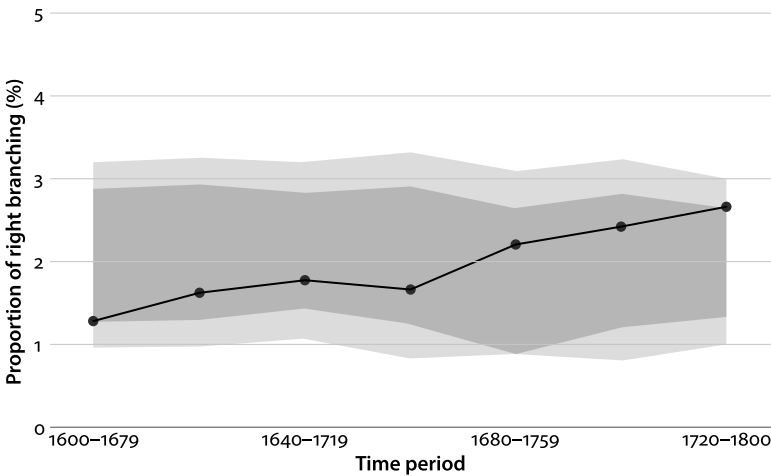


Figure 14. Proportion of right-branching *-ness* types out of all *-ness* types. Comparison of writers in each period with writers in all periods

4.3 Internal factors: Token-based measures

In our small sample of *-ity* tokens annotated for semantic type, we are not able to observe any gender variation. However, when we analyze change in the proportion of the ‘state’ sense in the sample as a whole, there appears to be an increasing trend over time (Figure 15). The peak around 1640–1679 is statistically insignificant ($p > 0.09$) and could be random variation owing to lack of data. Note that we use 40-year periods here to show that the initial time periods in the sample are indeed on the low side – an observation that would be lost if we used 80-year periods.

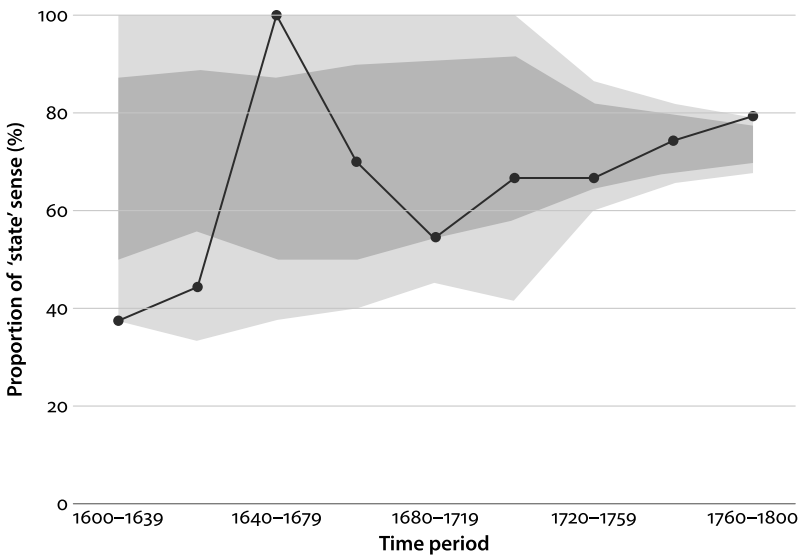


Figure 15. Proportion of *-ity* tokens denoting ‘state’ out of all *-ity* tokens in the sample. Comparison of writers in each period with writers in all periods. The proportion is somewhat high in the last period ($p < 0.047$)

Looking at variation in possessive constructions, several trends emerge. The most significant of these is the increase over time in the proportion of the *Xity of Y* construction (e.g., *the punctuality of my answers*; Figure 16).

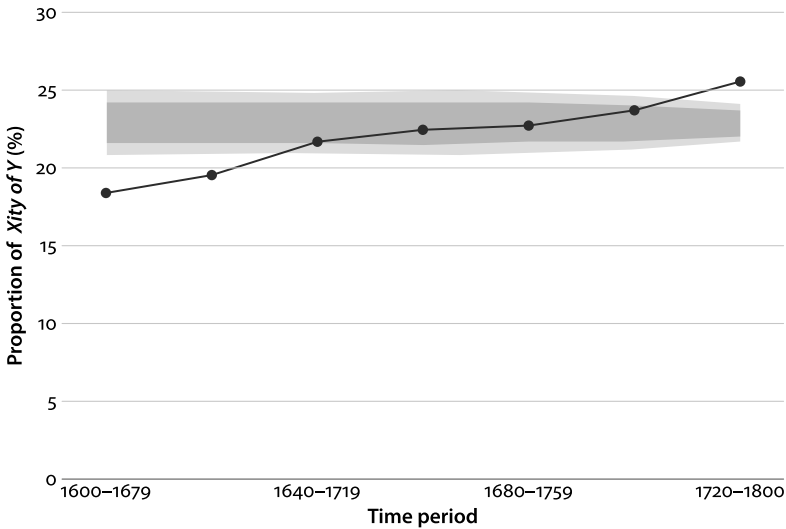


Figure 16. Proportion of *-ity* tokens occurring in the construction *Xity of Y* out of all *-ity* tokens (lexicalized types *society* and *university* excluded). Comparison of writers in each period with writers in all periods. The proportion is significantly low in the first two periods ($p < 0.0001$, $p < 0.0005$) and significantly high in the last period ($p < 0.0001$)

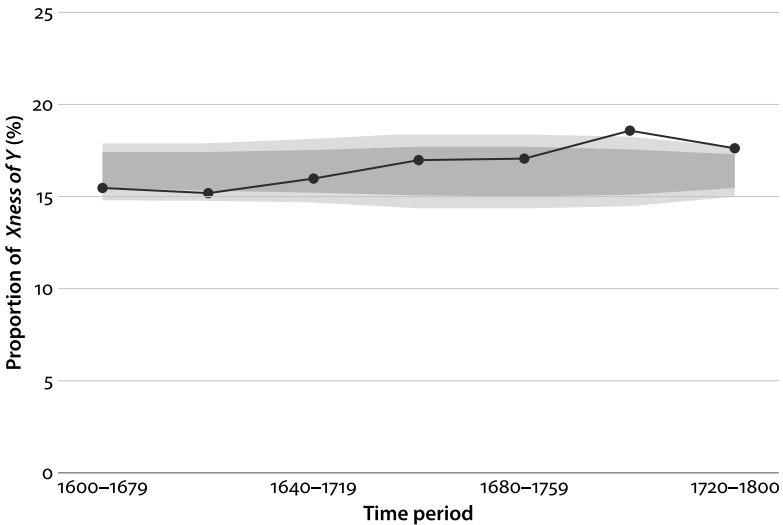


Figure 17. Proportion of *-ness* tokens occurring in the construction *Xness of Y* out of all *-ness* tokens (lexicalized types *business*, *highness* and *holiness* excluded). Comparison of writers in each period with writers in all periods. The proportion is significantly high in the last two periods ($p < 0.012$ and $p < 0.039$)

There is a similar but less significant increase in the proportion of the *Xness of Y* construction (e.g., *the wetness of the weather*; Figure 17). It may be that the frequency of the *of*-genitive in general is increasing over time, but this is not manifested quite equally in *-ness* and *-ity*, and it is not happening at the expense of the *s*-genitive in these contexts, as there are only a handful of instances like *my mother's unwillingness* in the corpus.

Interestingly, the proportion of *Xness of Y* tends to be lower with women (Figure 18), and it is significantly low in the periods 1600–1679 and 1640–1719 ($p < 0.021$ and $p < 0.023$). The gender difference is less consistent for *Xity of Y* and does not reach significance in any period ($p > 0.05$), but the tendency in most periods is in the same direction (Figure 19). The change therefore appears to be led by men.

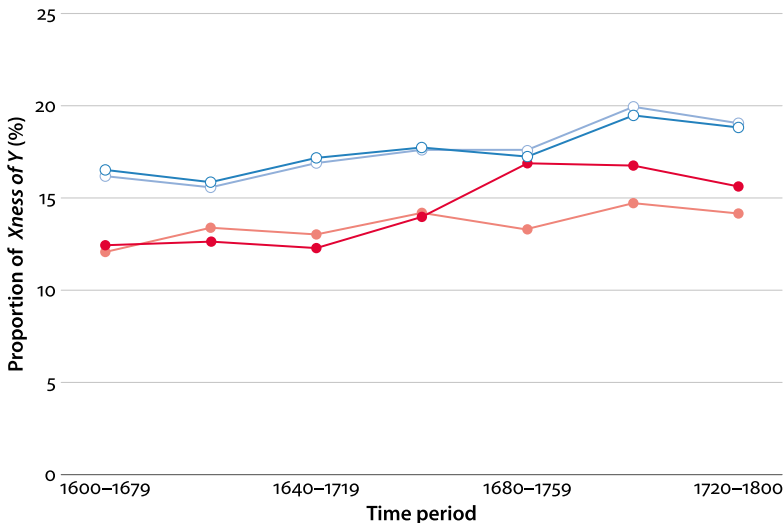


Figure 18. Change in the proportion of *-ness* tokens occurring in the construction *Xness of Y* over time by gender (*blue circles* = men, *red dots* = women; lexicalized types *business*, *highness* and *holiness* excluded). Trend lines (from lighter to darker): average proportion in randomly sampled subcorpora with a total of 100/200 *-ness* tokens

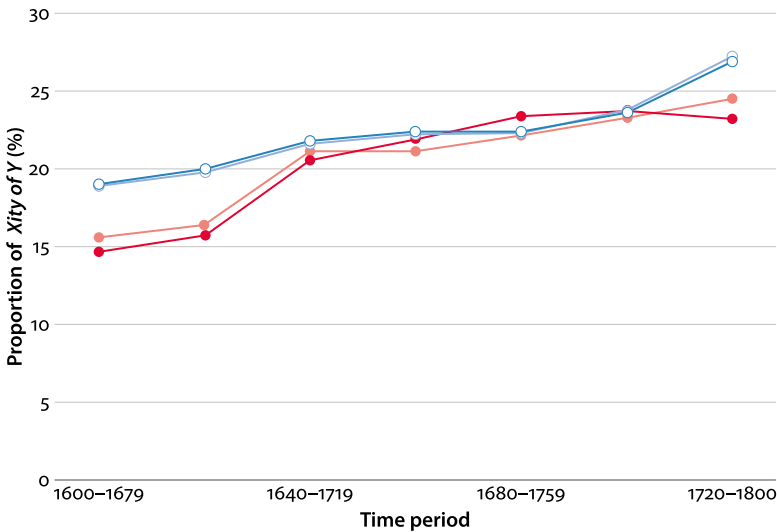


Figure 19. Change in the proportion of *-ity* tokens occurring in the construction *Xity of Y* over time by gender (blue circles = men, red dots = women; lexicalized types *society* and *university* excluded). Trend lines (from lighter to darker): average proportion in randomly sampled subcorpora with a total of 100/200 *-ity* tokens

The proportion of *-ity* tokens modified by a possessive pronoun, by contrast, seems to rather decrease than increase over time, and the change is mildly significant in the first and last periods (Figure 20). For *poss_pron Xness*, there is no clear pattern over time. Again, however, we observe a gender difference in the proportion of both constructions, which appear to be consistently favored by women (Figures 21–22). The differences are somewhat significant in three periods for *-ity* ($p < 0.046$ in 1660–1739, $p < 0.042$ in 1700–1779, $p < 0.027$ in 1720–1800) and in the last two periods for *-ness* ($p < 0.043$, $p < 0.025$); it seems that the gender difference increases over time for *-ness*.

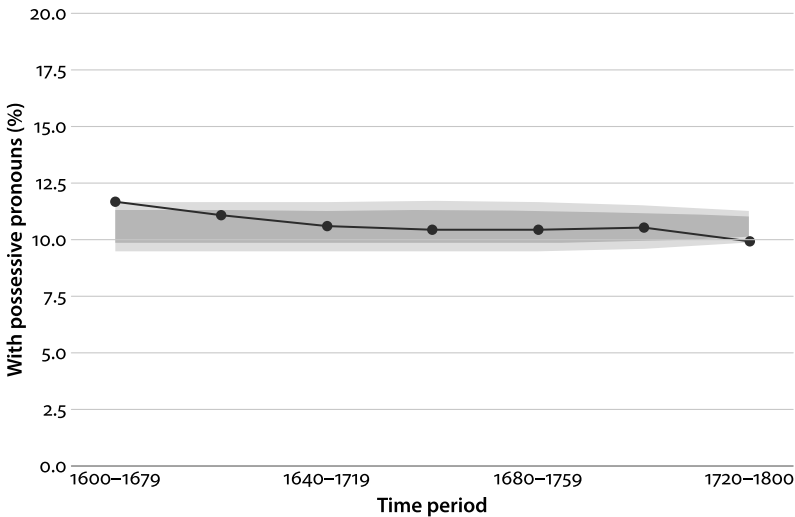


Figure 20. Proportion of *-ity* tokens occurring in the construction *poss_pron Xity* out of all *-ity* tokens (lexicalized types *society* and *university* excluded). Comparison of writers in each period with writers in all periods. The proportion is somewhat high in the first period ($p < 0.030$) and low in the last period ($p < 0.039$)

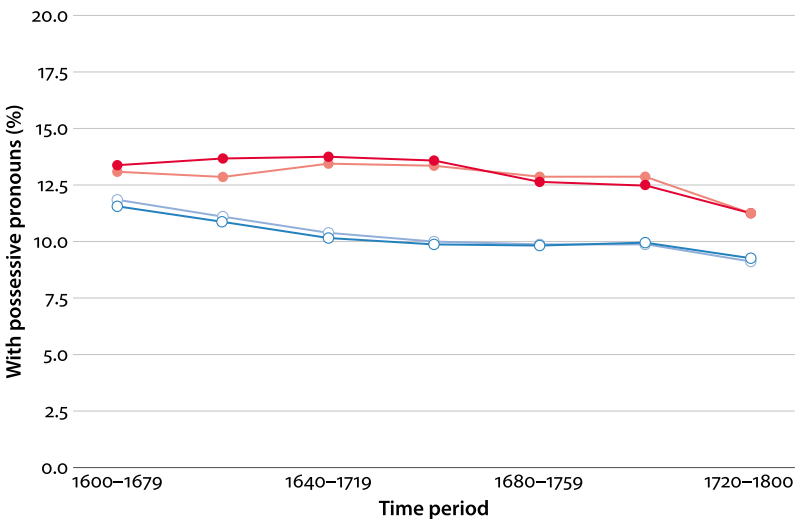


Figure 21. Change in the proportion of *-ity* tokens occurring in the construction *poss_pron Xity* over time by gender (*blue circles* = men, *red dots* = women; lexicalized types *society* and *university* excluded). Trend lines (from lighter to darker): average proportion in randomly sampled subcorpora with a total of 100/200 *-ity* tokens

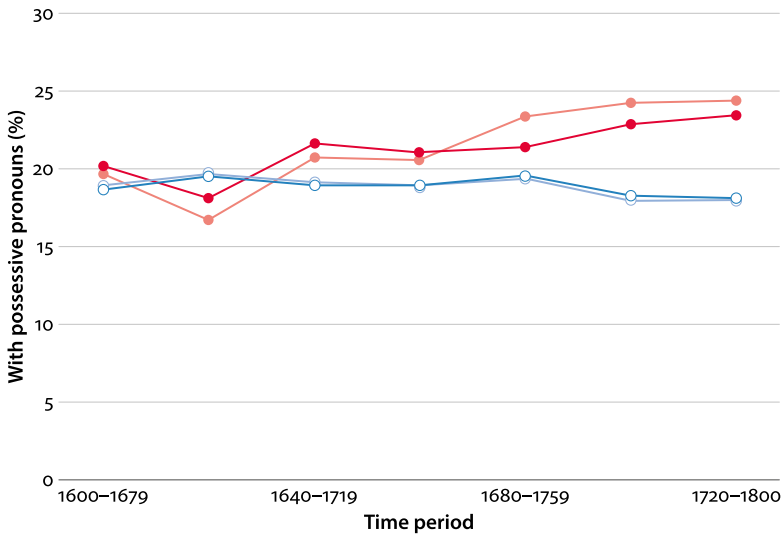


Figure 22. Change in the proportion of *-ness* tokens occurring in the construction *poss_pron Xness* over time by gender (blue circles = men, red dots = women; lexicalized types *business*, *highness* and *holiness* excluded). Trend lines (from lighter to darker): average proportion in randomly sampled subcorpora with a total of 100/200 *-ness* tokens

5. Discussion and conclusion

Our results support and refine earlier findings on *-ness* and *-ity*, but they also provide completely new information on the language-internal factors involved in the development of the two suffixes, and they shed new light on the interaction of the internal factors with the external factor of gender.

Säily (2014) found a male-led increase in the productivity of *-ity* in the *CEEC*; we have shown that the productivity of *-ity* also increases when we examine it in relation to *-ness*. In other words, *-ity* is gaining ground on *-ness* in seventeenth- and eighteenth-century correspondence, as was also the case in the study by Rodríguez-Puente et al. (2022), who analyzed different corpora and registers of Early and Late Modern English. Furthermore, we have provided more information on the timing of the gendered change: women are lagging behind in the seventeenth century, but then quickly catch up with men. It makes sense that men would be the leaders of this change, as Rodríguez-Puente (2020) and Rodríguez-Puente et al. (2022) showed that it spread from more formal written registers, to which women would have had little access, to more speech-related ones. This is of course also related to the fact that *-ity* is a borrowed, learned and prestigious suffix that would at least initially have been easier to use for men with a classi-

cal education and would have better matched their style of letter-writing, which tended to be more informational than that of women (Säily et al. 2017). As noted in Section 2.1 above, the convergence between genders in the eighteenth century has been explained through the development of a shared, polite style among the middle and upper classes, which could be linked to changing conceptions of gender and advances in women's education (Säily 2016, 2018).

In terms of internal factors, we have found that as the proportion of *-ity* types increases over time, so does the share of *-ity* types originally derived within English and those with adjectival bases; conversely, there is a decrease in the proportion of right-branching types formed by adding a prefix to an existing *-ity* word. The increase in derived types is again led by men, as women exclusively use borrowed types in the seventeenth century; however, the men's usage remains rather stable in the seventeenth century, and the proportion only starts to rapidly increase in the eighteenth century, when women join men in using derived types. As for token-based factors, there is an increase in the share of *-ity* tokens denoting 'state' and in those occurring in the possessive construction *Xity of Y*, whereas a decreasing trend is evident in the proportion of tokens occurring in the *poss_pron Xity* construction. The increase in the share of the more complex *Xity of Y* construction is led by men, as is the decrease in the more personally involved *poss_pron Xity* construction.

Regarding these factors as constructional subschemas, we could say that in earlier periods, compared to later periods, there is a higher proportion of *-ity* types that are used by men, are borrowed, have bases representing various parts of speech, and are right-branching, as well as a higher proportion of *-ity* tokens that express various meanings and occur in the *poss_pron Xity* construction. In later periods, compared to earlier periods, there is a higher proportion of types that are used by both men and women, are derived, have adjectival bases, and are binary or left-branching, as well as a higher proportion of tokens that denote 'state' and occur in the *Xity of Y* construction. We could also formulate different subschemas for men and women at different time periods to capture, for example, the phenomena that women's proportion of borrowed types is higher than men's in the seventeenth century and that the proportion of tokens in the *poss_pron Xity* construction is consistently higher with women, although the direction of change is the same for both genders in these contexts and the changes seem to be led by men.

The increase in the proportion of tokens denoting 'state' as well as that of types with adjectival bases is interesting because it means that *-ity* is being used less diversely than before in terms of meaning and the part of speech of the base. This could be related to the notions of coherence and coverage: the borrowed *-ity* types used in the earlier periods were less coherent or similar with each other, and

in the process of becoming more productive, the use of *-ity* focused on the variables with the most coverage, that is, ‘state’ and adjectival bases. (In 1600–1639 the ‘collectivity’ sense is in fact more frequent than ‘state’, but the numbers are very low and the ‘collectivity’ tokens represent just a single type, *laity*, whereas each of the ‘state’ tokens represents a different type.)

Also of interest is the finding that even though the overall gender difference in the use of *-ity* disappears, it remains in possessive constructions, and this difference is also evident in the use of the otherwise quite stable *-ness*. This reinforces the importance of a combined analysis of internal and external factors. Women consistently use both *-ity* and *-ness* with a possessive pronoun (as in *your kindness*) more often than men, supporting the hypothesis that women’s style of writing tends to be more personally involved (Biber & Burges 2000; Palander-Collin 1999, 2000; Säily et al. 2017). On the other hand, the hypothesis of a shared style in the eighteenth century does not seem to hold at this level of granularity. This could be examined more closely in future research.

In this paper we have focused on proportions of types of interest out of all relevant types, essentially considering the “realized productivity” or “extent of use” of the suffixes rather than their potential or expanding productivity (Baayen 2009). Future research on larger corpora could bring in other facets of productivity by considering proportions of hapaxes and new types; with a larger corpus it would also be feasible to use more robust type-based measures in addition to token-based measures to study the semantic types and possessive constructions. The amount of data permitting, it would be of interest to study other external factors besides gender as well as the more detailed social roles of men and women in the past, as for example Vartiainen et al. (2013) have shown that gender differences tend to decrease in contexts like spousal letters and that there is intra-gender variation in language use depending on the specific social role (e.g., men writing as husbands, fathers, and sons).

Supporting Rodríguez-Puente et al. (2022), this paper has demonstrated the importance of robust statistics for not only type frequencies but also proportions of types. We have developed the methods of Rodríguez-Puente et al. (2022) further to be able to better consider the significance of change over time as well as token-based internal factors. The methodological development was done in collaboration between linguists and computer scientists, and it also resulted in new software tools that enable us to explore linguistic phenomena with the help of modern large-scale supercomputers (Suomela 2023). Drawing on other disciplines to enrich the methods of corpus linguistics has proved to be a mutually beneficial strategy and is in line with practices in the digital humanities.

Funding

This work was supported in part by the Research Council of Finland, grant 323390.

Acknowledgements

The authors wish to acknowledge CSC – IT Center for Science, Finland, for computational resources. We would also like to thank the anonymous reviewers and the editors of this volume for helpful feedback, and Lassi Saario for coding the data for the type-based internal factors.

References

- Adamson, Sylvia. 1989. With double tongue: Diglossia, stylistics and the teaching of English. In *Reading, Analysing and Teaching Literature*, Mick Short (ed.), 204–240. London: Longman.
-  Baayen, R. Harald. 1992. Quantitative aspects of morphological productivity. In *Yearbook of Morphology 1991*, Geert Booij & Jaap van Marle (eds), 109–149. Dordrecht: Kluwer.
-  Baayen, R. Harald. 1993. On frequency, transparency and productivity. In *Yearbook of Morphology 1992*, Geert Booij & Jaap van Marle (eds), 181–208. Dordrecht: Kluwer.
-  Baayen, R. Harald. 2009. Corpus linguistics in morphology: Morphological productivity. In *Corpus Linguistics: An International Handbook*, Anke Lüdeling & Merja Kytö (eds), 899–919. Berlin: Mouton de Gruyter.
-  Barðdal, Jóhanna. 2008. *Productivity: Evidence from Case and Argument Structure in Icelandic* [Constructional Approaches to Language 8]. Amsterdam: John Benjamins.
-  Bauer, Laurie, Lieber, Rochelle & Plag, Ingo. 2013. *The Oxford Reference Guide to English Morphology*. Oxford: OUP.
-  Berg, Kristian. 2021. Productivity, vocabulary size, and new words. A response to Säily (2016). *Corpus Linguistics and Linguistic Theory* 17(1): 177–187.
-  Biber, Douglas & Burges, Jená. 2000. Historical change in the language use of women and men: Gender differences in dramatic dialogue. *Journal of English Linguistics* 28(1): 21–37.
- Biber, Douglas & Finegan, Edward. 1997. Diachronic relations among speech-based and written registers in English. In *To Explain the Present: Studies in the Changing English Language in Honour of Matti Rissanen* [Mémoires de la Société Néophilologique de Helsinki LII], Terttu Nevalainen & Leena Kahlas-Tarkka (eds), 253–275. Helsinki: Société Néophilologique.
-  Bolinger, Dwight L. 1948. On defining the morpheme. *WORD* 4(1): 18–23.
- Booij, Geert. 2010. *Construction Morphology*. Oxford: OUP.
- CEEC = *Corpora of Early English Correspondence*. Compiled by Terttu Nevalainen, Helena Raumolin-Brunberg, Samuli Kaislaniemi, Jukka Keränen, Mikko Laitinen, Minna Nevala, Arja Nurmi, Minna Palander-Collin, Tanja Säily & Anni Sairio at the Department of Modern Languages, University of Helsinki. <<https://varieng.helsinki.fi/CoRD/corpora/CEEC/>> (28 May 2024).

- doi** Cowie, Claire & Dalton-Puffer, Christiane. 2002. Diachronic word-formation and studying changes in productivity over time: Theoretical and methodological considerations. In *A Changing World of Words: Studies in English Historical Lexicography, Lexicology and Semantics* [Costerus New Series 141], Javier E. Díaz Vera (ed.), 410–437. Amsterdam: Rodopi.
- Culpeper, Jonathan & Kytö, Merja. 2010. *Early Modern English Dialogues: Spoken Interaction as Writing* [Studies in English Language]. Cambridge: CUP.
- doi** Gaeta, Livio & Ricca, Davide. 2006. Productivity in Italian word formation: A variable-corpus approach. *Linguistics* 44(1): 57–89.
- doi** Hilpert, Martin. 2013. *Constructional Change in English: Developments in Allomorphy, Word Formation, and Syntax* [Studies in English Language]. Cambridge: CUP.
- doi** Hilpert, Martin. 2017. Historical Sociolinguistics and Construction Grammar: From mutual challenges to mutual benefits. In *Exploring Future Paths for Historical Sociolinguistics* [Advances in Historical Sociolinguistics 7], Tanja Säily, Arja Nurmi, Minna Palander-Collin & Anita Auer (eds), 217–237. Amsterdam: John Benjamins.
- doi** Hilpert, Martin. 2018. Wie viele Konstruktionen stecken in einem Wortbildungsmuster? Eine Problematisierung des Produktivitätsbegriffs aus konstruktionsgrammatischer Sicht. In *Wortschätze: Dynamik, Muster, Komplexität*, Stefan Engelberg, Henning Lobin, Kathrin Steyer & Sascha Wolfer (eds), 91–106. Berlin: De Gruyter.
- doi** Hilpert, Martin. 2019. *Construction Grammar and its Application to English*, 2nd edn. Edinburgh: EUP.
- doi** Kaislaniemi, Samuli. 2018. The *Corpus of Early English Correspondence Extension* (CEECE). In *Patterns of Change in 18th-century English: A Sociolinguistic Approach* [Advances in Historical Sociolinguistics 8], Terttu Nevalainen, Minna Palander-Collin & Tanja Säily (eds), 45–59. Amsterdam: John Benjamins.
- Marchand, Hans. 1969. *The Categories and Types of Present-day English Word-formation: A Synchronic-diachronic Approach*, 2nd edn. Munich: C.H. Beck'sche Verlagsbuchhandlung.
- doi** McIntosh, Carey. 1998. *The Evolution of English Prose, 1700–1800: Style, Politeness, and Print Culture*. Cambridge: CUP.
- doi** McIntosh, Carey. 2008. British English in the long eighteenth century (1660–1830). In *A Companion to the History of the English Language*, Haruko Momma & Michael Matto (eds), 228–234. Chichester: Wiley-Blackwell.
- Nevalainen, Terttu. 1999. Early Modern English lexis and semantics. In *The Cambridge History of the English Language, III: 1476–1776*, Roger Lass (ed.), 332–458. Cambridge: CUP.
- Palander-Collin, Minna. 1999. *Grammaticalization and Social Embedding: I THINK and METHINKS in Middle and Early Modern English* [Mémoires de la Société Néophilologique de Helsinki LV]. Helsinki: Société Néophilologique.
- doi** Palander-Collin, Minna. 2000. The language of husbands and wives in seventeenth-century correspondence. In *Corpus Linguistics and Linguistic Theory: Papers from the Twentieth International Conference on English Language Research on Computerized Corpora (ICAME 20), Freiburg im Breisgau 1999* [Language and Computers: Studies in Practical Linguistics 33], Christian Mair & Marianne Hundt (eds), 289–300. Amsterdam: Rodopi.
- doi** Perek, Florent. 2016. Using distributional semantics to study syntactic productivity in diachrony: A case study. *Linguistics* 54(1): 149–188.

- doi** Plag, Ingo, Dalton-Puffer, Christiane & Baayen, R. Harald. 1999. Morphological productivity across speech and writing. *English Language and Linguistics* 3(2): 209–228.
- doi** Raumolin-Brunberg, Helena & Nevalainen, Terttu. 2007. Historical sociolinguistics: The *Corpus of Early English Correspondence*. In *Creating and Digitizing Language Corpora, Vol. 2: Diachronic Databases*, Joan C. Beal, Karen P. Corrigan & Hermann L. Moisl (eds), 148–171. Houndsmills: Palgrave Macmillan.
- doi** Rodríguez-Puente, Paula. 2020. Register variation in word-formation processes: The development of *-ity* and *-ness* in Early Modern English. *International Journal of English Studies* 20(2): 145–167.
- doi** Rodríguez-Puente, Paula, Säily, Tanja & Suomela, Jukka. 2022. New methods for analysing diachronic suffix competition across registers: How *-ity* gained ground on *-ness* in Early Modern English. *International Journal of Corpus Linguistics* 27(4): 506–528.
- doi** Romaine, Suzanne. 1985. Variability in word formation patterns and productivity in the history of English. In *Papers from the 6th International Conference on Historical Linguistics* [Current Issues in Linguistic Theory 34], Jacek Fisiak (ed.), 451–465. Amsterdam: John Benjamins.
- doi** Säily, Tanja. 2011. Variation in morphological productivity in the BNC: Sociolinguistic and methodological considerations. *Corpus Linguistics and Linguistic Theory* 7(1): 119–141.
- Säily, Tanja. 2014. *Sociolinguistic Variation in English Derivational Productivity: Studies and Methods in Diachronic Corpus Linguistics* [Mémoires de la Société Néophilologique de Helsinki XCIV]. Helsinki: Société Néophilologique.
- doi** Säily, Tanja. 2016. Sociolinguistic variation in morphological productivity in eighteenth-century English. *Corpus Linguistics and Linguistic Theory* 12(1): 129–151.
- doi** Säily, Tanja. 2018. Change or variation? Productivity of the suffixes *-ness* and *-ity*. In *Patterns of Change in 18th-century English: A Sociolinguistic Approach* [Advances in Historical Sociolinguistics 8], Terttu Nevalainen, Minna Palander-Collin & Tanja Säily (eds), 197–218. Amsterdam: John Benjamins.
- doi** Säily, Tanja & Suomela, Jukka. 2009. Comparing type counts: The case of women, men and *-ity* in early English letters. In *Corpus Linguistics: Refinements and Reassessments* [Language and Computers: Studies in Practical Linguistics 69], Antoinette Renouf & Andrew Kehoe (eds), 87–109. Amsterdam: Rodopi.
- Säily, Tanja & Suomela, Jukka. 2017. *types2*: Exploring word-frequency differences in corpora. In *Big and Rich Data in English Corpus Linguistics: Methods and Explorations* [Studies in Variation, Contacts and Change in English 19], Turo Hiltunen, Joe McVeigh & Tanja Säily (eds). Helsinki: VARIENG. <<https://urn.fi/URN:NBN:fi:varieng:series-19-10>> (28 May 2024).
- doi** Säily, Tanja, Vartiainen, Turo & Siirtola, Harri. 2017. Exploring part-of-speech frequencies in a sociohistorical corpus of English. In *Exploring Future Paths for Historical Sociolinguistics* [Advances in Historical Sociolinguistics 7], Tanja Säily, Arja Nurmi, Minna Palander-Collin & Anita Auer (eds), 23–52. Amsterdam: John Benjamins.
- Suomela, Jukka. 2023. *TypeRatio: Comparing Competing Suffixes* [Computer software]. <<https://github.com/suomela/type-ratio>> (28 May 2024).
- doi** Suttle, Laura & Goldberg, Adele E. 2011. The partial productivity of constructions as induction. *Linguistics* 49(6): 1237–1269.

Vartiainen, Turo, Säily, Tanja & Hakala, Mikko. 2013. Variation in pronoun frequencies in early English letters: Gender-based or relationship-based? In *Ex Philologia Lux: Essays in Honour of Leena Kahlas-Tarkka* [Mémoires de la Société Néophilologique de Helsinki XC], Jukka Tyrkkö, Olga Timofeeva & Maria Salenius (eds), 233–255. Helsinki: Société Néophilologique.

Wolfson, Nessa. 1990. The bulge: A theory of speech behavior and social distance. *Penn Working Papers in Educational Linguistics* 2(1): 55–83.