Appliction to Syntactic Argument Selection

- Similar phenomena can be observed in syntax:
  - Verbs can be more or less frequent (N)
  - Some verbs have more varied arguments than others (V)
  - are more or less likely to govern novel arguments (V1)
- We can measure N, V, V1, P, S for accusative objects:

<table>
<thead>
<tr>
<th>lemma</th>
<th>N</th>
<th>V</th>
<th>V1</th>
<th>P</th>
<th>S</th>
<th>(V_{V1})</th>
<th>(P_{V1})</th>
</tr>
</thead>
<tbody>
<tr>
<td>eat</td>
<td>3389</td>
<td>1561</td>
<td>984</td>
<td>0.868</td>
<td>1023</td>
<td>0.219</td>
<td>0.041</td>
</tr>
<tr>
<td>incur</td>
<td>3893</td>
<td>205</td>
<td>121</td>
<td>0.031</td>
<td>3506</td>
<td>0.444</td>
<td>0.041</td>
</tr>
<tr>
<td>harbor</td>
<td>787</td>
<td>456</td>
<td>248</td>
<td>0.152</td>
<td>1255</td>
<td>0.278</td>
<td>0.041</td>
</tr>
<tr>
<td>eat</td>
<td>9369</td>
<td>1561</td>
<td>984</td>
<td>0.868</td>
<td>1023</td>
<td>0.219</td>
<td>0.041</td>
</tr>
<tr>
<td>spn</td>
<td>3389</td>
<td>1561</td>
<td>984</td>
<td>0.868</td>
<td>1023</td>
<td>0.219</td>
<td>0.041</td>
</tr>
<tr>
<td>push</td>
<td>3389</td>
<td>1561</td>
<td>984</td>
<td>0.868</td>
<td>1023</td>
<td>0.219</td>
<td>0.041</td>
</tr>
<tr>
<td>sprn</td>
<td>3389</td>
<td>1561</td>
<td>984</td>
<td>0.868</td>
<td>1023</td>
<td>0.219</td>
<td>0.041</td>
</tr>
</tbody>
</table>

Productivity Rankings

We get different rankings based on different criteria:
- High P means high potential productivity (novelties expected)
- High V means high realized productivity (used often so far)
- High N means high usage (forms are central to language use)
- High S means low saturation (many new uses not explored yet)

Why is this important?
Or: Is Lexical Semantics Enough?

- In algebraic models categorical distinctions explain argument filling:
  - \{edible\} \(\Rightarrow\) possible object of eat
- In some cases, this leads to circular argument definition:
  - \{incurable\} \(\Rightarrow\) possible object of incur (???)
- Do we need to know how productive a construction is to use it right?

The case of wegen

- As an example, consider German wegen ‘because’ in 3 synonymous constructions (Petig 1997, Helbig & Buscha 2001:356):
  - Preposition with genitive: wegen des Vaters [standard, formal]
  - Preposition with dative: wegen dem Vater [colloquial nonstandard]
  - Postposition with genitive: des Vaters wegen [formal, archaic]
- Intuitively, the postposition is going out of use but still productive (novel arguments are found for all three variants)

DEWAC corpus, with case ambiguities

DEWAC corpus, no case ambiguities

Background from Morphological Theory

- Morphological processes can be more or less productive (Bauer 2001):
  - Neologisms in -tum are possible: Syntaxktetrum ‘syntactician-dom’
  - but not as likely as ones in -keit: Miniatursierbarkeit ‘miniaturizability’
- Different attempts have been made to measure productivity (Baayen 2009)
  - Based on token frequency (N) - e.g. nouns in -keit are very frequent
  - Type frequency or vocabulary (V) - there are many such nouns
  - Unique, potentially novel forms (hapax legomena, V1) – there are many nouns found only once in large corpora
- With increasing sample size N, it becomes harder to find new types, V rises more slowly, and the probability drops that V1 increases (P)
- We can chart the rise of V with growing N and estimate the limit S of V’s growth using statistical models (Evert 2004)
- A typical example: German adjectives in -sam-bar (cf. Lüdeling et al. 2000)

Development of accusative object vocabulary

Productivity Rankings

- Prepositional forms are more productive than the postposition in all respects: not just more frequent, but higher V, V1, P for the same N
- Algebraic grammar cannot explain why postpositive wegen takes novel objects less often (semantically compatible with same objects)
- Usage based approaches assume gradient productivity, predict lower S and explain how speakers know not use postpospositive wegen as readily with novel arguments
- Hypothesis for further study: rarity of hapax legomena leads to postposition being acquired as less productive since speakers reproduce the input frequency distribution

Literature: