Frequency effects and prosodic boundary strength

Tina Bögel and Alice Turk
University of Konstanz // University of Edinburgh

ICPhS 2019
Melbourne
Overview

- The *Smooth Signal Redundancy Hypothesis*
- Experiment on frequency effects with respect to
  - the placement of syntactic boundaries
  - the strength of the resulting prosodic boundaries
  - durational measurements of the boundary-related intervals
The smooth signal redundancy hypothesis (SSRH)

- Often no reliable cues to indicate prosodic boundaries in spoken language
- **Hypothesis**: prosodic boundary structure is planned to achieve SSR
  - make the recognition of each word in an utterance equally likely
  - prosodic boundary strength assumed to **inversely** relate to language redundancy, i.e., non-acoustic information:
    - likelihood of syntactic structure
    - lexical word frequency
    - word bigram frequency
    - ...
- More predictable elements require “less explicit signal information” than less predictable elements for successful recognition (Lindblom 1990)
Inverse relation

→ Inverse, complementary relationship between language redundancy and acoustic redundancy
→ Recognition likelihood spread evenly throughout an utterance
⇒ achieve maximal understanding with minimal effort
Previous work showed that increased

- lexical frequency (e.g., Jurafsky et al. 2001)
- syntactic predictability (e.g., Gahl and Garnsey 2004, Watson et al. 2006)

led to a reduction of word/segment duration, and influenced the placement of syntactic boundaries.

Clearly demarcating word boundaries $\rightarrow$ more salience
The smooth signal redundancy hypothesis

Hypothesis

- Inverse relationship between language redundancy and acoustic salience should hold for prosodic boundaries
- SSRH predicts greater final lengthening, initial lengthening, initial strengthening, F0 reset, etc., given low language redundancy
  → Stronger prosodic boundaries are expected to occur where language redundancy is low, e.g., within infrequent stretches of speech
- SSRH would further predict a (gradient) correlation between boundary strength and language redundancy (e.g. greater final lengthening, initial lengthening, initial strengthening, F0 reset, etc.),
  → Has not been tested experimentally!
Work presented here

- Investigates the relationship between language redundancy and prosodic boundary strength
- through the effect of:
  - syntactic frequency
  - word frequency
  - word bigram frequency

→ on the placement of intonational phrase boundaries
→ on durational measurements of boundary strength

Challenge:
Need to vary language redundancy, while using controlled material
  - with similar syntactic phrasing
  - with similar segments across boundaries (effects might be subtle)
Experimental design: syntactic ambiguities

When the cake was dropped flat plants stuck to its underside

- Syntax A: the cake was **dropped** .... **flat plants** stuck to its underside
  (= modifying construction, [V [A N]])

- Syntax B: the cake was **dropped flat** .... **plants** stuck to its underside
  (= resultative construction, [[V A] N])

⇒ Corpus study:
Syntax A (=modifying) is far more likely than Syntax B (=resultative)

<table>
<thead>
<tr>
<th></th>
<th>[V A]</th>
<th>[A N]</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICE-GB/Brown</td>
<td>~ 5/3%</td>
<td>~ 67/88.5%</td>
</tr>
</tbody>
</table>
Experimental design: placement of phrase boundaries

- Difference in syntax comes with difference in the placement of an intonational phrase boundary
  
  \[ V \% A N \]
  
  or
  
  \[ V A \% N \]

- Expect \texttt{V\%AN} to occur more often (if speakers are given a choice)
  
  \[\rightarrow\] corresponding syntactic structure is more frequent
Experimental Setup

**Experimental design: lexical frequencies I**

In order to determine:

1. **effects of frequency** on **syntactic choice**, the relevant syntactic sequence had to have four combinations:

<table>
<thead>
<tr>
<th>Verb</th>
<th>Adj.</th>
<th>Noun</th>
<th>Shortcut</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_{\text{frequent}}$</td>
<td>Adj.</td>
<td>$N_{\text{frequent}}$</td>
<td>ff</td>
</tr>
<tr>
<td>$V_{\text{frequent}}$</td>
<td>Adj.</td>
<td>$N_{\text{infrequent}}$</td>
<td>fi</td>
</tr>
<tr>
<td>$V_{\text{infrequent}}$</td>
<td>Adj.</td>
<td>$N_{\text{frequent}}$</td>
<td>if</td>
</tr>
<tr>
<td>$V_{\text{infrequent}}$</td>
<td>Adj.</td>
<td>$N_{\text{infrequent}}$</td>
<td>ii</td>
</tr>
</tbody>
</table>

2. **effects of frequency** on **boundary strength**, the four combinations above had to be comparable:
   - in the rhyme/coda of the verb
   - in the onset of the noun
   - in the onset and the rhyme/coda of the adjective
   → known to show the largest durational effects of boundary strength

**But:** had to allow for reliable measurements at the same time
Experimental design: lexical frequencies II

Estimation of lexical frequencies via WebCelex:

<table>
<thead>
<tr>
<th></th>
<th>Verbs</th>
<th>Nouns</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>frequent</strong></td>
<td>$&gt; 2000$</td>
<td>$&gt; 3000$</td>
</tr>
<tr>
<td><strong>infrequent</strong></td>
<td>$&lt; 200$</td>
<td>$&lt; 100$</td>
</tr>
</tbody>
</table>

Table: Raw number thresholds for lexical (in)frequencies

→ Matching of verbs/nouns with respect to the form

**ff:** When the cake was **dropped** flat **plants** stuck to its underside

**fi:** When the cake was **dropped** flat **planks** stuck to its underside

**if:** When the grass was **cropped** flat **plants** were able to grow again

**ii:** When the grass was **cropped** flat **planks** were laid across the lawn
Bigram frequencies

Determined **bigram frequencies** of Verb-Adj (V-A) and Adj-Noun (A-N) combination and their ratio: V-A/A-N

**Problem:** No corpus large enough to determine frequencies of infrequent combinations.

→ Google

→ ‘Noisy’, therefore just approximations

→ Great variance

⇒ Divided data into abstract categories:

<table>
<thead>
<tr>
<th></th>
<th>low</th>
<th>med (buffer)</th>
<th>high</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&lt; 40%</td>
<td>40% - &lt; 60%</td>
<td>&gt;= 60%</td>
</tr>
<tr>
<td>V-A</td>
<td>&lt; 13900</td>
<td>&lt; 314000</td>
<td>&gt;= 314000</td>
</tr>
<tr>
<td>A-N</td>
<td>&lt; 3180</td>
<td>&lt; 108000</td>
<td>&gt;= 108000</td>
</tr>
</tbody>
</table>

**Table:** Abstract representation of raw bigram frequencies
Data gathering

- **Data presentation:**
  - without commas (syntactic boundary placed according to choice)
  - several repetitions; only discuss first repetition here (58 sentences/speaker)

- **Subjects:** 23 participants
  (students at the University of Edinburgh, \( \Phi = 23.4 \) years, 14 females)

- **Recordings:** sound-treated studio at the University of Edinburgh with a high quality microphone
Frequency and syntactic choice: results I

Annotation of syntactic choice:
1 annotator (100%), 1 annotator (40%) – 100% agreement

Here: 23 speakers, repetition 1 → total of 1314 instances

... surprising given the results from the corpora ....
For the choice of syntax, the following factors were relevant:

- **Syntax A** (frequent syntax, V%AN) more likely with
  - highly frequent nouns ($p < 0.05$)
  - high A-N bigram frequency ($p < 0.001$)

- **Syntax B** (infrequent syntax, VA%N) more likely with
  - highly frequent verbs ($p < 0.001$)
  - high V-A bigram frequency ($p < 0.001$)
  - higher V-A in comparison to A-N bigram frequency ($p < 0.001$)
Strict selection:
- Only speakers that generally had a high consistency across repetitions (1 sentence - 1 choice - in both repetitions)
  → 10 speakers
- Only quadruplets that had the same syntactic choice across both repetitions
  → can measure frequency impact on duration — and later compare it to repetition 2
- Today: Discuss only repetition 1

<table>
<thead>
<tr>
<th>Annotated sentences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syntax A</td>
</tr>
<tr>
<td>124</td>
</tr>
</tbody>
</table>
Durational measurements: annotation

- Raw material, e.g.

<table>
<thead>
<tr>
<th>V-A</th>
<th>A-N</th>
<th>example</th>
</tr>
</thead>
<tbody>
<tr>
<td>roped_f</td>
<td>lat_t</td>
<td>dropped flat plants</td>
</tr>
<tr>
<td>k_f</td>
<td>ree_p</td>
<td>walk free people</td>
</tr>
</tbody>
</table>

- Problematic, a lot of segmental variation
- Abstract annotation scheme, three intervals per sequence (six in total)
**Durational measurements: annotation**

<table>
<thead>
<tr>
<th>Verb end</th>
<th>Adjective start</th>
<th>Adjective end</th>
<th>Noun start</th>
</tr>
</thead>
<tbody>
<tr>
<td>V-Rh</td>
<td>rhyme</td>
<td>A-On</td>
<td>A-Rh</td>
</tr>
<tr>
<td>V-Co</td>
<td>coda</td>
<td>A-C1</td>
<td>A-Co/Co1/Co2</td>
</tr>
<tr>
<td>V-ORh</td>
<td>with part of onset</td>
<td>A-ORh</td>
<td>nucleus, not coda</td>
</tr>
<tr>
<td>R-V-...</td>
<td>with onset release</td>
<td>A-Nu</td>
<td>R-A-...</td>
</tr>
</tbody>
</table>

**Intermediate (IM1 and IM2)**

| ...-R    | release         | Might include aspiration! |
| ...-P    | pause           | Missing pause (P) is only indicated if syntax requires it |
| ...-RP   | release and pause | Both -P/-RP are only indicated if there is no closure following |

**Supra-markers**

| ?        | insecurity      | Insecurity in annotation, mostly at preceding or following border |
| x        | connection      | Connection across word boundaries - e.g., V-Rh.IM1.A-On |
| ( )      | missing element | For elements that should be there, but are not (mostly R and P) |
| rel      | release         | Only on DurationSep level. Connected to other parts with + |
| NA       | release         | If a separation at word boundary in DurationSep (only!) is not possible |
| pause    | pause           | Same as release |
| glot     | glottalization  | Same as release |
| (breath) | non-expected release | Same as release |

→ Allows for grouping of similar patterns to get more reliable measurements!

**BUT:** If there was no clear boundary, intervals were connected via an underscore (_)

→ particular item then not part of analysis - further reduction of data
Syntax A (frequent, V%AN):

- When lexical frequency V is low: **increased verb coda interval duration**
  ($\beta=0.015$, SE=0.006, t=2.5, $p<0.05$)

- When bigram frequency AN is low: **increased noun onset interval duration**
  ($\beta=0.01$, SE=0.004, t=2.3, $p<0.05$)
Syntax B (infrequent, VA%N):

- When lexical frequency V is low or bigram frequency VA is low:
  increase overall VA duration
  \[ (\beta = 0.023, \ SE = 0.009, \ t = 2.65, \ p < 0.05 \text{ and } \beta = 0.029, \ SE = 0.01, \ t = 2.96, \ p < 0.01) \]

→ Same effect is found with the verb coda (but not with the adjective onset)

- When VA bigram frequency higher than AN frequency:
  - decrease of verb coda interval duration
    \[ (\beta = -0.029, \ SE = 0.007, \ t = -3.97, \ p < 0.001) \]
  - increase of noun onset interval duration
    \[ (\beta = 0.018, \ SE = 0.005, \ t = 3.27, \ p < 0.01) \]
Conclusion

**All of these results are consistent with the SSRH:**

→ inverse relationship between language redundancy (lexical frequencies, bigram frequencies, and their interaction) and durational measurements of the prosodic boundary-related intervals

**Outlook:**

- Compare repetitions
- Investigate F0
- Zoom in on bigram frequencies across boundaries
- ...

Bögel and Turk (Konstanz/Edinburgh)
Thank you!

... questions, comments...?
## Information on corpora

<table>
<thead>
<tr>
<th></th>
<th>Brown corpus</th>
<th>ICE-GB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Released</td>
<td>1964</td>
<td>1998</td>
</tr>
<tr>
<td>Tagging</td>
<td>Part of Speech (POS)</td>
<td>Syntactic (Treebank)</td>
</tr>
<tr>
<td>Tokens</td>
<td>~ 1 Million</td>
<td>~ 1 Million</td>
</tr>
<tr>
<td>English</td>
<td>BE</td>
<td>AE</td>
</tr>
<tr>
<td>Texts</td>
<td>Across all genres</td>
<td>Edited English prose</td>
</tr>
<tr>
<td>Citation</td>
<td>(Francis and Kučera 1964)</td>
<td>(ICE-GB corpus 1998)</td>
</tr>
</tbody>
</table>

**Table:** Information on the ICE-GB and the Brown corpus
Results corpus study

Frequency determination:

<table>
<thead>
<tr>
<th></th>
<th>Verb-Adj</th>
<th>Adj-Noun</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>main</td>
<td>copula</td>
</tr>
<tr>
<td><strong>ICE-GB corpus</strong></td>
<td>1771</td>
<td>8781</td>
</tr>
<tr>
<td>In %</td>
<td>~ 5%</td>
<td>~ 28%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>~ 33%</td>
</tr>
<tr>
<td><strong>Brown corpus</strong></td>
<td>1657</td>
<td>4562</td>
</tr>
<tr>
<td>In %</td>
<td>~ 3%</td>
<td>~ 8,5%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>~ 11,5%</td>
</tr>
</tbody>
</table>

Table: Frequency of syntactic combinations in the ICE-GB and the Brown corpus

**Conclusion:**
Syntax A (=modifying) is more likely than Syntax B (=resultative)
Experimental design: Material

Examples with four combinations:

<table>
<thead>
<tr>
<th>freq Verb</th>
<th>infreq Verb</th>
<th>freq Nouns</th>
<th>infreq Nouns</th>
</tr>
</thead>
<tbody>
<tr>
<td>dropped</td>
<td>cropped</td>
<td>plank</td>
<td>plant</td>
</tr>
<tr>
<td>buy</td>
<td>dye</td>
<td>paper</td>
<td>paisley</td>
</tr>
<tr>
<td>call</td>
<td>wall</td>
<td>door</td>
<td>dorm</td>
</tr>
<tr>
<td>made</td>
<td>shade</td>
<td>picture</td>
<td>pitcher</td>
</tr>
<tr>
<td>make</td>
<td>rake</td>
<td>field</td>
<td>fief</td>
</tr>
<tr>
<td>stayed</td>
<td>bayed</td>
<td>sister</td>
<td>sissy</td>
</tr>
<tr>
<td>play</td>
<td>slay</td>
<td>fish</td>
<td>fiend</td>
</tr>
<tr>
<td>shake</td>
<td>snake</td>
<td>boxes</td>
<td>bobbers</td>
</tr>
<tr>
<td>turned</td>
<td>churned</td>
<td>balls</td>
<td>baulks</td>
</tr>
<tr>
<td>wear</td>
<td>pare</td>
<td>farmers</td>
<td>farthings</td>
</tr>
<tr>
<td>works</td>
<td>lurks</td>
<td>markets</td>
<td>marshals</td>
</tr>
<tr>
<td>walk</td>
<td>stalk</td>
<td>people</td>
<td>peafowls</td>
</tr>
</tbody>
</table>
References I


Bell, Alan, Brenier, Jason M., Gregory, Michelle, Girand, Cynthia and Jurafsky, Dan. 2009. Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language* 60(1), 92–111.


