

A Corpus Study of the English Suffixes *-ness* and *-acy*: Productivity, Genre, and Implications for L2 Learning

Ben Naismith
University of Pittsburgh

Matthew Kanwit
University of Pittsburgh

Abstract

Despite substantial scholarship relating to word structure (Anderson, 2018), for English affixes the relationship between productivity, genre, and second language (L2) learning remains unclear. Analysis of the existing literature reveals that deadjectival noun suffixes (i.e., nouns derived from adjectives such as *appropriacy* or *goodness*) have been under-examined. To address this gap, we examine two rival suffixes, *-acy* and *-ness*, through the lens of Construction Morphology (Booij, 2010), considering numerous factors which might condition their varying usage. Critically, corpus data in the Corpus of Contemporary American English and the British National Corpus (Davies, 2008-) reveal the importance of considering these affixes' productivity in relation to genre, since *-acy* is especially frequent in academic texts, principally within certain social sciences. The implications for learners and teachers of English as a second language are discussed, particularly higher-level learners building communicative competence in academic contexts, along with a preliminary learner corpus comparison of the two variants.

Résumé

Malgré de nombreuses études vis-à-vis la structure des mots (Anderson, 2018), pour les affixes en anglais ; la relation entre productivité, genre et l'apprentissage d'une langue seconde (L2) reste floue. L'analyse documentaire met en évidence que les suffixes formant des noms désadjectivaux (i.e., des noms dérivés des adjectifs tels que *appropriacy* ou *goodness*) sont sous-étudiés. Pour combler cette lacune, l'on examine deux suffixes concurrents, *-acy* and *-ness*, du point de vue de la Morphologie Constructionnelle (Booij, 2010), tout en considérant de nombreux facteurs susceptibles de conditionner leur utilisation variable. Essentiellement, les données du *Corpus of Contemporary American English* et du *British National Corpus* (Davies, 2008-) font ressortir l'importance de considérer la productivité de ces affixes par rapport au genre, du fait que *-acy* est particulièrement fréquent dans des textes académiques, principalement dans certaines sciences sociales. L'on présente finalement une discussion des implications pour les enseignants ainsi que les apprenants d'anglais langue seconde, notamment ceux de niveau supérieur développant leur compétence de communication dans des contextes académiques, accompagné d'une comparaison préliminaire de deux variantes par le biais des corpus.

A Corpus Study of the English Suffixes *-ness* and *-acy*: Productivity, Genre, and Implications for L2 Learning

There is a long tradition of scholarship in linguistics in relation to word structure (Anderson, 2018), particularly English word structure (e.g., Bloomfield, 1933). To account for the observed patterns and variation in English, a language well known for its multilingual makeup, researchers from diverse areas of linguistics have proposed analyses which prioritize different domains of language (e.g., morphological, semantic, historical, sociolinguistic, etc.). After synthesizing the various findings in this field, we turn our focus to deadjectival abstract nouns in English, specifically those ending in the suffixes *-acy* and *-ness*. To do so, we rely on a Construction Morphology framework (Booij, 2010) to account for the various factors which affect morphological productivity, highlighting the importance of genre. From this literature review, and supplemented with our own corpus investigations, we attempt to answer the following three research questions:

1. Do past studies in English derivational morphology adequately describe the relative productivity levels of the deadjectival abstract noun suffixes *-acy* and *-ness*?
2. Do past studies in English derivational morphology explain when one rival form of *-acy* or *-ness* may be selected over the other?
3. To what extent is productivity in the deadjectival abstract noun suffixes *-acy* and *-ness* variable across different genres and modalities, such as spoken English, academic texts, or media?

Having addressed these questions, we center the subsequent discussion on the implications for learners and teachers of English as a second language (L2) because first language (L1) production data can “instantiate a linguistic standard or target” (Gries & Deshors, 2014, p.109) for which higher-level learners studying in academic contexts may strive. Finally, suggestions for future empirical studies are proposed with the aim of better understanding the effect of context on the productivity of affixes, especially those affixes which have been under-researched to date.

The Research Context

Productivity

In the field of morphology, affixes are considered productive if they can be used to form new words when attached to existing bases (Haspelmath & Sims, 2010). Given the usefulness of knowing which affixes remain productive, it is unsurprising that a great deal of research has investigated this issue. However, determining the best procedure for defining and measuring productivity has remained under debate (Aronoff & Lindsay, 2014; Baayen, 1992; Hay & Baayen, 2001).

Traditionally, studies measured productivity by counting dictionary entries that contained a particular affix (Aronoff, 1976), but research has shown that dictionaries are not reliable indicators of productivity because they do not include all possible words which could be produced through the use of productive affixes, such as attaching negative prefixes to nearly any adjective. As a result, productivity statistics derived from dictionary entries

will greatly underestimate productivity of affixes (Baayen & Renouf, 1996). As an alternative, Baayen's (1992) corpus-based measurement, the *Category-Conditioned Degree of Productivity* (P), has been the most commonly used formula since the 1990s (Aronoff & Lindsay, 2014), appearing in numerous studies (e.g., Hay & Baayen, 2002; Montero-Fleta, 2011; Plag et al., 1999). Using this method, productivity is calculated based on hapax legomena (i.e., tokens which appear exactly once in a corpus, annotated as $P = n1/N$). Here, $n1$ is the number of hapax legomena with a given affix, and N is the number of total tokens with the same affix. For example, in a study of rival affixes, Baayen (1992) calculated P for *-ity* and *-ness* using frequency counts from 18 million word-forms in the English Cobuild corpus (Renouf, 1987).

Table 1

P for de-adjectival abstract nouns -ity and -ness (Baayen, 1992)

affix	N (tokens)	V (types)	$n1$ (hapax)	P (productivity)
simplex nouns	2142828	5543	128	0.0001
<i>-ity</i>	45252	405	29	0.0007
<i>-ness</i>	17481	497	77	0.0044

As Table 1 shows, *-ity* occurs nearly three times more often than *-ness* in the corpus overall. However, as there are actually more hapax legomena for *-ness* than *-ity*, *-ness* is over six times more productive than *-ity*. Similarly, *-ity* is seven times more productive than the average simplex noun despite occurring nearly five times less frequently.

Although fair criticisms have been raised of P (e.g., that it is sensitive to corpus size [Baayen, 1992, 1993]), the measurement has allowed for effective testing of predictions (Aronoff & Lindsay, 2014) and is a strong predictor of a word's parsing ratio (Hay & Baayen, 2002; to be discussed in the "Learner Language and Productivity" section). Overall, hapax legomena can serve as a suitable proxy for neologisms (i.e., new words) in measuring productivity, since they are too infrequent to efficiently be stored whole as complex words in the lexicon, or a speaker's mental dictionary. Instead, they are stored as their component parts (e.g., base + suffix), at least until the complex word becomes encountered in the input with greater frequency. Rather than stored, complex hapax legomena are therefore likely to be built on the fly: they are compositional and formed using productive affixes (Baayen & Renouf, 1996). P thus provides a useful quantitative metric, a figure between 0 and 1, placing productivity on a nuanced spectrum rather than in a binary system of "productive" or "unproductive". Consequently, a more fine-grained discussion of productivity can be had, in-line with modern morphological conceptualizations of productivity (Aronoff & Lindsay, 2014), as well as other areas of morphology which have benefited from use of continuous rather than binary distinctions, including inflection and derivation (Manova, 2005), automatic and morphophonological alternations (Neumann & Kanwit, 2018), and cliticization and affixation (Zwicky & Pullum, 1983). As such, the Category-Conditioned Degree of Productivity is the formula which is used throughout the remainder of the paper for describing fine-grained levels of productivity.

Learner language and productivity

When considering productivity, one often overlooked variable is the user's language proficiency. And yet, research shows that "lexical depth", or "how well a learner knows a word" (Crossley et al., 2011, p. 182), including knowledge of affixation, can account for substantial variance in overall learner proficiency (Koizumi & In'ami, 2013). Furthermore, other aspects of overall lexical proficiency, such as lexical sophistication, i.e., how many infrequent words a learner knows (Jarvis, 2013), are predictive of overall proficiency (Daller & Xue, 2007; Naismith et al., 2018).

If we accept the premise then that lexical depth (including productive use of affixation) is related to overall proficiency, it is therefore important to consider the extent to which learners in fact demonstrate sensitivity to English morphology. Overall, this debate as to whether learners are in some way sensitive to morphology is notably one-sided. Taking the position that learners do not decompose complex word forms, Neubauer and Clahsen (2009) conclude from their research on advanced Polish-speaking learners of German, that L2 learners do not decompose inflected forms, instead relying on whole-word processing. However, while there is evidence that adults make less use of syntactic structures, most research refutes the strong claim that learners are not at all sensitive to L2 morphology (e.g., Bosch et al., 2017; Langman & Bayley, 2002; Rodríguez, 2013; Rehak & Juffs, 2010; Spinner & Thomas, 2014).

Often taking the form of masked priming tasks, numerous studies have revealed learner sensitivity to L2 morphology (Bosch et al. 2017; Deng et al., 2016a; Rehak & Juffs, 2010). Critically, to date most of these L2 processing studies have focused on inflection, not derivation, as Deng et al. (2016a) note. For example, in Rehak and Juffs (2010), participants with English, Spanish, and Chinese L1s ("advanced/proficient" L2 English)¹ completed lexical decision tasks in English involving words with derivational prefixes. The authors found that L1 effects existed, with Chinese learners having longer reaction times when processing morphology than their Spanish counterparts. Diependaele et al. (2011) reached similar overall conclusions with respect to participants with English, Spanish, and Dutch L1s ("relatively proficient" L2 English); in their masked morphological priming lexical decision task, bilinguals were seen to process derivation the same way as native speakers in terms of the types of complex words which would prime related simple words, though "there [appeared] to be some evidence for a general reduction in the use of morphology in L2 word recognition" (p. 346). However, unlike Rehak and Juffs (2010), Diependaele et al. (2011) did not find any significant difference between learners with different L1s, though this may be explained by the similarity of the learners' morphological systems, which were all Western-European languages.

Approaching morphological sensitivity from a slightly different angle, some research has used event-related potential (ERP) measurements in conjunction with masked priming tasks. In Bosch et al. (2017), late proficient bilingual (L1 Russian/L2 German) and monolingual (L1 German) speakers were primed with inflected adjectives. The findings indicated that while the two groups demonstrated similar performances, ERP results were significantly different, suggesting that L2 morphological processing is more demanding and less automatic. Using ERPs to test groups with differing morphological knowledge (L1 Chinese/L2 low-intermediate English), Deng et al. (2016b) noticed differences in the types of responses from the two groups; participants with higher morphological knowledge elicited a P600 response whereas the lower group elicited an N400 response, indicating that

greater morphological knowledge correlated to sensitivity to rule violations. In summary then, masked priming studies overall indicate that English learners are sensitive to English morphology, though less so than to syntax. Furthermore, the factors of L1, proficiency, and explicit knowledge likely impact morphological sensitivity.

As seen in the Rehak and Juffs (2010) and Diependaele et al. (2011) studies, the L1s of the participants had a direct impact on the extent to which L2 morphological sensitivity was observed. This correlation can be further observed in a number of related studies, with sensitivity corresponding closely to the degree of morphological richness of the L1. For example, Chinese is an isolating language lacking (comparatively) in morphology. Therefore, it is unsurprising that in Jiang (2004), Chinese learners of English (with “proficient” L2 English) were found to be insensitive to inflection (number agreement) or that Langman and Bayley (2002) discovered that Chinese learners of Hungarian (“beginner to intermediate” L2 Hungarian) were largely insensitive to morphological patterns. In contrast, Rodriguez (2013) obtained different results when testing English-speaking learners of Spanish (“intermediate” L2 Spanish), two languages with similar morphological systems. In this instance, Rodríguez found that the English-speaking learners of Spanish were sensitive to morphological mismatches during a reading test. Similarly, Spinner and Thomas (2014) found that English-speaking learners of Swahili (elementary to intermediate L2 Swahili) attend to morphophonological features when determining grammatical gender, even though English itself does not typically mark for gender. Addressing a wider range of L1s, Wu (2016) also uncovered priming differences between advanced learners of English with different L1s ranging from morphologically-rich Turkish to morphologically-poor Chinese. As expected, Wu found that there was a significant effect for the L1, with greater priming for Turkish learners than even native English speakers.

A corollary of this variable sensitivity to morphology is that the measures of productivity for learners must also be reexamined. In their work on relative frequency, Hay and colleagues (Hay, 2001; Hay & Baayen, 2002, 2003; Rácz et al., 2016) have shown that whether a person processes a word as a whole or decomposes the word into its component parts is dependent upon the relative frequency of the base and the complex word. For example, *insane* is more likely to be stored as a complex word than as *sane* + the prefix *in-* because *insane* is more frequent than the base *sane*. Conversely, *childlike* is more likely to be parsed as *child* + *-like* because the complex word *childlike* is less frequent than the base *child*. This finding is important in productivity studies, as Hay and Baayen (2002) found a link between parsing and productivity in that *P* is a strong predictor of the parsing ratio (i.e., the proportion of words with a given affix which are likely to be decomposed into their component morphemes based on their relative frequencies). If learners are in fact insensitive to varying degrees to the derivational morphology of words which would typically be considered decomposable, the parsing ratios and *P* values based on general corpora frequencies likely do not hold true for learner output. Hence, parsing ratios obtained from L2 corpora frequencies would be needed to determine productivity of affixes for learners, especially those which take into account learners’ L1s. Having considered the relationship between learner language and productivity, we turn our attention to our structure of focus.

English deadjectival abstract nouns

Cross-linguistically, derivational patterns often change the base lexeme from one word class to another, with those words derived from an adjective termed “deadjectival”. Within this category, a common distinction is between “quality” nouns (e.g., *purity*, *newness*) and person nouns (e.g., *purist*, *socialite*) (Haspelmath & Sims, 2010). In English, all such quality nouns are subsumed under the broader category of “abstract nouns” which refer not only to qualities, but to ideas and states (i.e., anything that is not concrete [Abstract noun, n.d.]). The list below (1-4) represents examples of four commonly cited deadjectival abstract noun suffixes:

- | | |
|--|---|
| (1) <i>good-ness</i>
good-N.CONDITION/STATE
“goodness” | (3) <i>appropri-acy</i>
appropriate-N.STATE/QUALITY
“appropriacy” |
| (2) <i>lucid-ity</i>
lucid-N.CONDITION/STATE OF
“lucidity” | (4) <i>relev-ance</i>
relevant-N.STATE/QUALITY
“relevance” |

Although in the current section only *-ness*, *-ity*, *-acy*, and *-ance* are considered, for a complete summary Reichl (1982) provides a more exhaustive list. The other affixes he cites were not included as they are generally considered archaic and unproductive (Lass, 2000; Quinion, 2008). It should also be noted that in our list, suffixes with spelling variants are grouped together because within each grouping the different spellings represent different allomorphs (i.e., the variant suffixes retain the same meaning) and thus should not be differentiated. These allomorphs may be phonologically or orthographically conditioned based on the sound or spelling of the base word, so that, for example, nouns ending in *-l*, *-n*, or *-r* typically take *-sion* whereas nouns ending in other consonants typically take *-tion*. In the groupings above, *-acy* includes *-asy*, *-ancy*, and *-ency*; *-ance* includes *-ence*; and *-sion* includes *-tion*, and *-cion* (Oxford Dictionaries, n.d.).

Considering the four aforementioned types, these affixes too have varying degrees of productivity and as such have garnered varying levels of interest from linguists. Notably, the suffixes *-ness* and *-ity* have received the most attention, which is understandable since they are “ostensibly synonymous affixes performing an identical function but differing with regard to their potential for affixation to particular bases” (Riddle, 1982, p. 435). As a result, they are ideal for comparative purposes, and their high frequency simplifies data collection because they are found with more regularity in different corpora. For example, focusing on these affixes from a more generative, historical perspective, Riddle (1982) argues for a semantic distinction between *-ness* and *-ity*, while also considering a range of potential factors affecting their relative productivity (which are explored further in the “Register” section). Taking a corpus-based approach to an analysis of these same suffixes, Baayen (1992) found that in terms of *P*, *-ness* was much more productive than *-ity*, at least when using the frequencies from the CELEX database, which is derived from the larger general reference corpus, the British National Corpus (BNC).

Other research, though not focusing exclusively on deadjectival abstract nouns, has investigated a range of derivational suffixes, including the most common deadjectival suffixes. In Baayen and Renouf (1996), the focus is on five common derivational affixes, *-ly*, *-ness*, *-ity*, *un-*, and *in-*. Through an analysis of a newspaper corpus, the authors

found that *-ness*, *un-*, and *-ly* were in fact becoming more productive. Exploring a comprehensive range of affixes, Hay and Baayen (2002) measured the productivity of 26 prefixes and 54 suffixes. Of those, there were nine deadjectival noun suffixes which demonstrated the following *P* values, again according to CELEX frequencies (though not all of these are restricted to adjectival bases, Table 2).

Table 2

P values of nine deadjectival noun suffixes (Hay & Baayen, 2002)

Suffix	<i>-ness</i>	<i>-ist</i>	<i>-hood</i>	<i>-ism</i>	<i>-dom</i>	<i>-ity</i>	<i>-ance</i>	<i>-ence</i>	<i>-ment</i>
<i>P</i> value	0.008	0.005	0.004	0.003	0.002	0.001	0.000	0.000	0.000

From this small sample of data, it is immediately apparent that whereas *-ness* appears to be considerably more productive than the other suffixes in this category, measuring productivity can pose challenges for less frequently occurring structures which may not appear with sufficient regularity in one type of corpus to provide substantial analyzable data. Therefore, to address our first research question, previous work on English derivational morphology does not appear to adequately describe the relationship between the deadjectival abstract noun suffixes *-acy* and *-ness*.

Comparison of the suffixes *-acy* and *-ness*

In the previous section we saw that English deadjectival abstract nouns affixes have been given varying levels of attention. Next, we consider the under-analyzed suffix pair of *-acy* and *-ness*, which share many traits despite their disparate frequencies and productivity levels. For example, although *-ness* is the clear default suffix for deadjectival abstract nouns, there are adjectives for which the affixes compete as possible rivals (Aronoff & Lindsay, 2014). As shown in Table 3, in addition to forming abstract nouns from adjective bases, both affixes attach to multi-syllabic words. In the case of these competing pairs, there may be subtle semantic differences as in the more specialized *appropriacy*, “the fact that a word or phrase sounds natural and is acceptable when used in a particular situation” (Appropriacy, n.d.), versus the more general *appropriateness*, “the quality of being suitable or right for a particular situation or occasion” (Appropriateness, n.d.). However, as will be seen, there are likely other differentiating factors in choosing between the suffixes beyond simply denotation.

Table 3

Sample rival pairs of nouns with both -acy and -ness suffixes

<i>-acy</i>	<i>-ness</i>
<i>accuracy</i>	<i>accurateness</i>
<i>adequacy</i>	<i>adequateness</i>
<i>appropriacy</i>	<i>appropriateness</i>
<i>delicacy</i>	<i>delicateness</i>
<i>efficacy</i>	<i>efficaciousness</i> ²
<i>intimacy</i>	<i>intimateness</i>
<i>immediacy</i>	<i>immediateness</i>
<i>obduracy</i>	<i>obdurateness</i>
<i>privacy</i>	<i>privateness</i>

In spite of their similarities, considerable differences between the two affixes are also present. Historically, *-acy* is of Latin origins, passing to French before adoption into English (Quinion, 2008); *-acy* is therefore said to possess the [+latinate] abstract feature value posited by Aronoff (1976). Such affixes are generally considered more decomposable and thus analyzable (Reichl, 1982). In contrast, *-ness* is of Germanic origin (Quinion, 2008), which may partially explain why *-ness* is most productive with monosyllabic bases (Baayen & Renouf, 1996), given the multi-syllabic nature of many words of Latin origin and the propensity for affixes to pair with bases of the same language of origin (Haspelmath & Sims, 2010).

A second stark difference is the status of the morphological conditioning of the two affixes. Like many other deadjectival nouns, *-acy* is morphologically conditioned; for example, *-acy* is much more productive for bases whose adjectival form ends in *-ate*, as demonstrated in yielding *adequacy* from *adequate* or *delicacy* from *delicate* (Aronoff & Lindsay, 2014). On the other end of the spectrum, *-ness* gains much of its productive power from its lack of morphological conditioning, consistent with the fact that *-ness* is a word-boundary affix (Chomsky & Halle, 1968), so that *-ness* will typically be the final morpheme of a word and will not affect the preceding phonology (e.g., the stress of *productive* being maintained in *productiveness*). In contrast, *-acy* is a morpheme- or formative-boundary affix, which may more often be followed by other morphemes (e.g., *delicacies*) and which may affect the phonology of the base (e.g., the vowel sound of *supreme* to *supremacy*). As a result, *-ness* is able to attach to a wider range of bases, capable of even extending to abstract, non-adjective bases, such as *togetherness* and *nothingness*, though such occurrences are not prototypical (Reichl, 1982). An alternative perspective which reaches the same conclusion in describing these affixes' ordering restrictions is through "complexity based ordering" (Hay & Plag, 2004), which claims that affixes align along a hierarchy of complexity based on how easily an affix can be parsed in processing. In this model, more easily parsed affixes are found to be more separable (Hay & Plag, 2004) as well as more productive (Plag & Baayen, 2009).

What remains to be accounted for, therefore (and in answer to our research questions 2 and 3), is the extent to which these two suffixes are productive, why one suffix or the other may be selected, and in what specific situations the more restrictive *-acy* continues to be used productively.

Theoretical framework

To answer these questions, the Construction Morphology framework is employed (Booij, 2010). As with other constructionist approaches, researchers focused on Construction Morphology seek to understand language in terms of speakers' experience with a language (e.g., in terms of cognition, pragmatics, and processing [Scott, 2012]). Starting at a young age (Tomasello, 2000), a speaker's language knowledge can be thought of in terms of an inventory of form-function pairings labelled constructions (Goldberg, 2006, 2013), which vary in strength depending on factors like frequency in the input and saliency. Through user knowledge of constructions (and their relative strengths), it is then possible to determine what constructions can be combined to produce new language (e.g., new words in the case of Construction Morphology). For example, in considering lexicogrammatical constructions, Bybee (2006, 2010, 2018) describes how grammatical

productivity is achieved through the repeated use of constructions in limited lexical contexts eventually being generalized to different contexts (i.e., syntactic generalization that accompanies semantic bleaching, or widening of application of meaning). Likewise, certain [base + affix] constructions with sufficient pattern strength may be generalized more easily (across word classes, genres, historical roots) than those [base + affix] constructions with less strength in the mental lexicon.

What is more, as this mental lexicon is intricately linked to frequency (Baayen 1992, 2009, 2010; Bien et al., 2011), productive word formation rules thereby increase the speed at which low frequency formations can be processed. Such an approach aligns well with earlier conclusions that many non-linguistic factors determine productivity, including social and historical factors, attitude, pragmatics, and register (e.g., Baayen, 1992; Riddle, 1982).

In the present study, Construction Morphology therefore allows us to consider the range of elements which may contribute to morphological productivity, and to synthesize the findings from research with disparate methodologies and theoretical perspectives. In doing so, we aim to present a cognitively realistic overview of these affixes and to apply this overview to language learners. We now look at this range of factors, focusing on genre, to determine how the *-ness* and *-acy* constructions may be represented in users' minds to allow for production under certain conditions.

Semantic and morphological conditioning

Recall from the comparison of *-acy* and *-ness* that in certain cases there is a link between semantics and productivity, a long-standing belief in morphology (Aronoff, 1976). For example, this link is argued for by Riddle (1982) in the cases of *-ness* and *-ity*, and is apparent in the dictionary definitions of *efficacy* and *effectiveness*. In both instances, one form, *-ness*, is seen to be more productive than its counterpart even when both forms are available. One explanation for this asymmetry is partial synonymy blocking. Whereas in traditional blocking one form prevents the existence of another feasible form (Aronoff, 1976) or set of forms (Aronoff & Lindsay, 2014), blocking need not be an all or nothing process. In some instances, for example with deadjectival *-nce/-ncyn* pairs, *-ncyn* is more specialized (e.g., *excellence/excellency* [Aronoff & Lindsay, 2014]). The same relationship can be argued for *-acy* and *-ness* pairs like *delicacy/delicateness*, where *delicacy* has come to refer to an expensive or desirable food in addition to its more general meaning. From a morphological perspective, in our comparison of *-acy* and *-ness* the differences in conditioning between the suffixes were also described, with the latter much less conditioned by the length or composition of the base. This difference in morphological condition is undoubtedly another key component in explaining the two affixes' differing productivity, as "the degree of productivity of an affix varies significantly with the morphological structure of the base word it attaches to" (Baayen & Renouf, 1996, p. 94), in line with complexity based ordering previously discussed (Hay & Plag, 2004; Plag & Baayen, 2009).

Genre

Language variation across genres is well attested in many areas of linguistic study. Specifically, the way language is used changes depending on the field (the content and

purpose), the tenor (the participants), and the mode (its format and function) (Halliday & Hassan, 1990). From Labov's (1972) seminal work on stylistic variation, to more recent research on phonetic variation (Medina-Rivera, 1999; Winter & Grawunder, 2012) and morphosyntactic variation (Geeslin & Gudmestad, 2008; Wilkinson, 2013), genre clearly affects our choice and production of words and grammatical structures. Additionally, a critical component of language acquisition is the ability to adapt one's language use according to linguistic, social, and contextual factors, as proficient L2 learners do in demonstrating communicative, rather than simply grammatical, competence (Bayley & Tarone, 2012; Canale & Swain, 1980; Geeslin, 2018; Kanwit, 2017). What is more, factors such as discourse topic have been shown to affect a broad array of linguistic phenomena, including the use of personal deixis (Blas Arroyo, 2000); evidential markers (Kanwit, 2015); and subject pronouns, temporal deixis, and parenthetic verbs (Flores-Ferrán, 2010). For example, we might predict certain linguistic variants (i.e., the use of one form with a similar meaning over another) based on the registers and discourse topics that are more likely to emerge in particular settings, such as the forms expected to be used by clients and therapists in counseling sessions (Flores-Ferrán, 2010) or by native speakers and learners in classroom discourse, in a particular text, or in academic advising sessions (e.g., Bardovi-Harlig, 1999; Bardovi-Harlig & Hartford, 2005). By extension then, genre should likewise affect our choice of new word production.

On the one hand, general claims have been made about the link between word origins and broad domains (e.g., that Latin/French words are associated with education in English [Riddle, 1982]). Overall, however, research regarding the role of derivational morphology in realizing register variation lags behind (Montero-Fleta, 2011). As a result, in studies making claims about the productivity of specific affixes, there is often no consideration of the discourse types under investigation (Plag et al., 1999). Instead, and particularly with respect to English derivation, other qualities of affixes have been prioritized including semantics, syntax, and phonology (Plag et al., 1999), with few studies (e.g. Baayen, 1994; Baayen & Nejat, 1997; Clark, 1998; Lüdeling & Evert, 2005; Montero-Fleta, 2011; Plag et al., 1999; Säily, 2014) investigating how genre interacts with morphological productivity (Baayen, 2009).

The effect of genre is, however, the focus of Montero-Fleta's (2011) study investigating the productivity of various affixes in different registers. The author found that in scientific registers, like medical and computer-science texts, there was a high overall prevalence of suffixation, particularly the abstract noun suffixes *-ity*, *-ion*, *-ness*, and *-ize*. In contrast, other suffixes such as *-free* and *-like* were completely unproductive in the specialized corpora, yet fully productive in the general BNC. As a result, Montero-Fleta concluded that both computer science and medical texts shared common characteristics in terms of affixation and productivity due to their scientific nature, and that productivity is in fact sensitive to register.

Corpus findings

In applying the conditioning factors of the prior section to *-acy* and *-ness*, there is currently more that can be said about *-ness* from the existing literature, and at first glance the difference between *-acy* and *-ness* may appear to be that the former is no longer productive. However, such a conclusion is likely premature. A careful examination of two large, general corpora, the COCA and the BNC (Davies, 2008-), reveals that of the nouns

ending in *-acy*, there are 226 types and 81,960 tokens in COCA and 86 types and 10,514 tokens in BNC (Table 4). Notably, of these types, 131 (COCA) and 26 (BNC) are hapax legomena, producing *P* statistics of 0.0016 and 0.0025, respectively. To arrive at these figures, the search results from the COCA and BNC interfaces were exported and manually checked in order to exclude misspellings (which otherwise often appear as hapax legomena) and types with non-adjectival bases, a “laborious and time-consuming enterprise” recommended by Lieber (2014, p. 90). What can be learned from the resulting figures is that *-acy* does appear to be productive, albeit to a limited degree, as confirmed by hapax legomena such as *deliracy* (to describe the state of being delirious), *innacy* (meaning *innateness*), and *obligacy* (the state of being obliged).³ In contrast, and as expected, in the same corpora *-ness* is far more frequent and productive; it accounts for about 450,000 more tokens in COCA and about 85,000 more in BNC, and is more than two times as productive (*P*) as *-acy* in COCA and 3 times as productive in BNC.

Table 4
Productivity of -acy and -ness in COCA and BNC

<i>Corpus</i>	<i>Affix</i>	<i>N (tokens)</i>	<i>V (types)</i>	<i>n1 (hapax)</i>	<i>P (productivity)</i>
COCA	<i>-acy</i>	81,960	226	131	0.0016
	<i>-ness</i>	531,637	4,610	2,080	0.0039
BNC	<i>-acy</i>	10,514	86	26	0.0025
	<i>-ness</i>	95,821	2,048	788	0.0082

Considering the distribution of these affixes in different domains and genres, Plag et al. (1999) discovered when comparing *-ness* and *-able*, that *-ness* was far more productive in written than spoken English. This finding matches our own searches of COCA and BNC for raw frequency counts across domains. In contrast, to our knowledge there is no study of *-acy* and its productivity across registers, though again, frequency counts may provide a rough estimate for potential productivity in different domains.⁴

Table 5
COCA token frequency counts and relative frequencies by domain for -acy and -ness

Section	Academic	Spoken	Newspaper	Magazine	Fiction	Total
Words (M)	111.4	116.7	113.0	117.4	111.8	577
Frequency	<i>-acy</i> (31%)	68,472 (24%)	24,644 (19%)	26,510 (17%)	27,365 (7%)	155,655 (21%)
	<i>-ness</i> (69%)	151,246 (76%)	77,214 (81%)	111,067 (83%)	131,099 (93%)	581,409 (79%)
Per Million	<i>-acy</i>	614.59	211.26	234.61	233.18	77.29
	<i>-ness</i>	1,357.56	661.37	982.93	1,117.12	990.50
						1,006.59

Table 6*BNC token frequency counts and relative frequencies by domain for -acy and -ness*

<u>Section</u>	<u>Acad.</u>	<u>Non-Acad.</u>	<u>News-paper</u>	<u>Mag.</u>	<u>Spok.</u>	<u>Misc.</u>	<u>Fiction</u>	<u>Total</u>
Words (M)	15.3	16.5	10.5	7.3	10.0	20.8	15.9	100
Frequency	-acy	6,310 (26%)	4,716 (22%)	1,586 (14%)	898 (13%)	591 (11%)	3,389 (11%)	18,736 (15%)
	-ness	18,144 (74%)	17,077 (78%)	9,586 (86%)	5,938 (87%)	4,603 (89%)	27,899 (89%)	103,669 (85%)
Per Million	-acy	411.57	285.90	151.53	123.66	59.32	162.66	78.32
	-ness	1,183.43	1,035.27	915.88	817.68	461.98	1,339.03	1,283.65
								1,036.69

Both corpora reveal that *-acy* appears at its highest rates relative to *-ness* in the academic domain (31% and 26% in COCA and BNC, respectively, in Tables 5 and 6), each of which eclipses the baseline relative rate of *-ness* in the corpora by 10-11% (21% and 15% overall, respectively). Alternatively, the fiction domain least lends itself to use of *-acy* in both corpora, with relative rates of 6-7% in each. Considering the strength of the relationship between *-acy* and academic texts, a look at the relevant sub-sections in COCA shows that within this domain (Table 7), in some academic genres in what might be termed the social sciences categories (History, Law/Political Science, Education), *-acy* appears at especially high rates relative to *-ness* (38-46%), exceeding its already elevated relative rate for the academic domain (31%). Turning to the academic subdomains of the BNC (Table 8), overall the two suffixes present a more balanced profile though there is no clear one-to-one correspondence with the subdomains of COCA. However, we do observe some of the same genre trends: in Law/Education, the largest academic subdomain, *-acy* has a higher rate relative to *-ness* than in all other subdomains except Engineering, an exception which may be a result of the small sample size of the latter domain which contains only .7M words, 165 *-acy* tokens, and 365 *-ness* tokens. Of note, in both corpora, there appears to be a dispreference for *-acy* in more scientific domains, especially in Medicine, Science/Technology, and Natural Sciences.

Table 7
COCA token frequency counts by academic sub-domain for -acy and -ness

Academic Section	Size (M)	Freq. -acy	Freq. -ness	Per M -acy	Per M -ness	% -acy relative to -ness
History	12.2	11,641	13,920	950.66	1,136.77	46%
Law/Political Science	8.6	8,119	11,365	944.03	1,321.45	42%
Education	9.4	13,560	21,726	1,435.94	2,300.68	38%
Misc.	4.3	3,089	7,140	725.69	1,677.38	30%
Geography/ Social Science	16.2	10,218	27,575	631.52	1,704.26	27%
Science/Tech.	14.1	3,816	12,385	271.11	879.91	24%
Humanities	11.9	4,979	17,970	417.47	1,506.73	22%
Medicine	6.7	3,111	11,848	464.29	1,768.23	21%
Philosophy/ Religion	6.7	3,659	14,947	542.86	2,217.56	20%
Total⁵	111.4	68,472	151,246	614.59	1357.56	31%

Table 8
BNC token frequency counts by academic sub-domain for -acy and -ness

Academic Section	Size (M)	Freq. -acy	Freq. -ness	Per M -acy	Per M -ness	% -acy relative to -ness
Engineering	0.7	165	365	243.14	537.6	31%
Law/Education	4.6	2,557	6,838	554.04	1,481.63	27%
Humanities/Arts	3.3	1,409	4,323	427.48	1,311.56	25%
Social Science	4.2	1,729	5,058	409.28	1,197.31	25%
Medicine	1.4	348	1,085	246.32	767.97	24%
Natural Science	1.1	102	475	92.35	430.05	18%
Total	15.3	18,736	103,669	187.36	1,036.69	26%

Discussion: Implications for language learning and teaching

Based on the above corpus findings, and the earlier discussion of learner language and productivity, it is clear that ESL teachers would benefit from considering how to increase learners' morphological sensitivity and awareness. For example, considering *-acy* and *-ness*, if a lower-level learner has not yet formed mappings for constructions like [ADJ + *-acy* = *abstract noun*], or that [ADJ + *-ness*] is associated more with writing than speaking, then we would predict different distributions of affixes in learner corpora and general corpora, with a more even distribution for learners that reveals less of an effect for genre. For learners at more advanced levels or in English for Academic Purposes (EAP) contexts, morphological awareness is necessary for mastery of academic genres because subtle differences in register are realized, as we have seen, in part through derivational suffix choices (Plag et al., 1999; Montero-Fleta, 2011), which contributes overall to variable form use across modalities, genres, institutional settings, and discourse topics (e.g., Bardovi-Harlig, 1999; Bardovi-Harlig & Hartford, 2005; Blas Arroyo, 2000; Flores-Ferrán, 2010; Geeslin & Gudmestad, 2008; Kanwit, 2015; Medina-Rivera, 1999). Moreover, morphological awareness also affects many systems and skills including vocabulary size and reading speed, both of which are strong predictors of academic achievement (Roche & Harrington, 2013), and to date, research on using morphemes to teach vocabulary has yielded positive results (Lems et al., 2010), though as Bosch et al. (2017) note, "morphosyntactic phenomena spelled out through affixation have been identified as one of the key challenging areas for L2 language acquisition and processing" (p. 438).

For language teachers, one final factor to consider is the effect of explicit (i.e., declarative) knowledge of morphology (DeKeyser, 2003; Ellis, 2005, 2009; Hulstijn, 2005). Although across theoretical approaches of language acquisition there is a longstanding and contentious debate about whether declarative knowledge can become procedural (cf. Krashen, 1982), a full discussion is beyond the scope of the current section (see the chapters of VanPatten & Williams [2015] for overviews of the treatment of knowledge types across different theoretical approaches of L2 acquisition). Nevertheless, we accept the assumption here that instruction and explicit knowledge can facilitate and speed up the natural learning progression and ultimate level of attainment (Larsen-Freeman & Long, 1991; Robinson et al., 2012/2014; Schmidt, 2012), by increasing learners' ability to notice salient patterns in the input. What is not as clear, is the extent to which this same principle applies to morphological knowledge, especially given that the majority of processing studies do not measure or address explicit knowledge.

We highlight, however, two notable exceptions of studies that measure participants' morphological knowledge prior to conducting processing research. In Deng et al. (2016a), the 40 Chinese-speaking learners of English were given a morphological knowledge test in which participants selected correct word forms. After controlling for overall language proficiency (B1 on the Common European Framework of Reference [CEFR] for both groups), it was found that greater priming occurred for the group with higher morphological knowledge, as well as slower reading times. Crucially, these findings suggest that aside from overall proficiency, explicit morphological knowledge is an important variable to consider in addition to subconscious processes. Nevertheless, in a contradictory study, Kraut (2015) also measured explicit morphological knowledge and did not find support for the idea that learners decompose words automatically. Importantly, in this study, though the groups are labelled "intermediate" and "advanced", their overall

proficiency levels were A2 and B2 on the CEFR; given that A2 corresponds to what most systems would designate “elementary” or “pre-intermediate”, it may be that learners below a certain threshold are as of yet unable to automatize their rudimentary explicit morphological knowledge. Considering both of these studies, and based on the general lack of research in this area, it seems that further data are needed to clarify the degree to which explicit knowledge plays a role in morphological processing, and how this factor interacts with both overall proficiency and learner L1s. Nevertheless, based on these findings, there are implications for language teaching, suggesting that time in the classroom should be devoted to improving learners’ explicit derivational morphological knowledge. In fact, since morphology plays an important role in demonstrating a learner’s ability to shift between registers (Roche & Harrington, 2013), a hallmark of advanced language use (Bayley & Tarone, 2012; Geeslin, 2018), and since morphological phenomena have shown to pose a learning problem (Bosch et al., 2017), the need to devote curricular attention to instruction on morphological patterns seems paramount, with preliminary evidence suggesting that such knowledge can indeed be deployed quite fruitfully (Deng et al., 2016a).

Future research

In her study of the suffixes *-ness* and *-ity*, Riddle (1982) found that these affixes’ productivity “cannot be accounted for in any simple way” (p. 457), and here we argue that the same can be said of *-acy* and *-ness*, and almost certainly of other competing pairs of affixes as well, pointing to the need to explain multiple factors that could trigger the use of one over the other. In attempts to tease apart these potential factors affecting productivity, future research will do well to consult corpora to obtain information about frequency, hapax legomena, and to see how new words are used in context across different domains. And, while Baayen and Renouf (1996) argue for the processing of large-scale corpora, we note that smaller specialized corpora or sub-corpora are particularly invaluable for investigating the interaction between productivity and genre (e.g., Montero-Fleta, 2011). As Linford et al. (2016) demonstrated in their research on lexical frequencies and methodological practices, results may be substantially affected depending on whether observed frequencies are based on an external measure (as in consulting some larger L1 corpus, like the BNC) or internal measure (as in using the very same L2 corpus as a point of comparison).

For example, although we are making no claims about learner productivity in the current paper, Table 9 presents basic learner production data from a learner corpus, the University of Pittsburgh English Language Institute Corpus (PELIC). This 4.2-million-word learner corpus consists of learner essays which are broadly academic in nature and thus comparable in domain to the academic sections of COCA and BNC. These figures, as in the earlier Table 4, show that for learners *-ness* is also more frequent and productive than *-acy*. Results also reveal that as learners achieve advanced proficiency, they use *-acy* at much larger relative rates than at previous levels, more greatly approximating native-like targets in the process, after beginning with nearly exclusive use of *-ness*.

Table 9*Productivity of -acy and -ness in PELIC*

<i>Affix</i>	<i>Level</i>	<i>N (tokens)</i>	<i>Rel Freq</i>	<i>V (types)</i>	<i>n1 (hapax)</i>	<i>P (productivity)</i>
<i>-acy</i>	Intermediate	8	8%	3	2	0.2500
	Upper Int.	161	9%	10	3	0.0186
	Advanced	221	24%	13	5	0.0226
	Total	390	14%	16	6	0.0154
<i>-ness</i>	Intermediate	89	92%	23	8	0.0899
	Upper Int.	1708	91%	107	40	0.0234
	Advanced	690	76%	115	38	0.0551
	Total	2487	86%	171	58	0.0233

Moreover, frequency may play a role as a main effect in predicting the occurrence of a particular variant (e.g., in predicting outright whether speakers use a subject pronoun as in Bayley et al. 2013) or it may play a mediating role for other variables (e.g., revealing significantly different distributions of pronoun use within tense-mood-aspect and semantic class when those variables are divided according to frequency, as in Erker & Guy, 2012; Linford et al., 2016). These considerations of frequency are likewise crucial in explaining why morphological constructions with *-acy* and *-ness* may reveal varying pattern strengths for different users or in different domains. Finally, as we have seen, affixes can differ in their productivity across registers (Plag et al., 1999), and as a result, one of the challenges for future research is “to study in greater detail how *context* and *cotext* [emphasis added] affect the use of complex words” (Plag et al., 1999, p. 226).

Conclusions

Returning to the original research questions, although we have advanced our knowledge of the characteristics of the affixes *-acy* and *-ness*, productive lines of research on the affixes remain. Overall, it seems likely that numerous different factors affect these affixes’ varying usage and productivity, including semantic conditioning, morphological conditioning, genre, and frequency (both absolute and relative). We argue, albeit from a more holistic perspective, that this amalgamation of factors can be suitably accounted for in a probabilistic fashion within a Construction Morphology framework, considering how each of these factors might increase these affixes’ pattern strengths within users’ minds. Focusing on register specifically, corpus findings suggest that genre plays an important role, with *-acy* occurring more regularly in fields in the social sciences. Should further investigations support these initial findings, the implications for English language learners and instructors are clear: as part of developing overall lexical proficiency and communicative competence, learners, especially those in, or planning to enter, English tertiary education, must develop awareness of competing word forms in order to better understand and produce genre-specific texts.

Correspondence should be addressed to Ben Naismith.
Email: bnaismith@pitt.edu

Acknowledgments

This research was supported by the Social Sciences and Humanities Research Council of Canada and is based in part upon work supported by the National Science Foundation under Grant No. SBE-0836012 (previously NSF Grant No. SBE-0354420). We express our gratitude to the editor, Eva Kartchava, and to the two anonymous reviewers for their helpful suggestions. All errors are, of course, our own.

Notes

¹ In the current section, proficiency levels in quotes reflect original author descriptions of participants.

² In the case of *efficacy/efficaciousness*, there is also a third competing form: *effectiveness*. However, although all three forms originally derive from the noun base *effect*, they do not derive from the same adjectival forms (*efficacious* and *effective*).

³ Some readers may wonder about the plausibility of these novel words. Although it is beyond the scope of the paper to assess the acceptability of these new forms, all three of these hapax legomena occurred in COCA in print publications (a novel, a magazine, and an academic journal, respectively) and were produced by native speakers.

⁴ Of note, unlike results for Table 4, the sectional breakdown of suffix frequencies does not allow for manual cleaning, and as such, these token counts are somewhat inflated by misspellings and words with non-adjectival bases. We note that the relative rate of *-acy* is slightly higher in both COCA and BNC when manual cleaning cannot occur. However, we would not expect these inclusions to affect the relative counts within *domains* (Tables 5-6).

⁵ The total values provided by COCA exceed the sum of the individual sub-domains that it reports, meaning that there is likely an additional “other” category (apart from “miscellaneous”) that the corpus does not report individually. Conversely, the sub-domains provided in BNC (Table 8) do sum to the total value reported.

References

- Abstract noun. (n.d.). In *Oxford dictionaries*.
https://en.oxforddictionaries.com/definition/abstract_noun
- Anderson, S. R. (2018). A Short History of Morphological Theory. In J. Audring & F. Masini, (Eds.), *The Oxford Handbook of Morphological Theory* (pp. 19-33). Oxford University Press.
- Appropriacy. (n.d.). In *Cambridge Dictionary*.
<https://dictionary.cambridge.org/us/dictionary/english/appropriacy>
- Appropriateness. (n.d.). In *Cambridge Dictionary*.
<https://dictionary.cambridge.org/us/dictionary/english/appropriateness>
- Aronoff, M. (1976). *Word Formation in Generative Grammar*. MIT Press.
- Aronoff, M. & Lindsay, M. (2014). Productivity, blocking, and lexicalization. In R. Lieber & P. Stekauer (Eds.), *The Oxford Handbook of Derivational Morphology* (pp. 67-

- 83). Oxford University Press.
<https://doi.org/10.1093/oxfordhb/9780199641642.001.0001>
- Baayen, R. H. (1992). Quantitative aspects of morphological productivity. In G. Booij & J. van Marle (Eds.), *Yearbook of Morphology 1991* (pp. 109-149). Kluwer.
https://doi.org/10.1007/978-94-011-2516-1_8
- Baayen, R. H. (1993). On frequency, transparency and productivity. In G. Booij & J. van Marle (Eds.), *Yearbook of Morphology 1992* (pp. 181-208). Kluwer.
https://doi.org/10.1007/978-94-017-3710-4_7
- Baayen, R. H. (1994). Derivational productivity and text typology. *Journal of Quantitative Linguistics*, 1, 16-34.
- Baayen, R. H. (2009). Corpus linguistics in morphology: Morphological productivity. In A. Lüdeling & M. Kyto (Eds.), *Corpus Linguistics. An international handbook* (pp. 900-919). Mouton De Gruyter.
- Baayen, R. H. (2010). Demythologizing the word frequency effect: A discriminative learning perspective. *The Mental Lexicon*, 5, 436-461.
- Baayen, R. H. & Nejit, A. (1997). Productivity in context: A case study of a Dutch suffix. *Linguists*, 35, 565-587.
- Baayen, R. H. & Renouf, A. (1996). Chronicling the Times: productive lexical innovations in an English newspaper. *Language*, 72, 69-96. <https://doi.org/10.2307/416794>
- Bardovi-Harlig, K. (1999). Examining the role of text type in L2 tense-aspect research: Broadening our horizons. In P. Robinson & N. Jungheim (Eds.), *Proceedings of the Third Pacific Second Language Research Forum* (Vol. 1, pp. 129-138). PacSLRF.
- Bardovi-Harlig, K. & Hartford, B. S. (2005). Institutional discourse and interlanguage pragmatics. In K. Bardovi-Harlig & B. S. Hartford (Eds.), *Interlanguage pragmatics: Exploring institutional talk* (pp. 7-36). Erlbaum.
- Bayley, R., Greer, K., & Holland, C. (2013). Lexical frequency and syntactic variation: A test of a linguistic hypothesis. *University of Pennsylvania Working Papers in Linguistics*, 19(2), 21-30.
- Bayley, R. & Tarone, E. (2012). Variationist perspectives. In S. Gass & A. Mackey (Eds.), *Handbook of Second Language Acquisition* (pp. 41-56). Routledge.
<https://doi.org/10.4324/9780203808184.ch3>
- Bien, H., Baayen, R. H., & Levelt, W. J. M. (2011). Frequency effects in the production of Dutch deverbal adjectives and inflected verbs. *Language and Cognitive Processes*, 26(4-6), 683-715. <https://doi.org/doi.org/10.1080/01690965.2010.511475>
- Blas Arroyo, J. L. (2000). Mire usted Sr. González ... Personal deixis in Spanish political-electoral debate. *Journal of Pragmatics*, 32(1), 1-27.
[https://doi.org/doi.org/10.1016/S0378-2166\(99\)00040-5](https://doi.org/doi.org/10.1016/S0378-2166(99)00040-5)
- Bloomfield, L. (1933). *Language*. Holt.
- Booij, G. (2010). Construction morphology. *Language and Linguistics Compass*, 4, 543-555. <https://doi.org/10.1111/j.1749-818X.2010.00213.x>
- Bosch, S., Krause, H., & Leminen, A. (2017). The time-course of morphosyntactic and semantic priming in late bilinguals: A study of German adjectives. *Bilingualism: Language and Cognition*, 20(3), 435-456.
<https://doi.org/10.1017/S1366728916000055>
- Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language*, 82(4), 711-733. <https://doi.org/10.1353/lan.2006.0186>

- Bybee, J. (2010). *Language, usage and cognition*. Cambridge University Press.
- Bybee, J. (2018). *Language change*. Cambridge University Press.
- Canale, M. & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1-47.
<https://doi.org/10.1093/applin/I.1.1>
- Chomsky, N. & Halle, M. (1968). *The sound pattern of English*. Harper & Row.
- Clark, H. H. (1998). Communal lexicons. In K. Malmkjaer & J. Williams (Eds.), *Context in Language Learning and Language Understanding* (pp. 63-87). Cambridge University Press.
- Crossley, S., Salsbury, T., McNamara, D., & Jarvis, C. (2011). What is lexical proficiency? Some answers from computational models of speech data, *TESOL Quarterly*, 45(1), 182-193. <https://doi.org/10.5054/tq.2010.244019>
- Davies, M. (2008-). *The Corpus of Contemporary American English (COCA): 560 million words, 1990-present*. <https://corpus.byu.edu/coca/>.
- Daller, H. & Xue, H. (2007). Lexical richness and the oral proficiency of Chinese EFL students. In H. Daller, J. Milton, & J. Treffers-Daller (Eds.), *Modelling and assessing vocabulary knowledge* (pp. 150-164). Cambridge University Press.
- DeKeyser, R. (2003). Implicit and Explicit Learning. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 313-348). Blackwell.
<https://doi.org/10.1002/9780470756492.ch11>
- Deng, T., Shi, J., Bi, H., Dunlap, S., & Chen, B. (2016a). The relationship between the morphological knowledge and L2 online processing of derivational words. *International Journal of Bilingualism*, 21(4), 402-418.
<https://doi.org/10.1177/1367006916629222>
- Deng, T., Shi, J., Dunlap, S., Bi, H., & Chen, B. (2016b). Morphological knowledge affects processing of L2 derivational morphology: An event-related potential study. *Journal of Neurolinguistics*, 37, 47-57.
<https://doi.org/10.1016/j.jneuroling.2015.09.001>
- Diependaele, K., Duñabeitia, J. A., Morris, J., & Keuleers, E. (2011). Fast morphological effects in first and second language word recognition. *Journal of Memory and Language*, 64(4), 2344-358. <https://doi.org/10.1016/j.jml.2011.01.003>
- Ellis, R. (2005). Planning and task-based research: Theory and research. In R. Ellis (Ed.), *Planning and task-performance in a second language* (pp. 3-34). Benjamins.
- Ellis, R. (2009). Implicit and explicit learning, knowledge, and instruction. In Ellis et al. (Eds.), *Implicit and explicit knowledge in second language learning, testing, and teaching* (pp. 3-25). De Gruyter.
- Erker, D. & Guy, G. R. (2012). The role of lexical frequency in syntactic variability: Variable subject personal pronoun expression in Spanish. *Language*, 88(3), 526-557. <https://doi.org/10.1353/lan.2012.0050>
- Flores-Ferrán, N. (2010). An examination of mitigation strategies used in Spanish psychotherapeutic discourse. *Journal of Pragmatics*, 42(7), 1964-1981.
<https://doi.org/10.1016/j.pragma.2009.12.005>
- Geeslin, K. L. (2018). Variable structures and sociolinguistic variation. In P. Malovrh & A. Benati (Eds.), *The handbook of advanced proficiency in second language acquisition* (pp. 547-565). Wiley.

- Geeslin, K. L. & Gudmestad, A. (2008). Comparing interview and written elicitation tasks in native and non-native data: Do speakers do what we think they do? In J. Bruhn de Garavito & E. Valenzuela (Eds.), *Selected proceedings of the 10th Hispanic linguistics symposium* (pp. 64-77). Cascadilla Proceedings Project.
- Goldberg, A. E. (2006). *Constructions at work: the nature of generalization in language*. Oxford University Press.
- Goldberg, A. (2013). Constructionist approaches. In T. Hoffmann & G. Trousdale (Eds.), *The Oxford handbook of construction grammar* (pp. 15-31). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780195396683.001.0001>
- Gries, St. Th. & Deshors, S. C. (2014). Using regressions to explore deviations between corpus data and a standard/target: Two suggestions. *Corpora*, 9(1), 109-136.
- Halliday, M. A. K. and Hasan, R. (1990). *Language, context, and text: aspects of language in a social-semiotic perspective*. Oxford University Press.
- Haspelmath, M. & Sims, A.D. (2010). *Understanding morphology*. London: Hodder Education.
- Hay, J. (2001). Lexical frequency in morphology: is everything relative? *Linguistics*, 39(6), 1041-1070. <https://doi.org/10.1515/ling.2001.041>
- Hay J. & Baayen H. (2002). Parsing and productivity. In G. Booij & J. van Marle (Eds.), *Yearbook of morphology 2001* (pp. 203-235). Kluwer Academic Publishers. https://doi.org/10.1007/978-94-017-3726-5_8
- Hay, J. & Baayen, H. (2003). Phonotactics, parsing and productivity. *Italian Journal of Linguistics*, 15, 99-130.
- Hay, J. & Plag, I. (2004). What constrains possible suffix combinations? On the interaction of grammatical and processing restrictions in derivational morphology. *Natural Language & Linguistic Theory*, 22(3), 565-596. <https://doi.org/10.1023/B:NALA.0000027679.63308.89>
- Hulstijn, J. (2005). Theoretical and empirical issues in the study of implicit and explicit second language learning. *Studies in Second Language Acquisition*, 27, 129-140. <https://doi.org/10.1017/S0272263105050084>
- Jarvis, S. (2013). Capturing the diversity in lexical diversity. *Language Learning*, 63, 87-106. <https://doi.org/10.1111/j.1467-9922.2012.00739.x>
- Jiang, N. (2004). Morphological insensitivity in second language processing. *Applied Psycholinguistics*, 25(4), 603-634. <https://doi.org/10.1017/S0142716404001298>
- Kanwit, M. (2015). The role of discourse topic in evidentiality marking: Variable (de)queísmo in Caracas. *eHumanista/IVITRA*, 8, 446-470.
- Kanwit, M. (2017). What we gain by combining variationist and concept-oriented approaches: The case of acquiring Spanish future-time expression. *Language Learning*, 67(2), 461-498. <https://doi.org/10.1111/lang.12234>
- Koizumi, R. & In’ami, Y. (2013). Vocabulary knowledge and speaking proficiency among second language learners from novice to intermediate levels. *Journal of Language Teaching and Research*, 4(5), 900-913. <https://doi.org/10.4304/jltr.4.5.900-913>
- Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Pergamon.
- Kraut, R. (2015). The relationship between morphological awareness and morphological decomposition among English language learners. *Reading and Writing*, 28(6), 873-890.

- Labov, W. (1972). *Sociolinguistic patterns*. University of Pennsylvania Press.
- Langman, J. & Bayley, R. (2002). The acquisition of verbal morphology by Chinese learners of Hungarian. *Language Variation and Change*, 14(1), 55-77.
<https://doi.org/10.1017/S0954394502141032>
- Larsen-Freeman, D. & Long, M. H. (1991). *An introduction to second language research*. Longman.
- Lass, R. (Ed.). (2000). *The Cambridge history of the English language*. Cambridge University Press.
- Lems, K., Miller, L. D., & Soro, T. M. (2010). Using morphemes to learn vocabulary. In *Teaching reading to English language learners: Insights from linguistics* (pp. 91-118). The Guilford Press.
- Lieber, R. (2014): Methodological Issues in Studying Derivation. In: R. Lieber & P. Štekauer (Eds.), *The Oxford handbook of derivational morphology* (pp. 84-94). Oxford University Press.
- Linford, B. G., Long, A. Y., Solon, M. E. & Geeslin, K. L. (2016). Measuring lexical frequency: comparison groups and subject expression in L2 Spanish. In L. Ortega, A. E. Tyler, H. I. Park, & M. Uno (Eds.) *The usage-based study of language learning and multilingualism* (pp. 137-154). Georgetown University Press.
- Lüdeling, A. & Evert, S. (2005). The emergence of productive non-medical -itis. Corpus evidence and qualitative analysis. In S. Kepser & M. Reis (Eds.), *Linguistic evidence. Empirical, theoretical, and computational perspectives*. Mouton de Gruyter.
- Manova, S. (2005). Derivation versus inflection in three inflecting languages. In W. U. Dressler et al. (Eds.), *Morphology and its demarcations: selected papers from the 11th morphology meeting, Vienna, February 2004* (Amsterdam Studies in the Theory and History of Linguistic Science. Series IV: Current Issues in Linguistic Theory, 264) (pp. 233–252). Benjamins.
- Medina-Rivera, A. (1999). Variación fonológica y estilística en el español de Puerto Rico. *Hispania*, 82(3), 529-541.
- Montero-Fleta, B. (2011). Suffixes in word-formation processes in scientific English. *Language for Special Purposes Journal*, 2(2), 3-14.
- Naismith, B., Han, N.-R., Juffs, A., Hill, B. L., & Zheng, D. (2018). Accurate measurement of lexical sophistication with reference to ESL learner data. In K.E. Boyer & M. Yudelson, (Eds.) (pp 259-265). *Proceedings of the 11th international conference on educational data mining*. Buffalo, NY.
- Neubauer, K. & Clahsen, H. (2009). Decomposition of inflected words in a second language: An experimental study of German participles. *Studies in Second Language Acquisition*, 31, 403-405. <https://doi.org/10.1017/S0272263109090354>
- Neumann, F. & Kanwit, M. (2018). New perspectives on automatic and morphophonological alternations: Harmonic processes in two Peninsular varieties of Spanish. *Borealis: An International Journal of Hispanic Linguistics*, 7(1), 97-112. <https://doi.org/10.7557/1.7.1.4150>
- Oxford Dictionaries. (n.d.). Words ending in -sion, -tion, and -cion.
<https://en.oxforddictionaries.com/spelling/words-ending-in-sion-tion-and-cion>
- Plag, I. & Baayen, R. H. (2009). Suffix ordering and morphological processing. *Language*, 85(1), 109-152.

- Plag, I., Dalton-Puffer, C., & Baayen, H. (1999). Morphological productivity across speech and writing. *English Language and Linguistics*, 33, 209-228.
- Quinion, M. (2008). *Affixes: the building blocks of English*. <http://www affixes.org>.
- Rácz, P., Papp, V., & Hay, J. (2016). Frequency and corpora. In A. Hippisley & G. Stump (Eds.), *The Cambridge handbook of morphology* (Cambridge Handbooks in Language and Linguistics, pp. 685-709). Cambridge University Press.
<https://doi.org/10.1017/9781139814720.024>
- Rehak, K. M. & Juffs, A. (2010). Native and non-native processing of morphologically complex English words: Testing the influence of derivational prefixes. In G. Granena et al. (Eds.) *Selected proceedings of the 2010 second language research forum* (pp. 125-142). Cascadilla Proceedings Project.
- Reichl, K. (1982). *Categorial grammar and word-formation: The de-adjectival abstract noun in English*. Series: Buchreihe der Anglia / Anglia Book Series 22.
- Renouf, A. (1987). Corpus development. In J. M. Sinclair (Ed.), *Looking up: An account of the Cobuild project in lexical computing* (pp. 1-40). Collins ELT.
- Riddle, E. (1982). A historical perspective on the productivity of the suffixes -ness and -ity. In J. Fisiak (Ed.), *Historical semantics - historical word-formation* (pp. 435-459). Gruyter Mouton.
- Robinson, P., Mackey, A., Gass, S., & Schmidt, R. (2012/2014). Attention and awareness in second language acquisition. In S. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 247-267). Routledge.
- Roche, T. & Harrington, M. (2013). Recognition vocabulary knowledge as a predictor of academic performance. *Language Testing in Asia*, 3(1), 1-13.
<https://doi.org/10.1186/2229-0443-3-12>
- Rodríguez, G. (2013). Are all L2 learners morphologically insensitive?: L1 influence and WM capacity as mitigating factors. In E. Voss, S.D. Tai, & Z. Li (Eds.), *Selected proceedings of the 2011 second language research forum: converging theory and practice* (pp. 70-81).
- Säily, T. 2014. *Sociolinguistic variation in English derivational productivity: Studies and methods in diachronic corpus linguistics*. Société Néophilologique.
- Schmidt, R. (2012). Attention, awareness, and individual differences in language learning. In W. M. Chan, K. N. Chin, S. Bhatt & I. Walker (Eds.), *Perspectives on individual characteristics and foreign language education* (pp. 27-50). Walter de Gruyter.
- Scott, A. K. (2012). A constructionist account of the Modern Dutch adnominal genitive. In F. Kiefer, M., Ladányi, & P. Siptár (Eds.) *Current issues in morphological theory: (ir)regularity, analogy and frequency* (pp. 83-104). Selected papers from the 14th international morphology meeting, Budapest, 13-16 May 2010.
<https://doi.org/10.1075/cilt.322.05sco>
- Spinner, P. & Thomas, A. J. (2014). L2 learners' sensitivity to semantic and morphophonological information on Swahili nouns. *International Review of Applied Linguistics in Language Teaching*, 52(3), 283-311.
<https://doi.org/10.1515/iral-2014-0013>
- Tomasello, M. (2000). The item-based nature of children's early syntactic development. *Trends in Cognitive Sciences* 4(4), 156-163.
- VanPatten, B. & Williams, J. (2015). *Theories in second language acquisition, 2nd edition*. Routledge.

- Wilkinson, E. (2013). Morphosyntactic variation in American Sign Language: Genre effects on the usage of SELF. In L. Meurant, A. Sinte, M. Van Herreweghe, & M. Vermeerbergen (Authors), *sign language research, uses and practices: crossing views on theoretical and applied sign language linguistics* (pp. 259-284). De Gruyter.
- Winter, B. & Grawunder, S. (2012). The phonetic profile of Korean formal and informal speech registers. *Journal of Phonetics*, 40, 808-815.
<https://doi.org/10.1016/j.wocn.2012.08.006>
- Wu, Z. (2016). Second language processing of English derived words: Effects of L1 typology, morphological awareness, suffix complexity, and L2 proficiency. Dissertation. University of Pittsburgh.
- Zwickly, A. M. & Pullum, G. K. (1983). Cliticization vs. inflection: English N'T. *Language*, 59(3), 502–513.