Quantifying Constructional Productivity with Unseen Slot Members

Amir Zeldes
Institut für deutsche Sprache und Linguistik
Humboldt-Universität zu Berlin
Unter den Linden 6, 10099 Berlin, Germany
amir.zeldes@rz.hu-berlin.de

Abstract

This paper is concerned with the possibility of quantifying and comparing the productivity of similar yet distinct syntactic constructions, predicting the likelihood of encountering unseen lexemes in their unfilled slots. Two examples are explored: variants of comparative correlative constructions (CCs, e.g. *the faster the better*), which are potentially very productive but in practice lexically restricted; and ambiguously attached prepositional phrases with the preposition *with*, which can host both large and restricted inventories of arguments under different conditions. It will be shown that different slots in different constructions are not equally likely to be occupied productively by unseen lexemes, and suggested that in some cases this can help disambiguate the underlying syntactic and semantic structure.

1 Introduction

Some syntactic constructions[1] are more productive than others. Innovative coinages like the CC: *The bubblier the Mac-ier* (i.e. the more bubbly a program looks, the more it feels at home on a Macintosh computer) are possible, but arguably more surprising and marked than: *I have a bubblier operating system with a Mac-ier look* in their respective construction, despite the same novel lexemes. The aim of this paper is to measure differences in the productivity of slots in such partially-filled constructions and also to find out if this productivity can be used to disambiguate constructions.

As one of the defining properties of language, productivity has received much attention in debates about the nature of derivational processes, the structure of the mental lexicon and the interpretation of key terms such as compositionality, grammaticality judgments or well-formedness. However in computational linguistics it is probably fair to say that it can be regarded most of all as a problem. Familiar items present in training data can be listed in lexical resources, the probabilities of their different realizations can be estimated from corpus frequency distributions etc. Thus using lexical information (statistically extracted or handcrafted resources) is the most successful strategy in resolving syntactic ambiguities such as PP-attachment (Hindle and Rooth, 1993; Ratnaparkhi, 1998; Stetina and Nagao, 1997; Pantel and Lin, 2000; Kawahara and Kurohashi, 2005), basing decisions on previous cases with identical lexemes or additional information about those lexemes. Yet because of productivity, even very large training data will never cover examples for all inputs being analyzed.

In morphological theory (and corresponding computational linguistic practice), the situation has been somewhat different: a much larger part of the word formations encountered in data can be listed in a lexicon, with neologisms being the exception, whereas in syntax most sentences are novel, with recurring combinations being the exception.[2] The focus in morphology has therefore often been on which word formation processes are productive and to what extent, with the computational counterpart being whether or not corresponding rules should be built into a morphological analyzer. Syntacticians, conversely, may ask which apparently regular constructions are actually lexicalized or have at least partly non-compositional properties (e.g. collocations, see Choueka, 1988, Evert, 2005,

[1] I use the term 'construction' in a construction grammar sense following Goldberg (1995, 2006) to mean mentally stored hierarchically organized form-meaning pairs with empty, partially-filled or fully specified lexical material. In this sense, both comparative adjectives and the pattern *The [COMP] the [COMP]* are constructions, and the productivity of such patterns is the quantity being examined here.

[2] Compounding represents an exception to this generalization, standing, at least for some languages, between syntax and word formation and often generating an unusually large amount of items unlisted in lexica (cf. Bauer, 2001:36-7).

Proceedings of the NAACL HLT Workshop on Computational Approaches to Linguistic Creativity, pages 47–54,
Boulder, Colorado, June 2009.

2009; multiword expressions, Sag et al., 2002; lexical bundles, Salem, 1987, Altenberg and Eeg-Olofsson, 1990, Biber et al., 1999, 2004).

In morphology, the realization that productivity is a matter of degree, rather than a binary trait of word formation processes (see e.g. Bauer, 2001:125-162), has lead to the exploration of quantitative measures to assess and compare different aspects of the fertility of various patterns (esp. the work of Baayen, 2001, 2009). Yet syntactic applications of these measures have only very recently been proposed, dealing with one slot of a pattern much like the stem operated on by a morphological process (cf. Barðdal, 2006; Kiss, 2007).

In this paper I will examine the application of measures based on Baayen's work on morphology to different variants of syntactic constructions with more or less variable slots. The goal will be to show that different constructions have inherently different productivity rates, i.e. they are more or less liable to produce new members in their free slots. If this view is accepted, it may have consequences both theoretically (novelty in certain positions will be more surprising or marked) and practically, e.g. for parsing ambiguous structures with novel arguments, since one parse may imply a construction more apt to novelty than another.

The remainder of this article is structured as follows: the next section introduces concepts underlying morphological productivity and related corpus-based measures following Baayen (2009). The following two sections adapt and apply these measures to different types of CCs (such as *the faster the better*) and NP/VP-attached PPs, respectively, using the BNC[3] as a database. The final section discusses the results of these studies and their implications for the study of syntactic productivity.

2 Morphological Productivity Measures

Productivity has probably received more attention as a topic in morphology than in syntax, if for no other reason than that novel words are comparatively rare and draw attention, whereas novel phrases or sentences are ubiquitous. The exact definition of a novel word or 'neologism' is however less than straightforward. For the present purpose we may use Bauer's (2001:97-98) working definition as a starting point:

> *[Productivity] is a feature of morphological processes which allow for new coinages, […] coining must be repetitive in the speech community […] Various factors appear to aid productivity: type frequency of appropriate bases, phonological and semantic transparency, naturalness, etc., but these are aids to productivity, not productivity itself.*

For Bauer, productivity is defined for a morphological process, which is ideally frequently and consistently found and coins ideally transparent novel forms. The word 'coining' in this context implies that speakers use the process to construct the transparent novel forms in question, which in turn means the process has a regular output. Yet novelty, transparency and regularity are difficult to judge intuitively, and the definitions of "new" vs. "existing" words cannot be judged reliably for any one speaker, nor with any adequacy for a speaker community (cf. Bauer, 2001:34-35).

This problem has led researchers to turn to corpus data as a sort of 'objective' model of language experience, in which the output of a process can be searched for, categorized and tagged for evaluation. Baayen (e.g. 2001, 2009) proposes three corpus-based measures for the productivity of word formation processes. The first measure, which he terms *extent of use*, is written V(C,N) and is simply the proportion of types produced by a process C in a corpus of size N, e.g. the count of different nouns in *-ness* out of all the types in N. According to this measure, *-ness* would have a much higher *realized productivity* than the *-th* in *warmth* since it is found in many more words. However, this measure indiscriminately deals with all existing material – all words that have already been generated – and hence it cannot assess how likely it is that novel words will be created using a certain process.

Baayen's other two measures address different aspects of this problem and rely on the use of *hapax legomena*, words appearing only once in a corpus. The intuitive idea behind looking at such words is that productively created items are one-off unique occurrences, and therefore they must form a subset of the hapax legomena in a corpus. Baayen uses V(1,C,N), the number of types from category C occurring once in a corpus of N words and V(1,N), the number of all types occurring once in a corpus of N words. The second measure, termed *hapax-conditioned degree of productivity* is said to measure *expanding productivity*, the rate at

[3] The British National Corpus (http://www.natcorp.ox.ac.uk/), with over 100 million tokens of British English.

which a process is currently creating neologisms. It is computed as V(1,C,N)/V(1,N), the proportion of hapax legomena from the examined category C within the hapax legomena from all categories in the corpus. Intuitively, if the amount of hapax legomena could be replaced by 'true' neologisms only, this would be the relative contribution of a process to productivity in the corpus, which could then be compared between different processes[4].

The third measure, *category-conditioned degree of productivity* measures the *potential productivity* of a process, meaning how likely it is to produce new members, or how saturated a process is. This measure is the proportion of hapax legomena from category C divided by N(C), the total token count from this category: V(1,C,N)/N(C). It intuitively represents the probability of the next item from category C, found in further corpus data of the same type, to be a hapax legomenon.

Baayen's measures (hence p1, p2 and p3 respectively) are appealing since they are rigorously defined, easily extractable from a corpus (provided the process can be identified reliably in the data) and offer an essential reduction of the corpus wide behavior of a process to a number between 1 and 0, that is, an item producing no hapax legomena would score 0 on p2 and p3, and an item with 100% hapax legomena would score 1 on p3, even if it is overall rather insignificant for productivity in the corpus as a whole (as reflected in a low score for p2). The measure p3 is the most important one in the present context, since it allows us to reason conversely that, given that an item is novel and could belong to one of two processes, it is more likely to have come from whichever process is more productive, i.e. has a higher p3 score.

Indeed the assumptions made in these measures do not necessarily fit syntactic productivity at a first glance: that the process in question has a clearly defined form (e.g. a suffix such as *-ness*) that it accommodates one variable slot (the stem, e.g. *good-* in *goodness*), and that each different stem forms a distinct type. Applying these measures to syntactic constructions requires conceptual and mathematical adaptation, which will be discussed in the next section using the example of comparative correlative constructions.

3 Measuring Productivity in CCs

Comparative correlatives are a complex yet typologically well attested form of codependent clauses expressing a corresponding monotonous positive or negative change in degree between two properties (see den Dikken, 2005 for a cross-linguistic overview). For example, in *the faster we go, the sooner we'll get there*, speed is monotonously correlated with time of arrival. A main reason for syntactic interest in this type of sentence is a proposed 'mismatch' (see McCawley, 1988, Culicover and Jackendoff, 1999) between its syntax, which appears to include two identically constructed paratactic clauses, and its semantics, which imply possible hypotaxis of the first clause as a sort of 'conditional' (*if and in so much as we go fast…*).

Two other noteworthy features of this construction in use (the following examples are from the BNC) are the frequent lack of a verb (*the larger the leaf the better quality the tea*) and even of a subject noun (*the sooner the better*) [5] and a tendency for the (at least partial) lexicalization of certain items. The verbless variant often houses these, e.g. *the more the merrier*, but also with verbs, e.g. *the bigger they come the harder they fall*. A context-free grammar might describe a simplified variant of such clauses in the following terms:

S_{cc} > *the* COMP (NP (VP))
S > S_{cc} S_{cc}

where S_{cc} is one of the comparative correlative clauses, COMP represents either English comparative allomorph (in *-er* like *bigger* or analytic with *more* or *less* in *more/less important*), and NP and VP are optional subjects and corresponding predicates for each clause.[6]

However like many CFG rules, these rules may be too general, since it is clearly the case that not

[4] This statement must be restricted somewhat: in items showing multiple processes, e.g. *bullishness*, the processes associated with the suffixes *-ish* and *-ness* are not statistically independent, creating a difficulty in using such cases for the comparison of these two processes (see Baayen, 2009). In syntax the extent of this problem is unclear, since even occurrences of NPs and VPs are not independent of each other.

[5] The latter form has been analyzed as a case of ellipsis of the copula *be* (Culicover and Jackendoff, 1999:554; similarly for German: Zifonun et al., 1997:2338). It is my position that this is not the case, as the bare construction has distinct semantic properties as well as different productive behavior, see below.

[6] These rules should be understood as agnostic with respect to the parataxis/hypotaxis question mentioned above. The parentheses mean NP may appear without VP but not vice versa.

all comparatives, nouns and verbs fit in this construction, if only because of semantic limitations, i.e. they must be plausibly capable of forming a pair of monotonously correlated properties. Corpus data shows that comparatives in CC clauses select quite different lexemes than comparatives at large, that the first and second slots (hence cc1 and cc2) have different preferences, and that the presence or absence of a VP and possibly a subject NP also interact with these choices. Table 1 shows comparatives in the BNC sorted by frequency in general, along with their frequencies in cc1 and cc2. Some frequent comparatives do not or hardly appear in CCs given their frequency[7] while others prefer a certain slot exclusively (e.g. *more likely* in cc2) or substantially (e.g. *higher* in cc1). Columns ø1 and ø2 show bare comparatives (no subject or verb) in cc1 or 2 and the next two columns show subsets of bare cc1 or 2 given that the other clause is also bare. The last columns show CCs with only NPs and no verb, either in one clause or both. In bare CCs we find that *better* selects cc2 exclusively, in fact making up some 88% of cc2s in this construction (*the* COMP *the better*) in the BNC.

word	comp	cc1	cc2	Ø1	Ø2	Ø1 (Ø2)	(Ø1) Ø2	n1	n2	n1 (n2)	(n1) n2
further	21371										
better	20727	15	143		89		51	9	22	5	15
higher	15434	97	39	4	2	3		84	23	44	21
greater	13883	82	171	1	1			75	92	35	80
lower	10983	20	27		2			18	12	7	12
older	8714	24	1	1		1		3		1	
…											
longer	3820	45	15	3	1	3		11	3	9	3
bigger	3469	43	13	4	1	3		30	8	15	8
more likely	3449		28						2		1
…											
more wholistic	1										
zanier	1	1						1			

Table 1. Comparative frequencies independently and in cc1/cc2, with or without nominal subjects/verbs in one or both clauses.

[7] Occurrences of items which cannot serve attributively, such as *more* with no adjective and *sooner*, have been excluded, since they are not comparable to the other items. Most occurrences of the most frequent item, *further*, should arguably be excluded too, since it is mostly used as a lexicalized adverb and not a canonical comparative. However comparative usage is also well-attested, e.g.: *he was going much further than that*.

A look at the list of lexemes typical to cc1 vs. cc2 shows that cc1 tends to express a dependent variable with spatiotemporal semantics (*higher*, *older*, *longer*), whereas cc2 typically shows an independent evaluative (*better*, *more likely*), though many common lexemes appear in both.[8]

Although the results imply varying degrees of preference and lexicalization in different constructions, they do not yet tell us whether or not, or better how likely, we can expect to see new lexemes in each slot. This can be assessed using Baayen's measures, by treating each construction as a morphological process and the comparative slot as the lexical base forming the type (see Kiss, 2007 for a similar procedure).[9] The results in Table 2 show that all constructions are productive to some extent, though clearly some yield fewer new types.

	toks	types	hpx	p1	p2	p3
comp	266703	5988	2616	0.00772	0.00651	0.0098
cc1	802	208	140	0.00026	0.00034	0.1745
cc2	802	181	126	0.00023	0.00031	0.1571
bare1	58	45	37	5.80E-05	9.22E-05	0.6379
bare2	58	7	5	9.03E-06	1.24E-05	0.0862

Table 2. Productivity scores for comparatives, cc-clauses in general and specifically for bare CCs

p1 and p2 show that CCs are responsible for very little of the productive potential of comparatives in the corpus. This is not only a function of the relative rarity of CCs: if we look at their rate of vocabulary growth (Figure 1), general comparatives gather new types more rapidly than CCs even for the same sample size[10]. Using a Finite Zipf Mandelbrot Model (FZM, Evert, 2004), we can extrapolate from the observed data to predict the gap will grow with sample size.

[8] I thank Livio Gaeta and an anonymous reviewer for commenting on this point.

[9] In fact, one could also address the productivity of the construction as a whole by regarding each argument tuple as a type, e.g. <*more ergonomic*, *better*> could be a hapax legomenon despite *better* appearing quite often. Since each slot multiplies the chances a construction has to be unique, the n^{th} root of the value of the measure would have to be taken in order to maintain comparability, thus the square root of p_k for 2 slots, the cube root for 3 slots and so on. Another option, if one is interested in the chance that any particular slot will be unique, is to take the average of p_k for all slots. However for the present purpose the individual score of each slot is more relevant.

[10] The comparative curve is taken from 2000 occurrences evenly distributed across the sections of the BNC, to correspond topically to the CCs, which cover the whole corpus.

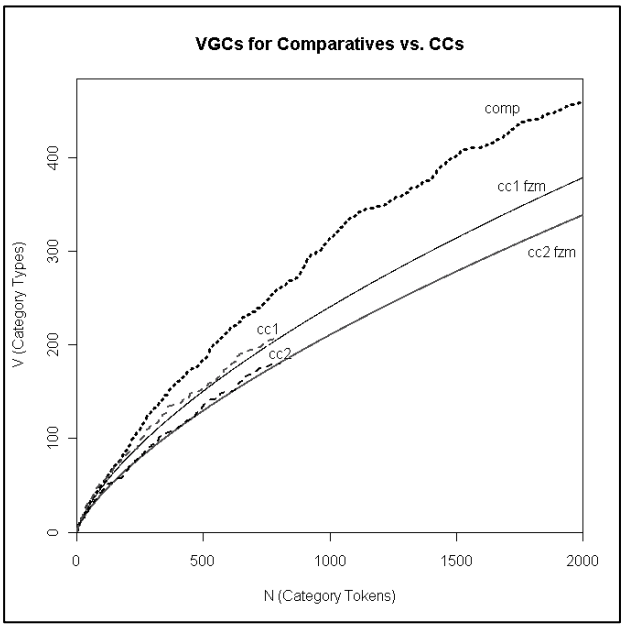

Figure 1. Vocabulary growth curves and FZM extrapolations for comparatives in cc1, cc2 and at large in the BNC.

However, p3 shows the surprising result that CCs have more potential productivity than comparatives in general, with the bare cc1 slot leading, both general CC slots somewhat behind, and the bare cc2 last. This means our data does not begin to approach covering this category – the next CC is much likelier to be novel, given the data we've seen so far.

With this established, the question arises whether a CFG rule like the one above should take account of the likelihood of each slot to contain novel vs. familiar members. For instance, if a PCFG parser correctly identifies a novel comparative and the input matches the rule, should it be more skeptical of an unseen bare cc1 than an unseen bare cc2 (keeping in mind that the latter have so far been *better* in 88% of cases)? To illustrate this, we may consider the output of a PCFG parser (in this case the Stanford Parser, Klein and Manning, 2003) for an ambiguous example.

Since CCs are rather rare, PCFGs will tend to prefer most other parses of a sentence, if these are available. Where no other reading is available we may get the expected two clause structure, as in the example in Figure 2.[11]

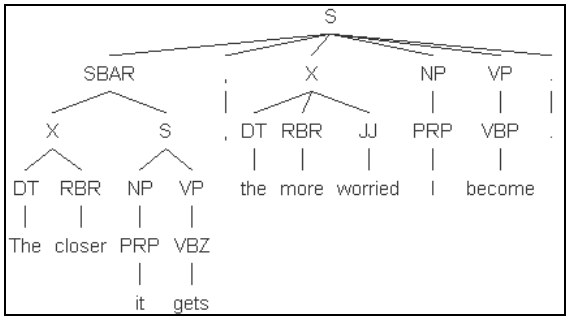

Figure 2. Stanford Parser tree for: *The closer it gets, the more worried I become.*

The Stanford Parser fares quite well in cases like these, since the pronoun (*it, I*) can hardly be modified by the comparative (*[$_{NP}$ *the closer it*] or *[$_{NP}$ *the more worried I*]), and similarly for NPs with articles (*[$_{NP}$ *the closer the time*]). Yet articleless NPs and bare CCs cause problems, as in the tree in Figure 3.

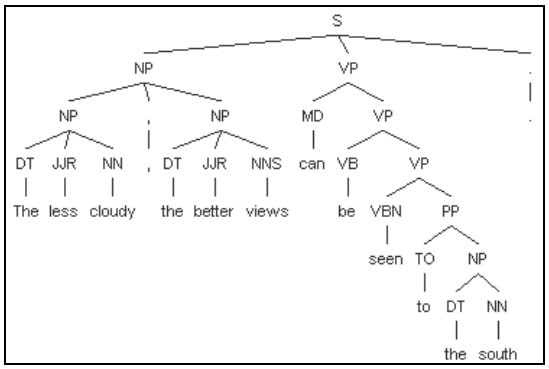

Figure 3. Stanford Parser tree for: *The less cloudy, the better views can be seen to the south.*

Here *The less cloudy* and *the better views* form one NP, separate from the VP complex. Such a reading is not entirely impossible: the sentence could mean 'less cloudy, better views' appositively. However despite the overall greater frequency of appositions and the fact that *less cloudy* has probably not been observed in cc1 in training data, the pattern of a novel form for cc1 and *better* in cc2 is actually consistent with a novel CC. With these ideas in mind, the next section examines the potential of productivity to disambiguate a much more prevalent phenomenon, namely PP attachment.

4 PP Attachment and Productivity

The problem of attaching prepositional phrases as sister nodes of VP or as adjuncts to its object nouns

[11] The X nodes conform to the Penn Treebank II Bracketing Guidelines for CCs (Bies et al., 1995:178).

is a classic case of syntactic ambiguity that causes trouble for parsers (see Hindle and Rooth, 1993; Manning and Schütze, 1999:278-287; Atterer and Schütze, 2007), e.g. the difference between *I ate a fish with a fork* and *I ate a fish with bones*[12], i.e. denoting the instrument or an attribute of the fish. There are also two further common readings of the preposition *with* in this context, namely attached either high or low in the VP in a comitative sense: *I ate a fish with Mary* and *I ate a fish with potatoes* respectively, though most approaches do not distinguish these, rather aiming at getting the attachment site right.

Already in early work on PP attachment (Hindle and Rooth, 1993) it was realized that the lexical identity of the verb, its object, the preposition and in later approaches also the prepositional object noun (Ratnaparkhi et al., 1994) are useful for predicting the attachment site, casting the task as a classification of tuples *<v, n1, p, n2>* into the classes V (VP attachment) and N (NP attachment). Classifiers are commonly either supervised, with disambiguated training data, or more recently unsupervised (Ratnaparkhi, 1998) using data from unambiguous cases where no *n1* or *v* appears. Other approaches supplement this information with hand-built or automatically acquired lexical resources and collocation databases to determine the relationship between the lexemes, or, for lexemes unattested in the tuples, for semantically similar ones (Stetina and Nagao, 1997; Pantel and Lin, 2000).

Although the state of the art in lexically based systems actually approaches human performance, they lose their power when confronted with unfamiliar items. For example, what is the likeliest attachment for the following BNC example: *I can always eat dim-sum with my dybbuk*? It is safe to assume that the (originally Hebrew) loan-word *dybbuk* '(demonic) possession' does not appear in most training datasets, though *dim-sum* is attested more than once as an object of *eat* in the BNC. Crucially, the triple (*eat*, *dim-sum*, *with*) alone cannot reliably resolve the attachment site (consider *soy-sauce* vs. *chopsticks* as n2). It is thus worth examining how likely a novel item is in the relevant slot of each reading's construction. The rest of this section therefore examines productivity scores for the slots in *eat NP with NP* and their correlation with different readings as an example.

Since these cases cannot be identified automatically in an unparsed text with any reliability, and since there is not enough hand-parsed data containing these constructions, a conservative proximity assumption was made (cf. Ratnaparkhi, 1998) and all occurrences of *eat* and related forms within ten words of *with* and with no intervening punctuation in the BNC were evaluated and tagged manually for this study. This also allowed for head-noun and anaphor resolution to identify the referent of a slot in the case of pronominal realization; thus all slot types in the data including pronouns are evaluated in terms of a single head noun.

Results show that out of 131 hits, the largest group of PPs (59 tokens) were object noun modifiers, almost all comitatives[13], justifying the prevalent heuristic to prefer low attachment. However verbal instrumentals and high comitatives (25 and 23 respectively) come at a very close second. The remaining 24 cases were adverbial modifications (e.g. *with enthusiasm*). Looking at hapax legomena in the respective slots we can calculate the measures in Table 3.

	n1 slot		n2 slot		total
	hapax	p3	hapax	p3	tokens
n	39	0.661	45	0.7627	59
v adv	15	0.625	21	0.875	24
v com	8	0.3478	20	0.8696	23
v inst	15	0.6	4	0.16	25

Table 3. p3 for the first and second head noun in nominal and three types of verbal PP attachment for *eat* n *with* n in the BNC.

The scores show that the verbal instrumental reading is the least likely to exhibit a novel head at the n2 slot, which is semantically plausible – the repertoire of eating instruments is rather conventionalized and slow to expand. The comitative reading is very likely to innovate in n2, but much less so in n1, fitting e.g. the "dim-sum with dybbuk"-scenario. This fits the fact that one may eat together with many distinct persons etc., but when

[12] Though in some cases the distinction is not so tenable, e.g. *we have not signed a settlement agreement with them* (Manning and Schütze, 1999:286), where *with them* can arguably be attached low or high. Incidentally, the 'fish' examples are actually attested in the BNC in a linguistic context.

[13] Only 4 hits were truely non-comitative noun modifiers, e.g. *<eat, anything, with, preservatives>*, where a comitative reading is clearly not intended. Since the group was so small, all noun modifiers have been treated here together.

these are specified, the exact nature of the meal or food is often left unspecified[14]. The adverbial reading is likely to innovate in both slots, since many ways or circumstances of eating can be specified and these hardly restrict the choice of object for *eat*. Interestingly, the choice of object maintains a very stable productivity in all but the high comitative construction. n2 innovation in nominal modifiers is actually lower than for adverbials and comitatives, meaning low attachment may not be the preferred choice for unknown nouns.

While these results imply what some reasonable expectations may be to find a novel member of each slot in each reading, they do not take the identity of the lexemes into account. In order to combine the general information about the slot with knowledge of a known slot member, we may simultaneously attempt to score the productivity of the construction's components, namely the noun or verb in question, for PP modifiers. This raises the problem of what exactly should be counted. One may argue that high-attached comitatives and adverbials should be counted separately, since they are almost always optional regardless of the verb (one can equally well eat or do anything else with someone in some way), unlike instrumentals which may be more closely linked to the verb. On the other hand, the exact constructional sense of such PPs is colored by the verb, e.g. eating a meal with someone has a rather particular meaning (as opposed to coincidentally performing the act of eating alongside another eater). If the decision is only between high and low attachment, then grouping all variants together may be sensible in any case.

Depending on the argument and verb, it is possible to make fine distinctions, provided enough cases are found. For *dim-sum*, for example, no cases of NP modifying *with* (novel or otherwise) are found, making the (correct) high comitative reading likely. By contrast, for the head noun *fish*, which is a common object of *eat*, 37 hits with *with*-PPs are found in the BNC, forming 32 prepositional object noun types of which 28 are hapax legomena in this slot. All high readings of *with*-PPs with *eat* (including intransitive *eat*) form 92 tokens, 68 noun types and 44 hapax legomena. Thus *fish + PP* scores p3=0.756 while *eat + PP* scores 0.478, corresponding to less productivity. This means novel prepositional objects are substantially less probable for the high attachment given that the direct object is *fish*.

5 Conclusion

The above results show that similar yet distinct constructions, which vary slightly in either constituent structure (high vs. low attachment), semantics (comitative or instrumental PPs), number of arguments (more and less bare CCs) or position (cc1 vs. cc2), show very different lexical behavior, exhibiting more or less variety in different slots and differing proportions of hapax legomena. The inference which should become apparent from the sharp contrasts in slot scores (especially in p3) given the size of the data, is that these differences are not coincidental but are indicative of inherently different productivity rates for each slot in each construction. These properties need not be attributed to system internal, linguistic reasons alone, but may also very well reflect world knowledge and pragmatic considerations.[15] However, from a construction grammar point of view, the entrenchment of these constructions in speakers and therefore in data is inextricably connected with interaction in the world, thus making syntactic productivity a plausible and relevant quantity both theoretically and potentially for NLP practice.

It remains to be seen whether or not productivity scores can help automatically disambiguate structures with unseen arguments (e.g. PP attachment with unencountered n2), or even distinguish semantic classes such as comitatives, instrumentals etc. for novel nouns, for which a classification into helpful semantic categories (animate, human and so forth) is not available. A large-scale evaluation of this question will depend on how easily and reliably productivity scores can be extracted automatically from data for the relevant constructions.

References

Bengt Altenberg and Mats Eeg-Olofsson. 1990. Phraseology in Spoken English. In: Jan Aarts and Willem Meijs, editors, *Theory and Practice in Corpus Linguistics*. Rodopi, Amsterdam: 1-26.

[14] In fact the non-food specific nouns *breakfast*, *lunch*, *dinner*, *dish* and *meal* cover 16 of the high comitative n1 tokens, almost 70%.

[15] In this context it is worth mentioning that similar ongoing examinations of German CCs reveal different lexical preferences, implying that some of this behavior is language dependent and to some extent language internally lexicalized.

Michaela Atterer and Hinrich Schütze. 2007. Prepositional Phrase Attachment without Oracles. *Computational Linguistics*, 33(4): 469-476.

R. Harald Baayen. 2001. *Word Frequency Distributions*. (Text, Speech and Language Technologies 18.) Kluwer Academic Publishers, Dordrecht / Boston / London.

R. Harald Baayen. 2009. Corpus Linguistics in Morphology: Morphological Productivity. In: Anke Lüdeling and Merja Kytö, editors, *Corpus Linguistics. An International Handbook, vol. 2*. Mouton de Gruyter, Berlin: 899-919.

Jóhanna Barðdal. 2006. Predicting the Productivity of Argument Structure Constructions. In: *The 32nd Annual Meeting of the Berkeley Linguistics Society*. Berkeley Linguistics Society, Berkeley. Available at: http://ling.uib.no/barddal/BLS-32.barddal.pdf.

Laurie Bauer. 2001. *Morphological Productivity*. (Cambridge Studies in Linguistics 95.) Cambridge University Press, Cambridge, UK.

Ann Bies, Mark Ferguson, Karen Katz and Robert MacIntyre. 1995. *Bracketing Guidelines for Treebank II Style Penn Treebank Project*. Technical report, University of Pennsylvania.

Douglas Biber, Susan Conrad and Viviana Cortes. 2004. If you look at…: Lexical Bundles in University Teaching and Textbooks. *Applied Linguistics*, 25(3): 371-405.

Douglas Biber, Stig Johansson, Geoffrey Leech, Susan Conrad and Edward Finegan. 1999. *The Longman Grammar of Spoken and Written English*. Longman, London.

Yaacov Choueka. 1988. Looking for Needles in a Haystack. In: *Proceedings of RIAO '88*. Cambridge, MA, 609-623.

Peter W. Culicover and Ray Jackendoff. 1999. The View from the Periphery: The English Comparative Correlative. *Linguistic Inquiry* 30(4): 543-571.

Marcel den Dikken. 2005. Comparative Correlatives Comparatively. *Linguistic Inquiry*, 36(4): 497-532.

Stefan Evert. 2004. A simple LNRE model for random character sequences. In: *Proceedings of JADT 2004*: 411-422.

Stefan Evert. 2005. *The Statistics of Word Cooccurrences: Word Pairs and Collocations*. PhD dissertation, University of Stuttgart.

Stefan Evert. 2009. Corpora and Collocations. In: Anke Lüdeling and Merja Kytö, editors, *Corpus Linguistics. An International Handbook, vol. 2*. Mouton de Gruyter, Berlin: 1212-1248.

Adele E. Goldberg. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. University of Chicago Press, Chicago and London.

Adele E. Goldberg. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford University Press, Oxford, UK.

Donald Hindle and Mats Rooth. 1993. Structural Ambiguity and Lexical Relations. *Computational Linguistics*, 19(1): 103-130.

Daisuke Kawahara and Sadao Kurohashi. 2005. PP-Attachment Disambiguation Boosted by a Gigantic Volume of Unambiguous Examples. In: *Proceedings of the 2nd International Joint Conference on Natural Language Processing (IJCNLP-05)*: 188-198.

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

Dan Klein and Christopher D. Manning. 2003. Accurate Unlexicalized Parsing. In: *Proceedings of the 41st Meeting of the Association for Computational Linguistics*: 423-430.

Christopher D. Manning and Hinrich Schütze. 1999. *Foundations of Statistical Natural Language Processing*. MIT Press, Cambridge, MA.

James D. McCawley. 1988. The Comparative Conditional in English, German and Chinese. In: *Proceedings of the Fourteenth Annual Meeting of the Merkeley Linguistics Society*. Berkeley Linguistics Society, Berkeley: 176-187.

Patrick Pantel and Dekang Lin. 2000. An Unsupervised Approach to Prepositional Phrase Attachment using Contextually Similar Words. In: *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics*: 101-108.

Adwait Ratnaparkhi. 1998. Statistical Models for Unsupervised Prepositional Phrase Attachment. In: *Proceedings of COLING-ACL98, Montreal Canada*: 1079-1085

Adwait Ratnaparkhi, Jeff Reynar and Salim Roukos. 1994. A Maximum Entropy Model for Prepositional Phrase Attachment. In: *Proceedings of the ARPA Human Language Technology Workshop*. Plainsboro, NJ: 250-255.

Ivan Sag, Timothy Baldwin, Francis Bond, Ann Copestake and Dan Flickinger. 2002. Multiword Expressions: A Pain in the Neck for NLP. In: *Proceedings of the Third International Conference on Intelligent Text Processing and Computational Linguistics (CICLING 2002)*. Mexico City, Mexico: 1-15.

André Salem. 1987. *Pratique des segments répétés*. Institut National de la Langue Française, Paris.

Jiri Stetina and Makoto Nagao. 1997. Corpus Based PP Attachment Ambiguity Resolution with a Semantic Dictionary. In: Jou Zhao and Kenneth Church, editors, *Proceedings of the Fifth Workshop on Very Large Corpora*. Beijing and Hong Kong: 66-80.

Gisela Zifonun, Ludger Hoffmann and Bruno Strecker, editors. 1997. *Grammatik der deutschen Sprache, Bd. 3*. (Schriften des Instituts für deutsche Sprache 7.) De Gruyter, Berlin / New York.