Corpus-based productivity measures of English *-er* agentives and instrumentals

Daniel R. McCloy
Department of Linguistics, University of Washington

Abstract

This paper investigates the claim that the agentive and instrumental forms of the English *-er* morpheme show different levels of productivity (Derwing, 1976). This study attempts to reproduce Derwing's productivity findings using modern corpus methods. Novel annotations for animacy and agentivity/instrumentality were created on the Brown corpus (Kučera & Francis, 1967). Agentive *-er* is found to be much more frequent than instrumental *-er* (>5× token frequency, >3× type frequency). Productivity is calculated as the probability of observing new types ($\mathscr{P} = n_1/N$, where n_1 is the number of *hapax legomena* and N is the total tokens; see Baayen & Lieber, 1991; Baayen, 1992, 1993). Exponential modeling suggests that the productivity of instrumental *-er* is at least as large as productivity of agentive *-er*, if not slightly larger (contra Derwing's findings). Agentive/instrumental annotations also revealed a number of difficult-to-classify cases. However, productivity values based on the agentive/instrumental split were nearly perfectly mirrored by those based on the animate/inanimate split. This parity arises from the high correlation between the agentivity and animacy categories, and suggests that future studies with larger corpora could safely rely on animacy as a proxy for agentivity.

1 Introduction

It is well known that the English nominalizing suffix *-er* has several meanings: an agentive interpretation (VERB+*er* = one who verbs), an instrumental interpretation (VERB+*er* = a thing with which one verbs), and many others. There is also no doubt that *-er* is a productive morpheme

for both agentive and instrumental uses (cf. *podcaster*, *blogger*, *transponder*, *recycler*). It is less clear that the two forms should count as lexically distinct. Derwing (1976) claims that the agentive and instrumental uses differ in their productivity, on the basis of sentence-completion tests and wug tests on child and adult subjects, suggesting that the two forms are in fact lexically distinct. However, some linguists argue that all *-er* nominalizations are of the same kind (see, e.g., Ryder 1999). This paper reports the results of a corpus study re-examining Derwing's claim about differential productivity of *-er* forms, and discusses some methodological considerations and theoretical implications for measuring and interpreting productivity differences.

1.1 Productivity in Derwing 1976

Derwing (1976) distinguishes between what he calls "formal" productivity and "psychological" productivity. For Derwing, "formal" productivity seems to mean the phonological and semantic similarity of forms across the lexicon (i.e., a measure related to the type frequency of the word form in question), and "psychological" productivity seems to mean the speaker's awareness of a word form and willingness or likelihood to generate new derived forms with it (roughly equivalent to the "unlimited" aspect of Schultink's definition).

Derwing's experiment is an early attempt to quantify his notion of psychological productivity using a between-subjects design, using a method similar to the wug test pioneered by Berko (1958). Based on the results of a sentence-completion task (in which elementary and secondary school students and adults gave morphologically derived forms from real and nonsense stems), Derwing claims that the agentive *-er* morpheme is more productive than the instrumental *-er* morpheme; degree of productivity is quantified as a percentage of respondents who produce the word form in question. A subset of his results relevant to the *-er* morpheme are given in

Table 1; note that there was only one word form tested per category, and Derwing found no attested irregular instrumental form. There were 95 respondents (40 children, 28 adolescents, 27 adults); percentages are the portion of respondents that gave the correct target form.[1]

	Agentive		Instrumental	
Regular	sing/singer	100%	point/pointer	94%
Irregular	cook/cook	71%		
Wug	yurse/yurser	86%	cume/cumer	55%

Table 1: Partial results of Derwing's sentence completion task (reprinted from Derwing 1976). Wug forms were assigned to agentive or instrumental based on the supplied sentence frames.

One problem with Derwing's findings is that his conclusions about psychological productivity (as a speaker's willingness or likelihood of generating new derived forms) seem radically undersupported given that he only tests one regular and one wug stem for each form, opening the door to lexical bias in his results (cf. the discussion of *cume* in Note 1). One could argue that his results reflect properties of the individual words that he tested, rather than the word formation rules that they exemplify. For instance, some evidence shows that word frequency is an important factor in a child's phonological awareness (e.g., Hogan et al., 2011), and by extension may impact a child's ability to morphologically decompose words. Other studies emphasize the effect of lexical neighborhood and age of acquisition on various lexical tasks (e.g., Garlock et al., 2001), further supporting the idea that Derwing's choice of a single exemplar for each morphemic condition (a total of just five word forms) may have biased his findings.

On a related note, it is unclear whether extant forms are even relevant in quantifying productivity under Derwing's definition, since the regular stems he tested (*singer* and *pointer*) can

[1] Although his study includes respondents from three different age groups, it is unclear how differences across groups are supposed to be integrated into a single metric for productivity. Therefore I present Derwing's results collapsed across age groups. The overall trend is mirrored within the elementary and secondary school age groups, but the adult respondents scored at ceiling (>95%) for all items except *cume*, perhaps due to similarity with the irregular agentive *groom* "one who grooms horses".

scarcely be considered "new derived forms" (*pointer* dates from 1500 and *singer* from 1330; OED Online, 2011d,c). This question of how to quantify productivity in terms of "new derived forms" is discussed extensively by Baayen & Lieber (1991), and is addressed in the next section.

1.2 Other approaches to quantifying productivity

As Baayen and Lieber point out, some scholars have defined productivity as the ratio of the number of *possible* forms to the number of *attested* forms, where the number of possible forms is quantified as the number of bases which could potentially combine with the morpheme in question (Aronoff, 1976, p.36, quoted in Baayen & Lieber, 1991). Baayen and Lieber show that this conception of productivity is ill-formed, on the grounds that (1) the number of attested forms does not often correlate well with native speaker intuitions about productivity, and (2) the number of possible forms is in many cases impossible to estimate, especially in circumstances where one productive affix generates forms that are input for another affix. In such cases, they argue, looking at larger and larger corpora leads to *unproductive* affixes approaching a productivity index of 1, while *productive* affixes approach an index of zero (Baayen & Lieber, 1991, p.5).

Baayen and Lieber also critique later developments in productivity theory that emphasize token frequency rather than type frequency, particularly the notion that productivity is related to the ratio of the token frequency of the derived form to that of the base form (e.g., Aronoff, 1983; Anshen & Aronoff, 1988). These "derivation ratios," they point out, vary radically across the spectrum of possible values for both productive and unproductive forms. In other words, there are a number of words bearing productive affixes that happen to be much more frequent than their base forms (e.g., *recently* vs *recent*) while other forms in the same affix show the op-

posite pattern (e.g., *blackly* vs *black*); the same patterns can be seen among unproductive affixes (Baayen & Lieber, 1991).

Regardless of whether derivations based on extant or wug stems are relevant to measuring productivity, the point stands that Derwing's experiment does nothing to characterize the productivity of word formation rules *in general*, i.e., as they apply to a multitude of inputs (attested or otherwise). Baayen and Lieber's solution to this problem is to count *hapax legomena*[2] in a corpus instead of attempting to count new derived forms (Baayen & Lieber, 1991; Baayen, 1992, 1993). Their metric of productivity is the ratio of *hapax legomena* (n_1) to total tokens (N) for a given word formation rule, as seen in Equation 1. This metric is meant to express the likelihood of encountering a new type given the number of tokens so far observed. In other words, affixes that are highly productive ought to apply to a wide variety of bases and thus have a high number of types, many of which will to appear only once in the corpus (thereby giving them a high $\mathscr{P}$value).

(1) $$\mathscr{P} = \frac{n_1}{N}$$

The emphasis on type frequency relates to a definition of productivity that Baayen and Lieber attribute to Schultink: productive word-formation rules are used by speakers and accepted by listeners *unconsciously* and their application is in some sense *unlimited* (Schultink, 1961; Baayen & Lieber, 1991, 9). This definition, Baayen and Lieber point out, explains why conscious misapplication of a word formation rule carries special meaning (e.g., humor, as in a soft drink marketed as *uncola*) and it is precisely the limitations on those word formation rules that make hu-

[2] *Hapax legomena* are forms that appear only once in a corpus.

morous misapplication possible. One advantage of such an approach is that it accounts for the influence of word frequency, which is a serious point of concern for Derwing's data (as discussed above). Of course, a Baayen-style metric of productivity cannot help us sidestep these aspects of Derwing's data, since they come from sentence completion tasks, not from texts or running speech. For this reason, the study described in this paper investigates Derwing's claims using corpus methods.

1.3 Etymology and typology of *-er*

Another problem with Derwing's study is that he seems to have assumed *a priori* that the agentive and instrumental forms of *-er* represent distinct lexical items, i.e., distinct entries in the mental lexica of English speakers. There is nothing terribly controversial about an assertion that homophonous forms are lexically distinct, after all, few would quibble with the idea that *lye* and *lie* are distinct lexemes. However, it seems prudent to remain agnostic as to whether the agentive and instrumental *-er* forms are lexically distinct, and to seek converging evidence (not limited to measures of productivity) to support such an assertion. To that end, a discussion of the etymology of *-er* forms follows.

Regarding the productivity of *-er* forms, the Oxford English Dictionary (OED) seems at least consistent with Derwing's claims, stating that the agentive *-er* can nominalize all verbs except those already lexically specified for other agentive endings (chiefly *-ent* as in *correspondent*, *claimant*; OED Online, 2011a).[3] Whether the agentive *-er* is truly applicable to *all* verbs is an empirical matter that need not concern us here; it suffices that the OED counts the agentive

[3] *-er* is of course highly productive as a denominal affix for both agentive and instrumental meanings, though the OED does not comment on its productivity as such.

form as highly productive, and (as mentioned in the introduction) recent neologisms confirm this (e.g., *blogger, podcaster*). But merely looking for the presence or absence of neologisms will shed no light on the supposed difference in productivity between the two, since similar neologisms testify to the productivity of the instrumental form as well (e.g., *transponder*, *recycler*).

Insight into their lexical distinctiveness is similarly elusive if we turn to the etymology of *-er* forms. The *-er* suffix denoting profession allegedly derives from Old English *-ęre* (Germanic **-ârjo-z*) and has thus been a part of English since its inception (examples of such ancient forms being *singer*, *hatter*, *bowyer*, OED Online, 2011a). This usage is thought to have expanded during the Middle English period into a few non-professional (but still agentive) words such as *bencher* or *cottager*, and further expanded in the Modern English period to two additional meanings: a "resident, origin or native" usage in words like *outsider*, *southerner*, *villager*, or *Londoner*, and a "things and actions" usage in such words as *header*, *double-decker*, and *fiver* (OED Online, 2011a). A further two classes of words were borrowed into Middle English from Old French and Latin and in the modern language are indistinguishable from *-er* forms of Germanic origin. These also chiefly involve occupations (e.g., *draper*, *mariner*, *grocer* and the like from Old French, and *geographer*, *chronologer*, *astronomer* and the like from Latin) (OED Online, 2011b). Notably absent from the OED's discussion is any specific mention of *graters*, *slicers*, or *bill-acceptors*, i.e., instrumental *-er* forms, except in the brief mention of "function" in a note about orthography: "when the sense is purely agentive, without any added notion such as that of office, trade, or profession, function, etc., *-er* is often used [instead of *-or*]" (OED Online, 2011a).

This lack of explicit mention of instrumental *-er* in the OED underscores the question of whether it makes sense to treat the agentive and instrumental forms as distinct lexemes, though based

on the discussion above we might guess that the agentive interpretation is somewhat older in the language, and the instrumental interpretation emerged through semantic extension. Such speculations aside, however, there is certainly no consensus establishing a lexical difference between the agentive and instrumental forms. As mentioned above, Marchand's monumental work on English derivational morphology groups the agentive and instrumental interpretations together under one entry, with the rather broad meaning of "someone or something connected with what the basis denotes," and likewise describes both forms with the word "agent," saying "[d]everbal derivatives are chiefly agent substantives with the meaning 'animate or inanimate substantive denoting the performer of an action, occasional or habitual'" (Marchand, 1969, 273). Baayen & Lieber (1991) also do not distinguish among the various forms of *-er* in their corpus study of English morphological productivity, and Ryder (1999) argues that all *-er* nominalizations are of the same kind, its application constrained on purely pragmatic grounds and its variability due to the inherent ambiguity of *-er* nominals (cf. the discussion in Bauer, 2001, 199–203).

If, then, the two forms are not lexically distinct (i.e., do not comprise separate entries in the mental lexica of English speakers) then any difference in productivity must reflect something else about language (e.g., the relative type or token frequency of agents vs. instruments). What then, to make of Derwing's results? His finding could be dismissed given that it is based on one word form per condition, and pays little or no attention to issues of lexical frequency, neighborhood density, or phonotactic plausibility of the nonce forms. This is not to be overly critical of Derwing, except to say that psycholinguistic methodology has come a long way since 1976. Simply dismissing Derwing's study brings us no closer to answering the question of whether a

genuine productivity difference exists between agentive and instrumental forms of *-er*, or how to interpret such a difference if we find one.

2 Methods

To address the question of whether agentive and instrumental *-er* show differential productivity, the Brown Corpus (Kučera & Francis, 1967) was selected because of its broad sampling of linguistic domains and manageable size for the hand-annotation required for this study. Target forms were defined as orthographic wordforms ending in {*er*, *ers*, *ar*, *ars*, *or*, *ors*}, and were extracted via python script using the Natural Language Toolkit (Bird et al., 2009). This script generated a list of 20321 tokens of 2112 types. The type list was manually edited to remove forms that were deemed universally spurious, e.g., types like *finger* were rejected because they will never mean "one who fings" or "a thing with which one fings." In contrast, types like *poster* were retained because they could conceivably mean "one who posts" (e.g., to a forum or blog). This initial step reduced the data to 5102 tokens of 1196 types. These remaining tokens were examined within their occurrence context and coded for one of five *-er* meanings, given in Table 2.

Classification	**Example**	**Tokens**
Agentive	*voter* = one who votes	3964
Instrumental	*starter* = device used to start a car	765
Appointee	*juror* = one appointed to a jury (not "one who jures")	58
Locative association	*treasurer* = one associated with a treasury (not "one who treasures")	45
Abbreviation	*homer* = "home run" (baseball jargon)	74
None of the above	all cases of *sewer*, which turned out to all be sanitation-related, but were not excluded initially in case they turned out to be tailoring-related	193

Table 2: Classification scheme for *-er* types

In most cases a context of five words preceding and following was sufficient to determine proper coding, though in some cases unclear tokens were marked for re-extraction from the corpus with a longer stretch of context to disambiguate interpretation. This manual curation of the data resulted in a further reduction to 4729 tokens of 1028 types, of which 3964 tokens comprising 803 types were coded as agentive and 765 tokens comprising 263 types were coded instrumental. It is worth noting the strong asymmetry in frequency between the agentive and instrumental types, which will be discussed further below.

An important hurdle in conducting this study was handling a lot of "corner cases" that arose in the coding process. For example, it is quite clear that a screwdriver is an instrument, but a number of cases were not so clear. Chemical terms like *binder*, *softener* and *emulsifier* do not seem to be obviously instruments, since their binding or emulsifying action is a consequence of their inherent chemical nature, rather than something that they do strictly as a result of human action and as a means to some human end. A variety of other non-animate entities seem more like agents than instruments, e.g., a hot day (*scorcher*), the first game of a season (*opener*), an economic trend (an *indicator* of change), or a strip-mall development (a *contributor* to growth). To clarify this picture, each token was additionally coded for animacy. Overall, 137 inanimate agent tokens were noted, versus 3827 animate agents, 765 inanimate instruments, and a single animate instrument (see Table 3). To investigate the effect of these "hard cases" on measures of productivity, the measures were calculated both on the basis of agentive/instrumental split (thereby including the inanimate agents with the other agentives), as well as on the basis of the animate/inanimate split (thereby including the inanimate agents with the other inanimates, i.e., the instrumentals).

Classification	Example	Tokens
animate agent	people and animals, e.g., *banker*	3827
animate instrument	a rider and his horse that were "used as a marker"	1
inanimate agent	chemical agents, e.g., *emulsifier*	137
inanimate instrument	*calculator*, *elevator*, *typewriter*	765

Table 3: Additional codes for *-er* types: ±animate

3 Results

The productivity values calculated from the Brown Corpus are given in Table 4. Three features bear noting in the results of the productivity analysis. First, as mentioned above, the instrumental usage of *-er* is drastically lower in frequency than the agentive usage, in both token and type frequency.

Class	Types	Tokens (N)	Hapax legomena (n_1)	Productivity ($\mathscr{P}$)
agentive	803	3964	400	0.1009591
instrumental	263	765	128	0.1675393
animate	769	3828	384	0.1003659
inanimate	299	901	146	0.1622222

Table 4: Productivity of *-er* nominals calculated on the Brown corpus

Second, it would appear that the instrumental usage exhibits *greater* productivity than the agentive, in direct contradiction to Derwing's findings. However, consider Figure 1, which shows the value of $\mathscr{P}$ as a function of number of tokens thusfar encountered as the corpus is read sequentially. It is clear that although the agentive productivity calculation finds a stable asymptote somewhere around the 1000th token, the instrumental calculation appears never to reach a stable value — it is still downtrending when the end of the corpus is reached, suggesting that a larger corpus is needed to fully gauge the productivity of instrumental *-er* forms. Finally, it is striking how similar the agentive/animate lines are to one another, and similarly for the instru-

mental/inanimate lines. This in itself is a significant methodological finding, since (as discussed above) agentive/instrumental judgments are substantially more difficult to make than judgements of animacy.

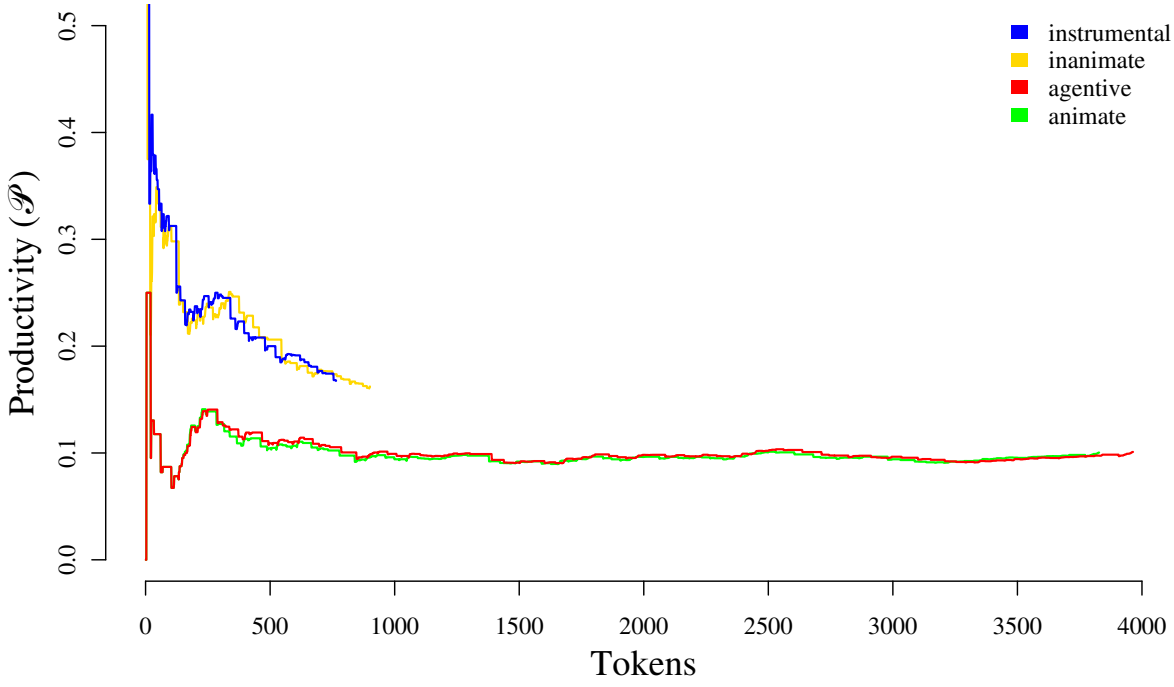

Figure 1: Plot of productivity ($\mathscr{P}$) for agentive, instrumental, animate, and inanimate *-er* forms

4 Discussion

Considering the trajectory of the instrumental line in Figure 1, it is unclear whether it will asymptote above, even with, or below the agentive line, thus we cannot be certain whether this study supports or refutes Derwing's original finding. However, the trend of the instrumental points is quite well modeled by an exponential curve, as seen in Figure 2, suggesting that in a larger corpus the instrumental line would fall even with or slightly above the agentive line (thus re-

futing Derwing's findings). Nonetheless, in the absence of a larger corpus with the necessary annotations, our conclusions about the relative productivity must be somewhat tentative. What is clear is that there is a radical difference in frequency between agentives and instrumentals (a more than fivefold difference in token frequency and more than threefold difference in type frequency). These differences in frequency are almost certainly relevant to Derwing's findings, as mentioned above.

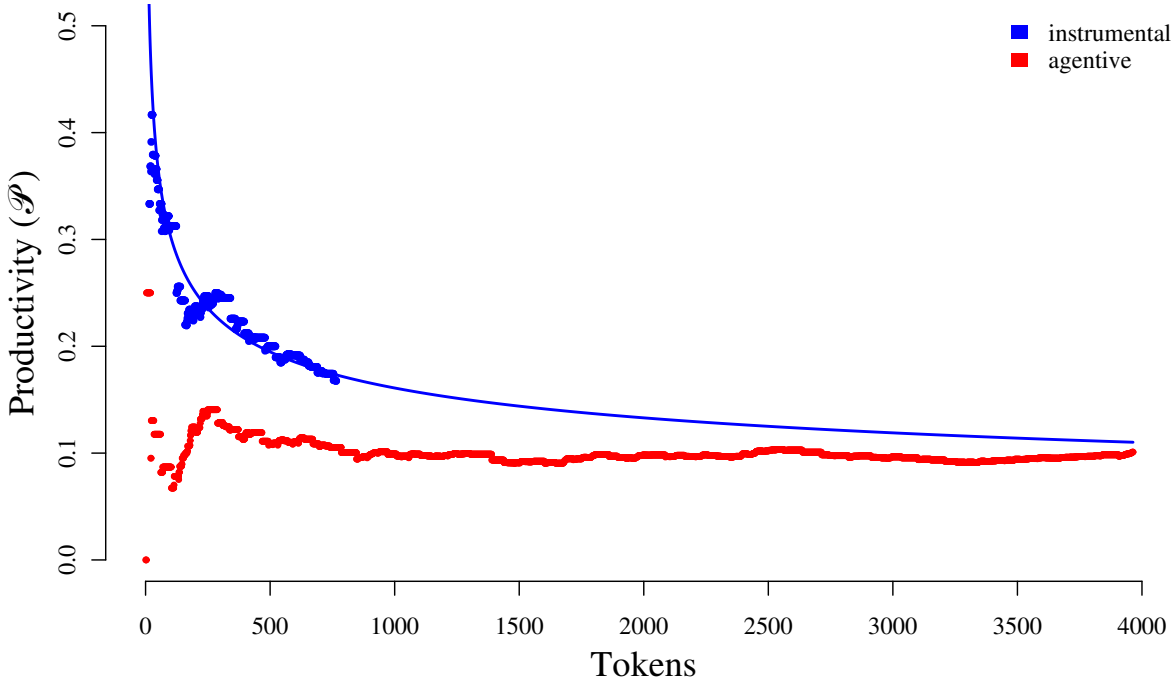

Figure 2: Exponential fit of productivity values for instrumental *-er* forms

These issues underscore one advantage of corpus-based methods for productivity research, and one shortcoming. The advantage is that variation due to specific lexemes or specific subjects is minimized or can be more easily corrected for; the disadvantage is the need for high-quality annotation that in cases such as this were too complicated and subtle to be generated automat-

ically. As such, a difference in the productivity of agentive and instrumental *-er* forms cannot be definitively established until such annotations are available on more extensive corpora. That said, this study's finding that animacy is a fair proxy for agentiveness in sufficiently large corpora does make that annotation task somewhat more achievable.

References

Anshen, F., & Aronoff, M. (1988). Producing morphologically complex words. *Linguistics*, *26*, 641–655.

Aronoff, M. (1976). *Word formation in generative grammar*. Cambridge, MA: MIT Press.

Aronoff, M. (1983). Potential words, actual words, productivity and frequency. In S. Hattori, & K. Inoue (Eds.) *Proceedings of the XIIIth International Congress of Linguists*, (p. 163–171). Tokyo.

Baayen, R. H. (1992). Quantitative aspects of morphological productivity. In G. E. Booij, & J. van Marle (Eds.) *Yearbook of morphology 1991*, (p. 109–149). Dordrecht: Kluwer Academic Publishers.

Baayen, R. H. (1993). On frequency, transparency and productivity. In G. E. Booij, & J. van Marle (Eds.) *Yearbook of morphology 1992*, (p. 181–208). Dordrecht: Kluwer Academic Publishers.

Baayen, R. H., & Lieber, R. (1991). Productivity and English derivation: A corpus-based study. *Linguistics*, *29*(5), 801–844.

Bauer, L. (2001). *Morphological Productivity*. Cambridge: Cambridge University Press.

Berko, J. (1958). The child's learning of English morphology. *Word*, *14*, 150–177.

Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python*. Sebastopol, CA: O'Reilly.

Derwing, B. L. (1976). Morpheme recognition and the learning of rules for derivational morphology. *Canadian Journal of Linguistics/Revue Canadienne de Linguistique*, *21*, 38–66.

Garlock, V. M., Walley, A. C., & Metsala, J. L. (2001). Age-of-acquisition, word frequency, and neighborhood density effects on spoken word recognition by children and adults. *Journal of Memory and Language*, *45*(3), 468–492.

Hogan, T. P., Bowles, R. P., Catts, H. W., & Storkel, H. L. (2011). The influence of neighborhood density and word frequency on phoneme awareness in 2nd and 4th grades. *Journal of Communication Disorders*, *44*(1), 49–58.

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

Marchand, H. (1969). *The categories and types of present-day English word-formation: A synchronic-diachronic approach*. Handbücher für das Studium der Anglistik. München: Beck, 2 ed.

OED Online (2011a). -er, suffix1. http://www.oed.com/view/Entry/63869.

OED Online (2011b). -er, suffix2. http://www.oed.com/view/Entry/63870.

OED Online (2011c). pointer, n. http://www.oed.com/view/Entry/146629.

OED Online (2011d). singer, n.1. http://www.oed.com/view/Entry/180116.

Ryder, M. E. (1999). Bankers and blue-chippers: An account of -er formations in present-day English. *English Language and Linguistics*, *3*(2), 269–297.

Schultink, H. (1961). Produktiviteit als morfologisch fenomeen. *Forum der Letteren*, *2*, 110–125.