Productive argument selection: Is lexical semantics enough?

AMIR ZELDES

Abstract
This paper is concerned with the ability of lexical semantic classes to predict the behavior of argument structures in productively selecting novel lexical material. It will be argued that while lexical semantic classes can explain a great deal about the spectrum of potential arguments and resulting entailments for those arguments, constructions which are synonymous from a formal semantic perspective exhibit significant differences in their empirically observed propensity for argument realization. Using corpus data and methods from Baayen's morphological productivity paradigm, I will show that such differences are not restricted to lexicalized, collocational preferences, but also extend to the readiness with which different constructions admit novel arguments.

Keywords: *productivity; argument selection; syntax; semantic classes; specific thematic roles; argument realization; usage-based grammar*

1. Introduction

Argument realization stands at an interface between linguistic and pragmatic knowledge in the description of language usage.[1] On the one hand, the PATIENT we choose for a verb like *drink* depends on our communicative needs and our experience with drinkable and undrinkable things: we often drink *water*, leading to that noun commonly occurring as the object of *drink*. Another fact of world knowledge is that we generally drink liquids, corresponding to a conceptual semantic class [+liquid] (going back to Katzian semantics, cf. Katz and Fodor 1963, Jackendoff 1990:51-55), which can be used to extend *drink* to previously unused objects which are members of that class. This leads to the formulation of lexical semantic representations which contain more specific argument roles than merely PATIENT, either through decompositional semantics as in (1), taken from Jackendoff (1990:53), or using explicit reference to 'individual thematic roles' (cf. Dowty 1991), as in the HPSG representation in (2) adapted from Pollard and Sag (1994).

(1)
$$\left[\begin{array}{l} \text{drink} \\ \text{V} \\ \text{— } \langle \text{NP}_j \rangle \\ [_{\text{Event}}\ \text{CAUSE}\ ([_{\text{Thing}}\ ]_i,\ [_{\text{Event}}\ \text{GO}\ ([_{\text{Thing}}\ \text{LIQUID}]_j, \\ \qquad\qquad [_{\text{Path}}\ \text{TO}\ ([_{\text{Place}}\ \text{IN}\ ([_{\text{Thing}}\ \text{MOUTH OF}\ ([_{\text{Thing}}\ ]_i)]\ )]\ )]\ )]\ )] \end{array}\right]$$

(2)

$$\begin{bmatrix} \text{PHON} & \langle \textit{drink} \rangle \\ \text{SYNSEM} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \textit{category} & \\ \text{HEAD} & \textit{verb} \\ \text{SUBCAT} & \langle \text{NP[nom]}\boxed{1}, \text{NP[acc]}\boxed{2} \rangle \end{bmatrix} \\ \text{CONTENT} & \begin{bmatrix} \text{RELN} & \textit{drink} \\ \text{DRINKER} & \boxed{1} \\ \text{DRUNK} & \boxed{2} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

The latter representation is problematic if we wish to use semantic classes for a predictive model of argument realization, since trivially, a role definition DRUNK or for that matter a semantic class like [+drinkable] will by definition capture exactly the objects of the verb *drink* (this has been called the 'individual-thematic-role escape hatch', cf. Dowty 1991:551). In either case, non-membership in the semantic class or role representation can be used to account for the infelicity of (3).

(3) # *John drank the train*$_{\text{[-liquid]}}$

On the other hand, whether we choose to designate a certain real world event as a case of drinking, some other near synonym or another predicate altogether is in part due to linguistic convention. In English we would more often say we *eat*, rather than *drink soup* (though the latter might also be possible), but in Japanese the equivalent of *drink* is preferred for *soup* (4). Similarly, in Japanese one drinks medicine (including pills), as in (5), whereas in English we *take medicine*.

(4) *suupu o nonda*
soup ACC drank
'(I) drank soup.'

(5) *kusuri o nonda ka?*
medicine ACC drank QST
'Did (you) take your medicine?'

While the semantic class [+liquid] does not capture the preference of *eat* rather than *drink* for the ingestion of *soup*, this can be seen as a case of language-specific lexicalization, and is therefore of limited interest in assessing the capacity of semantic classes to explain productive argument choice in language use.[2] However, what I will suggest in this article is that there are also idiosyncratic effects in the selection of novel, non-lexicalized arguments, or more particularly, that some argument structures are more or less available for novel argument realization in a way that is not predictable from conceptual semantic classes of the type [+liquid], however their scope is defined.

In order to demonstrate this, I will use the following reasoning. If semantic classes are reliable predictors for the spectrum of novel arguments that is realized in an argument slot, then there must be some constructions taking exactly the same argument class (otherwise semantic classes reduce to individual thematic roles and their predictiveness becomes circular). If we accept that two constructions take the same semantic class of arguments, then they must behave in 'the same way' to some extent. Three possible ways in which constructions can show similar selectional behavior are:

i. Similar frequency – the constructions are selected equally often to describe the same kind of real world events and/or the same arguments are selected as often in each construction.
ii. Similar vocabulary – the constructions exhibit the same argument members of the semantic class, regardless of relative frequency.
iii. Similar extensibility – the constructions are equally available for use with novel members of the class, not experienced before in that slot by a speaker.

While it may come as little surprise that similar constructions may violate i., since synonym or near-synonym sets will often have a marked member which is less frequent, and ii., since they may lexicalize different preferred arguments as shown above, I will show that iii. also does not hold for argument selection. In fact, significant differences in the extensibility of slots appearing to take the same class permeate usage data.

With this goal in mind, the next section briefly discusses the operationalization of the criteria in i.-iii. in the vein of Baayen's (1992, 2001, 2009 i.a.) morphological productivity paradigm as it applies to syntactic argument selection. Section 3 presents several corpus-based case studies of candidates for synonymous argument classes in English and German, in which distinct argument selection processes reveal significant, idiosyncratic differences in usage. It is also shown that translational pairs from these languages also differ significantly despite comparable meanings. Section 4 discusses the empirical findings and proposes a cognitively founded Hebbian model for the representation of argument class usage in the mental lexicon that aims to reconcile the observed differences with the evidence for semantic classes, and Section 5 draws the conclusion.

2. Baayen's morphological productivity paradigm and novel argument selection

In morphology, the phenomenon of different sized vocabularies for competing constructions has been discussed extensively under the heading of morphological productivity (see Bauer 2001 for a detailed overview). Some morphological processes, such as affixation of a certain morpheme, are more common than other comparable processes (i.e. have more tokens in corpus data), exhibit a wider variety of distinct forms (i.e. types) in corpora or lexica, or possibly have a higher probability of generating novel forms, as assessed either by elicitation experiments or the appearance of corpus items deemed novel or simply observed to be rare. These aspects of productivity can, but need not correspond: for example, the Latinate English nominalization suffix *-ment* appears in a wide variety of words, but is hardly used to create new words at present (cf. Plag 2006:540).

Since it is arguably impossible to determine for each case we encounter whether it represents a novel type (neither for hearers nor for speakers of a given form, barring deliberate conscious coining processes, cf. Bauer 2001:34-43), there has been much work on the development of criteria to determine the extent of productive use for morphological processes. Beyond qualitative approaches, which attempt to find categorical constraints limiting processes to a certain class of bases (e.g. phonological, morphological, semantic or even etymological constraints, such as restriction to Latinate stems in English), there has been much quantitative work on determining the extent to which usage within the conceivable productive domain actually takes place. The most influential paradigm is found in the works of Harald Baayen and colleagues (e.g. Baayen and Lieber 1991, Baayen 1992, 2001, 2009), who have formulated a variety of productivity measures based mostly on the token frequency, denoted *N(C)* for some category *C*, the type frequency, denoted by *V* (for vocabulary), and the prevalence of rare types, particularly items with a frequency of 1 or *hapax legomena*, whose count in a sample corpus is denoted by *V1* (in general, *Vm* can be used to refer to vocabulary with the frequency *m*, see more below). Use of the latter quantity is justified by the heuristic assumption that neologisms, being novel, should occur only once, though it is clear that neologisms are sometimes coined and used a few times in succession, and that some hapax legomena occur only once in a corpus by chance, despite being lexicalized forms familiar to most speakers and/or being listed in lexica. Baayen uses the proportion of hapax legomena in a sample of *N(C)* tokens as an estimate for the likelihood of the occurrence of the next hapax legomenon, and designates this proportion as the productivity measure $\mathscr{P}$:

[1] $$\mathscr{P} = \frac{V1}{N(C)}$$

To briefly illustrate the utility of these measures, we can consider how the above measurements behave in a sample of German adjectives formed with the suffixes *-bar*, equivalent to English '-able' (e.g. *machbar* 'doable'), and *-sam*, akin to English '-some' (e.g. *biegsam* 'flexible'), with various meanings.[3] While *-bar* is intuitively felt to be highly productive, *-sam* is considered hardly productive or not productive at all (cf. Evert and Lüdeling 2001). Charting the development of *V* with growing sample size *N(C)* in Figure 1A plots a so-called vocabulary growth curve (VGC). The VGC shows that the vocabulary of *-bar* grows much more quickly, as expected, and that *-sam* quickly reaches a plateau due to the exhaustion or near exhaustion of the types in which it is used.[4] For both processes, *V* grows increasingly slowly as the sample grows, since it becomes progressively difficult to find previously unseen forms. The measurements turn out to be quite robust (cf. similar results for these suffixes in Evert and Lüdeling 2001), at least within the same genre, as seen in Panel B, which shows a 95% confidence interval for both processes based on 10 equal sized samples for each process. It is also possible to use statistical models to estimate the asymptotic limit of each VGC, such as parameterized finite Zipf-Mandelbrot models (fZMs, Evert 2004), which can be evaluated based on their fit with the data and tested for their predictiveness using further data. To calculate such a model we require a sample frequency spectrum containing *Vm* for every attested *m*, that is the size of the vocabulary with a frequency of 1 (hapax legomena), 2, 3… *m* up to the highest frequency attested in our data (see ibid. for details). Once the parameters of the optimal model function have been found, its limit of *V* can be used as an estimate for the total vocabulary size, designated by *S*:

[2] $$S = \lim_{N(C)\to\infty} V$$

Mathematically *S* corresponds to the maximal value of *V* we can expect to find if we keep reading texts from the same genre for an unlimited amount of time. This limit is given graphically as a horizontal dashed line for the *-sam* data in Panel A below.

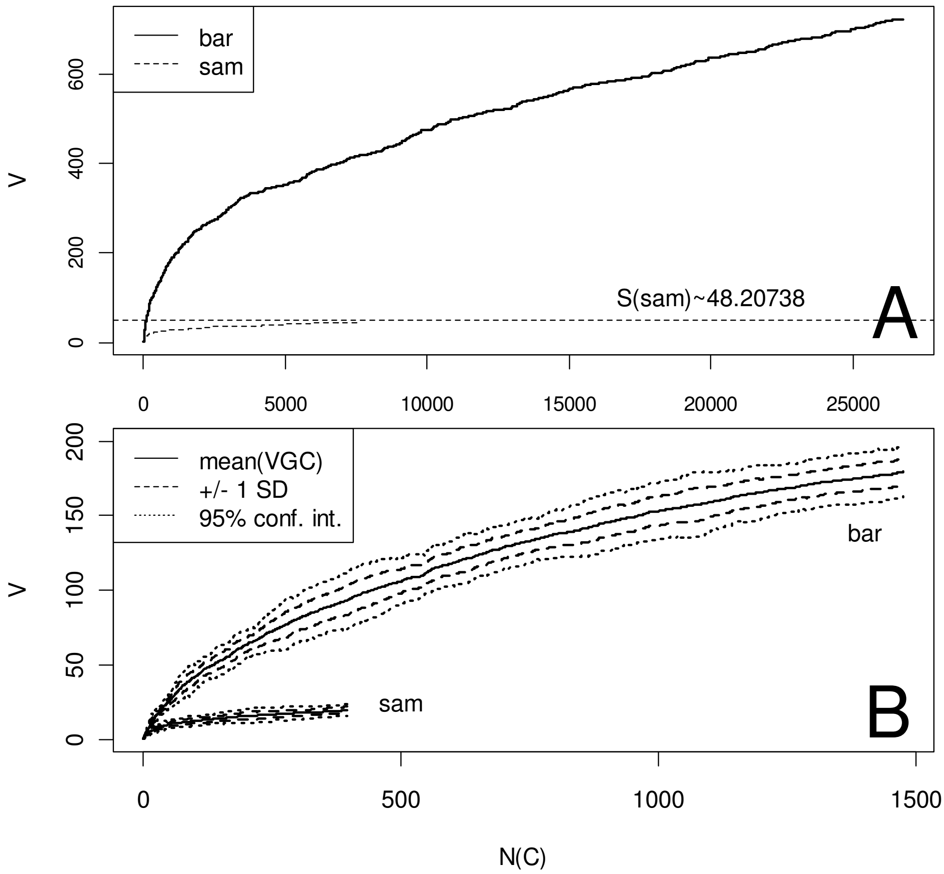

Figure 1. Vocabulary growth curves (A), standard deviations and 95% confidence intervals (B) for *-sam* and *-bar*. The line S(sam) gives the asymptotic vocabulary limit for *-sam*, predicted by an fZM model.

Comparing these measures at a fixed sample size can thus give an empirical correlate for our intuitions about productivity. The fact that vocabulary size (how large a vocabulary has been realized to date) and growth rate (how often do we expect novel items at this point) are not necessarily related can be illustrated by considering a third adjective derivation process in German, suffixation of *-lich* as in *erhältlich* 'obtainable', which exceeds *-bar* in realized vocabulary, but has a much flatter slope (Figure 2A). The difference can be quantified using the measure $\mathscr{P}$, showing that the proportion of hapax legomena for *-lich* is much lower, which implies that most of its types are in fact repeated, and unlikely to include productively formed neologisms. The black circles in Figure 2B show that *-lich* is closer to *-bar* in realized vocabulary, but closer to *-sam* in its generation of unique forms. However, since it becomes more difficult to find hapax legomena at a larger sample size, a fairer comparison should be drawn at an equal sample size for all processes (cf. Gaeta and Ricca 2006 and Säily 2011 for discussion), marked by the grey dashed vertical line in Panel A. This would result in the grey circles in Panel B, which reveal that *-bar* generates more types than *-lich* at an equal sample, and only loses out to *-lich* on account of its lower frequency (compare the length of the *-bar* and *-lich* curves in Panel A). It is thus evident that productivity is a multidimensional phenomenon and different aspects of it (particularly frequency, realized vocabulary and extensibility) can develop independently along individual paths with the growth in sample size.

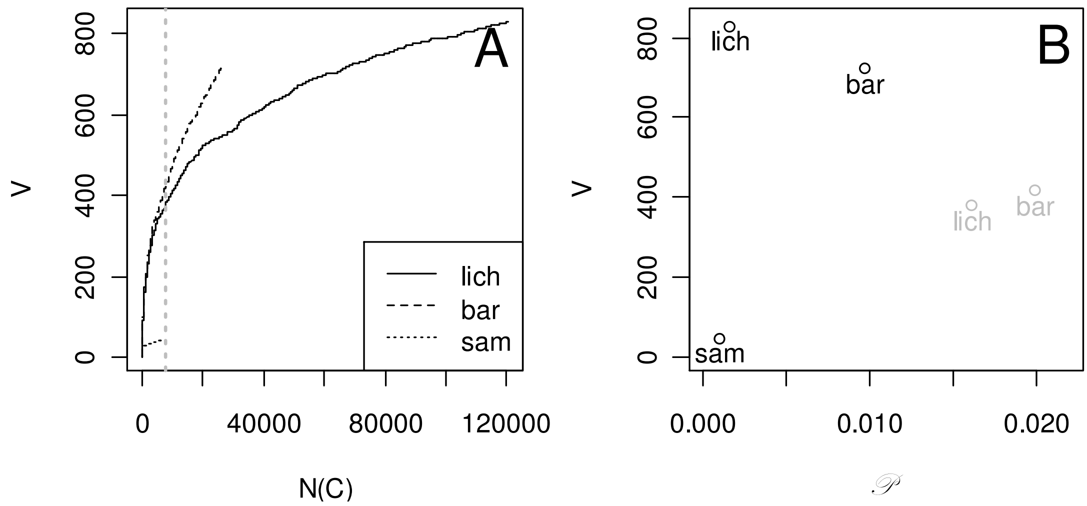

Figure 2. Vocabulary growth and measurements of $\mathscr{P}$ and V for *-bar*, *-sam* and *-lich*.

With this quantitative methodology to evaluate morphological processes in mind, it is now time to consider whether the same techniques can be applied to the classes of lexemes taken by syntactic argument slots. Are there more repetitive object positions? Are some verbs quantitatively more "collocation-happy" in their argument realization than others? In order to measure and estimate frequency, V, $\mathscr{P}$ and S, we need operationalizations for tokens, types and hapax legomena. Kiss (2007), in a study demonstrating that prepositional phrases taking articleless noun phrases as arguments are not a closed class, but a productive construction in German, operationalizes prepositional argument tokens and types based on the identity of the head noun. This approach draws an analogy between morphological affixes in Baayen's methodology and lexical heads in syntax, while the place of morphological bases is taken by the lexical head of the argument phrase. In other words, in both cases the invariable part of the construction in question is seen to identify a potentially productive process, while the variable part is seen as a selected argument:

	Morphology		**Syntax**
Invariable	Process	:	Process
	(e.g. affix requiring a base)		(a construction with one(?)[5] slot)
Variable	Morphological base	:	Lexical head

Using this type definition, we can test some intuitions about the sizes of argument classes. For example, it is my intuition that a verb like *drink* has a reasonably productive class of objects, but it is not as productive (at least in English) as *eat*: I imagine one speaks of drinking the same things quite often (*water*, *coffee*…), whereas the class of edible foods is probably much more varied in its usage (*bread*, *salad*, *sushi*…). By contrast, a verb like *incur* is probably rather repetitive and conventionalized, immediately bringing to mind collocated objects like *costs*, *problems* or *losses* and little else. In order to confirm these intuitions, we can examine vocabulary growth in a large corpus. In the following, data for English will be extracted from the UKWaC corpus (approx. 2.25 billion tokens of Web data) and for German from the DEWaC corpus (approx. 1.63 billion tokens of Web data). Both corpora have been tagged for part-of-speech and lemmatized using the TreeTagger (Schmid 1994); see Baroni et al. (2009) for more details. Figure 3 shows the results for the direct object slot of some English verbs.[6]

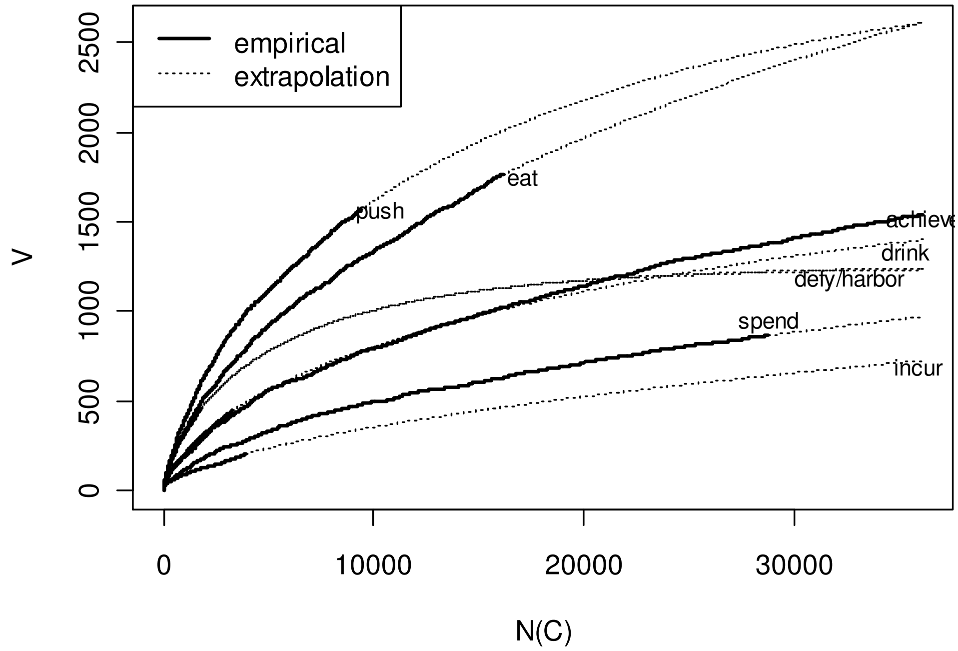

Figure 3. Vocabulary growth curves and fZM extrapolations for the direct objects of some English verbs. Extrapolations are less reliable with increasing distance from empirical data.[7]

Though the projected development of vocabulary according to the statistical model's extrapolation may not be reliable, differences are obvious and significant already within the empirical data for an equal sample size. Table 1 gives rankings for the different verbs according to four criteria: frequency, observed object vocabulary *V*, projected object vocabulary *S*, and the estimated probability of novel types based on $\mathscr{P}$ at an equal sample of 1000 object tokens from each verb.

Table 1. Productivity rankings for verbal object selection according to different criteria.

rank	frequency	*V*	*S*	$\mathscr{P}_{N(C)=1000}$
1	*achieve*	*eat*	*eat*	*push*
2	*spend*	*push*	*achieve*	*eat*
3	*eat*	*achieve*	*incur*	*harbor*
4	*push*	*spend*	*push*	*defy*
5	*incur*	*drink*	*spend*	*achieve*
6	*drink*	*harbor*	*drink*	*drink*
7	*harbor*	*defy*	*harbor*	*spend*
8	*defy*	*incur*	*defy*	*incur*

As we can see, the verbs differ in all aspects and each ranking is different: transitive *achieve* is more frequent than *eat* in this dataset, but *eat* still reaches a larger vocabulary. *Eat* has more objects than *drink*, as expected. *Spend* has a larger realized vocabulary than *incur*, but the statistical model predicts that we will run out of objects for *spend* before we do for *incur*. Finally, the likelihood of encountering novel types after an equal sample of 1000 tokens sees *push* at the top, and the otherwise rather low ranked *harbor* and *defy* are surprisingly liable to innovate (the latter two verbs are very similar in all aspects of productivity).[8]

Thus productivity measures seem to perform rather similarly for argument selection and morphological word formation. They correspond to intuitively understandable differences (more edible objects than drinkable ones, *spend*, which is rather limited to objects referring to time and currency, runs out of types quickly, etc.), and they confirm the multidimensional nature of productivity in the syntactic domain. However, so far we have looked at verbs taking rather different object classes, so that differences can be explained by means of semantic classes. We now turn to search for intuitive differences between argument structures selecting the same or similar classes, and to see whether these too can be captured by the above measures.

3. Differential productivity in synonymous argument slots

With productivity measures to quantify the behavior of argument selection processes at hand, we can examine some concrete cases of synonymous or nearly synonymous argument structures using corpus data (taken again from the UKWaC and DEWaC corpora). The most intuitive point of entry for a comparison of two selectional slots is to examine the realized direct objects of near-synonym verbs.[9] Figure 4 shows that it is possible to find verbal object slots with similar behavior for near synonyms such as *hate*, *dislike*, *despise*, *detest*, *loathe* and *abhor* (Panels A1-A2), which do not differ significantly except in frequency.[10] In fact, the most frequent object noun lemmas of these verbs also overlap substantially:

Table 2. Top 5 objects for *hate* and semantically similar verbs. Recurring objects are italicized, those shared by 5 members or more are underlined.

hate		*despise*		*abhor*		*loathe*		*dislike*		*detest*	
<u>*people*</u>	434	<u>*people*</u>	51	<u>*people*</u>	5	stranger	19	<u>*people*</u>	89	<u>*people*</u>	16
<u>*thing*</u>	116	<u>*thing*</u>	38	crime	5	<u>*people*</u>	10	<u>*thing*</u>	30	convention	6
man	101	other	28	vacuum	4	yourselves	7	practice	14	*man*	5
woman	59	*man*	18	idol	4	<u>*thing*</u>	5	aspect	11	dictator	5
child	57	*woman*	14	form	4	bureaucrat	4	cat	10	<u>*thing*</u>	4

Panels B1-B2 demonstrate that unrelated verbs show the intuitively expected differences based on world-knowledge between e.g. *eat* and *drink*: more varied objects occur in the corpus, among other things, since more varied foods are consumed in daily life. Rows C and D show cases of differences that seem less motivated by any conceptual or real world differences.

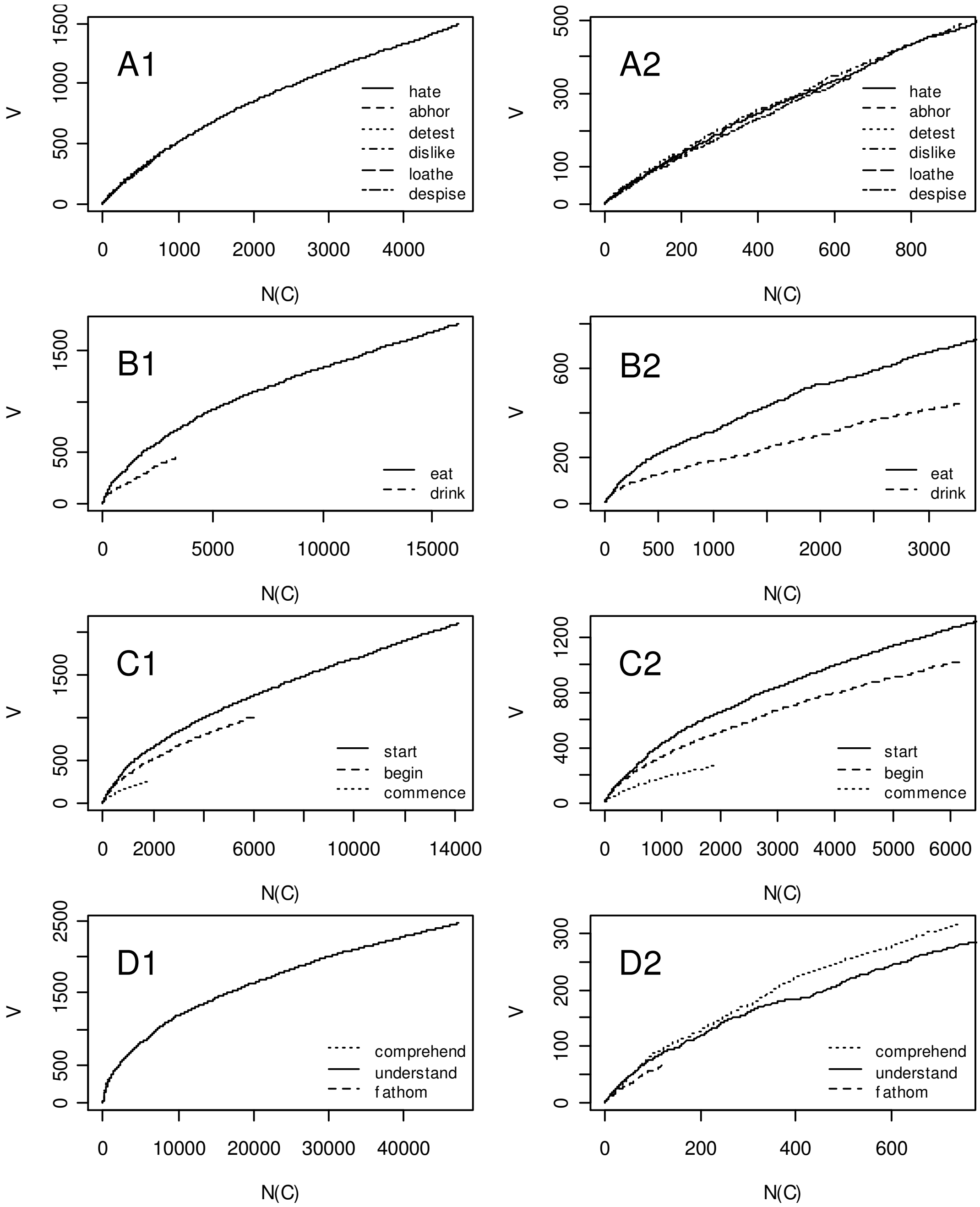

Figure 4. Vocabulary growth curves for English verbal objects. In each row, the panel on the left is scaled to show the entire data while the panel on the right is scaled to fit the second-most frequent category, to better show the behavior of the nearest competitors.

The verbs *start*, *commence* and *begin* exhibit, for the largest common sample size, very significant differences in vocabulary, frequency and vocabulary growth rates (all at $p<0.0005$ in a χ^2 test of equal proportions) despite being interchangeable in most contexts *salva veritate*. Their most common objects also overlap substantially, as shown in Table 3.[11] Similarly *understand*, *comprehend* and *fathom*, show very significant differences in frequency and vocabulary (p<0.0005) and marked differences in vocabulary growth rates (p<0.001). For *comprehend* and *understand* the most frequent objects are still similar (though not quite as much), but as far as this relatively small dataset goes, *fathom* stands apart. However arguably, all arguments in the table for each of these three verbs are compatible with the other two as well.

Table 3. Top 5 objects for the *start* and *understand* sets of verbs. Recurring objects are italicized, those shared by all members are underlined.

start		**begin**		**commence**	
proceeding	527	*operation*	497	*proceeding*	540
negotiation	325	*negotiation*	381	*operation*	268
fire	278	discussion	285	*negotiation*	167
business	276	*proceeding*	247	study	76
thing	275	year	210	work	57
understand		**comprehend**		**fathom**	
issue	2151	text	30	mystery	28
need	1587	thing	28	depth	13
principle	1173	*implication*	18	reason	6
implication	1136	*issue*	15	emotion	3
process	1039	*principle*	15	*implication*	2

Yet in the cases depicted in Table 3, there is no doubt that there is some contextual or pragmatic difference in meaning among the members of the near-synonym sets. Even if almost any argument compatible with one verb seems compatible with another, the forms belong, at least in part, to different registers and intuitively have slightly different meanings.[12] This is to be expected in light of the long-standing tradition in linguistics that there are no true synonyms in language (e.g. Bloomfield 1935:145), and more recent formulations to the effect that two distinct forms must always be accompanied by a difference in either semantic or pragmatic meaning (the "Principle of No Synonymy" in Goldberg 1995:67).

It would therefore be preferable to find even more similar sets, ideally involving the same lexemes to ensure a maximal proximity in form. Such cases may be found by looking at syntactic alternations that do not alter formal semantic meaning. One difficulty in finding such examples is that many well studied alternations, such as the English locative alternation (see Levin 1993:49-55, Iwata 2008), are accompanied by subtle differences in meaning. A well-known example is the verb *spray*, which in (6a) implies that the entire wall is covered with paint, while the variant in (6b) only describes some paint being sprayed either on some part of the wall or on the entire wall, i.e. the so called 'holistic/partitive' distinction.

(6) a. *Jack sprayed the wall with paint*
b. *Jack sprayed paint on the wall*

Taking a naïve approach, the thematic roles of the verb's argument structure should be inherited by both alternants and accept the same semantic classes. However given the subtle difference in meaning, a difference in productivity could be attributed to pragmatic reasons, e.g. that many things get sprayed on other things, but spraying an entire surface or object is usually done with only a smaller subset of substances. The argument for productivity in argument selection will therefore be more convincing if we find alternations which truly do not alter the meaning of the verb or its semantic interaction with its arguments.[13]

Taking this difficulty into account, examples (7)-(9) suggest three sets of variants that to my mind exhibit no difference in meaning *salva veritate* (some register differences notwithstanding): the German adposition *wegen* 'because of', which can appear as either a postposition governing the genitive or a preposition governing the genitive or more colloquially the dative (see Helbig and Buscha 2001:356, Zifonun et al. 1997:2080ff, Petig 1997), the English choice of verbal complement for *start* (a gerund or a *to*-infinitive, see Řeřicha 1987, Mair 2002, 2003) and the optional use of a *to* or bare infinitive as a complement of *help* (see Mair 1995, 2002; McEnery and Xiao 2005).

(7) a. *wegen des Vaters*
because-of the.GEN father.GEN
b. *des Vaters wegen*
the.GEN father.GEN because-of
c. *wegen dem Vater*
because-of the.DAT father.DAT
'because of the father'[14]

(8) a. *The leaves started to fall.*
b. *Rosie started crawling at 5 p.m.*
[examples adapted from Řeřicha (1987)]

(9) a. *Savings can help finance other Community projects*
b. *Perhaps the book helped to prevent things from getting even worse*
[examples from McEnery and Xiao (2005:162)]

In all three of these cases, it is essentially impossible to find contexts in which formal semantic differences ensue from the choice of construction, and in the latter two cases it is also not clear that there is a register difference. While other contextual factors which increase the probability of one construction or the other have been discussed in the above literature, there has been no evidence to indicate that, given that one construction has been chosen over the other, there should be for example a lower probability of finding neological or rare complement lexemes. However, as Figure 5 reveals, this is in fact the case.

For *wegen*, Panel A shows the prepositional constructions have a significantly larger vocabulary than the postpositional construction (over 500 more types or ~38% larger vocabulary for the maximal common sample size, at p<2.2e-16 in a test of equal proportions), and similarly the proportion of hapax legomena for the equal sample is significantly higher (~59% at p<2.2e-16). Panel B shows results for non-feminine, singular arguments, for which the prepositional constructions' case can be identified. Differences between dative (pre-d) and genitive (pre-g) are not significant, but the postposition (post-g) is again significantly less productive (p<2.2e-16 for *V*, p=1.221e-15 for *V1*). This corresponds to a readily available intuition that the postpositional construction is an archaism that is going out of use, while prepositional genitive is the formal standard and dative the colloquial form, both very much in active use. Note however that the relative rarity of the postpositional construction and its marked register are not at issue: the results indicate that it has a smaller, more repetitive vocabulary *given that it has been found appropriate and selected for use*. fZM extrapolations (dotted grey curves) also indicate that gaps between the constructions are set to grow if we were to increase the sample size. In total we can say that postpositional *wegen* is more repetitive and is co-selected more strongly with its collocational objects.[15]

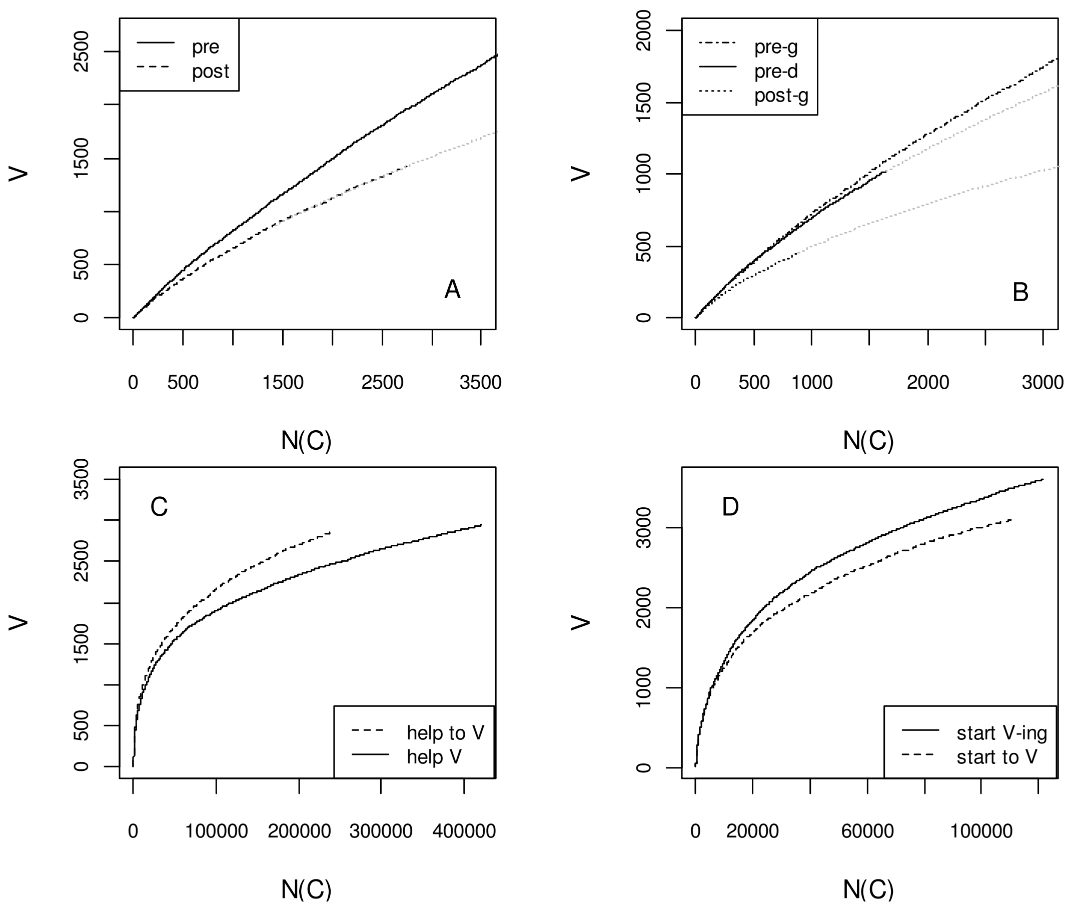

Figure 5. Vocabulary growth curves for arguments of synonymous syntactic constructions.

Panels C and D reveal equally significant differences between forms with no noticeable register differences. In the case of verbal complements for *help* with bare or *to* infinitives, the latter are rarer, but have a significantly larger vocabulary (~16% more items, p=1.078e-07) which is also less repetitive (~25% more hapax legomena, p=0.00009676). For verbal complements of *start* we also find significant differences in vocabulary (~13%, p=7.569e-07) and hapax legomena (~19%, p=0.0001075). Since the variants with less vocabulary are still quite frequent (or even more frequent for *help*), it appears that in each case one of the constructions is more prone to select a collocated object, which also results is a lower proportion of rare items. This can be confirmed by examining the odds ratio bias of the most common complements for either construction, as shown in Table 4 for infinitive complements of *help*. Of the five overall most common complements, four show a marked preference for the bare infinitive, while only *develop* shows an almost complete indifference to the choice of construction.

Table 4. Frequencies and odds ratio for the most common verbal complements of *help* in either construction.

verb	N_{to}	$total_{to}$	$freq_{to}$	N_{bare}	$total_{bare}$	$freq_{bare}$	*odds r*	N_{total}
make	10465	238067	0.043958	20470	421066	0.048615	0.899812	30935
develop	8004	238067	0.033621	13769	421066	0.0327	1.029127	21773
get	5189	238067	0.021796	15457	421066	0.036709	0.584706	20646
find	4533	238067	0.019041	14546	421066	0.034546	0.542468	19079
improve	5613	238067	0.023577	10648	421066	0.025288	0.930714	16261

In these cases there seems to be no obvious semantic reasons for the effects found in the data, suggesting an idiosyncratic tendency to select the more productive construction when communicative needs call for a non-collocational object.[16]

Before moving on to the discussion of these results, I would like to offer one final set of data to elucidate the language-specific and world knowledge-independent nature of this type of idiosyncrasy. If knowledge about the kinds of objects operated on by a verb or construction were the determining factor in its tendency to embed novel lexemes then we might expect translational pairs from different languages to exhibit similar productivity. Though different collocations may apply, as we have seen in Section 1, at least the proportion of hapax legomena, which cannot be lexicalized, and perhaps also vocabulary size, should be comparable. Taking data from UKWaC and DEWaC, Table 5 shows productivity measures for English *harbor/harbour* with mental-state objects, and its translational equivalent in German *hegen*.[17]

Table 5. Productivity measures for *harbo(u)r* and its German translational equivalent *hegen*. Differences in *V* and *V1* are significant at p<0.0005 for the smaller, equal sample size of *N(C)*=883.

verb	*N(C)*	*V*	*V1*	$\mathcal{P}$	$V_{N(C)=883}$	$V1_{N(C)=883}$	$\mathcal{P}_{N(C)=883}$
hegen	1115	59	15	0.013453	56	15	0.019355
harbo(u)r	883	103	53	0.060022	103	53	0.060022

Unless we assume that English speakers harbor more different mental states than German speakers (or rather that they write about more states in these corpora), we may simply surmise that German speakers use other verbs or constructions to express the same meanings as their English speaking counterparts, whereas the lexical verb most closely matching *harbor* is rather restricted and has a stronger preference for its collocations. Such inequalities are also not limited to lexical verbs, but can be found in more abstract constructions, such as comparative correlatives (CCs, see Zeldes 2009, 2011). Figure 6 plots vocabulary growth for the adjective in the apodosis of bare CCs of the form *the Xer the Yer* (e.g. *the faster the better*, or with the analytic comparative *the riskier the more profitable*) and German equivalents (*je Xer desto Yer*).

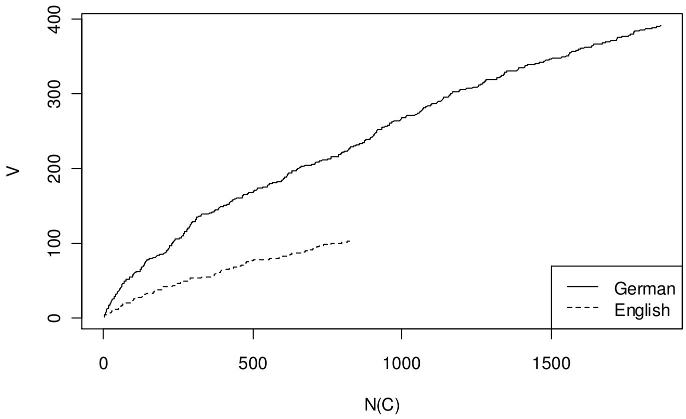

Figure 6. Vocabulary for comparatives in the apodosis of German and English bare CCs.

As we can see, English vocabulary in CCs is much more limited, owing mainly to a much starker preference for the most common lexemes *better* and lexicalized *merrier* (in *the more the merrier*). Though German *besser* 'better' is also very frequent, the proportion of unique lexemes in the German data is significantly greater.

4. A Hebbian model of productivity in argument selection

The results above show that lexical semantic classes of the Katzian or decompositional/property based kind are not sufficient for the prediction of selectional behavior in argument structure constructions. This unpredictability relates not only to idiosyncratic collocations, but also to idiosyncratic productivity in unlexicalized cases, which leads to highly significant differences in frequency, vocabulary and the probability of encountering unfamiliar types in data. Although a large part of such differences between semantically unrelated argument slots is due to extralinguistic differences that can be explained by world knowledge (e.g. the greater repetitiveness and smaller repertoire of beverages as compared to foods), the same cannot be said for cases of near or even complete synonyms explored in the previous section. We must therefore surmise that there is a language and construction specific factor involved in constraining the different aspects of productivity which is to some extent arbitrary and unpredictable: there is no lexical semantic reason why one member in a set of (near) synonyms is more productive and not another.

In interpreting these results I assume a usage-based model of the sort posited by Construction Grammar (CxG, see Goldberg 2006, i.a.), which assumes that any linguistic sign can be stored in the mental lexicon alongside usage information such as entrenchment. In a usage-based model, we must assume that the differences in productivity described above form a part of speakers' linguistic knowledge, i.e. they must be specified as part of the usage information stored for each argument structure construction in order to account for the lexical choices we observe in practice.[18] This opens up the pivotal question of what knowledge about productivity is stored exactly and how it is acquired. Here I would like to suggest that productivity information for argument slots is acquired by storing entrenchment values for argument types from the slots' input distributions themselves. While entrenchment for a construction in itself is insufficient to give rise to the observed differences in productivity (very common constructions can also be repetitive), it is here suggested that a combination of entrenchment for each observed argument type and the construction itself, together with the general principle of learning referred to as Hebb's Law directly allows productive distributions to be acquired as such in a self-perpetuating fashion.

Hebb's Law (Hebb 2002[1949], see also Pulvermüller 1996 for the linguistic domain) states that when a neuron A activates a neuron B repeatedly, their connection strengthens such that A can more easily activate B in subsequent cases. Its converse, wherein cells that consistently do not coactivate, or even inhibit each other, develop an increasing inhibitory relationship, has also been frequently postulated (e.g. Marshall 1995, Kohonen 2001). Applied to representations of heads and arguments, the law would suggest that frequent arguments, and especially collocations, would develop very strong connections with their selecting head representation, while infrequent arguments would only develop weak links, as expressed schematically for arguments of two constructions with German *wegen* 'because of' in Figure 7.

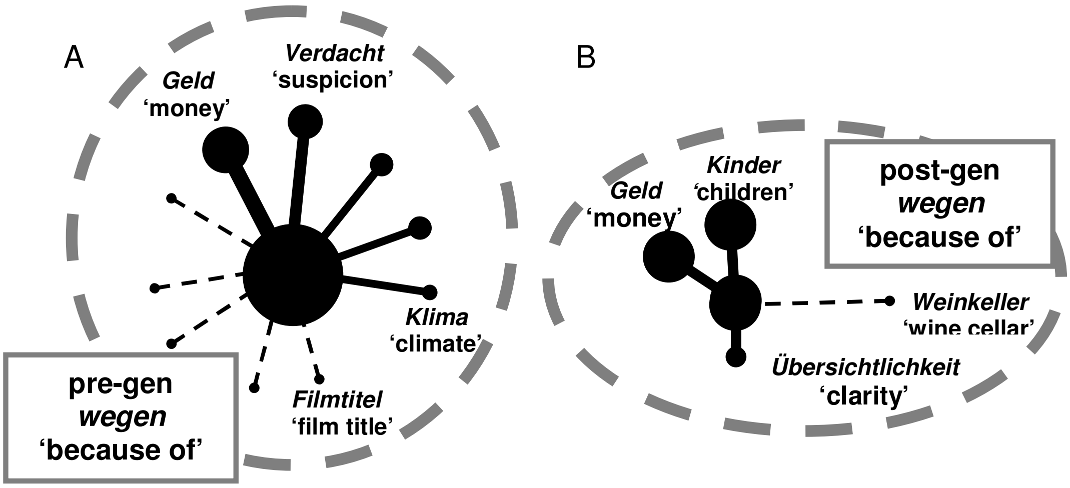

Figure 7. Hebbian network representation of argument slot entrenchment information for objects of the more productive prepositional *wegen* 'because of' (A) and the less productive postpositional *wegen* 'because of' (B).

In argument slots with lower levels of productivity, such as objects of the postposition *wegen* (post-gen) in Panel B, the most frequent arguments have a very high relative entrenchment, since they are coactivated with the verb proportionally much more often than the most frequent arguments of the preposition *wegen* (pre-gen). As a result, the entire construction comes to be identified chiefly with these argument choices and its activation strongly excites those argument head nouns, causing them to be selected. If we also assume inhibitory relations, then once communicative needs activate the representation of a potential novel argument of post-gen such as *Weinkeller* 'wine cellar' in Figure 7B, the inhibitory link between the novel argument and the established ones suppresses post-gen as a choice, since both the more powerful arguments and the construction itself, which have not coactivated with this argument before will weaken the activation of the construction.[19] Argument slots which are attested with a high proportion of hapax legomena (or rare arguments in general), such as pre-gen in Panel A, have an advantage in attracting novel arguments in this model, since it is easy for the overarching slot representation to be activated without dominant arguments taking over the activation pattern as easily. This is because the abstract representation of the slots, represented by the node in the middle, is strengthened (i.e. entrenched) by each argument type it cooccurs with, while the rare arguments themselves are not sufficiently strong to be identified with the construction more exclusively. Thus the central node in Panel A is proportionally larger and its most frequent argument nodes smaller, giving novel nodes a greater chance to compete and extend the argument structure's repertoire.

With regard to the construction as a whole, the present analysis suggests a possible dissociation, but at the same time a likely correlation between entrenchment and productivity. Argument slots with a high type count can activate the representation of their containing construction exactly as often as argument slots with fewer types, as long as the token count remains the same in both cases. However in the former slots, the admission of novel material will be substantially more likely. In practice, slots with more argument types are often also the more common ones, and conversely some rare verbs allow very few arguments, but in more than a few cases constructions that are somewhat more frequent are nevertheless more repetitive (note the similarities and differences in the rankings based on frequency and vocabulary size in Table 1).[20]

This mechanism makes it possible for the representation of entrenchment to lead directly to the productivity effects observed earlier: the vocabulary size is directly modeled by the members of the network and the probability of novel items corresponds to the proportional entrenchment strength of the construction compared to that of its argument types. Hapax legomena within data for the entire category serve as good estimators of this proportion since they are hardly entrenched themselves and form the greatest contribution to constructional entrenchment that is not biased towards lexicalized prototypes ($V1$ is usually the largest Vm for any m). This does not necessarily require that every hapax legomenon is retained long term in the mental lexicon, since the effects of proto-type independent constructional entrenchment are retained in the representation of the construction as a whole (the central nodes in Figure 7 are strengthened even if the weak connection with a coactivating hapax argument node is later forgotten). In this way, productivity can be perpetuated to remain similar in a grammar over time and can be acquired from the input using the very general notions of entrenchment and Hebb's Law.

If this explanation of productivity effects in usage is accepted, it still remains unclear what the status of such effects in a categorial theory of grammar should be, or how and where this kind of knowledge should be represented. Taking the usage-based approach seriously, I believe it should be uncontroversial that prototypical, collocational lexemes or lexicalized phrases must be referred to explicitly in the representation of argument slots in the mental lexicon. Such storage cannot be limited to non-compositional cases if we wish to account for unpredictable preferences or bias even in compositional cases. What the present findings add to the latter suggestion, which is by no means new, is that specific degrees of entrenchment within each relevant slot must be stored not only for specific lexemes, but also for traces of the amount of unlexicalized arguments witnessed (corresponding to the proportion of rare items). Put more formally, the mental lexicon of a specific language would have not only a weighted list of language-specific collocated prototypes and generalized semantic classes based on those prototypes (in order to account for entailments etc.), but also information corresponding to Vm for m, that is a representation of the entrenchment strength of each frequency band of arguments (recall this is also the information required to compute the fZM approximations used in Section 2, currently the best predictive mathematical model of productivity phenomena). These will vary from highly familiar lexicalized arguments (high values of m), to collections of occasional arguments, i.e. unlexicalized items from low values of m, including hapax legomena (m=1) and up to some frequency threshold θ, above which all arguments are lexicalized (cf. Baayen 1993 for a similar use of threshold frequencies). An example for the structure of a list of this type is given in Figure 8.

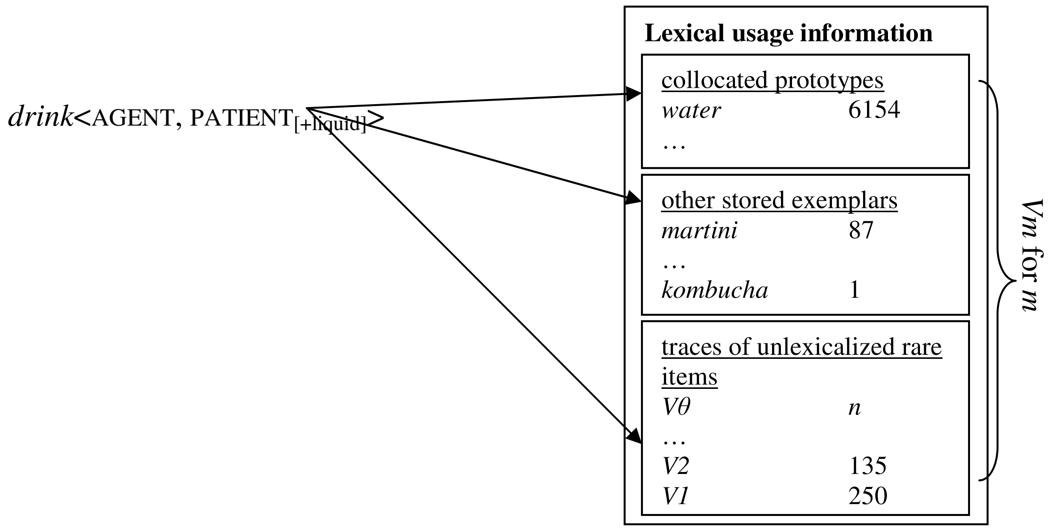

Figure 8. Lexical usage information in the mental lexicon for the object slot of *drink*.

The numbers in Figure 8 can be interpreted as raw frequencies for the sake of simplicity but they are more properly seen to represent entrenchment, as it is understood that salience can modulate the impact of usage experience. Additionally, it is quite possible that some low frequency information, and even hapax legomena, will be stored, at least temporarily, within the middle tier of as yet non-prototypical lexical experience. Finally, though it may seem unrealistic to assume discrete differences between the effects of $V1$, $V2$… $V\theta$, this rough approximation aims to leave the option open that the weight of different unlexicalized frequency bands may differ in unpredictable ways. For the purpose of ordinal rankings of potential productivity, i.e. the probability of encountering a novel type, we have already seen that $V1$, or rather its proportion $\mathscr{P}$, is a good estimator for argument slots by itself, as has been generally established for productivity in word formation.

The representation of the argument structure of transitive *drink* on the left is meant to be more or less theory neutral, but the same usage information could be applied to an entry from any formalism which accepts usage information, including the DRUNK slot of the CONTENT feature structure in the HPSG entry for *drink* in (2). A mental lexicon containing this type of information provides a grammar formalism with a means of modeling and predicting productivity phenomena based on usage experience. Lexical usage information thus determines not only collocational behavior, but also degrees of extensibility. Note that although the information in Figure 8 might be overly detailed, and possibly difficult to keep track of, it is also a direct result of the representation in Figure 7: by storing entrenchment strength for lexicalized exemplars and modifying the representation of the constructional slot during any form of usage (including exemplars which are later forgotten), this information is implicitly stored in the network representation. If we assume mental representations of constructions, slots and corresponding lexemes, all of this information can be learned via entrenchment and Hebb's Law, mechanisms that can generally be assumed to exist in almost any usage-based model.[21]

5. Conclusion

In this article I have argued that lexical semantic classes do not suffice for the description of restrictions on the usage of argument selection in practice. Instead, semantic classes only delineate a spectrum of semantically conceivable or possible arguments for a certain slot, while actual selection is also modulated by productivity effects which favor or disfavor certain constructions particularly in the case of unfamiliar material. Viewed from the morphological perspective on productivity, this state of affairs corresponds to the behavior of word formation processes discussed in Section 2: as in morphology, restrictions on the class of possible bases for a process do not dictate the productive profile realized for potential items within that class.

The data presented here supports the existence of idiosyncratic effects in the propensity of argument slots to admit unfamiliar lexical material that cannot be explained in purely semantic terms. By using lexical near synonyms and synonymous constructional alternants as minimal pairs, productive behavior has been isolated and dissociated from semantic meaning, showing that there are consistent and language-specific tendencies in the choice of competing constructions which pertain to unfamiliar and hence non-lexicalized material. I have suggested that these differences between possible linguistic options must at least partly be traced to some aspect of speakers' knowledge about usage, which I have suggested may be acquired by means of Hebb's Law and the storage of degrees of entrenchment in the mental lexicon.

Further aspects that could not be discussed within the scope of this article but seem to merit additional attention relate to differences between morphology and syntax, such as the simultaneous treatment of multiple syntactic slots and the definition of types in cases where argument phrases themselves contain embedded arguments or modifiers. In the former case, interdependence between productivity in multiple slots (i.e. productivity in one slot significantly increasing or decreasing the chances of productivity in another) would require the examination of multi-slot innovation and repetitiveness using multi-slot type definitions. In the case of material embedded within an argument, again concessions may have to be made such that the same head with different modifiers or internal arguments may constitute a different type. The use of different type definitions is expected to lead to differences in productivity ratings in many cases (e.g. it is possible that a verb selects a very small number of lexical head nouns for an argument slot, but that these in turn head a wide variety of compounds; see Zeldes, to appear, chapter 4 for some discussion). Another direction alluded to in note 16 but not explored here in detail is the use of productivity information in multifactorial models for the prediction of syntactic alternations. The results above suggest that productivity may be able to complement knowledge about the preferences of frequent, lexicalized arguments with more informed decisions in those cases where novel types occur that are not familiar from training data. In such cases, the findings above predict that the more productive construction should be preferred (cf. Zeldes 2009).

There are also further theoretical aspects which need to be explored, perhaps most urgently regarding the connection between different arguments. The representations in Figures 7 and 8 say nothing about possible links between the different argument lexeme nodes in the representations, although it is clear that novel arguments are often found that draw on existing, prototypical arguments by way of analogy. The network structure of semantically related arguments vying for selection is doubtless connected to productivity and its inclusion will substantially complicate the model. Although vocabulary information from corpora includes these effects (since they are inherent in corpus data as well), it is likely that a model explicitly dealing with such connections will be more accurate and serve as a fuller explanation of the facts. As the topic of productivity in syntax is only starting to receive attention (see also Barðdal 2008), it will require time before this and further challenges are worked into a usage-based theory of argument selection in the mental lexicon.

Finally, the model presented in Section 4 requires corroboration from outside of corpus data, ideally in an experimental setting. If speakers indeed retain degrees of productivity for syntactic slots, it may be expected that they react with more surprisal to innovation in less productive slots. Many approaches are possible in order to test this hypothesis, such as obtaining differences in reaction times in lexical decision tasks, differential ERP measurements when perceiving productive usage in more or less productive slots, or possibly also corresponding differences in reading behavior in eye tracking studies. Such findings would support the suggestion presented here for productivity in argument selection and complement experimental results on morphological productivity which have already been established.

Acknowledgements

I would like to thank Stefan Th. Gries, Anke Lüdeling and two anonymous reviewers for valuable comments on previous versions of this article. For discussions about various parts of the research described here I am also indebted to Peter Adolphs, Jasmine Bennöhr, Hans C. Boas, Peter Bosch, Stefan Evert, Livio Gaeta, Felix Golcher, Hagen Hirschmann, Marc Reznicek, Sören Schalowski and Stefanie Wulff. All errors and oversights are of course my own.

Bionote

Amir Zeldes is a post-doc researcher working and teaching at Humboldt-Universität zu Berlin. He is a member of the Corpus Linguistics team at the Institute for German Language and Linguistics and part of Project D1: Linguistic Database for Information Structure, working on corpus and computational linguistics methods within Collaborative Research Centre 632 on Information Structure in Potsdam and Berlin. E-mail: amir.zeldes@rz.hu-berlin.de.

Notes

1 The perspective taken in this article is a usage-based one (cf. Langacker 1987, Goldberg 2006 i.a.). For reasons of space I will forgo an exposition of the tenets of usage-based approaches and assume that conventionalized, language-specific systematicities in the observed spectrum of realized lexical arguments of an argument structure constitute facts requiring a linguistic explanation.

2 At the same time, lexicalized cases like (5) are also prototypes for open classes of arguments in specific languages. For example, in Japanese we may find unlexicalized arguments as in *ekisedorin o nomu*, literally ‘drink Excedrin’ (a migraine pill), meaning the object slot’s vocabulary is open to novelties not available to the nearest English counterpart, in this case any specific type of pill or orally ingested medicine.

3 Needless to say, it is possible to annotate and take measurements for distinct meanings of formally identical affixes, but for the purpose of this illustration we can forego this procedure here.

4 These results come from a corpus of approx. 14.5 million tokens from 5 years of a German computer journal, c't Magazin (for access to the corpus see http://www.linguistik.hu-berlin.de/institut/professuren/korpuslinguistik/institutkorpora).

5 The question of processes filling multiple slots at the same time cannot be discussed within the scope of this article. It can be observed, however, that under some process definitions, the problem is also found in morphology, e.g. if compounding is seen as a process selecting a head and a modifier. In many cases, a hierarchical model of selection can be used such that each choice is considered to be made independently, so that it can receive its own productivity ratings. In cases where statistical independence between choices in multiple slots can be shown not to hold, the measures should be calculated for multiple slots simultaneously, with some adjustments (see Zeldes, to appear, chapter 4 in detail).

6 Since accuracy was preferred to recall, only simple cases of post verbal objects with possible adjective modifiers were retrieved, excluding cases where the noun was followed by possessive *'s* or *'* and/or any further nouns (in order to avoid compounds, in which case the first noun following the verb is not the argument's lexical head). A subset of 1000 random hits for each verb was filtered manually and error rates were observed to remain under 1%.

7 Logarithmic scaling has been avoided for comparability with common practice in morphological studies, but see details for $N(C)$=1000 in Table 1.

8 An anonymous reviewer has noted that the data suggests the search for a single scale of productivity is doomed, since a reduction of dimensionality necessarily loses pertinent information, and has advocated introducing a further dimension in the VGC plots to explicitly model $\mathscr{P}$. I am inclined to agree that productivity is a multidimensional phenomenon, though it is still possible to use ordinal scales when explicitly modeling a single dimension (e.g. amount of types a speaker might know or which are likely to occur in a certain text size and genre). For three dimensional VGC plots with $\mathscr{P}$ see Zeldes (to appear), chapter 3, though it should be noted that $\mathscr{P}$ is already represented indirectly in two dimensional VGCs as the slope of the curve at each sample size $N(C)$. This is because $\mathscr{P}$ gives the chance of a hapax legomenon, which corresponds to the rate at which V will increase by one item, i.e. the derivative function of the VGC curve, cf. Baayen (1992:115).

9 It is notoriously difficult to find examples of true synonyms, in the sense of universal interchangeability in any sentence *salva veritate*, i.e. without altering truth values. I therefore refer here to near-synonyms, which can at least be expected to require the same argument classes. Further on we will also see some examples of alternating constructions that can be held to be truly synonymous and interchangeable *salva veritate* in all contexts.

10 Logarithmic scaling has again been avoided in keeping with the broad practice in morphological productivity studies, in order to retain the familiar comparability of steepness, characteristic of more productive processes. Instead, a zoomed-in view is offered for each plot on the right hand side.

11 To the exclusion of different collocations, e.g. *start/?begin a fire*, and with the different martial meaning *commence/*start/*begin fire*.

12 Interactions between register and productivity are also well documented for morphological processes, cf. Plag et al. (1999), Grabar and Zweigenbaum (2003). Whether or not register differences and other subtle differences in meaning that do not lend themselves to feature based decomposition should be part of lexical semantics proper is a matter of some contention, see Jackendoff (1990:32ff).

13 There are some further examples of seemingly equivalent constructions not amenable for the present purposes for different reasons. Wulff (2006) uses collostructional and distinctive collexeme analysis (Stefanowitsch and Gries 2003, Gries and Stefanowitsch 2004) to show that despite a very similar interpretation, English *go VERB* and *go and VERB* prefer semantically distinct verbs in the second position. However, the argument status of the second, seemingly paratactic verb in this case is not sufficiently clear for the present discussion. By contrast, the English verb particle alternation between *pick up the book* and *pick the book up*, which has been referred to as a case of 'allostructions' with the same meaning (Cappelle 2006), shows little in the way of semantic bias, but since length of the direct object is known to interact with the choice of construction (see Gries 2003 i.a.), a bias can be expected depending on the length of novel arguments and the phrases they head. Similar factors apply to the dative alternation, see Bresnan et al. (2007), i.a., and more below on multifactorial approaches in general.

14 The two prepositional variants cannot be distinguished for all genders and numbers, since, e.g. dative and genitive feminine singular are identical due to syncretism. Below separate results will be given for masculines and neuters, which can be categorized conclusively as either construction b. or c.

15 Notably with *Kinder* 'children', *Liebe* 'love', *Geld* 'money' and the lexicalized expression *der Übersichtlichkeit wegen* 'for the sake of clarity'. If the construction becomes even less productive, it may someday become limited to such collocations, much like its etymological precursor the circumposition *von … wegen*. The latter is now completely limited to the lexicalized expression *von Amts wegen* 'officially, ex officio', lit. 'from ways of office' (see Paul 1959: vol. 4, 43-44; Braunmüller 1982:200-207).

16 This is in no way meant to suggest that productivity is the only, or the most important factor in determining which alternant is selected in usage: multifactorial modeling has shown that for most alternations a wide variety of factors is involved (e.g. Gries 2003 on verb particle placement, Bresnan et al. 2007 on the dative alternation, and Lohmann 2011 specifically on *help* with bare or *to* infinitive). The present results merely suggest that productivity is another factor that significantly distinguishes the usage profile of competing syntactic constructions, much as is the case in morphological word formation.

17 Results had to be manually filtered to rule out concrete objects such as *harboring fugitives* etc., and in order to find sufficient English examples to approach the German sample size, all cases of *harbo(u)r* tagged as a verb had to be retrieved and the object head noun manually extracted, even if the object NP did not directly follow the verb. In the interest of cross-linguistic comparability, transparent German compounds were reduced to their head nouns (in English compounds only the head noun was considered as well).

18 Whether this knowledge forms part of grammar (knowledge about rules) or part of the lexicon (knowledge about lexical heads, or even abstract constructions stored in the mental lexicon) can be viewed differently depending on one's theoretical framework. CxG and related frameworks cast the rules of grammar as lexical entries with open slots,

relegating the information to the mental lexicon (or "constructicon"). However, the findings above could just as well be applied to the grammar component of some other theories which aim to describe and predict usage in cases of lexically unspecified constructions.

19 Alternatively we can simply assume that positive activation for another construction, e.g. pre-gen, is stronger and wins out in a shouting match without having to assume explicit inhibition in our model (cf. Rumelhart and Zipser 1985, Roelofs 1992).

20 An anonymous reviewer has asked whether this model of entrenchment implies that frequency information is stored outside the linguistic/contextual information proper. The present suggestion is in fact primarily meant to be understood within the context of knowledge about linguistic usage and how it is stored, i.e. the sizes of the nodes and links in the figure above represent usage of linguistic constructions at all levels (argument structure constructions, specific slot positions and specific lexemes). However I generally accept Elman's (2009) view that a great portion of the knowledge used for linguistic selectional processes is shared with other cognitive processes accessing the same type of e.g. encyclopedic knowledge, without separate stores for strictly linguistic and non-linguistic conceptual knowledge. On the strength of Elman's evidence and the related studies discussed by him it could therefore be expected that the frequency distribution of extralinguistic categories could also play a role in linguistic productivity, and indeed the statements in Section 2 about the relative extralinguistic frequencies of eating and drinking different substances refer precisely to knowledge of this kind, which normally has a much greater effect on argument selection than the often subtle impact of arbitrary, language-specific productivity effects discussed here.

21 Another reviewer has noted that an interactive activation model with lateral inhibition could equally explain the phenomena observed here without resorting to neural network representations. This is absolutely true: in fact, any model somehow representing both lexicalized and non-lexicalized entrenchment information for argument selection in usage has the necessary information for modeling productivity phenomena. The appeal of a neurally oriented model in particular is largely due to the usage-based commitment to cognitive plausibility (cf. Lakoff's 1990 'cognitive commitment'). I suggest that a Hebbian approach is particularly suitable since it requires only a relatively simple, already well-established mechanism, which could account for the observed productivity phenomena naturally. However I do not mean to exclude other formal representations of the facts, nor do I believe that this explanation is a complete one (see Section 5 for some specific desiderata).

References

Baayen, R. Harald. 1992. Quantitative aspects of morphological productivity. In Geert E. Booij & Jaap van Marle (eds.), *Yearbook of Morphology* 1991, 109–149. Dordrecht: Kluwer.

Baayen, R. Harald. 1993. On frequency, transparency and productivity. In Geert E. Booij & Jaap van Marle (eds.), *Yearbook of Morphology* 1992, 181–208. Dordrecht: Kluwer.

Baayen, R. Harald. 2001. *Word frequency distributions*. (Text, Speech and Language Technologies 18). Dordrecht, Boston & London: Kluwer.

Baayen, R. Harald. 2009. Corpus linguistics in morphology: Morphological productivity. In Anke Lüdeling & Merja Kytö (eds.), *Corpus linguistics. An international handbook*. Vol. 2, 899–919. Berlin: Mouton de Gruyter.

Baayen, R. Harald & Rochelle Lieber. 1991. Productivity and English derivation: A corpus-based study. *Linguistics* 29(5). 801–843.

Barðdal, Jóhanna. 2008. *Productivity: Evidence from case and argument structure in Icelandic*. Amsterdam & Philadelphia: John Benjamins.

Baroni, Marco, Silvia Bernardini, Adriano Ferraresi & Eros Zanchetta. 2009. The WaCky Wide Web: A collection of very large linguistically processed Web-crawled corpora. *Language Resources and Evaluation* 43(3). 209–226.

Bauer, Laurie. 2001. *Morphological productivity*. Cambridge: Cambridge University Press.

Bloomfield, Leonard. 1935. *Language*. London: George Allen & Unwin.

Braunmüller, Kurt. 1982. *Syntaxtypologische Studien zum Germanischen*. Tübingen: Gunter Narr.

Bresnan, Joan, Anna Cueni, Tatiana Nikitina & R. Harald Baayen. 2007. Predicting the dative alternation. In: Gerlof Bouma, Irene Kraemer & Joost Zwarts (eds.), *Cognitive foundations of interpretation*, 69–94. Amsterdam: Royal Netherlands Academy of Science.

Cappelle, Bert. 2006. Particle placement and the case for 'allostructions'. *Constructions* SV1(7). Available at: http://www.constructions-online.de/articles/specvol1/683.

Dowty, David R. 1991. Thematic proto-roles and argument selection. *Language* 67(3). 547–619.

Elman, Jeffrey L. 2009. On the meaning of words and dinosaur bones: Lexical knowledge without a lexicon. *Cognitive Science* 33(4). 547–582.

Evert, Stefan. 2004. A simple LNRE model for random character sequences. *Proceedings of JADT 2004*, 411–422. Louvain-la-Neuve, Belgium.

Evert, Stefan & Anke Lüdeling. 2001. Measuring morphological productivity: Is automatic preprocessing sufficient? In Paul Rayson, Andrew Wilson, Tony McEnery, Andrew Hardie & Shereen Khoja (eds.), *Proceedings of Corpus Linguistics 2001*, 167–175. Lancaster.

Gaeta, Livio & Davide Ricca. 2006. Productivity in Italian word formation: A variable-corpus approach. *Linguistics* 44(1). 57–89.

Goldberg, Adele E. 1995. *Constructions: A construction grammar approach to argument structure*. Chicago & London: University of Chicago Press.

Goldberg, Adele E. 2006. *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.

Grabar, Natalia & Pierre Zweigenbaum. 2003. Productivité à travers domaines et genres: dérivés adjectivaux et langue médicale. *Langue française* 140. 102–125.

Gries, Stefan Th. 2003. *Multifactorial analysis in corpus linguistics: A study of particle placement*. London & New York: Continuum.

Gries, Stefan Th. & Anatol Stefanowitsch (2004), Extending collostructional analysis: A corpus-based perspective on 'alternations'. *International Journal of Corpus Linguistics* 9(1). 97–129.

Hebb, Donald O. 2002[1949]. *The organization of behavior*. New York: Wiley & Sons.

Helbig, Gerhard & Joachim Buscha. 2001. *Deutsche Grammatik: Ein Handbuch für den Ausländerunterricht*. Berlin & Munich: Langenscheidt.

Iwata, Seizi 2008. *Locative alternation: A lexical-constructional approach*. Amsterdam & Philadelphia: John Benjamins.

Jackendoff, Ray. 1990. *Semantic structures*. Cambridge, MA: MIT Press.

Katz, Jerrold. J. & Jerry A. Fodor. 1963. The structure of a semantic theory. *Language* 39(2). 170–210.

Kiss, Tibor. 2007. Produktivität und Idiomatizität von Präposition-Substantiv-Sequenzen. *Zeitschrift für Sprachwissenschaft* 26(2). 317–345.

Kohonen, Teuvo. 2001. *Self-organizing maps*. 3rd edition. Berlin, Heidelberg, New York: Springer.

Lakoff, George. 1990. The invariance hypothesis: Is abstract reason based on image-schemas? *Cognitive Linguistics* 1(1). 39–74.

Langacker, Ronald W. 1987. *Foundations of cognitive grammar. Vol. 1: Theoretical prerequisites*. Stanford: Stanford University Press.

Levin, Beth. 1993. *English verb classes and alternations*. Chicago: University of Chicago Press.

Lohmann, Arne. 2011. *Help* vs *help to*: A multifactorial, mixed-effects account of infinitive marker omission. *English Language and Linguistics* 15(3). 499–521.

Mair, Christian. 1995. Changing patterns of complementation, and concomitant grammaticalisation, of the verb *help* in present-day British English. In Bas Aarts & Charles F. Meyer (eds.), *The verb in contemporary English. Theory and description*, 258–272. Cambridge: Cambridge University Press.

Mair, Christian. 2002. Three changing patterns of verb complementation in Late Modern English: A real-time study based on matching text corpora. *English Language and Linguistics* 6(1). 105–131.

Mair, Christian. 2003. Gerundial complements after *begin* and *start*: Grammatical and sociolinguistic factors, and how they work against each other. In Günther Rohdenburg & Britta Mondorf (eds.), *Determinants of grammatical variation in English*, 329–345. Berlin: Mouton de Gruyter.

Marshall, Jonathan A. 1995. Adaptive perceptual pattern recognition by self-organizing neural networks: Context, uncertainty, multiplicity, and scale. *Neural Networks* 8(3). 335–362.

McEnery, Anthony & Zhonghua Xiao. 2005. *HELP* or *HELP to*: what do corpora have to say? *English Studies* 86(2), 161–187.

Paul, Hermann. 1959. *Deutsche Grammatik*. Halle: Max Niemeyer Verlag.

Petig, William E. 1997. Genitive prepositions used with the dative in spoken German. *Unterrichtspraxis* 30(1). 36–39.

Plag, Ingo. 2006. Productivity. In Bas Aarts & April M. S. McMahon (eds.), *The handbook of English linguistics*, 537–556. Malden, MA: Blackwell.

Plag, Ingo, Christiane Dalton-Puffer & R. Harald Baayen. 1999. Productivity and register. *English Language and Linguistics* 3(2). 209–228.

Pollard, Carl & Ivan A. Sag. 1994. *Head-driven phrase structure grammar*. Chicago: University of Chicago Press.

Pulvermüller, Friedemann. 1996. Hebb's concept of cell assemblies and the psychophysiology of word processing. *Psychophysiology* 33(4). 317–333.

Řeřicha, Václav. 1987. Notes on infinitival and *-ing* complements of the verbs *begin* and *start*. *Philologica Pragensia* 30(3). 129–133.

Roelofs, Ardi. 1992. A spreading-activation theory of lemma retrieval in speaking. *Cognition* 42(1–3). 107–142.

Rumelhart, David E. & David Zipser. 1985. Feature discovery by competitive learning. *Cognitive Science* 9(1). 75–112.

Säily, Tanja. 2011. Variation in morphological productivity in the BNC: Sociolinguistic and methodological considerations. *Corpus Linguistics and Linguistic Theory* 7(1). 119–141.

Schmid, Helmut. 1994. Probabilistic part-of-speech tagging using decision trees. In: *Proceedings of the Conference on New Methods in Language Processing*. Manchester, UK, 44–49. Available at: http://www.ims.uni-stuttgart.de/ftp/pub/corpora/tree-tagger1.pdf.

Stefanowitsch, Anatol & Stefan Th. Gries. 2003. Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8(2). 209–243.

Wulff, Stefanie. 2006. *Go*-V vs. *go-and*-V in English: A case of constructional synonymy? In: Stefan Th. Gries & Anatol Stefanowitsch (eds.), *Corpora in cognitive linguistics: Corpus-based approaches to syntax and lexis*, 101–125. Berlin & New York: Mouton de Gruyter.

Zeldes, Amir. 2009. Quantifying constructional productivity with unseen slot members. In *Proceedings of the NAACL HLT Workshop on Computational Approaches to Linguistic Creativity, June 5, Boulder, CO*, 47–54. Stroudsburg, PA: Association for Computational Linguistics.

Zeldes, Amir. 2011. On the productivity and variability of the slots in German comparative correlative constructions. In Marek Konopka, Jacqueline Kubczak, Christian Mair, František Štícha & Ulrich H. Waßner (eds.), *Grammar & Corpora, Third International Conference, Mannheim, 22.–24.09.2009*, 429–449. Tübingen: Narr.

Zeldes, Amir. To appear. *Productivity in argument selection. From morphology to syntax*. Berlin & New York: De Gruyter Mouton.

Zifonun, Gisela, Ludger Hoffmann & Bruno Strecker (eds.). 1997. *Grammatik der deutschen Sprache*. Berlin & New York: De Gruyter.