Homework

- Read chapter 11
- Final paper (graduate students)
  - talk to me about a topic
  - may be a computational implementation
  - 10–12 pages
  - due May 18
- Final exam (undergraduates)
  - same format as midterm
  - in class part May 11
  - take-home part due May 16
Frequency

- Frequency effects are very important in morphology, perhaps more than any other subfield of linguistics.

- What is frequency?
  - Relative and absolute frequency
  - Word, stem, and morph frequency
  - Type and token frequency

- Speakers are very aware of relative word frequencies.
Frequency

- Frequency has consequences for processing
  - more frequent items are (generally) processed faster
- Baayen, Dijkstra, and Schreuder (1997) compare regular plural recognition times, varying stem frequency and number dominance (*tong* vs. *gast*)
  - high frequency nouns are recognized faster than low frequency nouns (regardless of number)
  - plural dominant plurals are recognized faster than singular dominant plurals
  - singular dominant singulars are not recognized faster than plural dominant singulars
Zipf’s law

- Token frequencies follow a characteristic pattern
- Zipf’s law (1949): Frequency times rank equals a constant
  \[ f \times r = c \]
- Brown corpus

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>the</td>
<td>69,868</td>
</tr>
<tr>
<td>of</td>
<td>36,426</td>
</tr>
<tr>
<td>and</td>
<td>28,891</td>
</tr>
<tr>
<td>to</td>
<td>26,215</td>
</tr>
<tr>
<td>a</td>
<td>24,564</td>
</tr>
<tr>
<td>in</td>
<td>21,398</td>
</tr>
<tr>
<td>that</td>
<td>10,779</td>
</tr>
<tr>
<td>is</td>
<td>10,182</td>
</tr>
<tr>
<td>was</td>
<td>9,968</td>
</tr>
</tbody>
</table>
Zipf’s law
Zipf’s law
Zipf’s law

- Principle of Least Effort
- Bee dances
- But: budgets, random text
- Benford’s Law: the probability that the first digit of a number is 1 equals about 30%, not 11.1% as expected
- Zipf’s Law is not unique to language, but still has important consequences for linguistics
Productivity

• “Productivity” in glossary:
  A morphological pattern or rule is productive if it can be applied to new bases to create new words.

• Productivity isn’t really an all-or-nothing concept

• Some observations
  • Though many things are possible in morphology, some things are more likely than others (cf. walked and ran)
  • Though there are infinitely many potential words in a language, some are more likely to become actual words than others (cf. mini-burger, burgerlet, burgerette)
Productivity

- The suffix -th (as in warmth, width, death) is generally considered unproductive

- But, WWW searches turn up many citations:

  Coolth, once a nonce word made up on analogy with warmth, is now tiresomely jocular. (1923)

  Increase the capacity of your house to store coolth. (Yes, it is a real word.) Using the mass in your house...

  The team developed a strategy to capture night-time coolth and store it for release during the following day.

  Do we see the whiteness of the snow, but only believe in its coolth.
Productivity

• If productivity is a gradient concept, how can we define or measure it?

• Absolute type frequency
  • There are 3,604 words ending in -able in Webster’s 2nd
  • Some, though, are French loans: acceptable, changeable, desireable
  • Some might be French loans: payable, regrettable
  • Calques: understandable

• Cryptotypes
Productivity

- Aronoff and Schvaneveldt (1978) compared speakers’ perceptions of words in -*ity* and -*ness*

- Speakers were given three sets of words
  - Actual words like *activity* and *assertiveness* (from Webster’s Collegiate Dictionary)
  - Possible words like *effervescivity* or *affirmativeness* (base is in dictionary, but neither suffixes form)
  - Non-words like *remortiveness* and *lugativity* (neither base nor suffixed form in dictionary)

- Subjects preferred actual word in -*ity* and possible words in -*ness* (but showed no preference with non-words)
Productivity

• Anshen and Aronoff (1988) compared -iveness and -ivity, -bleness and -ibility

• They asked subjects to list all the words they could think of with the given endings

• They expected more words in the productive pattern, but:

<table>
<thead>
<tr>
<th></th>
<th>-ness</th>
<th>-ity</th>
</tr>
</thead>
<tbody>
<tr>
<td>X-ible</td>
<td>30</td>
<td>101</td>
</tr>
<tr>
<td>X-ive</td>
<td>61</td>
<td>86</td>
</tr>
</tbody>
</table>

• And:

<table>
<thead>
<tr>
<th>Words</th>
<th>Non</th>
</tr>
</thead>
<tbody>
<tr>
<td>-ity</td>
<td>57</td>
</tr>
<tr>
<td>-ness</td>
<td>26</td>
</tr>
</tbody>
</table>
These results suggest that productive rules are formed ‘on the fly’ and not stored in the mental dictionary.

This may account for the fact that highly productive rules can be lost completely from the language ($\alpha$-, $u$-, $be$-, $with$-).

- $of$- vanished in 1400’s ($ofask$, $ofstand$, $ofswink$)
- $over$- and $out$- are still productive

Words in -$ibleness$ may have once been productive, but now that the rule is blocked.
Measuring productivity

- Instead, we could look at the ratio of actual to possible words to get an index of productivity (Aronoff 1976, Baayen and Lieber 1991)

\[ I = \frac{V}{S} \]

- This depends on being able to compute \( S \), the number of word which a WFR ought to apply to (infinity?)
- And, we need to compute \( V \), the number of words which it does apply to
- Compounding is productive, but will have close to \( I = 0 \)
- Even if we can get past this, at best we get an index of past productivity, not current productivity
Measuring productivity

• Token frequency can also be relevant

• Average token frequency in Brown corpus

<table>
<thead>
<tr>
<th>Base word</th>
<th>Derived word</th>
</tr>
</thead>
<tbody>
<tr>
<td>Xivity</td>
<td>27.261</td>
</tr>
<tr>
<td>Xiveness</td>
<td>13.117</td>
</tr>
</tbody>
</table>

• Words formed by productive rules tend to have lower token frequencies than word formed by non-productive rules

• Non-productive rules are lexicalized, and so can be semantically specialized
Measuring productivity

- Words which occur only once in a corpus (*hapax legomena*) are likely to be coinages, made up on an ad hoc basis.

- E.g., *handsome-dandsome* follows a productive reduplication pattern (*zigzag, chitchat, jingle-jangle, flip-flop*), and is likely to occur only once in a corpus.

- Memorized words, like *monitor, third, argues*, or *get*, are likely to occur many times in a sample of text.
Measuring productivity

- Baayen (1989) uses token frequency as a surrogate for semantic complexity

- New productivity index

  \[ P = \frac{n_1}{N} \]

  where \( N \) is the total token frequency of words derived by a rule, and \( n_1 \) is the number of those word types which occur exactly once
Measuring productivity

- A test: compare the two ways of analyzing ordinal -th
  - 67,586th ends in 6th (unproductive)
  - 67,586th is a different word from 6th (productive)
- In the Wellington Corpus, there are 434 ordinal numbers and fractions
- Under the first hypothesis, there are only 2 hapaxes, so
  \[ P = \frac{2}{434} = 0.005 \]
- Under the second, there are 13 hapaxes, so
  \[ P = \frac{13}{434} = 0.03 \]
Measuring productivity

- Doesn’t work for small counts ($N < 50$)
  - *-iana* occurs only in the hapax *Victoriana*

- Generalization vs. productivity
  - *-ment* is general, but not productive
  - From Wellington corpus, $P = 0.0083$, less than clearly productive processes

- Problems
  - Very few hapaxes in *-ly*, though *-ly* is very productive
  - Many hapaxes in *-ate* are centuries old
  - Ignores number of potential bases
Measuring productivity

- Baayen’s “hapax-conditioned degree of productivity” $P^*$ is the number of hapaxes of the appropriate category, divided by the total number of hapaxes in the corpus.

- Contribution of a word to the growth rate of the vocabulary as a whole.

- For example:

<table>
<thead>
<tr>
<th></th>
<th>-th</th>
<th>-ness</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P$</td>
<td>$3.1 \times 10^{-18}$</td>
<td>0.0049</td>
</tr>
<tr>
<td>$P^*$</td>
<td>$2.8 \times 10^{-17}$</td>
<td>0.04</td>
</tr>
</tbody>
</table>

- $P$ only considers a single morphological category, $P^*$ considers all morphemes together (Dutch -ster has high $P$ but low $P^*$).