

Second Occurrence Focus and the Acoustics of Prominence

Jonathan Howell
Cornell University

1. Introduction

What happens to the phonetic prominence of a constituent when it is expected to have semantic focus, but occurs in a repeated or given context? Recent experimental results have suggested that some degree of prominence survives, in favor of a grammaticalized view of association with focus. I will present results that question the current methodology and cast doubt on the perceptual relevance of an inconsistent but statistically significant production of prominence (Sections 2-4). I conclude that pragmatic reasoning is a necessary ingredient in the interpretation of association with focus (Section 5).

1.1. *Second occurrence focus and semantic theory*

Since at least Partee (1991), a debate in the semantics literature has centered around the phonetics and phonology of prominence. Partee challenged the significance of the observation that certain adverbs appear to “associate” with prominent words in an utterance to truth-conditional effect. The minimal pair in (1), taken from Rooth (1992), illustrates the observed “association with focus”.

- (1) a. Mary only introduced BILL to Sue.
b. Mary only introduced Bill to SUE.

Many analyses of this phenomenon, including the structured meaning semantics of Jackendoff (1972), or the intermediate theory of Rooth (1992), maintained that the association with focus operators is *grammatically* mediated. Partee (1991, 1999) adduced (2) and similar examples as evidence against these theories. She, and others including Dryer (1994), Roberts (1996) and Büring (1997), favor some variety of *pragmatic* association.

- (2) Eva only gave xerox copies to the GRADUATE STUDENTS.
No, PETR only gave xerox copies to the graduate students.

Although *graduate students* associates with *only* in both sentences of (2), Partee made the impressionistic observation that the “second occurrence” of *graduate students* lacks phonological prominence. Under the grammatically mediated accounts of association with focus however, it is precisely phonological prominence which licenses the syntactic/semantic annotation (i.e. “F-marking”) of focus. Indeed, more recent grammatical accounts (e.g. Büring 2006, Rooth 2007, Selkirk 2007) have incorporated the explicit interface constraint *Stress F*, due to Truckenbrodt (1995):

* I was fortunate to present different versions of this work at several venues. I wish to thank audiences at the Cornell Linguistics department, 4th Joint ASA/ASJ Meeting, 2007 LSA Annual Meeting, MOT 2007, WCCFL 26, SPINE 2007 and OnLI 2007. Thanks also to the participants of the experiments for their patience and to the following individuals for discussion: Johanna Brugman, Abby Cohn, Adam Cooper, Effi Georgala, Carlos Gussenhoven, Hyun Kyung Hwang, Florian Jaeger, Dan Kaufman, Bob Ladd, Amanda Miller, Mats Rooth and Michael Wagner.

(3) Stress F

Let β be an F-marked phrase with scope ϕ . Then the strongest stress in the phonological realization of ϕ falls within the realization of β .

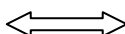
Such theories predict that cases of second occurrence focus (henceforth SOF) will have some level of prominence.

1.2. Previous investigations of prominence

Several experimental investigations followed Partee's observation, beginning with Rooth (1996). Investigating the acoustics of his own speech, Rooth found that, while the SOF words lacked F0 movement, an SOF word had a measurably longer duration and greater intensity than the adjacent unfocused word (e.g. *named* had a greater duration than *Manny* in 4aB). The direction of comparison (syntagmatic) is indicated by the horizontal arrow. Perceptually, he judged it possible to identify the correct association with focus in the absence of the context-supplying "first occurrence" sentence.

- (4) a. A: Paul only NAMED Manny today.
 B: So what. Even EVA only named Manny today.

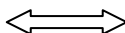
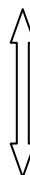
- b. A: Paul only named MANNY today.
 B: So what. Even EVA only named Manny today.



Beaver, Clark, Flemming, Jaeger & Wolters (2007)/Jaeger (2005) tested naïve speakers with a larger number of stimuli, in this case objects and adjuncts (cf. 5-6). The authors compared F0, word duration, RMS intensity and acoustic energy, both syntagmatically (cf. horizontal arrow) and paradigmatically (cf. vertical arrow). Again, F0 was not significant for SOF. With respect to production, they found a statistically significant 6 ms (paradigmatic) and statistically significant 10.1 ms and -8.1 ms (syntagmatic) duration difference. Intensity was also significant or approaching significance: 0.13 dB (paradigmatic) and .8519-.6354 dB (syntagmatic).

- (5) a. Both Pete and Edward are suffering from the flu.
 b. But the nurse only gave PETE a pill today.
 c. Even THE DOCTOR only gave Pete a pill today.

- (6) a. Pete really needed an injection to ease the pain.
 b. But the nurse only gave Pete A PILL today.
 c. Even THE DOCTOR only gave Pete a pill today.



While statistically significant, these values fall short of some published just noticeable differences (JNDs) for speech: 10-40 ms (Lehiste 1970), ~25 ms (Klatt 1976); 1-4 dB (Stevens 1998). The authors, therefore, conducted a perception experiment.

Beaver *et al.* presented naïve listeners with minimal pairs (e.g. 5c and 6c) and asked them to identify the utterance in which the second probe (e.g. *a pill*) was more prominent. Subjects performed above chance in this discrimination task, but averaged only 63% accuracy. The authors speculate that the less than perfect performance may be due to reader disfluencies in the production stimuli, or to the inherent unnaturalness of laboratory elicitation. Further, it is impossible to discern from these results alone whether listeners actually exploit this discrimination in interpretation, or even whether the discrimination reflects purely linguistic competence.

Finally, for considerations of space, I gloss over production studies by Bartels (1996) and Féry & Ishihara (2005). The former compared "first occurrence" focus (henceforth FOF) with SOF, and the latter investigated SOF in the pre- and postnuclear domains in German. Both found duration to be a

significant cue to SOF; F0 was significant only prenuclearly in German.

2. First production experiment

2.1. Method

2.1.1. Subjects

I ran a small production study with three male speakers, including myself. Sophisticated (i.e. non-naïve) speakers were used both to complement the results of Beaver *et al.* (2007)/Jaeger (2005) based on naïve speakers, and in answer to the speculation that speakers' naiveness weakened the results.¹ Recall that Rooth (1996) reported clear results in his own speech.

2.1.2. Recording

Subjects were recorded in a sound-attenuated room in the Cornell Phonetics Laboratory, using a Plantronics DSP-500 headset to control for head movement. Recording and analysis were both conducted with Praat 4.2.29 (Boersma & Weenink 2004). Subjects repeated each discourse five times for a total of 60 tokens per speaker. The subjects were asked to read the stimuli as naturally as possible, and without exaggeration. One token produced with a pitch accent on a SOF word was deemed unnatural by one speaker, who asked to re-record the token without prompting.

2.1.3. Stimuli

Following the established methodology, the elicited discourses consisted of a context sentence with FOF and a target sentence with SOF, allowing paradigmatic and syntagmatic comparison (see above). In this case, the target words were noun/verb homophones (*peddles/pedals, patches, labels*), in order to avoid *post-hoc* normalization and to control for vowel quality. Another innovation was the inclusion of both pre- and postnuclear SOF.² Finally, with the intent of controlling for possible isochronic effects (disruptions to regular rhythm), I limited my target words to bisyllabic trochees (cf. Lehiste 1980). Some examples follow;³ (7) is a postnuclear pair and (8) is prenuclear.

- (7) a. A: Johnson only PATCHES patches for Microsoft. (He doesn't create them.)
 B: That's right. Even THOMPSON only patches patches for Microsoft.
- b. A: Johnson only patches PATCHES for Microsoft.
 (He doesn't patch the programs themselves.)
 B: That's right. Even THOMPSON only patches patches for Microsoft.
- (8) a. A: Johnson only PATCHES patches for Microsoft. (He doesn't create them.)
 B: That's right. Johnson only patches patches even for APPLE.
- b. A: Johnson only patches PATCHES for Microsoft.
 (He doesn't patch the programs themselves.)
 B: That's right. Johnson only patches patches even for APPLE.

2.1.4. Measurements

The target words were manually annotated in Praat for stop closure, aspiration, first vowel and

¹ Another methodological approach would try to elicit stimuli in a communicative task. See the conclusion and footnote 8 for some comments on why this approach may not be useful.

² Rather than creating a third discourse to elicit the unfocused target word (cf. Féry & Ishihara 2005), I used the unfocused expressions in the PF and SOF sentences, following Rooth (1996) and Beaver *et al.* (2007)/Jaeger (2005).

³ Capitalization has been added for expository purposes.

second syllable. The following values were then automatically extracted using Praat scripts⁴: stressed syllable duration, spectral balance and acoustic energy. Spectral balance (i.e. spectral tilt) measures relative intensity at different harmonics rather than over the entire spectrum and has been claimed to be an acoustic correlate of stress in English and Dutch (Campbell & Beckman 1997; Sluijter, van Heuven & Pacilly 1997). This was calculated as H1-H3 and H1-H2. Acoustic energy is the product of root mean square (RMS) intensity and vowel duration.

2.2. Results

2.2.1. Paradigmatic comparison

Beginning with duration, paradigmatic differences were generally not statistically significant between SOF and unF, with the exception of Speaker 3 whose difference between SOF and unfocused (unF) was marginally significant, but only for nouns. FOF differed significantly from other conditions for all three speakers. This is illustrated in Figure 1 for the postnuclear condition.

Figure 1A: Paradigmatic mean stressed syllable durations (**verb**, postnuclear condition, n=15).
L to R: FOF, SOF, unF (before SOF), unF (following SOF)

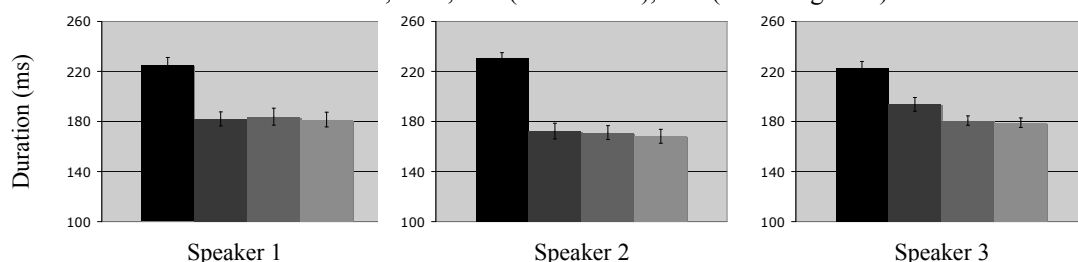
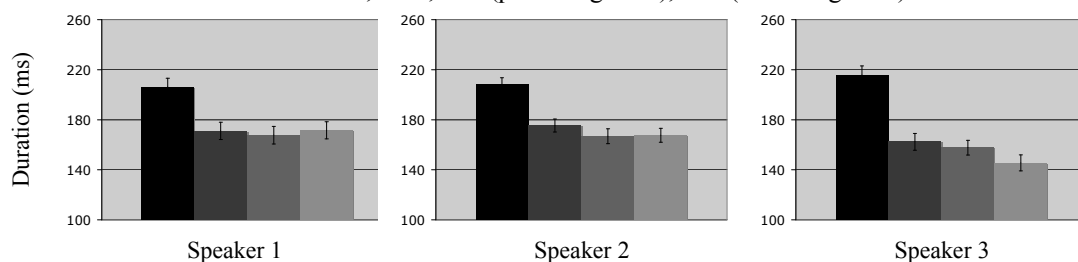


Figure 1B: Paradigmatic mean stressed syllable durations (**noun**, postnuclear condition, n=15).
L to R: FOF, SOF, unF (preceding SOF), unF (following SOF)



A similar pattern emerged for energy.⁵ For Speaker 1, SOF did not differ significantly from unF. For Speaker 2, there was a strongly significant difference for nouns ($p=.0008$), but only a marginal difference for verbs ($p=.0606$). For Speaker 3, there was a strongly significant difference for nouns ($p<.0001$), but no significant difference for verbs ($p=.5272$).

Neither measure of spectral balance was significant in differentiating SOF from unF (although interestingly the p -values were lower for nouns than for verbs). Does this mean that spectral balance is not a relevant cue? Looking at FOF, H1-H2 was significant for nouns in Speaker 1 ($p=.0262$) and Speaker 3 ($p=.0488$), but H1-H2 was insignificant for verbs.

2.2.2. Syntagmatic comparison

Turning to a syntagmatic comparison, a FOF verb had significantly greater duration than the adjacent unF noun (Figure 2A). Similarly, a SOF verb had significantly greater duration than an adjacent unF noun, although the difference was of a smaller magnitude (roughly 20 ms) (Figure 2B).

⁴ Thanks to Amanda Miller for assistance, and to Marc Brunelle and Mietta Liennes for related scripts.

⁵ I report significance values for the postnuclear case, but the same pattern holds in the prenuclear case.

Figure 2A: Mean stressed syllable durations (verb focus, postnuclear condition, n=15 for each).

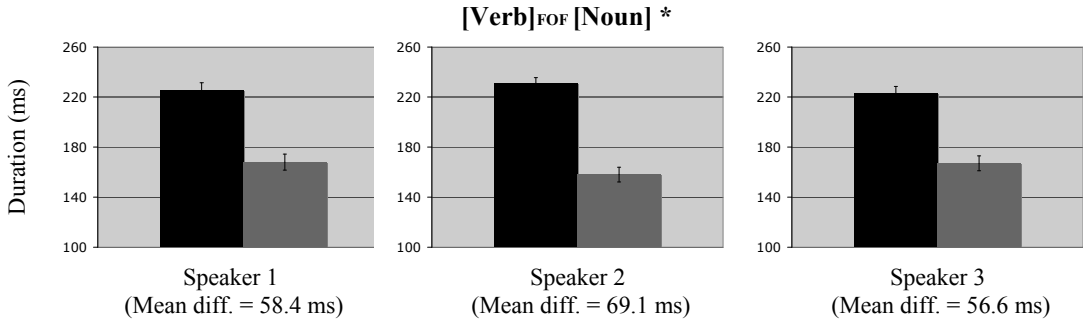
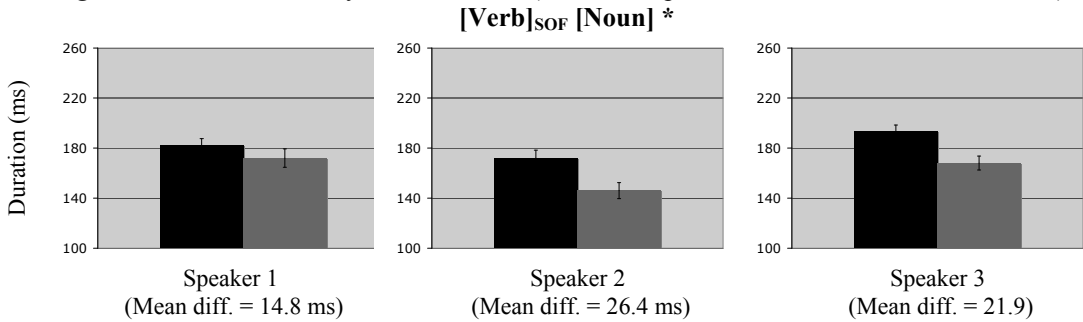


Figure 2B: Mean stressed syllable durations (verb focus, postnuclear condition, n=15 for each).



A FOF noun, too, had significantly greater duration than the adjacent unF verb (Figure 3A), although the difference was noticeably less than that between a FOF verb and the adjacent unF noun. A SOF noun, surprisingly, did not have significantly different duration from the adjacent verb (Figure 3B). Note that the averages in fact favour the verb.

Figure 3A: Mean stressed syllable durations (verb focus, postnuclear condition, n=15 for each).

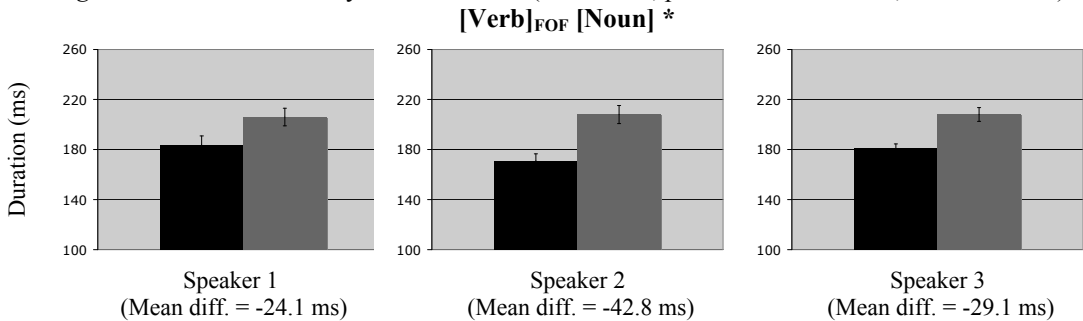
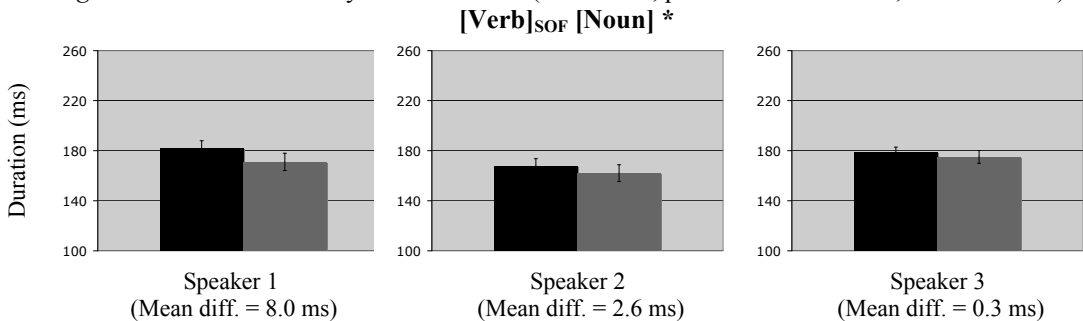


Figure 3B: Mean stressed syllable durations (verb focus, postnuclear condition, n=15 for each).



Similar results obtained for energy. There was a significant effect for position (i.e. verb vs. noun; $p < .0001$), but only a very marginal effect for focus (i.e. SOF vs. unF; $p = .0889$). This also held for spectral balance in terms of H1-H3 (position, $p < .0001$; focus, $p = .2400$) and H1-H2 (position, $p < .0001$; focus, $p = .6258$).

2.3. Discussion

The data do not bear out the expectation that a semantically focused constituent will have increased duration and acoustic energy absolutely. Rather, there is an effect of position which produces the opposite of the expected prominence for SOF nouns. Whatever the cause of the unexpected prominence, it seems clear that these acoustic measures, at least considered absolutely, are not reliable cues to SOF. This leaves us with (at least) two remaining questions. What is this unexpected prominence, and can listeners perceive focus despite it? These questions are addressed with follow-up experiments in Sections 3 and 4, respectively.

3. Second production experiment

3.1. Methods

Two of the speakers from the first production experiment were recorded under the same conditions. In this case, the associating adverb was replaced with a non-associating adverb in one discourse (e.g. 9) and no adverb in the other discourse (e.g. 10). This meant that neither the verb nor the noun had semantic focus. A wh-question in the context forced the verb/noun pair to be semantically given. The recordings were manually annotated, as before.

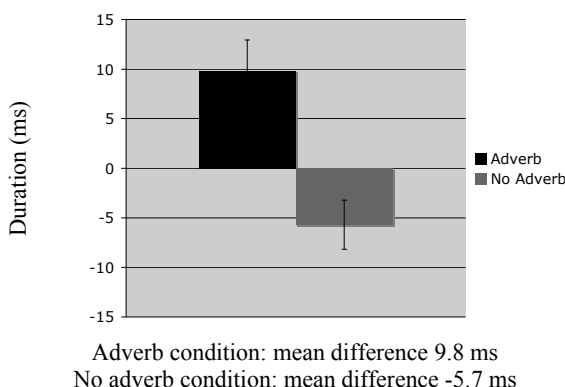
- (9) A: Who **poorly** patches patches for Microsoft?
B: JOHNSON **poorly** patches patches for Microsoft.

- (10) A: Who patches patches for Microsoft?
B: JOHNSON patches patches for Microsoft.

3.2. Results

Considering the difference in stressed syllable duration between the noun and verb in the given (i.e. B) sentences, the presence vs. absence of the adverb had a significant effect ($p = .0002$) on duration, while the choice of homophone, V vs. N (= 1st vs. 2nd probe), had no significant effect ($p = .6297$).⁶ The V-N stressed syllable duration differences are shown in Figure 4 for the adverb and no-adverb conditions.

Figure 4: Mean stressed syllable duration difference (V-N, $n = 30$)



⁶ Other acoustic measures are not reported at this time.

3.3. Discussion

The presence of the adverb had a clear effect (although an overall phonetic effect, rather than a strictly observed categorical effect). A likely explanation for the difference is rhythm.⁷ Note that, as in the first production experiment, the stimuli up to and including the target words consisted of bisyllabic trochees. In case the adverb intervenes between FOF *Johnson* and the verb, it is the verb which has greater duration than the noun. This was the case for the first production experiment. In case no adverb intervenes, it is the noun that has greater duration than the verb. This effect is strikingly analogous to results on word stress (e.g. *import* vs. *impórt*) in the postnuclear domain reported by Huss (1978). He found in speakers' production that "syllables which fit into a regular pattern of stressed and unstressed syllables are emphasized", although listeners do not perceive this rhythmic "emphasis". Let us turn, then, to perception.

4. Perception experiment

4.1. Methods

Six linguistically trained listeners participated in a forced-choice and acceptability listening experiment. Stimuli were presented in the form of a context retrieval task (cf. Gussenhoven 1983). Subjects chose between two discourses: the SOF sentence was the same in both; in one of the two discourses the context sentence matched in terms of intended focus association; in the other discourse the context sentence did not match (cf. 11). Stimuli were taken from the *pedals* recordings of all three speakers in the first production experiment, and the pairs of discourses were presented twice each in random order (n=60). After choosing the "better" discourse, listeners would then judge them: (i) only the selected discourse is acceptable; (ii) both are acceptable; or (iii) neither is acceptable.

(11) *Matching discourse*

A1: Johnson only PEDDLES pedals lately.

B1: Even THOMPSON only peddles pedals lately.

Non-matching discourse

A2: Johnson only peddles PEDALS lately.

B1: Even THOMPSON only peddles pedals lately.

4.2. Results

Figure 5 tabulates the percentage of stimuli correctly identified (i.e. the appropriately matching discourse was selected as "better") and the percentage of stimuli judged as "both acceptable" for each listener. The average success rate was 57.5%. L2, a trained phonetician, reported guessing for each stimulus and judged all discourses acceptable. L3, a trained musician as well as linguist, performed best (65% correct), although this listener judged nearly all discourses acceptable. Not surprisingly, duration had a significant effect on overall correct identification ($p=.0054$).

Figure 5: Rates of correct identification and rates of acceptability of both discourses (n=60)

	L1	L2	L3	L4	L5	L6	Average
% correct	55	50	68	65	47	60	57.5
% both acceptable	61	100	93	28	37	100	

4.3. Discussion

Overall, the sophisticated listeners were not particularly successful at identifying the matching discourse.⁸ Although anecdotal, the fact that a musician outperformed a trained phonetician leads one

⁷ Thanks to Bob Ladd for the initial suggestion that rhythm may be involved. Christian DiCanio (p.c.) points out that this result is also consistent with an earlier hypothesis that syllable duration decreases over the course of an utterance. I have not yet explored this possibility.

⁸ Regarding the complexity of the task, a simple matching task was also run (omitted here for space) where a matching/not matching judgment was elicited for a single discourse only. Results were largely the same: low success and reports of "guessing".

to speculate whether the discrimination is purely linguistic. It is also important to note that even for the best performing listener, an association with focus could be “coerced” by the context sentence.

5. Deciding among semantic theories

How do these results inform our choice between a grammatically-mediated or purely pragmatic account of association with focus? Let us consider the possibilities.

First, suppose that SOF is (always) semantically marked. Given that the SOF sentences are functionally ambiguous in most (if not all) cases, listeners must disambiguate by pragmatic reasoning.⁹ Speakers have a representation of the focus in their grammar, but the production of it may be “washed out” by other factors (e.g. rhythm). On this view, any notion of *Stress F* (cf. 3) must apply at a very abstract level, since prominence is neither faithfully realized nor used in interpretation. Another issue is economy: if pragmatics is doing the work, why posit a role for focus marking?

Second, suppose that SOF is never semantically marked. This would be something along the lines of Krifka (1996), who suggested that SOF amounts to a reduction anaphor without internal compositional semantics. The weak but observable production effects, we might argue, are due to phonetic copying, motor planning or some other “low-level” phenomenon.

Third, suppose that SOF is only sometimes semantically marked. It is not marked when pragmatically recoverable; it is marked otherwise. Again the effects are due to phonetic copying. One test for this would be the cases of bridging discussed by Rooth (1996) (cf. 12). If prominence on *younger* is necessary to license the implication that *Susan and Harold* are younger candidates, that would suggest that SOF is indeed sometimes marked.

- (12) The provost and the dean aren't taking any candidates other than Susan and Harold seriously.
Even the CHAIRMAN is only considering younger candidates.

Despite the promising results of Beaver *et al.* and others, it appears that the debate over SOF is still open. While we have not yet determined the precise role of pragmatics in the interpretation of focus, the present results argue that pragmatics are indeed a key ingredient given the unreliable acoustic prominence.

References

- Bartels, Christine. 2004. Acoustic correlates of ‘second occurrence’ focus: towards an experimental investigation. In H. Kamp and B. Partee (eds.), *Context-Dependence in the Analysis of Linguistic Meaning*. Amsterdam: Elsevier, 354-61.
- Beaver, David, Brady Clark, Edward Flemming, Florian Jaeger, and Maria Wolters. 2007. When semantics meets phonetics: acoustical studies of second occurrence focus. To appear in *Language* 83(2).
- Boersma, Paul and David Weenink. 2004. Praat: doing phonetics by computer (Version 4.2.29) [Computer program].
- Büring, Daniel. 2006. Been there, marked that – a tentative theory of second occurrence focus. Ms., UCLA.
- Büring, Daniel. 1997. *The Meaning of Topic and Focus – The 59th Street Bridge Accent*. London: Routledge.
- Campbell, Nick, and Mary Beckman. 1997. Stress, prominence, and spectral tilt. In A. Botinis et al. (eds.), *Intonation: Theory, Models and Applications* (Proceedings of an ESCA Workshop, September 18-20, 1997, Athens, Greece). ESCA and University of Athens Department of Informatics.
- Dryer, Matthew. 1994. The pragmatics of association with *only*. Paper presented at the 1994 Winter Meeting of the LSA. Ms., Boston, Massachusetts.
- Féry, Caroline, and Shinichiro Ishihara. 2005. Interpreting Second Occurrence Focus. Ms., University of Potsdam.
- Huss, Volker. 1978. English word stress in the postnuclear position. *Phonetica* 35, 86-105.
- Gussenhoven, Carlos. 1983. Testing the reality of focus domains. *Language and Speech* 26, 61-80.
- Jaeger, Florian. 2005. Only always associates audibly. Even if only is repeated. The prosodic properties of second occurrence focus in English. Ms., Stanford University.

⁹ Given that SOF occurs by definition in a given (and therefore disambiguating) context, this should not be surprising. This also suggests that the audience design hypothesis (speakers only disambiguate for the benefit of the listener) will be irrelevant for cases of SOF. That is, if the listener knows that SOF occurs in a disambiguating context and the speaker knows that the listener knows that it occurs in a disambiguating context, there should be no motivation for the speaker to disambiguate prosodically.

- Jackendoff, Ray. 1972. *Semantic Interpretation in Generative Grammar*. Cambridge: MIT Press.
- Klatt, Dennis H. 1976. Linguistic uses of segmental duration in English: acoustic and perceptual evidence. *Journal of the Acoustical Society of America* 59, 1208-1221.
- Krifka, Manfred. 1996. Focus and/or context: a second look at second occurrence focus. In H. Kamp and B. Partee (eds.), *Proceedings of the Prague/Tainach workshops*. IMS, University of Stuttgart.
- Lehiste, Ilse. 1980. Phonetic manifestation of syntactic structure in English. *Annual Bulletin of the Research Institute of Logopaedics and Phoniatrics* 14, 1-27.
- Partee, Barbara. 1991. Topic, focus and quantification. *Semantics and Linguistic Theory* 1, 159-87.
- Partee, Barbara. 1999. Focus, quantification, and semantics-pragmatics issues. In P. Bosch and R. van der Sandt (eds.), *Focus: Linguistic, Cognitive, and Computational Perspectives*, 213-231. Cambridge University Press.
- Roberts, Craige. 1996. Information structure in discourse: towards an integrated formal theory of pragmatics. *OSU Working Papers in Linguistics 49: Papers in Semantics*.
- Rooth, Mats. 2007. Second occurrence focus and relativized stress F. Ms., Cornell University.
- Rooth, Mats. 1996. On the interface principles for intonational focus. In T. Galloway and J. Spence (eds.), *Proceedings of SALT VI*, 202-226, Ithaca, NY: Cornell University.
- Rooth, Mats. 1992. A theory of focus interpretation. *Natural Language Semantics* 1, 75-116.
- Selkirk, Elizabeth. 2007. Contrastive focus, givenness and phrase stress. Ms., UMass-Amherst.
- Sluijter, Agaath, and Vincent van Heuven. 1996. Spectral balance as acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America* 100, 2471-85.
- Sorenson, John M., William E. Cooper, and Jeanne M. Paccia. 1978. Speech timing of grammatical categories. *Cognition* 6, 135-153.
- Stevens, Kenneth. 1998. *Acoustic Phonetics*. Cambridge: MIT Press.
- Truckenbrodt, Hubert. 1995. *Phonological Phrases: Their Relation to Syntax, Focus and Prominence*. Unpublished Ph.D. dissertation, MIT.

Proceedings of the 26th West Coast Conference on Formal Linguistics

edited by Charles B. Chang
and Hannah J. Haynie

Cascadilla Proceedings Project Somerville, MA 2008

Copyright information

Proceedings of the 26th West Coast Conference on Formal Linguistics
© 2008 Cascadilla Proceedings Project, Somerville, MA. All rights reserved

ISBN 1-57473-423-2 library binding

A copyright notice for each paper is located at the bottom of the first page of the paper.
Reprints for course packs can be authorized by Cascadilla Proceedings Project.

Ordering information

Orders for the library binding edition are handled by Cascadilla Press.
To place an order, go to www.lingref.com or contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, fax: 1-617-776-2271, e-mail: sales@cascadilla.com

Web access and citation information

This entire proceedings can also be viewed on the web at www.lingref.com. Each paper has a unique document # which can be added to citations to facilitate access. The document # should not replace the full citation.

This paper can be cited as:

Howell, Jonathan. 2008. Second Occurrence Focus and the Acoustics of Prominence. In *Proceedings of the 26th West Coast Conference on Formal Linguistics*, ed. Charles B. Chang and Hannah J. Haynie, 252-260. Somerville, MA: Cascadilla Proceedings Project.

or:

Howell, Jonathan. 2008. Second Occurrence Focus and the Acoustics of Prominence. In *Proceedings of the 26th West Coast Conference on Formal Linguistics*, ed. Charles B. Chang and Hannah J. Haynie, 252-260. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #1679.