Novel Argument Realization: Semantic, Pragmatic and Conventional Productivity Effects
Linguistic Evidence 2012, Tübingen, 9-11 February 2012

Research question and intuitions
- Some verbs are more restricted than others in allowing novel or unlexicalized arguments:
  1. pose/represent a challenge
  2. pose/represent a provocation
- Verbs like pose preferably appear with collocated objects
- They occur with much fewer arguments in corpus data

Questions:
- What determines the spectrum of realized arguments?
- Can lexical semantic classes predict argument realizability?
- Are there differences between verbs motivated by pragmatics/world knowledge?
- Are there idiosyncratic effects that cannot be derived from verb meaning?

Semantic classes and argument selection
- Lexical semantics regards argument slots as realizing semantic classes (Katz & Fodor 1963, Jackendoff 1990), as in (3).
  
  (3) `drink(agent, patient)`
  
  (4) `/phon <drink>`
  
  SYNSEM CAT
  
  RELN DRINKER drink
  
  CONTENT DRINK DRUNK
  
  Using such classes it is possible to account for any argument spectrum:
  
  (5) `pose(agent, patient)`
  
  • Risk of circular logic: turning semantic classes into a tautology (cf. Dowty 1991)
  • In order for semantic classes to predict novel arguments, classes should be:
    1. Cognitively plausible
    2. General, i.e. applying to as many predicates as possible
    3. As specific as necessary, cf. McCawley 1968: diagonalize[patient]`\{\text{matri}\}`
  
  Classes we define should be preserved under decomposition, cf. Jackendoff (1990):
  
  (6) `\langle \text{drink} \rangle`
  
  `\langle \text{event} \rangle` `\langle \text{cause} \rangle` `\langle \text{go} \rangle` `\langle \text{thing} \rangle` `\langle \text{liquid} \rangle` `\langle \text{mouth of (\text{thing} \rangle} \rangle` `\langle \text{place} \rangle`

Methodology – productivity in argument selection
- Multiple measures are useful for comparing argument productivity (cf. Baayen 2001 on morphological productivity):
  - Frequency N(C)
  - Hapax legomena V
  - Vocabulary V
  - Estimated total vocabulary S (s. Evert 2004)
- Different rankings depending on the measure selected:

<table>
<thead>
<tr>
<th>Rank</th>
<th>Frequency N(C)</th>
<th>Type frequency V</th>
<th>Hapax frequency V</th>
<th>Total vocabulary S</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>achieve</td>
<td>36121</td>
<td>398</td>
<td>5377.586</td>
</tr>
<tr>
<td>2</td>
<td>spend</td>
<td>28748</td>
<td>323</td>
<td>4343.072</td>
</tr>
<tr>
<td>3</td>
<td>eat</td>
<td>16201</td>
<td>319</td>
<td>3506.464</td>
</tr>
<tr>
<td>4</td>
<td>push</td>
<td>9380</td>
<td>307</td>
<td>3022.019</td>
</tr>
<tr>
<td>5</td>
<td>incur</td>
<td>3893</td>
<td>190</td>
<td>2588.051</td>
</tr>
<tr>
<td>6</td>
<td>drink</td>
<td>3293</td>
<td>148</td>
<td>2101.245</td>
</tr>
<tr>
<td>7</td>
<td>harbour</td>
<td>1781</td>
<td>100</td>
<td>1255.090</td>
</tr>
<tr>
<td>8</td>
<td>defy</td>
<td>1705</td>
<td>74</td>
<td>1245.033</td>
</tr>
</tbody>
</table>

What does this mean? 3 case studies
1. Translational pairs
   - Rankings = real world differences? (e.g. [+edible] > [+drinkable])
   - If so, we expect similar productivity cross-linguistically
   - Counter examples can be found in:
     - Lexeme pairs: e.g. En. `harbor` > De. `hegen`, with [+mental state]
     - Lexically unspecified constructions, e.g. De. `je Xer desto Yer` > En. `the Xer the Yer`

2. Near synonyms
   - Synonyms are defined as interchangeable salva veritate
   - Should select same semantic class, but again we find differences:

3. Syntactic alternations
   - Same head, different constructions: again significant differences

Extensible slots in the mental lexicon – A Hebbian account
- Lexical semantics cannot explain differences in these minimal pairs
- Speakers somehow reproduce input distributions:
  - Argument slots with large V. V1 produce more arguments in unseen data
  - I assume Hebb’s Law strengthens connected representations of constructional slots and attested arguments
  - Activation by hapax legomena strengthens almost only the construction itself, without creating an entrenched argument
  - Slots with too frequent prototypes and too few rare items become identified with those arguments and tend to activate only with them

References