Overview: What is grammar and what is processing?

1. Actually, that’s way too big a question
   - I’ve seen a lot of criteria proposed (or just used) that don’t convince me:
     - if it’s frequency-sensitive, it’s not grammar
     - if it’s variable, it’s not grammar
     - if it’s phonetically gradient, it’s not grammar
   - These all rely on a-priori assumptions about what grammar can be, but if we’re trying to figure out what grammar is, we can’t make those assumptions.
   - Instead, in the first half of today let’s look at some frequency effects and how they could fit in to our model of language; in the second half, we’ll look at consequences of speech planning and lexical access occurring in real time.

2. Classic frequency effect: English irregular verbs
   - There are only about 200 of them, but they are disproportionately likely to be frequent (e.g., Bybee & Slobin 1982).
   - Top 25 most frequent verbs (Oxford English Corpus)—irregulars are in bold:
     1. be
     2. have
     3. do
     4. say
     5. get
     6. make
     7. go
     8. know
     9. take
     10. see
     11. come
     12. think
     13. look
     14. want
     15. give
     16. use
     17. find
     18. tell
     19. ask
     20. work
     21. seem
     22. feel
     23. try
     24. leave
     25. call
   - Locus of explanation?

Diachrony
   - In order to learn an irregular past tense form, you have to be exposed to it enough times → low-frequency verbs will tend to regularize from one generation to the next (bode > bided).
   - Kirby 2001: simulation study
Processing
Dual-route model (see Pinker 2000 for overview and application to this case)
- When you want to say a past tense, there’s a race between retrieving a stored form (which might be irregular) and creating the form via the –ed rule.
- The more frequent stored form → higher resting activation → more likely to win the race.
  - low-frequency verbs may get pronounced as regular, even if speaker knows irregular form.

Grammar? (I don’t think anyone has proposed it for this case, but it’s a logical possibility)
- Some constraints are sensitive to frequency.

<table>
<thead>
<tr>
<th>/bowd/, cf. [bajd]</th>
<th>I-O FAITH(hi freq)</th>
<th>O-O FAITH</th>
<th>I-O FAITH(lo freq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>bowd</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>bajdid</td>
<td></td>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>

- Or there’s just one I-O FAITH constraint, but its ranking is a function of frequency

3. Ng 2010: Singapore English prosodic boundaries
- Singapore English has strong glottalization at prefix-stem, stem-stem, but not stem-suffix boundary
  - *mis-understand* [misʔandastæn], *stop-over* [stɔpʔovə], *magic-al* [mædʒikʊəri](p. 8)
    - Ng analyzes this in terms of p-word structure: let’s sketch it out

- Tone pattern is roughly \( L^* (\ 'M M^* ) H \) (p. 11)
- Domain of tone assignment \( \approx \) p-word
  - tone pattern generally re-starts in compounds: *century egg* (\( MH \),\( H \)) (p. 13)
  - tone pattern may or may not restart at prefix-stem boundary: *un-install* (\( H \),\( L'H \) ~ (\( L-(L'H) \)) (p. 12)
  - tone doesn’t restart at stem-suffix boundary: *remove-able* (\( L'MMH \)) (p. 12)
- Initialisms show varying degrees of prosodic merger:

<table>
<thead>
<tr>
<th>Least merger</th>
<th>Most merger</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. SPCA</td>
<td>((((H),H),H),H)</td>
</tr>
<tr>
<td>b. ACIC</td>
<td>((((H),H),H),H) ((H),MH)</td>
</tr>
<tr>
<td>c. NRIC</td>
<td>((((H),H),H),H) ((H),MH) ((H),MMH)</td>
</tr>
<tr>
<td>d. NTUC</td>
<td>((((H),H),H),H) ((H),MH) ((H),MMH) (MMMH)</td>
</tr>
</tbody>
</table>

(p. 23)
- Ng finds correlation between which group an initialism belongs to and its # of Google hits.

**Why?**

- Frequency determines speed of production, perhaps because of faster access:

  ![Figure 4: Frequency and duration](image)

  (p. 31)

- Constraints are then sensitive to speed, e.g. “Grammatical word accessed at speed $n$ allows only $n$ levels of stress”
  - Result is a prosodification of higher-frequency words that results in fewer stresses.
  - This is an interesting way of removing the need for the grammar to refer to frequency
    - Predicts that if we can manipulate speaking rate or retrieval speech independent of word frequency, we’ll get similar effects.

  (58) *Do not restore faithful stress to distressed initialisms*

<table>
<thead>
<tr>
<th>$[[M][O][E]]_3$</th>
<th>$\text{WRAP} : *\text{SCLASH}$</th>
<th>$\text{STRESS}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. $((M),O,E)$</td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>b. $((M),O,E)$</td>
<td>$W^*$</td>
<td>L</td>
</tr>
<tr>
<td>c. $((M)(O),E)$</td>
<td>$W^*$</td>
<td>L</td>
</tr>
<tr>
<td>d. $(M)(O)(E)$</td>
<td>$W^*$</td>
<td>L</td>
</tr>
</tbody>
</table>

(p. 33)

4. **Hammond 1999: English rhythm rule**

*thirteen mén* or *thirtèen mén?*

- In survey, shift is more likely if adjective is more frequent: *nàive friend* vs. *obèse child*
- Hammond proposes morpheme-specific faithfulness constraints, whose ranking depends on the word’s frequency.
5. Löfstedt 2010: frequency-specific constraints

- Famous paradigm gaps in Swedish result when vowel shortening produces too much of a quality change.

<table>
<thead>
<tr>
<th>STEM</th>
<th>NEUTER</th>
<th>GLOSS</th>
<th>ALLÉN (PL)</th>
<th>GOOGLE (-A)</th>
</tr>
</thead>
<tbody>
<tr>
<td>bl[ɔːː:]</td>
<td>bl[ɔː] +ː</td>
<td>‘blue’</td>
<td></td>
<td></td>
</tr>
<tr>
<td>v[ːː]t</td>
<td>v[ːː] +ː</td>
<td>‘white’</td>
<td></td>
<td></td>
</tr>
<tr>
<td>v[ːː]d</td>
<td>v[ːː] +ː</td>
<td>‘wide’</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(p. 152)

- But! Sufficiently frequent words don’t have a gap

<table>
<thead>
<tr>
<th>STEM</th>
<th>NEUTER</th>
<th>GLOSS</th>
<th>ALLÉN (PL)</th>
<th>GOOGLE (-A)</th>
</tr>
</thead>
<tbody>
<tr>
<td>gr[ɑːː]</td>
<td>INEFFABLE</td>
<td>‘straight’</td>
<td>0</td>
<td>7,140</td>
</tr>
<tr>
<td>l[ɑːː]t</td>
<td>INEFFABLE</td>
<td>‘lazy’</td>
<td>0</td>
<td>581,000</td>
</tr>
<tr>
<td>gl[ɑːː]d</td>
<td>gl[ɑː] +ː</td>
<td>‘happy’</td>
<td>29</td>
<td>2,110,000</td>
</tr>
</tbody>
</table>

(p. 154)

- For each of the vowels that can show a gap, there seems to be a frequency cut-off above which there’s no gap. (Löfstedt shows this for some phenomena in other languages too) E.g.,

<table>
<thead>
<tr>
<th>STEM</th>
<th>NEUTER</th>
<th>GLOSS</th>
<th>ALLÉN (PL)</th>
<th>GOOGLE (-A)</th>
</tr>
</thead>
<tbody>
<tr>
<td>gr[ɑːː]</td>
<td>INEFFABLE</td>
<td>‘straight’</td>
<td>0</td>
<td>7,140</td>
</tr>
<tr>
<td>l[ɑːː]</td>
<td>INEFFABLE</td>
<td>‘lazy’</td>
<td>0</td>
<td>581,000</td>
</tr>
<tr>
<td>gl[ɑːː]d</td>
<td>gl[ɑː] +ː</td>
<td>‘happy’</td>
<td>29</td>
<td>2,110,000</td>
</tr>
</tbody>
</table>

(p. 154)

- Löfstedt’s solution: faithfulness constraints penalizing vowel changes are indexed to frequency:

<table>
<thead>
<tr>
<th>/glad +ː/</th>
<th>‘happy’ neut.</th>
<th>Cf. [glåd]</th>
<th>(Freq /glad /= 2,110,000)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. glåːt:</td>
<td>*!</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. glåːt:</td>
<td>*!</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. glåːt:</td>
<td>*!</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d. &gt; glåːt:</td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>e. ○</td>
<td></td>
<td></td>
<td>*!</td>
</tr>
</tbody>
</table>

(p. 167)
6. **Boersma 1999: lexical-access constraints**
   - The problem: in Dutch, you want to be able to recognize [rat] as either /rɑt/ or /rɑd/.
   - If you try to use a standard grammar to map perceived form to underlying form, you’ll always pick the faithful one:

   (7) *Failure to recognize the wheel*

<table>
<thead>
<tr>
<th>[rat]</th>
<th>*VOICEDCODA</th>
<th>MAXVOI</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>rɑ</em></td>
<td>[rat] ‘rat’</td>
<td></td>
</tr>
<tr>
<td>[rad] ‘wheel’</td>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>

   So, Boersma proposes a family of constraints *LEX(x) “don’t recognize any utterance as lexical item x” (one for each lexical item).

   - Ranking depends on word’s frequency:

   (10) *A strong tendency to recognize the rat*

<table>
<thead>
<tr>
<th>[rat]</th>
<th>*LEX ([rad] ‘wheel’)</th>
<th>*VOICED CODA</th>
<th>MAXVOI</th>
<th>*LEX ([rat] ‘rat’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rɑ</td>
<td>[rat] ‘rat’</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[rad] ‘wheel’</td>
<td>*!</td>
<td></td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>

   Actually, it’s a bit more complex: *LEX(x/context=y) to allow for semantic context to matter

7. **More proposals in which grammar refers to frequency (if we have time)**

   Can we think of ways to determine whether grammar makes direct reference to frequency, or sees only to the outcome of lexical access?

   - Coetzee 2008: a lexical item’s frequency determines how likely it is to be assigned to a given lexical class on any production occasion
   - Myers 2005: how can lenition be both postlexical and sensitive to lexical frequency?
     - proposes a diachronic solution, where high frequency results in a more lenited lexical entry over time (exemplars? see Pierrehumbert 2001), but plays no synchronic role
     - diachronic and synchronic explanations should make different predictions about effects of priming on production...
   - Alcántara 1998 (English): high-frequency exceptions can be protected by high-ranking idiosyncratic constraints
• Carlson & Gerfen 2011 (not a proposal about grammar, but a cool case): when a Spanish diphthong loses stress (say, because of suffixation), it should monophthongize. But it’s variable:

Carlson & Gerfen 2011 (not a proposal about grammar, but a cool case): when a Spanish diphthong loses stress (say, because of suffixation), it should monophthongize. But it’s variable:

<table>
<thead>
<tr>
<th>STEM</th>
<th>DERIVED FORM</th>
</tr>
</thead>
<tbody>
<tr>
<td>n[je]ve</td>
<td>n[e]voso</td>
</tr>
<tr>
<td>verg[we]nzə</td>
<td>verg[o]nzoso</td>
</tr>
<tr>
<td>v[je]jo</td>
<td>v[je]jeco</td>
</tr>
<tr>
<td>p[wé]blo</td>
<td>p[we]blio</td>
</tr>
<tr>
<td>cal[jé]nte</td>
<td>cal[e]ntito</td>
</tr>
</tbody>
</table>

The more productive the suffix (by corpus measures), the more likely to keep the diphthong.

• Gouskova & Roon 2008: in Russian compounds, the constraint requiring each stem to bear a prominence is ranked low, but there’s a higher-ranked version of the constraint for low-frequency stems, forcing a secondary stress:

Gouskova & Roon 2008: in Russian compounds, the constraint requiring each stem to bear a prominence is ranked low, but there’s a higher-ranked version of the constraint for low-frequency stems, forcing a secondary stress:

Figure 1: Effect of frequency on secondary stress realization

8. Lexical information becomes available in real time

• How does that affect phonology that needs the lexical information? Student presentations of Wagner 2012.

References


Ng, E-Ching. 2010. Reduction, frequency and morphology in Singaporean English prosody. Manuscript. Yale University, ms. ROA #1102.


