It’s not size that counts\textsuperscript{1}

An even more tentative theory of second occurrence focus

David I. Beaver

Department of Linguistics
UT Austin

WCCFL, USC Feb 2010

\textsuperscript{1}This presentation is based on collaborative work with: Jason Brenier, Sasha Calhoun, Edward Flemming, Florian Jaeger, Dan Jurafsky, Anubha Kothari, Annie Nenkova, Dan Velleman, Laura Whitton, and Maria Wolters
Second occurrence focus

Examples

(1) a. Everyone already knew that Mary only eats [vegetables]_F.

b. If even [Paul]_F knew that Mary only eats [vegetables]_{2F}, then he should have suggested a different [restaurant]_F. (Partee 1999)

Configuration:

...even_j ( ... [NEW+ASSOC_j]_F ) ... only_i ( ... [GIVEN+ASSOC_i]_{2F} )
Research question:

How are semantic and pragmatic features of an expression reflected in its prosodic prominence?
My intuition (similar to Watson et al 2009)

- Stress happens when stuff is important and/or unpredictable.
- More important or less predictable $\implies$ more stress
What is predictable?

- First approximation: predictable $=$ Given. (Prince 1981, etc.)
- But in joint work with Dan Jurafsky and others, we found that while Given/New is correlated with presence of accent, the effect is weaker than many other factors, e.g. unigram frequency.

A. Nenkova, J. Brenier, A. Kothari, S. Calhoun, L. Whitton, D. Beaver, & D. Jurafsky, *To Memorize or to Predict: Prominence Labeling in Conversational Speech*, NAACL-HLT 2007

The goal was to build a model of prominence for use in e.g. TTS systems, and at the same time test some of the ideas in the linguistic literature.

14,555 words from the Switchboard corpus were annotated for information status, contrast and concrete/non-concrete distinctions, all features that linguistic literature suggests are predictive of prominence.

When building decision-tree classifiers using leave-one-out selection, the best accent classifiers did not include information status.

The best classifiers (>77% accuracy) made limited use of some linguistic features (animacy, contrast).

Performance was near a ceiling of inter-speaker variation identified in a separate study: for six subjects reading the same text, a speaker on average agreed on accents for only 82% of words, <5% above the accuracy of our best classifier.
Further problems with standard Given-New

- The Synthesis project gives one type of evidence that standard Givenness is not what we need.
- But note also examples like the following:

**Examples**

(2) Lots of people only had fun at Judith’s party (and not at any others). Even \([\text{Judith}]_F\) only had fun at \([\text{Judith’s}]_{2F}\) party.

**Configuration:**

\[\text{...even}_j (\ldots [\text{GIVEN+ASSOC}_j]_F) \ldots \]
\[\text{only}_i (\ldots [\text{GIVEN+ASSOC}_i]_{2F})\]
Based on such examples, Buring (2008) concludes:

*The difference between primary focus and 2OF can not generally be reduced to the latter being focussed and Given, while the former is focussed and non-Given.*
If predictability is not standard Givenness, what is it?

- Predictability could be e.g. Schwarzschild’s (1999) variant: loosely, an expression is given if it doesn't add new information to any of the constituents that contain it (modulo a special existential closure operation).

- Or predictability could be a general information theoretic measure: how probable is the expression given everything you know about the rest of the context?

- I remain agnostic between these here, but note that a possible operationalization: one expression is more predictable than another if hearers can fill in the blank more easily when you chop it out.
What is important?

- The answering constituent for a salient question.
- Contrastive elements.
- The associate of a focus sensitive particle.
- Politeness.
- Things you have strong emotions about (c.f. expressives)
- Winning moves in games. (Watson et al 2009)
Is importance important?

- In many cases, e.g. answers, it seems that importance might be reduced to a special case of predictability, so I’ve wondered about eliminating it from the account.
- But I’m so far unable to reduce some cases in this way, e.g. contrast, being polite, expressing your emotions, all seem to yield prominence even when there is very high predictability.
Background assumptions on prominence realization

- **Surface prominent**: Prominence is established partly as a result of phonological stress assignment. Other ways to mark prominence include word choice, word order, use of intensive modifiers, gesture.

- **Stress**: Sentential stress in English is realized through a combination of acoustic factors, including pitch excursions in proximity to syllables bearing primary word-stress, duration, intensity, spectral tilt, and vowel quality.

- **Nuclear stress**: The most prominent element is the one bearing nuclear stress within an intonational phrase. Typically English realizations have major pitch excursion, then a reduction in pitch range for following expressions.

- **Non-nuclear stress**: A second tier of prominence is realized through non-nuclear stress in pre-nuclear or post-nuclear positions.
The P+I model

- **Prominence Principle:** If one expression is more *communicatively significant* than another, then the first should be more *surface prominent* than the second.

- **Communicative significance:** The communicative significance of one expression relative to another may be affected independently by (i) how *predictable* expressions are for hearers, and (ii) how *important* they are as regards the speaker’s interests.
Example 1: A simple question answer

Examples

(3)  
  a. Who drank David’s vodka?
  b. \([\text{Judith}]_F\) drank David’s vodka.

“Judith” is both more important than “David”, and less predictable, and so gets NS.
Example 2: A simple question answer

<table>
<thead>
<tr>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>(4)</td>
</tr>
<tr>
<td>a.</td>
</tr>
<tr>
<td>b.</td>
</tr>
</tbody>
</table>

“He” and “did” are both quite predictable as words, but as the answering constituent “he” is more important, and gets NS.
Example 3: SOF

(5) a. Everyone already knew that Mary only eats [vegetables]$_F$.

b. If even [Paul]$_F$ knew that Mary only eats [vegetables]$_{2F}$, then he should have suggested a different [restaurant]$_F$. (Partee 1999)

- Now we have two important elements, “Paul” and “vegetables”. But the first is less predictable (by Schwarzschild Givenness or fill-in-the-blank Givenness), and so gets NS.
- The account also predicts a secondary stress on “vegetables”. Is there one?
The realization of second occurrence foci

- How are second occurrence foci realized?
- I’ll comment on why this has been an important question, using experimental results from Beaver et al (2007), and Beaver & Clark (2008).


Why second occurrence focus has been studied

The argument from second occurrence focus

1. Weak, grammaticized theories of focus sensitivity require the semantic focus to be F-marked.

2. These theories also predict that the semantic focus should contain a phonological focus.

3. (a) Second occurrence foci are semantic foci, (b) but second occurrence foci contain no phonological focus marking.

4. Therefore we should prefer a strong, pragmatic account of focus.
The SOF mission

To use a systematic, objective instrumental methodology to answer the following questions:

1. Are there any intonational correlates of second occurrence focus?
2. What are those correlates?
3. Are hearers sensitive to the effects?
A multi-speaker multi-discourse production experiment was performed. The goal was to examine whether words in second occurrence focus are intonationally distinct from non-focal occurrences of the same words. A number of possible intonational correlates of focus were examined, including maximal and minimal ($f_0$) pitch, pitch range, word duration and r.m.s. intensity.
Experimental details

- The production experiment was based on 14 written texts made up of 7 minimal pairs.
- Each member of a minimal pair ends with the same sentence in written form, but prior context means that different elements are in second occurrence focus.
20 native U.S. English speakers read all the texts twice, along with filler discourses.

The positions of onsets and offsets of relevant expressions were ‘hand’ annotated, and pitch and intensity information was automatically extracted for statistical analysis.
A minimal pair

Examples

(6)  
   a. Both Sid and his accomplices should have been named in this morning’s court session.
   b. But the defendant only named Sid in court today.
   c. Even the state prosecutor only named Sid in court today.
Second version of SOF example

(7)  
  a. Defense and Prosecution had agreed to implicate Sid both in court and on television.  
  b. Still, the defense attorney only named Sid in court today.  
  c. Even the state prosecutor only named Sid in court today.
Production results (minF’ repeated measures analysis)

- There are significant main effects for duration ($p = 0.039$) and energy ($p = 0.027$): SOF expressions average 6 msecs longer than non-focal expressions and receive more energy.
- There are marginal main effects for r.m.s. intensity ($p = 0.07$), minimum pitch ($p = 0.08$), and pitch range ($p = 0.06$).
- No main effect of maximum and mean pitch could be found.
Sample pitch track
Production is only half the story

- The production experiment yielded significant effects, but they are small. Are these effects large enough to be perceptible?
- If not, the marking of second occurrence foci obviously could not influence interpretation, and the argument from second occurrence focus might survive.
A perception experiment

- To test perceptual sensitivity, 15 native speakers judged prominence of words in 40 sentence pairs differing only in SOF.
- The pairs were spliced from the production data without discourse context (i.e. last sentence only).
- The production pairs used were chosen without regard to the acoustic measurements above.
Subjects distinguished SOF significantly above chance (mean=64%; p < .001).

We investigated whether acoustical measures from the production analysis could predict perception results using stepwise multiple linear regression.

Energy is a predictor (44% of variation), intensity and duration can also be used, but none of the f0 measures are predictors.
Fery and Ishihara (2009) studied SOF in German. For the standard SOF configuration, their results echoed those for English. They also tested a variant condition where the SOF preceded the primary focus, and found that in this case while the SOF was less prominent than the primary focus, it tended to be both accented and lengthened.
There is an acoustic correlate of SOF (broadly as claimed by Rooth 1996).

The effects are small but perceptible. So the information status of SOF is marked grammatically.

Thus, SOF does not demonstrate optionality of focus, and this argument against semantic theories of focus is bogus.

The results are as expected on the P+I model of prominence.
Focus Domains

- Buring (ms:2006, ms:2008), having argued against using Givenness to account for prominence, proposes instead to use the notion of Focus Domain (following Jacobs 1991).
- Rooth (2010) uses a variant of this approach.
- In Rooth’s terms, (8)a is analyzed as in (8)b, with the ∼ operator marking the domain of the focus, and providing a set of focus alternatives for “only”.

Examples

(8)  
  a. Judith only drank vodka
  b. only_i (Judith drank [vodka]_F) \sim_i
Focus Domains and SOF

- A SOF example involves two focus domains, and two $\sim$ operators.

Examples

(9) $\text{even}_j \ ([\text{Paul}]_F \ \text{knew that Mary only}_i \ (\text{eats} \ [\text{vegetables}]_{2F} \ \sim_i) \ \sim_j)$

- The domain size principle: one focus is more prominent than another just in case it has a strictly bigger domain, and a focus must be the most prominent element within its own domain.

- This principle correctly predicts the relative prominence of the two foci in the above SOF case.
As Buring himself notes, his account leaves a classic case unexplained, Rooth’s rice-grower examples.

Here “eat” is more prominent than “rice”, but “only” is the focus associating with “only”, contradicting the domain size principle.

**Examples**

(10) People who grow rice usually only [eat]$_F$ [rice]$_{2F}$.
(Rooth 1992)

Configuration:
only$_i$ ( ([NEW+CONTRAST]$_F$ [GIVEN+ASSOC]$_{2F} \sim j) \sim i$)
Buring was prepared to accept the rice-grower examples as exceptions, because of a closely related type of data, based on answers to questions like the following:

**Examples**

(11) What movie have you only watched in IMAX 3d?  
(Full sentence answer please.)
### Examples

(12) a. What movie have you only watched in IMAX 3d?

b. I only watched \([\text{Avatar}_F]\) in \([\text{IMAX 3d}]_{2F}\).

**Configuration:**

\[(\text{only}_i) \ (\left[\text{NEW} + \text{ANSWER}\right]_F \ \left[\text{GIVEN} + \text{ASSOC}\right]_{2F} \sim_i) \sim j\)
Buring (2006) had taken examples with this configuration to be “ineffable”, since Schwarzschild judged the following to be infelicitous:

<table>
<thead>
<tr>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>(13)</td>
</tr>
<tr>
<td>a. Where do you only eat crepes?</td>
</tr>
<tr>
<td>b. I only eat crepes in Paris.</td>
</tr>
</tbody>
</table>
Secondary foci following answers — better crepes

- Rooth suggested that some variants on the crepe example (e.g. with partial answer intonation) were ameliorated.
- In fact, the configuration occurs naturally, at least in written form. Here’s one of many web examples (thanks to my student Dan Velleman):

<table>
<thead>
<tr>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>(14) a. Q: What do you only eat out at a restaurant and, vice versa, what do you only eat at home?....</td>
</tr>
<tr>
<td>b. A: I only eat fried chicken at home, I do not like eating with my hands when I go out.</td>
</tr>
</tbody>
</table>

- In fact, consultants either produced the examples in spoken form without a problem, or said they were fine when enough context was given.
Rooth’s relativized rule

- It’s clear that the oddity for some speakers of the original crepe example does not result from the focus configuration alone.
- Rooth (2010) concluded that Buring’s original proposal was not tenable, implying that the Domain size principle must be modified.
- **Modified domain size principle:** one focus is more prominent than another just in case it has a strictly bigger domain, and a focus must be the most prominent element within its own domain apart from elements which are associated with bigger domains.
A problem with Rooth’s relativized rule

- It is hard to counterexemplify Rooth’s rule without some independent constraint on when a focus can be associated with a sentential domain.
- But (15) is problematic if the contrastive “eat” has “eat rice” as its domain, conforming to his analyses of contrast elsewhere.
- (15) is fine for the P+I model, since “eat” is less predictable than “rice”.

Examples

(15) People who grow rice usually only want to $[\text{eat}]_F \ [\text{rice}]_{2F}$.

Configuration:
only $; (\ \{[\text{NEW}+\text{CONTRAST}]_F \ [\text{GIVEN}+\text{ASSOC}]_{2F} \sim j) \sim i\}$
The P+I theory seems empirically stronger than extant versions of domain size models, leading me to suggest: it's not size that counts.

But maybe the domain size theory can be repaired given a clever enough set of constraints on when a domain has sentential scope, and anyway, the domain size and P+I theories are not so far apart...
Allow me to graffiti Buring (2008):
“The difference between primary focus and 2OF cannot generally be reduced [is] reduced to the latter being focussed and [Schwarzschild] Given, while the former is focussed and non-[Schwarzschild]-Given.”

In effect, domain size is a way of representing givenness, and hence predictability, at syntax (or phonology).

So the differences between the models are not huge, although there is one big issue I haven’t discussed: domain size models are inherently categorical, while the P+I model is non-committal.

If prominence is a matter of degree, the P+I model may be greatly preferable. And if you’ve spent a lot of time looking at naturally occurring speech data, you might find this plausible.
Finally, I’d like to briefly mention one striking empirical issue that comes up when the data I’ve studied are compared to the empirical results of Watson et al (2009), who performed an experiment based on a spoken version of Tic Tac Toe.

They found that while both unpredictable information (unpredictable moves) and important information (game changing moves) were made prominent, they were made prominent in different ways.

This seems like excellent support for the P+I theory. But...
In SOF, important, unpredictable information is accented, and the important but predictable information (i.e. the SOF itself) is merely lengthened.

Watson et al’s results went almost exactly the other way: unpredictable moves tended to be lengthened, and important moves accented.

This contrast seems to me both important, and unpredicted, with obvious consequences. (I’m stressed.)