Predicting the Productivity of Argument Structure Constructions

JÓHANNA BARDDAL

University of Bergen / University of California, Berkeley

0. Introduction

Productivity has been a much-debated concept within morphology and word formation for years, while syntactic productivity has received little attention in the research community. In mainstream syntax the concept of syntactic productivity is focused on speakers’ ability to generate and understand sentences never encountered before. This notion of productivity captures what has been labeled ‘linguistic competence’ within the generative paradigm, but does not address the interesting question of how case and argument structure constructions are extended to new verbs. A usage-based construction grammar that takes argument structure constructions to exist at different levels of schematicity offers an account of that. I argue that productivity is a function of type frequency and coherence (i.e. internal consistency) and that the productivity of argument structure constructions is predictable from their type frequency and semantic coherence, and an inverse correlation between the two. Such a view offers a unified account of the complexity of the productivity concept, including the relation between the different subconcepts of productivity, i.e. those of regularity, generality, and extensibility. Also, facts of different degrees of productivity as well as analogy fall out directly from this approach.

1. The Concepts of Productivity

In my work on the productivity of argument structure constructions, I have found that the concept of productivity is a highly complex and multi-faceted one (Barðdal 2000:23-5, 2008). Not only are the terms productive and productivity used in many different senses in the linguistic literature, but there are at least three different subconcepts of the productivity concept around (cf. Barðdal 2008:Ch. 2). Consider the following statements, all from Bybee (1995):

In this section and the next, I will argue that default status or productivity is not necessarily associated with source-oriented rules. (Bybee 1995:444)
Clahsen and Rothweiler present considerable evidence that the productive or regular affix for past-participles in German is the -t affix … Thus they claim that the determinant of productivity or regularity is not type frequency. (Bybee 1995:435)

… certain classes of verbs within the irregulars show some degree of productivity – that is, the ability to be extended to new items. (Bybee 1995:426)

The first quote above illustrates that productivity has been equated with default status in the literature, but default status is generally considered a consequence of how open or general a pattern is. In the second quote, regular is used synonymously with productive. In the third quote, productivity is defined as the ability of a pattern to be extended to new items. The use of the terms productive and productivity in the three quotes above are by no means unique in the literature. I have found these terms used in at least nineteen different senses in the linguistic literature alone (Barðdal 2008:Ch. 2), and all nineteen can be related in one way or another to the concepts of generality, regularity, and extensibility. This suggests that the productivity concept consists of at least three different subconcepts, namely the following:

1. Productivity as GENERALITY
2. Productivity as REGULARITY
3. Productivity as EXTENSIBILITY

The reason that these three concepts have become associated with productivity is presumably that they all co-occur in the largest and most productive categories discussed in the literature. Consider, for instance, regular past-tense formation in English, materialized with the dental suffix -ed. Approximately 96% of all English verbs form past tense with this suffix (Ragnarsdóttir, Simonsen, and Plunkett 1999). This past-tense formation is also the one extended most to both new and nonce verbs, and the one that has attracted most verbs from other patterns in the history of English. This pattern is totally regular, with the suffix -ed attached to the stem. There are, moreover, no semantic, morphological, or phonological constraints on which verbs can form past tense in this way. Hence, this is a completely general pattern. Therefore, in the category of regular verbs in English all three subconcepts of productivity – REGULARITY, GENERALITY, and EXTENSIBILITY – are combined.

Figure 1. The three subconcepts of productivity
The relation between the subconcepts can be modeled as in Figure 1 above. The most regular patterns in a language are usually also the most general patterns. Hence, these two properties are often concomitant with each other, represented with \( Co \) in Figure 1 for ‘concomitance relation’. The most regular pattern is also the one that is most extended to either new or existing items in a language, exactly as the most extensible pattern is usually the most regular pattern. Regularity and extensibility are therefore concomitant with each other, even though they are not logically dependent on each other. The most general pattern is usually also the most extensible pattern, whereas extensible patterns need not be the most general ones. This is indicated with \( Ent \) for ‘entailment relation’ in Figure 1. In other words, generality entails extensibility, whereas extensibility does not presuppose generality, as restricted patterns can also be extended to new or existing items, although they are extended much less than non-restricted patterns. In the next section, I will show that the regularity and the generality subconcepts can be treated as derivatives of the extensibility concept of productivity.

2. Predicting Productivity
Following Bybee (1995) and Clausner and Croft (1997), I assume that productivity is a function of type frequency and coherence, i.e. internal consistency, and that there is an inverse correlation between the two. For a category that is high in type frequency, this means that only a low degree of internal consistency across its items is needed for the category to be extended. Categories of intermediate type frequency have to be more coherent than high type-frequency categories in order to be extended to new items. Finally, categories that are low in type frequency must show the highest degree of internal consistency across their items in order to be extended to new types. This can be represented graphically as in Figure 2 below, where the vertical axis stands for type frequency and the horizontal one for coherence, in this case semantic coherence, as argument structure constructions are regarded as semantically defined (form-meaning correspondences) in a usage-based construction grammar (Goldberg 1995; Barðdal 2001a, 2001c, 2004, 2006; Barðdal and Molnar 2003; Croft 2001, 2003). Observe that high type-frequency patterns are located at the upper leftmost end of the productivity cline in Figure 2, as the patterns highest in type frequency also show the lowest degree of coherence, exactly because of their high type frequency. Patterns lower in type frequency will be located lower on the vertical axis, and the higher the degree of coherence found across their instantiating items, the closer they are to the productivity cline going from the upper leftmost corner to the lower rightmost corner in Figure 2.

What, then, about the lower rightmost end of the productivity cline in Figure 2? How does productivity manifest itself there? Obviously, the lowest possible type frequency is only one, and the highest possible degree of semantic coherence is full synonymy. In other words, only one item will be needed as a model for the extensibility of a pattern to a new lexical item, if the degree of semantic coherence
is high enough. Such a process is usually labeled ‘analogy’ in the linguistic literature and considered as fundamentally different from productivity (cf. Bauer 2001:93). On the present approach, however, full productivity and analogy are two sides of the same coin, not different in ontological status but in degree. I return to this issue in section 3 below.

**Figure 2.** The inverse correlation between type frequency and coherence

Observe that at the top of the productivity cline where we find the most extensible categories, we also find the most general and the most regular categories in languages. Hence, not only does the present model of productivity based on the EXTENSIBILITY concept provide us with the cline from full productivity via different degrees of productivity to analogical extension, the subconcepts of REGULARITY and GENERALITY can also be derived from it.

I assume, moreover, that constructions exist at different levels of schematicity in the schematicity-lexicality continuum (Croft 2003; Barðdal 2000, 2001a, 2004, 2006). The lowest level consists of concrete, lexically filled instantiations of the construction; the next level above that contains verb subclass-specific constructions; the next level contains verb class-specific constructions; and the level above contains event-type constructions, etc. Hence, every level above the lowest, most concrete level is an abstraction across the items of the level below. The lowest level contains the most lexical and semantic information, whereas the highest, most schematic level contains the least lexical and semantic information. This is the most abstract level, hence the term ‘schematic’. This organization of argument structure constructions is given in Figure 3 below for two argument structure constructions differing in type frequency and hence in the number of intermediate levels, as well as in their highest level’s degree of schematicity.
Observe that on this bottom-up approach, the highest level of schematicity is derived from the lower levels. That is, as every level is an abstraction of its subordinate level, each construction’s highest level is dependent on the number of intermediate levels. These in turn are highly dependent on the type frequency of each construction. To give an example, compare the type frequency of the ditransitive construction in English with that of the transitive construction. Goldberg (1995: 129) lists 69 ditransitive verbs divided across nine semantic verb classes. This means that there are at least 69 verb-specific constructions of the ditransitive construction in English and at least nine verb subclass-specific constructions. Some of the nine verb classes belong more closely together semantically than others. For instance, verbs of giving, ballistic motion, and bringing all entail that the agent successfully causes the recipient to receive the moved object. Verbs of creation and obtaining both entail intended transfer (Goldberg 1995:37-9). This means that the verb class-specific constructions are only six, as some of the verb subclass-specific constructions are subsumed under the more abstract verb class-specific constructions. All the six verb class-specific constructions are based on the notion of transfer; hence, the ditransitive construction’s highest and most schematic level only contains information about the event type ‘transfer’.

Compare now the ditransitive construction with the ordinary transitive. Although I am not aware of any counts of the type frequency of the transitive construction in English, I have carried out a count of the type frequency of the transitive construction in Icelandic. A simple count based on a bilingual Icelandic-English dictionary yields 2,119 verbs (see Table 1 below). This is probably the lowest possible estimate of the type frequency of the transitive construction in Icelandic, and there is no reason to assume that there are any fewer transitive verbs in English. A type frequency of approximately 2,100 yields around 2,100

---

1 See, however, Barðdal (2007) for a comparative analysis of the ditransitive construction in Germanic, where it is argued that the semantic scope of the ditransitive stretches far beyond the notion of transfer, not only in Icelandic, Mainland Scandinavian, and German, but also in English.
verb-specific constructions at the bottom of the schematicity-lexicality continuum; hence, the verb subclass-specific constructions are bound to be considerably more than nine for the transitive. This also means that there are many more intermediate levels of schematicity found for the transitive than the ditransitive, and the event-type level will denote many more event types than only transfer. Thus, the transitive construction’s highest level of schematicity is bound to be considerably higher and more schematic than the highest level of the ditransitive. The proportion between the transitive and the ditransitive may thus correspond approximately to the proportions between the two hypothetical constructions in Figure 3. As a consequence, the transitive construction should be much more productive than the ditransitive, as the former exists at a much higher level in the schematicity-lexicality continuum. In other words, the productivity of each construction is a derivative of the construction’s highest level of schematicity.

Clausner and Croft (1997) take productivity to be predictable not from a construction’s highest level of schematicity, but from its most entrenched level. They suggest a representation of gradient productivity as in Figure 4 below:

**Figure 4.** Gradient productivity on Clausner and Croft’s (1997:271) approach

To give an example, consider English past-tense formation again. The regular past-tense formation with the dental suffix -ed is definitely the highest one in type frequency, which in practice means a highly entrenched schema and full productivity (4a). The irregular past-tense formation with vowel change in the stem is very low in type frequency, although the instances are very high in token frequency (4b). This means that it is not the schema that is entrenched but the instances, resulting in low productivity. Finally, suppletive forms, like go-went, are completely idiosyncratic; thus, they do not form a schema and are as such unproductive (4c). Hence, it is the most entrenched level of schematicity that is the best predictor of productivity; if it is the schema the pattern will be productive, whereas if it is the instances the pattern will be unproductive.

While I agree with Clausner and Croft, and Bybee, in principle, I disagree with them on which factor is predictive of productivity. They argue that it is each construction’s most entrenched level, while I claim that it is each construction’s highest level of schematicity. On their view, the irregular past-tense formation in English is only semi-productive because the instances are so high in token frequency, which in turn means that it is the lowest verb-specific level that is en-
Predicting the Productivity of Argument Structure

trenched and not the higher schematic level. In contrast, I argue that the irregular past-tense formation is low in productivity because its highest level of schematicity is much lower than that of the regular past-tense formation. Hence, I believe that the gradience of productivity is better represented as in Figure 5.

**Figure 5.** Gradient productivity on the present approach

This figure better captures the fact that the differences in type frequency found for highly productive vs. semi-productive patterns also entail the existence of intermediate schematic levels for the more productive pattern, which again results in a higher level of schematicity for highly productive constructions, but a lower level of schematicity for semi-productive constructions. Hence, the difference between highly productive and semi-productive patterns is not necessarily a difference in the entrenchment of a super-ordinate level, but rather a difference in both the type frequency of the two constructions and in the levels of schematicity at which they exist. This last fact, that a schema’s type frequency and its degree of schematicity are to a large extent concomitant with each other, at least for high type-frequency constructions, is not adequately represented in Figure 4.

I now turn to the productivity of case and argument structure constructions in Icelandic.

3. **Productivity of Argument Structure Constructions**

In order to study the productivity of argument structure constructions, I have collected verbs from the area of information technology borrowed into Icelandic. I mostly confine my material to the borrowings I found at the discussion forum for Icelandic Mac users. Only transitive argument structure constructions are included, but as is well known, Icelandic has several case-marking patterns for two-place predicates, including Nom-Acc, Nom-Dat, and Nom-Gen (cf. Barðdal 2001b:51-6). These borrowings are listed in (1) and (2) below (spelling original):
Jóhanna Barðdal


It is worth pointing out that the verbs assigned the Nom-Dat construction are not verbs whose object is necessarily human, as is the case with many dative-marked arguments in the world’s languages. In fact, only two of the borrowed Nom-Dat verbs select for a human argument, namely msna ‘communicate with somebody through MSN’ and neimdroppa ‘namedrop’:


The remaining Nom-Dat verbs select for ordinary non-human objects. Also, since most of these verbs are borrowed from English, which does not have a morphological dative, these verbs have not been borrowed together with their source language’s verb-specific argument structure constructions, which would otherwise be a possibility (cf. Barðdal 2001b:132-3). Prior research on 15th-century Icelandic has shown the same tendency for borrowed verbs, namely that the dative object is
not necessarily human (Barðdal 1999). Hence, dative case-marking of ordinary non-human objects is certainly not new in Icelandic.

Of the 107 borrowed verbs listed in (1)-(2) above, Icelandic speakers assign the Nom-Acc construction to 68 verbs and the Nom-Dat construction to 39. The proportion between the Nom-Acc and Nom-Dat is thus less than 2:1, which is certainly unexpected given that the dative is not reserved for human arguments. A question arises as to what the proportion between the type frequencies of Nom-Acc and Nom-Dat constructions is otherwise in Icelandic. I have elsewhere (Barðdal 2008:Ch. 3) carried out two counts of the transitive construction in Icelandic. One is based on the lexical entries in a recent bilingual Icelandic-English dictionary (Hólmarsson, Sanders, and Tucker 1989), and the other is based on a corpus of 40,000 running words from six different genres of Modern Icelandic texts (cf. Barðdal 2001b). The results are presented in Table 1 below:

<table>
<thead>
<tr>
<th></th>
<th>Dictionary Count</th>
<th>Corpus Count</th>
<th>Borrowed Verbs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>f</td>
<td>N</td>
</tr>
<tr>
<td>Nom-Acc</td>
<td>1,381</td>
<td>65.2%</td>
<td>303</td>
</tr>
<tr>
<td>Nom-Dat</td>
<td>738</td>
<td>34.8%</td>
<td>188</td>
</tr>
<tr>
<td>Total</td>
<td>2,119</td>
<td>100.0%</td>
<td>491</td>
</tr>
</tbody>
</table>

Observe that the figures for the borrowed verbs are in conformity with both the absolute figures from the dictionary count and the relative figures from the corpus count. Nom-Acc verbs are around 62-65%, and Nom-Dat verbs around 35-38%. The type frequency of both the Nom-Acc and the Nom-Dat construction is thus reflected in the type frequency of borrowed verbs. In other words, the productivity of the two constructions matches their type frequency in Modern Icelandic.

An analysis of the semantic range of the transitive construction in Icelandic only based on the 303 Nom-Acc and 188 Nom-Dat predicates occurring in the corpus of Modern Icelandic texts reveals that Nom-Acc verbs divide across 46 semantic verb classes and the Nom-Dat across 33 verb classes. Due to space limitations I cannot lay out the semantic structure of these constructions in a graphically coherent way here, but I refer the reader to Barðdal (2008:Ch. 3). At the highest schematic level, i.e. the event-type level, the Nom-Acc construction consists of at least six basic humanly relevant notions, i.e. MAKING, MOVEMENT, AFFECTEDNESS, COGNITION/EMOTION, CHANGE, and LOCATION. Each of these in turn consumes between two and twelve verb class-specific constructions, and one of the verb class-specific constructions under COGNITION/EMOTION can be divided into several verb subclass-specific constructions. In contrast, the Nom-Dat construction only subsumes three event-type constructions, i.e. CHANGE, COGNITION/EMOTION, and SUPERIORITY. There are also several Nom-Dat verb class-specific constructions which do not fall under any of these labels, nor do they make up an event type of their own. The Nom-Dat construction thus does not exist at as high a schematic level as the Nom-Acc construction.
Notice that some of the general event-type categories are shared across the Nom-Acc and the Nom-Dat, e.g. COGNITION/EMOTION and CHANGE, which are found for both the Nom-Acc and the Nom-Dat construction. However, the subclasses are not necessarily the same for both constructions. The Nom-Acc verbs in (1) above belong to the following narrowly circumscribed semantic classes: verbs of attaching, attempting, building, cognition and mental activity, creation and reshaping, (means of) cutting, (de)limitation, destruction, displaying, producing, dwelling and keeping in place, emotion, feeding and consumption, (means of) gaining, increasing and strengthening, human manipulation, meeting and uniting, non-translational motion, physical affectedness, possession, catching and termination, (means of) traveling, and (interactive) verbal behavior.

The Nom-Dat verbs in (2) above belong to the following narrowly circumscribed semantic classes: verbs of destruction, division, starting and finishing, increase and decrease, losing, (caused) motion, non-translational motion, organizing, stealing, (means of) verbal communication, and changing. As evident from the lists in (1)-(2), some verbs are in fact listed twice, i.e. as taking both Nom-Acc and Nom-Dat. In those cases there is a clear difference in both semantic construal and subconstruction of the relevant argument structure construction. Consider the following two examples of bomba ‘bomb’:

(5) þá voru Kanarnir að bomba Júgóslava.  
affectedness
‘Then the Americans were bombing the Yugoslavians.’

(6) Mér tókst að bomba tannkremi í augað á mér.  
caused-motion
‘I managed to bomb toothpaste into my eye.’

In (5) bomba denotes affectedness, whereas in (6) it is used in the caused-motion construction, where the moved object is in the dative case (cf. Barðdal 2003).

Finally, consider the following example:

(7) Vörubíllinn donaði uppi í brekkunni.  
truck the  dawned up in slope the
‘The truck ended up/was forgotten on the slope.’

In this example the English verb dawn has been borrowed into Icelandic and used as a particle verb dona uppi on the basis of the Icelandic particle verb daga uppi, which here means ‘end up somewhere/be forgotten’. The predicate daga uppi consists of the simple verb daga meaning ‘dawn’ and the locative particle uppi ‘up’, originally used in reference to night trolls who were caught by daylight. Obviously, if one is a night troll and is caught by daylight, one does not constitute a threat to the environment anymore, and hence one may simply end up somewhere and gradually be forgotten. Observe, moreover, that the predicate daga uppi ‘be
caught by daylight, be forgotten’ is a lexicalized idiom whose meaning is non-compositional and not derivable from the meaning of the parts. This is an idiosyncratic set phrase; hence, the formation of the predicate *dona uppi* is not based on a pattern, rule, or a super-ordinate schema, but on the already existing *daga uppi* and that alone. In other words, *dona uppi* is formed on the basis of the Icelandic expression *daga uppi*, where we find the lowest possible type frequency, i.e. *one*, and the highest possible degree of semantic coherence, i.e. *full synonymy*. This is therefore a clear-cut example of analogy, located at the bottom of the productivity cline in Figure 2 above.

In this section I have presented examples of three argument structure constructions in Icelandic: the Nom-Acc construction, which is a semantically open, high type-frequency construction; the Nom-Dat construction, which is a semantically more restricted construction of intermediate type frequency; and finally a specific example of a new particle verb with a very specific meaning, based on only one model verb. These three cases represent three points on the productivity cline, i.e. full productivity, intermediate productivity, and low productivity.

4. Conclusions

The extensibility of argument structure constructions is a function of their type frequency and semantic coherence, and an inverse correlation between the two. The current approach to syntactic productivity focusing on the EXTENSIBILITY concept offers a unified account of productivity since it captures different degrees of productivity ranging from highly productive patterns to various intermediate degrees of productivity to low-level analogical modeling. The REGULARITY and the GENERALITY concepts can also be derived from this analysis.

References


