Interfaces in Linguistics

New Research Perspectives

Edited by
RAFFAELLA FOLLI AND CHRISTIANE ULBRICH

OXFORD UNIVERSITY PRESS
Second Occurrence Focus and the Acoustics of Prominence*

JONATHAN HOWELL

16.1 Introduction

The interpretation of a variety of different linguistic expressions show sensitivity to phonological prominence, among them quantificational adverbs, determiner quantifiers, counterfactuals, generics, modals, comparatives, superlatives, and negation. Jackendoff (1972) called this phenomenon ‘association with focus’. The constituent which receives prominence and associates with a focus-sensitive expression is known as the focus associate, focus argument, or simply the focus. Often, the choice of focus associate yields truth-conditional effects, as illustrated by the minimal pair in (1) with the focus-sensitive adverb only. We can imagine a scenario on which (1a) is true and (1b) false, and vice versa. Prominence on the focus associate is marked with capital letters.

(1) a. Mary only offered CAKE to Sue.
   b. Mary only offered cake to SUE.

Second occurrence focus (SOF) is the label given to the repeated occurrence of a particular association with focus. One of the first examples in the formal semantics literature comes from Partee (1991).  

* 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, MOT2007, WCCFL26, SPINE2007 and OnLI2007. 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. And last but not least, thanks to two anonymous reviewers.

1 She notes that similar examples are found in Gussenhoven (1983), Roberts (1990), and Krifka (1991).
(2) A: Eva only gave xerox copies to the GRADUATE STUDENTS.
   B: No, PETR only gave xerox copies to the graduate students.

Rooth (1996, 2004) also includes in the category of SOF those cases in which an association with focus has not previously been uttered, but may be implicated. We assume in his example (3) that Susan and Harold are among the set of younger candidates.

(3) A: The provost and the dean aren’t taking any candidates other than Susan and Harold seriously.
   B: Even the CHAIRMAN is only considering younger candidates.

The empirical question addressed in this chapter is whether a potential associate (e.g. graduate students in 2B or younger in 3B) has phonological prominence. The answer has important consequences for our understanding of focus and the modules of language that conspire to produce it. First, syntactic and semantic theories of association with focus predict that the underlined constituents in (2) and (3) should have prominence; pragmatic theories of association with focus predict that they do not. Second, we also want to understand whether focus phenomena can be modelled in terms of a single notional category focus or givenness, or whether it is necessary to posit two distinct mechanisms. SOF plays an important role in arguments for both views. In the rest of this section, we’ll briefly review the role of SOF in these debates and the growing but not yet conclusive experimental research on SOF, which has found small but controversial phonetic differences between SOF associates and their unfocused counterparts.

In section 16.2, I report on a new production study of the acoustics of SOF. The results do not support the view that the associate of a focus-sensitive expression in general, or of only in particular, must always have phonological prominence. Indeed, a conflicting prominence pattern is observed, which in section 16.3 I attribute to rhythm based on the results of a second, follow-up production experiment. In section 16.4, I present a perception experiment using a subset of the production data. The perception results do not support the view that listeners use prominence to determine an SOF associate.

16.1.1 Second Occurrence Focus and Semantic Theory

According to a grammaticized theory of association with focus (cf. Jackendoff 1972; Rooth 1985; von Stechow 1991) the focus sensitivity of an expression is achieved by a lexical rule that makes direct reference to syntactic annotation, usually known as F(ocus)-marking, or to focus-determined semantic objects,
such as Rooth’s (1985, 1992) focus semantic value (FSV). By way of illustration, consider the following simplified definition for only.²

\[(\text{only } \Phi) \text{ is true iff } \Phi, \text{ and for all } p \in \text{FSV}(\Phi), \text{ if } p \text{ is true then } p = \Phi\]

Roughly, (3) states that the constituent of only and its semantic argument \(\Phi\) is true if and only if \(\Phi\) is true and there is no other true alternative in its FSV. The FSV is determined by making substitutions for an F-marked constituent. In (4a), the FSV of \(\text{John eats [dinner]}\) is the set in (4b).

\begin{enumerate}
\item a. John only eats [dinner]_{\Phi}.
\item b. FSV = \{’John eats breakfast’, ’John eats lunch’, ’John eats dinner’, \ldots\}\]
\end{enumerate}

These theories are considered ‘weak’ (Rooth 1992) because it must be stipulated for each focus-sensitive expression how it operates on its focus.

A pragmatic account holds that a more general extra-grammatical principle is responsible, indirectly, for association with focus.⁴ In particular, the ‘domain selection’ type of pragmatic account (cf. Rooth 1992; von Fintel 1994; Kadmon 2001; Marti 2003; Kriika 2004) seeks to analyse association with focus by capitalizing on extant theories of other context-sensitive expressions. Consider the domain of the universal quantifier everyone in (5).

(5) Mary had a party. Everyone danced.

Since in context everyone does not refer to every individual in the world, some pragmatic mechanism is responsible for specifying its domain (e.g. to the set of attendees at Mary’s party). Similarly, a focus-sensitive expression such as only also has a domain. Following Rooth (1992), let’s assign that domain a variable \(C\) and suppose that the definition of only refers to \(C\) (rather than to the FSV as above).

\[(\text{only } \Phi) \text{ is true iff } \Phi, \text{ and for all } p \in C, \text{ if } p \text{ is true then } p = \Phi\]

The role of focus is to constrain the domain \(C\) to a subset of the FSV.⁵ Suppose that \(C\) is fixed to (6a) in one context and to (6b) in another. \(C_1\) is an appropriate domain for only in (4a) since \(C_1\) is a subset of the FSV (4b); \(C_2\) is not an appropriate domain, since it is not a subset of the FSV.

² See Kadmon (2001: 299) for a more formally explicit definition of only within a rule-based theory of focus association.
³ For presentational purposes, I assume that only takes a sentential argument (e.g. John eats dinner).
⁵ We must also stipulate that the subset includes \(\Phi\) and at least one other alternative.

On this type of analysis, the meaning of a focus-sensitive operator like only does not refer directly to focus-dependent objects like F-marked syntactic constituents or FSVs. Consequently, the pragmatic account allows for occurrences of focus-sensitive expressions without an associating focus (and therefore without prominence on a potential associate). Grammaticized accounts, on the other hand, require an associating focus, and thus predict that an associate must always have prominence.

Beaver and Clark (2008) propose an intermediate or hybrid account (see also Rooth 1992) according to which the semantics of some but not all focus-sensitive expressions refer directly to focus, among them only and even. On this intermediate account, prominence is only predicted for the associates of these particular focus-sensitive expressions.

A second, related debate concerns the mapping between the semantic representation of focus and the phonological representation of prominence. On the semantics side, it is often debated whether focus phenomena should be modelled in terms of a single notional category ‘focus’ (e.g. Rooth 1992), a single notional category ‘givenness’ (e.g. Schwarzschild 1999), or two distinct notional categories ‘contrastive focus’ and ‘givenness’ (e.g. Selkirk 2008). On the phonology side, there is general agreement that prominence is hierarchical, with relative prominence existing in English at the word level (e.g. Chomsky and Halle 1968), at the sentence level, and at various other levels of stress in between (e.g. Selkirk 1980; Nespor and Vogel 1986; Beckman and Pierrehumbert 1986). We also know, since Liberman (1979) and Pierrehumbert (1980) that intonational events (i.e. tones or pitch accents) belong to a semi-autonomous, highest level of prominence.

The existence of different levels of phonological prominence means there need not be a single phonological correlate of semantic focus, such as pitch accent. Indeed, most formulations of the mapping between the scope of a focus-sensitive expression and phonology allow for this, referring to relative stress or prominence, rather than one particular level of stress or prominence (e.g. Truckenbrodt 1995; Rooth 2008).

(7) Stress F (Rooth 2009)
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.

(8) Destress Given (Féry and Samek-Lodovici 2006)
A given phrase is prosodically non-prominent.

Selkirk (2008) (see also Féry and Ishihara 2009) advances SOF as one piece of evidence for the coexistence of two distinct categories of contrastive focus and discourse-givenness in the grammar, the former mapped to phonology according to a principle like (7) and the latter according to a rule Destress Given (cf. 8). Since in SOF examples both the focus-sensitive expression and everything in its scope are already given (or may be taken to be so, cf. (3)), they must be phonologically less prominent relative to the rest of the sentence. For example, in (2B), *only gave xerox copies to the graduate students* is given, and is therefore predicted to have less prominence (e.g. reduced or no pitch accenting) than *Petr*. However, the focus associate within *gave xerox copies to the graduate students*, namely *graduate students*, will still have greatest relative prominence (e.g. stress) within that constituent. Selkirk's theory is therefore motivated in part by putative prominence in SOF.

Rooth (2006, 2009) and Büring (2008), who are inclined towards a single-notion theory of focus, offer proposals relating the phonological realization of focus to the relative scope of different focus-sensitive operators. As schematized in (9), a focus-sensitive operator semantically embedded under another focus will have a focus associate that is realized with SOF phonology (e.g. lacking a pitch accent), while the associate of the widest-scope focus operator will have regular focus phonology (e.g. with a pitch accent). In (3b), for example, *only* is embedded under *even* and therefore the associate of *only* will have second occurrence focus.

(9) Configurational SOF (adapted from Rooth 1996)

\[ \ldots F \ldots [[ \ldots SOF \ldots ] \ \text{Op}_2] \ldots ] \ \text{Op}_1 \]

These analyses allow for the phonological realization of SOF to differ from regular focus, but the semantic notion of relative scope itself is independently motivated and does not by itself necessitate that the focus associate in an SOF utterance have prominence.

16.1.2 Previous Investigations of Prominence

Several experimental investigations followed Partee's (1991) first observations, beginning with Rooth (1996). Investigating the acoustics of his own speech, Rooth compared three different acoustic measures. The first, $f_0$ movement, is the change in fundamental frequency, the physical correlate of pitch; the second is syllable duration; and the third, root mean squared (RMS) amplitude, is calculated from sound wave and measures acoustic intensity. Rooth found in his data, uncontroversially, that that regular focus (henceforth 'first occurrence focus' FOF) showed large $f_0$ maxima, while SOF lacked any significant $f_0$ movement. However, in comparing an expected SOF associate and an adjacent unfocused
(unF) word, Rooth found that the SOF associate had a measurably longer syllable duration and greater RMS amplitude. This is illustrated in (10), with the direction of comparison illustrated with a horizontal arrow: in (10aB), named had a greater syllable duration and amplitude than Manny in Rooth’s data, and vice versa in (10bB). Perceptually, he judged it possible to identify the correct association with focus listening only to an SOF utterance (i.e. 10aB or 10bB), in the absence of the context-supplying FOF sentence (i.e. 10aA or 10bB).

(10)  
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.

Bartels (2004) later determined in a production experiment with six native English speakers that the acoustics of FOF and SOF were statistically different. She measured relative f₀ maxima, RMS amplitude, and word duration, calculated as ratios of a focus associate and its preceding words or syllables. The results support the intuition that FOF and SOF have different realizations—overall, the SOF associates had reduced relative f₀ maxima, RMS amplitude, and word duration compared to FOF associates. Unlike Rooth (1996), Bartels’s experiment did not test whether an SOF constituent had greater prominence compared to an unF constituent.

Beaver et al. (2007) and Jaeger (2005) conducted a large production study with twenty native English speakers. In addition to f₀ maxima, RMS amplitude, and word duration, the authors also measured f₀ minima, f₀ mean, f₀ range and acoustic energy, a function of amplitude and duration, within a target utterance. The direction of comparison was both within a given utterance (‘syntagmatic’ comparison), as in Rooth (1996), and across utterances (‘paradigmatic’ comparison). Two minimal discourse pairs from the study are given in 11–12. The (b) sentence contains an FOF and the (c) sentence a SOF. In (5c), Pete is the potential associate; in (12c), a pill is the potential associate. The syntagmatic comparison (cf. horizontal arrow) contrasts Pete with a pill in the same utterance. The paradigmatic comparison (cf. vertical arrow) contrasts Pete in (11c) with Pete in (12c), and a pill in (11c) with a pill in (12c). Note that the inclusion of both syntagmatic and paradigmatic comparisons is important, since it remains an open question whether speakers and listeners produce and perceive focus prominence syntagmatically or paradigmatically.
In confirmation of previous studies, Beaver et al. and Jaeger found that $f_0$ was not a significant predictor of SOF. However, the authors report a statistically significant difference in duration between SOF and unF, both syntagmatically (10.1ms and 8.1ms) and paradigmatically (6ms). Intensity and energy were also significant or approaching significance, both syntagmatically (.8519/.6354 dB and 0.0049/0.0027) and paradigmatically (0.13 dB and 0.0011).

(11)  
a. Context sentence  
Both Pete and Edward are suffering from the flu.  
b. FOF sentence  
But the nurse only gave PETE a pill today.  
c. SOF sentence  
Even THE DOCTOR only gave Pete a pill today.

(12)  
(11c)  
a. Context sentence  
Pete really needed an injection to ease the pain.  
b. FOF sentence  
But the nurse only gave Pete A PILL today.  
c. SOF sentence  
Even THE DOCTOR only gave Pete a pill today.

While statistically significant, these acoustic values fall short of some published just noticeable differences (JNDs) (i.e. smallest perceivable differences) for speech sounds: 10-40ms (Lehiste 1980) and ~25ms (Klatt 1976) for duration; 1–4dB (Stevens 1998) for amplitude. The authors, therefore, conducted a perception experiment.

Beaver et al. presented native English-speaking listeners with minimal SOF pairs (e.g. (11c) and (12c)) from a subset of their production data and asked them to identify in which of the two renditions the second target (e.g. a pill) was more prominent. Subjects performed above chance in this discrimination task, but averaged only 63 per cent accuracy. The authors speculate that the less than perfect performance may be due to reader disfluencies in the laboratory-elicited production stimuli. Further, as the authors admit, it is impossible to conclude from these results alone whether listeners actually

---

7 The two syntagmatic differences reported correspond to different linear orderings: SOF-unF and unF-SOF, respectively.
exploit this discrimination in interpretation, or even whether the discrimination reflects purely linguistic competence.

Féry and Ishihara (2009) investigated SOF in German. In a 29-subject production experiment, they compared mean $f_0$ and word duration paradigmatically (cf. (13a–c)), although unlike Beaver et al. (2007) they measured unF targets in an unrelated discourse (cf. (13c)). A major concern in this study is the phonological notion of nuclear accent (see Newman 1946: 176; Chomsky and Halle 1968: 90), namely that the last pitch accent in an utterance is perceived as most prominent and any following material in the utterance is reduced in pitch. This reduction phenomenon in English and German is known as post-nuclear deaccenting. Since an FOF associate is typically a nuclear accent, Féry and Ishihara investigated both SOF associates which preceded an FOF (cf. (13b)), as well as SOF associates which followed the nuclear accent (cf. all previous studies).

(13) Context sentence:
Die meisten unserer Kollegen waren beim Beriebsausflug lässig angezogen.
‘Most of our colleagues were dressed casually at the staff outing.’

a. FOF sentence
Nur PETER hat eine Krawatte getragen.
‘Only Peter wore a tie.’

b. SOF sentence
Nur Peter hat sogar einen ANZUG getragen.
‘Only Peter even wore a suit.’

The authors found a significant paradigmatic difference in $f_0$ and word duration between SOF and unF in both pre-nuclear and post-nuclear positions, although they report that the difference in $f_0$ is ‘radically reduced’ in the post-nuclear position.

8 The authors also investigated sentence position and found that SOF has a higher mean pitch and a longer word duration sentence-initially than sentence-medially. They attribute this to the phonological phenomena of intonational downstepping and phrase-final lengthening.
Finally, Bishop (forthcoming) tests pre-nuclear SOF and unF in English in a five-speaker production experiment, using the methodology of Rooth (1996) and Beaver et al. In a paradigmatic comparison, Bishop found, contra Féry and Ishihara (2009) for German, that the pre-nuclear SOF associates did not have significantly greater $f_0$ maxima than the related unF targets. It is not clear whether this contrast is due to differences in methodology or due to differences between English and German. Bishop did confirm, however, that the duration of SOF and unF differed significantly, although as in Beaver et al. the durational differences were small and hovered around the JNDs mentioned above.

In summary, previous investigations confirmed that potential SOF associates are less prominent than FOF associates. Small, but statistically significant acoustic differences were observed between SOF associates and their unF counterparts, both syntagmatically and paradigmatically, and both pre-nuclearly and post-nuclearly. Presented with a minimal pair of SOF sentences, listeners identified prominence poorly, but above chance.

16.2 First Production Experiment

16.2.1 Method

16.2.1.1 Subjects I ran a small production study with three male speakers, including myself. Sophisticated, non-naïve speakers were chosen in order to complement the results of Beaver et al. (2007)/Jaeger (2005) who used naïve speakers, and in answer to the speculation that speakers’ naïveté and reading disfluencies weakened their results. Recall that Rooth (1996) reported clear results in his own speech.

16.2.1.2 Recording Subjects were recorded in a sound-attenuated room, using a Plantronics DSP-500 headset to control for head movement. Recording and analysis were both conducted with Praat 4.2.29 (Boersma and Weenink 2008). Subjects repeated each discourse five times for a total of sixty 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 rerecord the token without prompting.

16.2.1.3 Stimuli Following the methodology of Rooth (1996) and Beaver et al. (2007), the elicited discourses consisted of a context sentence with FOF and a target sentence with SOF, allowing paradigmatic and syntagmatic comparison.

---

9 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.
The target words were noun/verb homophones (peddles/pedals, patches, labels), used in order to avoid post-hoc normalization and in order to control for vowel quality. Following Féry and Ishihara (2009) and Bishop (forthcoming), I included SOF and unF targets which preceded FOF as well as those which followed FOF. Finally, with the intent of controlling for possible isochronic effects (disruptions to regular rhythm, cf. Lehiste 1980), I limited my target words to bisyllabic trochaes (two-syllable words with first-syllable word stress). Some examples follow; (14) contains a post-nuclear SOF and (15) a pre-nuclear SOF.

14. Post-nuclear
   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.

15. Pre-nuclear
   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.

16.2.1.4 Measurements

   The target words were manually annotated in Praat for stop closure, aspiration, first vowel duration, and second syllable duration. The following values were then automatically extracted using Praat scripts: maximum $f_0$, minimum $f_0$, mean $f_0$, stressed syllable duration, mean RMS intensity, energy, power, and spectral balance. Spectral balance (aka spectral tilt), not used in previous investigations of SOF, measures relative amplitude at different harmonic frequencies, rather than over the entire spectrum and has been claimed to be an acoustic correlate of stress in English and Dutch (Campbell and Beckman 1997; Sluijter and van Heuven 1996). The following spectral-balance values were calculated from differentials of harmonic frequencies (H) and amplitudes (A) using a 10ms window centred at the time of the first formant maximum: $H_1-H_2$, $H_1-H_3$, $H_1-A_1$, and $H_1-A_2$.

---

$^{10}$ Thanks to Amanda Miller for assistance, and to Marc Brunelle and Mietta Liennes for related scripts.
16.2.1.5 Statistical Analysis  

For the syntagmatic comparison, a paired t-test was conducted to compare the means of adjacent verb and noun targets in the FOF sentences, and then in the SOF sentences. For the paradigmatic comparison, a t-test was also conducted to compare the means of the focused targets (FOF or SOF) to those of the unfocused targets among verbs, and the means of focused targets to those of the unfocused targets among nouns. The differences are summarized in Tables 1 and 2; significance is indicated by (*) and by shading.

Finally, a 2x2 analysis of variance (ANOVA) was conducted for focus value (focused vs unfocused) and word position (verb vs noun, or equivalently first position vs second position), both for the FOF condition, as a kind of baseline, and for the SOF condition of interest. The ANOVA tests the hypothesis that focus value and word position are significant predictors ('main effects') of the observed differences of an acoustic measure.

Statistical tests were also performed by speaker, by nuclear position (pre-nuclear or post-nuclear) and by word type, but did not pattern differently from the complete dataset and so are omitted here for space. Transformations of the data were performed for non-linear measurements, including the natural log of duration, the square of energy, and root mean square of

---

**Table 1.** Summary of mean syntagmatic differences between focused and unfocused by focus type (verb focus vs noun focus) for FOF and SOF. Significance of $\alpha < 0.05$ on a paired t-test indicated by (*).

<table>
<thead>
<tr>
<th></th>
<th>FOF Verb Focus</th>
<th>Noun Focus</th>
<th>SOF Verb Focus</th>
<th>Noun Focus</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[verb]$_{FOF}$</td>
<td>[noun]$_{unF}$</td>
<td>[verb]$_{SOF}$</td>
<td>[noun]$_{unF}$</td>
</tr>
<tr>
<td>Duration (ms)</td>
<td>N</td>
<td></td>
<td>N</td>
<td></td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>31.0*</td>
<td>21.0*</td>
<td>21.0*</td>
</tr>
<tr>
<td>Mean $f_0$ (Hz)</td>
<td>83</td>
<td>50.3*</td>
<td>0.75</td>
<td>0.81</td>
</tr>
<tr>
<td>Max $f_0$ (Hz)</td>
<td>83</td>
<td>57.4*</td>
<td>38.0*</td>
<td>2.3</td>
</tr>
<tr>
<td>Min $f_0$ (Hz)</td>
<td>83</td>
<td>39.6*</td>
<td>19.1*</td>
<td>2.1*</td>
</tr>
<tr>
<td>RMS Intensity (dB)</td>
<td>90</td>
<td>7.66*</td>
<td>4.70*</td>
<td>1.14*</td>
</tr>
<tr>
<td>Energy</td>
<td>90</td>
<td>0.00586*</td>
<td>0.00569*</td>
<td>0.000005</td>
</tr>
<tr>
<td>Power</td>
<td>90</td>
<td>0.038*</td>
<td>0.029*</td>
<td>0.001*</td>
</tr>
<tr>
<td>$H_1$-$H_2$ (Hz)</td>
<td>90</td>
<td>-0.867*</td>
<td>-0.998*</td>
<td>-0.482</td>
</tr>
<tr>
<td>$H_1$-$H_3$ (Hz)</td>
<td>90</td>
<td>-1.227*</td>
<td>-1.68*</td>
<td>-0.672</td>
</tr>
<tr>
<td>$H_1$-$A_1$ (Hz)</td>
<td>90</td>
<td>-7.747*</td>
<td>-5.655*</td>
<td>-0.728</td>
</tr>
<tr>
<td>$H_1$-$A_2$ (Hz)</td>
<td>90</td>
<td>-7.542*</td>
<td>-5.737*</td>
<td>-1.146</td>
</tr>
</tbody>
</table>
energy. All tests were performed in the statistical computing environment R (R Development Core Team 2008).

16.2.2 Results

16.2.2.1 Syntagmatic Comparison Unsurprisingly, results of the syntagmatic comparison revealed that an FOF target was more prominent than its adjacent unfocused target. All measures in the FOF condition were significant, as summarized in Table 1. Few of the measures were statistically significant for the SOF condition duration, minimum \( f_0 \), intensity and power were significant, but only in the verb focus comparison.

Looking specifically at duration, we can see that an FOF verb had significantly greater duration than the adjacent unF noun (Figure 1A), and an FOF noun had significantly greater duration than the adjacent unF verb (Figure 1B). It is also worth mentioning that, intuitively, the magnitude of difference is much greater for the verb-focused condition than for the noun-focused condition. This will be relevant for the discussion of rhythm in section 3.

A SOF verb did have significantly different duration from the adjacent noun (Figure 2B); however, an SOF noun was not significantly longer than the adjacent verb (Figure 2A). In fact, the direction of difference, like most of the other measures, favoured the verb.

**Table 2.** Summary of mean paradigmatic differences between focused and unfocused by word position (verb position vs noun position) for FOF and SOF. Significance of \( \alpha < 0.05 \) on a paired t-test indicated by (*).

<table>
<thead>
<tr>
<th></th>
<th>FOF Verb</th>
<th>FOF Noun</th>
<th>SOF Verb</th>
<th>SOF Noun</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duration (ms)</td>
<td>180</td>
<td>48.4*</td>
<td>45.1*</td>
<td>10.4*</td>
</tr>
<tr>
<td>Mean fo (Hz)</td>
<td>173/4</td>
<td>35.5*</td>
<td>44.1*</td>
<td>2.47</td>
</tr>
<tr>
<td>Max fo (Hz)</td>
<td>173/4</td>
<td>48.6*</td>
<td>50.8*</td>
<td>2.58</td>
</tr>
<tr>
<td>Min fo (Hz)</td>
<td>173/4</td>
<td>24.5*</td>
<td>34.2*</td>
<td>2.74</td>
</tr>
<tr>
<td>RMS Intensity (dB)</td>
<td>180</td>
<td>5.78*</td>
<td>6.57*</td>
<td>0.57</td>
</tr>
<tr>
<td>Energy</td>
<td>180</td>
<td>0.00353*</td>
<td>0.00402*</td>
<td>0.00015</td>
</tr>
<tr>
<td>Power</td>
<td>180</td>
<td>0.033*</td>
<td>0.034*</td>
<td>0.002</td>
</tr>
<tr>
<td>H1-H2 (Hz)</td>
<td>180</td>
<td>-0.7900+</td>
<td>-1.0742*</td>
<td>-0.0269</td>
</tr>
<tr>
<td>H1-H3 (Hz)</td>
<td>180</td>
<td>-1.0500</td>
<td>-1.8599*</td>
<td>-0.395</td>
</tr>
<tr>
<td>H1-A1 (Hz)</td>
<td>180</td>
<td>-5.4910*</td>
<td>-7.9117*</td>
<td>-2.089</td>
</tr>
<tr>
<td>H1-A2 (Hz)</td>
<td>180</td>
<td>-5.1442*</td>
<td>-8.1350*</td>
<td>0.002</td>
</tr>
</tbody>
</table>
Acoustics of Second Occurrence Focus

Previous investigations of SOF, in particular Jaeger (2005), would lead us to expect parallel behaviour between the FOF condition and the SOF condition, i.e. that a focused target will have greater prominence than the adjacent unfocused target, regardless of whether the focus is FOF or SOF, at least for some measures. Instead, we observe some significant differences among verb-focused SOF cases, and no significant differences among noun-focused SOF cases.

16.2.2.2 Paradigmatic Comparison As seen in Table 2, paradigmatic difference between FOF and unfocused verb targets and between FOF and unfocused noun targets were statistically significant for almost all measures. Those same differences between SOF and unF targets lacked statistical significance for nearly all measures, with the notable exception of duration.

The mean durations of FOF, SOF, and unfocused targets (adjacent to FOF and adjacent to SOF given separately) are illustrated in Figures 3A, B. The contrast between the FOF targets and the other targets is clear; as in the syntagmatic comparison above, the contrast between the SOF targets and the unfocused targets, while statistically significant, is much less striking.

**Figures 1A and 1B.** Mean stressed syllable durations for verb FOF and adjacent unfocused noun, and noun FOF and adjacent unfocused verb, respectively (with 95% confidence interval).

**Figure 2A and 2B.** Mean stressed syllable durations for verb SOF and adjacent unfocused noun, and noun SOF and adjacent unfocused verb, respectively (with 95% confidence interval).
16.2.2.3 Analysis of Variance  In the FOF condition, there was a main effect of focus value (focused vs unfocused) for all measures. There was also a main effect of word position (verb vs noun, or 1st vs 2nd position) for all but the energy and spectral balance measures. No interaction effects were observed.

In the SOF condition, there was a main effect of focus value and of position on duration only, mirroring the pattern observed in the paradigmatic comparison. From Beaver et al. (2007), we would predict main effects of both focus value and position on several of the measures.

16.2.3 Discussion
The results of the first production experiment partly confirm previous experimental results. Paradigmatically, there is indeed a small but significant difference in duration between SOF and unF. The ANOVA also revealed a main effect of focus on duration.

Surprisingly, however, none of the other measures showed a main effect. Furthermore, there was also a main effect of position. Syntagmatically, the results strongly suggest a tendency towards greater relative prominence on the verb, independent of which target was the intended focus associate. In the next section, I explore this tendency towards prominence on the verb with a second production experiment and hypothesize that it is due to rhythm. Whether or not this hypothesis is correct, the results of the main production experiment reported in this section caution against absolute syntagmatic comparison as a reliable correlate of focus in SOF.

Section 16.4 explores whether listeners can perceive SOF in these data, either given the admittedly weak paradigmatic information or in spite of the seemingly misleading syntagmatic information.
292  Acoustics of Second Occurrence Focus

Table 3a. Analysis of Variance (ANOVA) by focus value (focus vs unfocus) and word position (verb vs noun) for FOF. Significance of $\alpha < 0.05$ on a paired t-test indicated by (*).

<table>
<thead>
<tr>
<th>FOF</th>
<th>N</th>
<th>Focus Value</th>
<th>Word Position</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>F = 337.49, p &lt; 0.001*</td>
<td>F = 34.60, p &lt; 0.001*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F = 189.88, p &lt; 0.001*</td>
<td>F = 11.74, p &lt; 0.001*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F = 115.58, p &lt; 0.001*</td>
<td>F = 1.03, p = 0.31</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F = 277.65, p &lt; 0.001*</td>
<td>F = 19.29, p &lt; 0.001*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F = 341.66, p &lt; 0.001*</td>
<td>F = 14.06, p &lt; 0.001*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F = 138.27, p &lt; 0.001*</td>
<td>F = 17.07, p &lt; 0.001*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F = 318.41, p &lt; 0.001*</td>
<td>F = 4.60, p &lt; 0.01*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F = 9.81, p &lt; 0.01**</td>
<td>F = 0.05, p = 0.83</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F = 7.94, p &lt; 0.01**</td>
<td>F = 0.19, p = 0.66</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F = 28.57, p &lt; 0.001*</td>
<td>F = 0.70, p = 0.40</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F = 25.79, p &lt; 0.001*</td>
<td>F = 0.48, p = 0.49</td>
</tr>
</tbody>
</table>

Table 3b. Analysis of Variance (ANOVA) by focus value (focus vs unfocus) and word position (verb vs noun) for SOF. Significance of $\alpha < 0.05$ indicated by (*).

<table>
<thead>
<tr>
<th>SOF</th>
<th>N</th>
<th>Focus</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>F = 13.05, p &lt; 0.001*</td>
<td>F = 20.78, p &lt; 0.001*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F = 1.19, p = 0.28</td>
<td>F = 1.75, p = 0.19</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F = 1.14, p = 0.29</td>
<td>F = 1.46, p = 0.29</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F = 0.22, p = 0.64</td>
<td>F &lt; 0.01, p = 0.99</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F = 0.02, p = 0.90</td>
<td>F = 1.30, p = 0.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F = 0.28, p = 0.59</td>
<td>F = 0.63, p = 0.43</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F = 2.91, p = 0.09</td>
<td>F = 0.37, p = 0.54</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F = 0.37, p = 0.55</td>
<td>F = 1.18, p = 0.28</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F = 0.11, p = 0.74</td>
<td>F = 1.01, p = 0.32</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F = 0.15, p = 0.70</td>
<td>F = 0.03, p = 0.87</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F = 0.01, p = 0.90</td>
<td>F = 0.57, p = 0.45</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F = 0.01, p = 0.90</td>
<td>F = 0.57, p = 0.45</td>
</tr>
</tbody>
</table>

16.3 Second Production Experiment

16.3.1 Methods

Two of the speakers from the first production experiment were recorded under the same conditions. The stimuli were constructed to closely resemble the SOF sentences in the first production experiment, but without the
focus-sensitive adverb only. Half of the stimuli contained instead no adverb at all (e.g. (16b)); the other half contained a non-focus-sensitive adverb (e.g. (17b)). The target sentence was the answer to a wh-question, which meant that the subject, Johnson, was always an FOF, and the target words (e.g. patches and patches) were always unfocused. Any prominence found on the target words would not, therefore, be attributable to SOF. Each discourse was repeated five times in each condition, for a total of sixty discourse tokens. Recordings were manually annotated, as before.

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

(17) A: Who poorly patches patches for Microsoft?
    B: JOHNSON poorly patches patches for Microsoft.

16.3.2 Results
To test whether adverb condition (presence vs absence of the adverb) or word position (being the verb or the noun) was a predictor of any of the acoustic measures, I carried out a two-way ANOVA. A main effect of adverb condition emerged for duration, as well as two of the spectral measures. There were no main effects of word position for any measure except minimum \( f_0 \): the nouns were a significant 2.08 Hz lower than verbs (paired t-test: \( t(45) = 2.59, p < 0.05 \)). A possible hypothesis for the main effect of position on minimum \( f_0 \) is the phenomenon of declination, the tendency for \( f_0 \) to decrease over the course of an utterance, all else being equal.
Among the utterances with the adverb, verbs had significantly longer stressed syllable duration than the nouns (mean difference 9.8 ms, paired t-test: t(29) = 2.87, p < 0.01) (Figure 4A). Among the utterances without an adverb, verbs had significantly shorter stressed syllable duration than the nouns (mean difference 5.7 ms).

### 16.3.3 Discussion

Perhaps surprisingly, the presence of the adverb had a clear effect on acoustic prominence in the rest of the utterance. The utterances with the adverb are relevant because they closely resemble the stimuli in the first production experiment, with the difference that the adverb in this experiment is not a focus-sensitive expression. In the first production experiment, we also noted a tendency towards a more prominent verb, both paradigmatically and syntagmatically. The tendency is not categorical, i.e. present in every production, but it is overall statistically significant.

The important point for our discussion of SOF and acoustic prominence is that we have observed a statistically significant difference that is not attributable to semantic focus. Notably, the difference for duration is on the same order of magnitude as that attributed to SOF in previous investigations.

I also want to speculate that the particular source of the observed prominence is eurhythmny, or regular rhythm. I will discuss two studies. First, according to the Phonetic Spacing Hypothesis of Hayes (1984), speakers attempt to produce phonetic prominence at regular intervals, either with respect to actual physical time or to abstract phonological timing units. Hayes proposes the hypothesis as an alternative to purely phonological analyses of a stress-shift phenomenon known as the ‘rhythm rule’. For example, many speakers who would consistently utter *fourteen* in isolation demonstrate a preference for *fourteen wômen* over *fourteên wômen*, the latter...
of which involves a ‘clash’ of stress on two adjacent syllables (stress indicated by accent marks). However, Hayes notes (see also Liberman and Prince 1977) that another strategy employed is to lengthen a syllable to create greater phonetic distance between adjacent stresses (i.e. lengthening of the -téen in *fourtéen*).

Second, while little study has been undertaken on the phonetics of prominence in the postnuclear domain, an experiment by Huss (1978) is of particular note. Huss compared the production and perception of stress-sensitive noun-verb pairs like *increase* and *incrédase* following a nuclear accent, or what we could call an FOF (e.g. (18) vs (19)). He found that the stressed syllable of the *noun* had longer relative duration and intensity than the unstressed syllable, but the two syllables of the verb were largely even. Further, when listeners listened to the target sentence without the preceding context, they were most likely to identify nouns.

(18) Noun rendition
(Whereas formerly the workers’ increase used to benefit from inflation)
now the EMPLÓYERS’ *increase benefits*.

(19) Verb rendition
(Whereas formerly government used to increase benefits)
now the EMPLÓYERS incrédase benefits.

Huss attributed this behaviour to the tendency towards regular rhythmic intervals: speakers’ productions tended towards first-syllable prominence on *increase* in (18)–(19) in order to achieve even spacing between the stress on *increase* the preceding stress on *employers*.

In a second experiment, Huss recorded the same target words in sentences with a rhythmic context favouring second-syllable prominence; cf. (20)–(21). The opposite effect was observed: the second syllable had greater duration and intensity in the verb rendition, and even the noun rendition.

(20) Noun rendition
(Whereas formerly the French increase was worse)
now the DÚTCH increase is worse

(21) Verb rendition
(Whereas formerly the French increased their stock)
now the DÚTCH incrédase their stock.

The targets *pátches pátches* in the present experiment have word stress at equal levels. If, however, we imagine the phonetic spacing hypothesis also holds
at a higher level between phonological words (bisyllabic trochees in our data),
the observed tendency parallels the results of Huss at the syllable level. In
(22)–(23), the adverb leaves exactly one word or trochee between the FOF and
the verb. By hypothesis, the verb therefore receives greater prominence in
order to maintain eurhythmy. I leave more rigorous investigation of this
phenomenon to future study.

(22)  [JÓHNSON poorly] [páčhes patches] for Microsoft.
(23)  Even [THÓMPSON only] [páčhes patches] for Microsoft.

16.4 Perception Experiment

16.4.1 Methods
Six linguistically trained listeners participated in a forced-choice and ac-
ceptability 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 discourses; in one of the
two discourses the context sentence matched in intended focus association;
in the other discourse the context sentence did not match (cf. (24)). 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). In a forced-choice judgement, listeners were
asked to choose one of the two discourses as more felicitous. Then in a
rating task, listeners would also choose one of the following three categories:
(i) only the selected discourse is acceptable; (ii) both are acceptable, or; (iii)
neither is acceptable.

(24)  Matching discourse
  FOF_v: Johnson only PEDDLES pedals lately.
  SOF_v: Even THOMPSON only pedals pedals lately.
  Non-matching discourse
  FOF_n: Johnson only peddles PEDALS lately.
  SOF_v: Even THOMPSON only pedals pedals lately.

16.4.2 Results
Table 5 gives the percentage of stimuli correctly identified (i.e. the appropri-
ately 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 per cent, and listeners judged the stimuli as both acceptable at 69.8
per cent, with three of the listeners judging all or nearly all discourses as
acceptable. Not surprisingly, duration had a significant effect on overall correct identification (p = .0054). Finally, listeners’ successes were normally distributed over the different stimuli: listeners did not perform consistently better for any particular stimuli.

16.4.3 Discussion

While the context retrieval task is somewhat more complex than the discrimination task in Beaver et al. (2007)/Jaeger (2005), it more closely approaches listeners’ intuitions about the semantic/pragmatic felicity of prominence, and the use of sophisticated listeners was intended to mitigate the complexity. Additionally, by including both FOF and SOF sentences, listeners have access to explicit paradigmatic comparison, in addition to syntagmatic comparison.

If the statistically significant paradigmatic differences observed in the production data are indeed perceptually significant, listeners ought to be able to exploit this for correct identification. This prediction was not borne out.

If the statistically significant syntagmatic trend for greater prominence on the verb in the production data were perceptually significant, we might expect listeners to hear SOF on the verb more often than the noun. Alternatively, we might imagine listeners are able to ‘filter out’ the non-semantic prominence in order to identify the intended focus associate. These predictions were also not borne out.

Overall, the sophisticated listeners were not particularly successful at identifying the matching discourse. I believe it is telling that two of the listeners (L2 and L6) were willing to accept all discourse pairs as prosodically felicitous, that even for the best performing listener (L3), an association with focus in the SOF sentence could be essentially ‘coerced’ by the FOF context sentence.

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.

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

Table 5. Rates of correct identification, and rates of acceptability of both discourses (n = 60)

<table>
<thead>
<tr>
<th></th>
<th>L1</th>
<th>L2</th>
<th>L3</th>
<th>L4</th>
<th>L5</th>
<th>L6</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>%correct</td>
<td>55</td>
<td>50</td>
<td>68</td>
<td>65</td>
<td>47</td>
<td>60</td>
<td>57.5</td>
</tr>
<tr>
<td>%both acceptable</td>
<td>61</td>
<td>100</td>
<td>93</td>
<td>28</td>
<td>37</td>
<td>100</td>
<td>69.8</td>
</tr>
</tbody>
</table>
16.5 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 a SOF associate is always syntactically F-marked. The results of the first production experiment suggest, however, that the focus associate is not always realized as most prosodically prominent. Listeners could in principle disambiguate a SOF sentence, as they do in Rooth’s bridging examples (cf. (3)). This possibility, criticized by Partee (1999) as ‘phonologically invisible focus’, certainly undermines the interface principle Stress-F, either on its own or in conspiracy with Destress-Given. Either a revision of the principle or an account of its neutralization under certain conditions would be required.

Second, consider an account on which SOF is never semantically marked. On this scenario, what remains to be explained is the documented phonetic differences that sometimes arise. For example, Krifka (2004) suggests that SOF amounts to an anaphor (cf. he, her, it) without internal compositional semantics. The weak but observable production effects are due to phonetic copying, motor planning, or some other ‘low-level’ phenomenon. This provides a way of maintaining a grammaticized theory of focus, and one which observes Stress-F.

Third, consider an account according to which SOF is only sometimes marked. SOF is not marked when contextually recoverable; it is marked otherwise. Since all experimental investigations elicited SOF productions by using contexts in which the phonetically identical associate is identified, we would again require an explanation for any production effects, such as phonetic copying. A way of testing this account against the others would be to investigate the production of SOF in bridging examples. If prominence is required to license bridging inferences (e.g. that Harold and Susan are among the younger candidates in (3)), that may suggest that SOF is indeed sometimes marked. Selective or contextually dependent marking of SOF would potentially be consistent with Stress-F. As a theory in which SOF is sometimes realized with relative prominence, it is also consistent with a multi-notional theory of focus (cf. Selkirk 2008).

Despite the impressive and promising results of Beaver et al. and others, it appears that the debate over SOF is still open. There is indeed an observable and repeatable effect of focus in the acoustic signal. From this study, however, it seems unlikely that English speakers and listeners consistently exploit these acoustic contrasts in the perception of second occurrence focus in all contexts.
References


References

References


References


References


References


—— (forthcoming). ‘Will they keep on starting their sentences the German way?’. Ms.


References

—— (2008). 'Been there, marked that – a tentative theory of second occurrence focus'. Ms. UCLA.
References

392 References

References

References


References

References


400 References

—— (2007b). ‘The bipartite structure of verbs cross-linguistically, or why Mary can’t exhibit John her paintings’. Write-up of a talk given at the 2007 ABRALIN Congres in Belo Horizonte, Brazil.
References


References


References


References


References

References


References


References


O’Leary, C. and Crain, S. (1994). ‘Negative polarity items (a positive result) positive polarity items (a negative result)’. Paper presented at the Boston University Conference on Language Development.


References


References


References

References


References


References


References


References


References


References


