Only always associates audibly. Even if only is repeated
The prosodic properties of second occurrence focus in English
T. Florian Jaeger, Stanford University

Abstract - The apparent absence of pitch-marking on second occurrence focus expressions has been taken as a strong argument, often referred to as the ‘argument from second occurrence focus’ (e.g. by Partee 1999) against semantic theories of focus in which focus is computed compositionally based on prosodic cues (e.g. Rooth 1985). Several studies (Bartels 1997; Beaver et al. 2003; Rooth 1996) have raised doubt on the empirical validity of this argument. Beaver et al. (2003) show that second occurrence focus is marked prosodically. However, the prosodic correlates of second occurrence focus discovered so far are very subtle (e.g. very small duration and intensity differences). This paper has two purposes. First, I present new results showing that second occurrence focus-marking is perceptible and that the phonetic marking of second occurrence focus is less subtle than assumed in Beaver et al. (2003) if one also considers the possibility of syntagmatic marking. This provides strong evidence against the argument from second occurrence focus. Second, I present several studies comparing ordinary (non-repeated/first occurrence) focus-marking to second occurrence focus-marking. The results put second occurrence focus-marking in the context of prosodic marking in general. Most importantly, it is shown that, unlike standard focus-marking, second occurrence focus-marking can but does not have to be realized by a pitch accent. Of the features investigated, energy-marking (or duration and intensity) seems to be the most reliable predictor of second occurrence focus. The possibility of systematic non-pitch-based marking (in this case of focus) is in conflict with almost all current models of focus-marking and sentence prosody, which almost exclusively rely on pitch-marking (e.g. Bolinger 1958; Kadmon 2001; Erteschik-Shir 1999; Jackendoff 1972; Pierrehumbert 1980; Selkirk 1995; Welby 2002).
1 Introduction *

The interpretation of focus sensitive operators, such as ‘only’ in (1), is dependent on the position of focus in their syntactic scope. De-contextualized and in the absence of intonational cues, the sentence in (1) has two possible interpretations. In one reading the only-operator is associated with ‘Mary’, informally represented in (2a), resulting in a reading where it is only Mary and not someone else who was hit by a frozen chicken. This reading is compatible with the continuation in (3a). In the second reading, informally represented in (2b), the only-operator associates with ‘in the back’. In that reading, it is the case that the chicken hit Mary only in the back and nowhere else. This interpretation is compatible with the continuation in (3b). The only-operator is therefore prima facie potentially ambiguous.

(1) The frozen chicken only hit Mary in the back.
(2) a. The frozen chicken only hit [Mary]F in the back.
   b. The frozen chicken only hit Mary [in the back]F.
(3) a. … luckily it didn’t hit Peter. He doesn’t like frozen chicken at all.
   b. … luckily the chicken didn’t hit her on the head.

Pre-theoretically, I will say that in both interpretations the only-operator associates with the focus (indicated by […]F in (2) above) and the two different interpretations of (1) result from two different focus assignments, the output of which is informally represented in (2a) and (2b). A focus assignment is a mapping from a chain of sounds/string of words to information structure (cf. ‘focus-structure’ in Erteschik-Shir 1999:209). Throughout this paper, I will refer to any expression that is identified as a focus by a focus assignment, as ‘in focus’, ‘focused’, or as being ‘the focus’. In the literature, this information structural (or semantic) notion of focus is often confused or at least not clearly distinguished from the phonological form correlating with it. To avoid confusion, I use the terms ‘focus-marking’ when referring to the phonological marking or phonetic correlates of focus.

It is important to note that for a focus sensitive operator like ‘only’, it is not just subtle meaning differences that depend on what the operator associates with. The two readings of (1), i.e. (2a) and (2b), differ truth-conditionally. While (2a) is true in a scenario where Mary is being hit all over by frozen chickens, (2b) would be false. Vice versa, (2b) would be true if it was raining frozen chickens and basically everyone gets hit by one but Mary got hit only in the back (regardless who else has been hit wherever). In the same scenario (2a) would be false. In sum, it should be clear that understanding focus sensitivity is an issue that can decide between

---

* I am grateful to Edward Flemming and David Beaver for countless discussions on various aspects of this paper and for originally making me aware of this topic. Many of the ideas presented here are heavily influenced by those discussions (though they, of course, cannot be held responsible for any of the ideas presented below). Special thanks are also due to the three remaining QP committee members Dan Jurafsky, Elizabeth Traugott, and Arnold Zwicky whose comments were – as always – right on the point and made me rethink some of the issues touched on in this paper as well as restructure the presentation of my thoughts. I also owe many thanks to Maria Wolters (for providing scripts that made part of this work a lot easier), Brady Clark, Elisabeth Nordcliffe, and Philip Hofmeister (all of whom listened to long and often unclear ramblings on second occurrence focus and statistics). None of the above mentioned researchers necessarily agrees with the views presented in this paper. Parts of this paper have been presented at the LSA 2004 meeting, and to Stanford audiences. I am thankful for the financial support given to me as part of the Edinburgh-Stanford LINK grant “Sounds of Discourse”.

- 2 -
life and death. Besides focus sensitive operators like ‘only’ and ‘even’, other focus sensitive elements include questions, counterfactual conditionals, adverbs of quantification, frequency adverbs, generics, emotive factives and attitude verbs, superlatives and modals (see Hajicová et al. 1998; Rooth 1996).

As already hinted at above, a sentence like (1) would usually not be ambiguous in spoken language. In English focused expressions are typically marked by some type of pitch accent (Ladd 1996:45-46; Bolinger 1972; Cohan 2000; Jackendoff 1972; Selkirk 1984; Welby 2003), most frequently by a high pitch accent.¹ That is, typically (2a) would be realized with some sort of high pitch accent on ‘Mary’, whereas (2b) would be realized with a high pitch accent on ‘back’.

In combination, the fact that focus assignment changes the truth conditions and therefore the meaning of a sentence and the fact that focused expressions seem to be prosodically marked by pitch, have led some researchers to adopt a theory of focus in which focus is marked prosodically and this focus-marking contributes compositionally to the overall semantics of sentences. Theories that analyze the impact of focus on the meaning of a sentence as part of semantics include, among others, Jacobs (1983), Krifka (1992), Rooth (1985), and von Stechow (1985/1989). These theories contrast with theories in which focus is a pragmatic phenomenon (e.g. Asher 1995; Dryer 1994; Roberts 1996; Schwarzschild 1997; Vallduví 1990; Williams 1997).² In pragmatic theories of focus, focus-assignment is inferred from the context rather than being (obligatorily) encoded in the prosody. Following Beaver et al. (2003, 2004, to appear), I will refer to the former type as semantic theories of focus and to the latter type as pragmatic theories of focus. Semantic and pragmatic theories of focus differ in whether they assume a special semantic module for the interpretation of focus assignments or whether the truth-conditional differences between e.g. (2a) and (2b) are derived by more general pragmatic principles.

One standard argument against semantic theories of focus is ‘the argument from second occurrence focus’ (cf. Beaver et al. 2003, 2004, to appear, who are not themselves proponents of it). It is based on the observation by Partee (1999) that in examples like (4) the second occurrence of the focused expression ‘vegetables’ does not seem to be focus-marked (i.e. prosodically prominent). Whereas, ‘vegetables’ would typically bear a pitch accent in A’s utterance, the main pitch in B’s response will typically be on ‘Paul’ and not on ‘vegetables’.

(4) A: Everyone already knew that Mary only eats [vegetables]F.
    B: If even [Paul]F knew that Mary only eats [vegetables]SOF, then he should have suggested a different restaurant.

While the presence of a focus on ‘Paul’ does not posit a problem to semantic theories of focus (it follows from the fact that the focus sensitive operator ‘even’ associates with ‘Paul’), the absence of pitch accent on ‘vegetables’ is prima facie problematic. Since – just as the first occurrence of ‘vegetables’ – the second occurrence of ‘vegetables’ is associated with the focus sensitive only-operator (hence the name ‘second occurrence focus’), semantic theories of

¹ A detailed discussion of the relation between focus-marking and pitch-marking is provided in section 5.2.
² Here I do not summarize the literature on focus sensitivity or go into detail about the just mentioned debate since this has been done elsewhere (e.g. Beaver et al. 2003, 2004, to appear; Beaver & Clark 2003).
focus would predict that the second occurrence of ‘vegetables’ is focus-marked. Partee’s argument can be summarized as follows (cf. Beaver et al. 2003, 2004):

<table>
<thead>
<tr>
<th>The Argument from second occurrence focus</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Semantic theories require focus marked elements in the scope of focus sensitive expressions.</td>
</tr>
<tr>
<td>(2) In the case of second occurrence focus there is no such element.</td>
</tr>
<tr>
<td>(3) Therefore focus sensitivity is optional and requires a pragmatic account.</td>
</tr>
</tbody>
</table>

Crucially, in the argument of second occurrence focus (henceforth SOF), absence of a pitch accent is taken to imply lack of focus-marking. While, surprisingly, even opponents of the argument from second occurrence focus accept this premise (e.g. Rooth 1985, 1996), the argument from second occurrence focus only holds if there aren’t any systematic perceivable differences between the prosodic realization of SOF and unfocused expressions.

In this paper, I provide evidence that SOF is marked prosodically different from unfocused expressions as well as, at least in some cases, from OFC (i.e. ordinary first occurrence focus; henceforth OFC in order to distinguish it from the super-category focus, which I take to include SOF and OFC). Taken together with earlier empirical investigations (Bartels 1997; Beaver et al. 2003, to appear; Rooth 1996), the current paper thus supports what I will refer to as the Second Occurrence Focus-marking Hypothesis (henceforth SOF-marking Hypothesis):

<table>
<thead>
<tr>
<th>The Second Occurrence Focus-marking Hypothesis (SOF-marking Hypothesis)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) The phonetic realization of second occurrence focus (SOF) differs from the phonetic realization of unfocused expressions (i.e. SOF is prosodically marked).</td>
</tr>
<tr>
<td>(2) SOF-marking is perceptible (and used in processing of focus assignments).</td>
</tr>
<tr>
<td>(3) SOF-marking shares features with ordinary focus-marking (OFC-marking; i.e. marking of non-repeated/first occurrence focus) but is not identical to it.</td>
</tr>
</tbody>
</table>

The results and discussion presented in the remainder of this paper address two audiences that traditionally do not overlap. The first part of the paper deals with SOF-marking Hypothesis 2 and therefore bears on the debate between semantic and pragmatic theories of focus. The second part, starting with the investigation of SOF-marking Hypothesis 3 has consequences for theories of focus-marking and of intonation.

Section 2 provides a short introduction to earlier experimental work on second occurrence focus. The perception experiment in section 3 shows that native speakers of English can distinguish SOF expressions from unfocused expressions in naturally elicited speech. These findings support SOF-hypothesis 2 and argue against the argument from SOF. Section 4 presents several statistical and in-depth studies of the phonetic realization of SOF as compared with OFC (e.g. ‘Paul’ in B’s utterance in (4) above). The results support a view in which SOF-marking optionally employs pitch-marking (and is in this regard like OFC-marking) but does not have to. This provides evidence for SOF-marking Hypothesis 3. Specifically, section 4 supports a view in which there are two focus-marking ‘allomorphs’,\(^3\) one

\(^3\) A discussion of the adequacy of the term ‘allomorph’ in this context is provided in section 4, footnote 14.
that includes pitch-marking and one that doesn’t. While the OFC-marking seems to always employ the former, SOF-marking can employ either. A general discussion of the results and their consequences for theories of theories of focus, theories of focus-marking, and theories of intonation is presented in section 5. In particular, the results are connected to well-established facts about non-pitch correlates of accent (e.g. Beckman 1986; Beckman & Edwards 1994; Fry 1955, 1958; Kohler 1991; Ladefoged 2003:Ch. 4; Lehiste 1970; Liberman & Pierrehumbert 1984:172; Shattuck-Hufnagel et al. 1994) and prominence-marking in the absence of pitch (e.g. Ladd 1996:58, 226f.; Hockey 1998; Sluijter & van Heuven 1996a,b, 1997). The results also bear on the highly relevant distinction between (pitch) accent and sentence stress (e.g. Ladd 1996:226f). Finally, the last part of section 5 draws all parts of the discussion together and provide the outline of an account of the distribution of the two proposed allomorphs of focus-marking.
2 Background

To the best of my knowledge, there have been three studies investigating the empirical validity of the argument from second occurrence focus (Bartels 1997; Beaver et al. 2003; Rooth 1996). The goal of those studies was to test whether it is true that SOF is not prosodically marked. In the following paragraphs, I discuss the main results of each of the studies. I conclude that, although previous studies have provided sufficient evidence to doubt the validity of the argument from second occurrence focus, none of the studies has addressed the pivotal question: Do hearers have access to prosodic information that distinguishes second occurrence expressions from unfocused expressions? (cf. SOF-marking Hypothesis 2).

Rooth (1996) was the first to challenge the validity of the argument from second occurrence focus. He recorded himself uttering several examples like (5) and (6) and argued that SOF seems to be marked.

(5) A: Do you want Sue to only [name] \(_f\) Manny today?
   B: No, I only want [Eva] \(_f\) to only [name] \(_{SOF}\) Manny today.

(6) A: Do you want Sue to only name [Manny] \(_f\) today?
   B: No, I only want [Eva] \(_f\) to only name [Manny] \(_{SOF}\) today.

Rooth (1996:12) showed that while the SOF in his examples was indeed not marked by pitch movement, the rhyme of ‘name’ in (5b) was longer and realized with more intensity than the rhyme of ‘Manny’. The opposite was the case for (6b). Rooth took this and his intuition that he perceived the SOF phrase to be more prominent than the unfocused phrase (e.g. ‘Manny’ in (5b)) as evidence that SOF is prosodically marked. However, Rooth’s study was rather impressionistic – he was the researcher, experimenter, and subject at once. Also, and more problematically, Rooth compared the expression in SOF in each example to the unfocused expression in the same sentence. This is methodologically questionable since it is well known that the concrete phonetic realization of features such as pitch depends on the phonological environment. Without appropriate control for such factors as the phonemic structure and word length, comparing the length of two vowels that occur in two different words is not an adequate comparison. To address the question whether SOF is phonetically marked, the crucial comparison is between the phonetic realization of a SOF expression and the phonetic realization of the same expression in the same context, but being unfocused (i.e. not in the scope of a focus sensitive operator or otherwise focused or contrastive). In other words, the relevant contrast is between the realization of ‘Manny’ in B’s utterance in (5) and (6) (ditto for ‘name’).

Beaver et al. (2003, to appear) did just this and in a methodologically more advanced way. In their production experiment 20 participants read two repetitions of 14 short texts like those in (7) and (8). Each context began with an introductory sentence setting the stage for the remainder of the context, (a). This introductory sentence also contextually determined which expression was intended to be associated with the only-operator in the second sentence of the

---

4 In this specific case, this methodological deficit is somewhat ameliorated by the fact that Rooth found that the length of the two rhymes was inversed under the two conditions. The methodological concerns discussed above would have been much more important in case the comparison had not yielded an inverse relationship between the relative length of the rhymes and the two conditions.
context, (b). Finally, the third sentence contained the same expression as the SOF that was associated with the only-operator in the second sentence. The third sentence also always contained an OFC expression (e.g. ‘the state prosecutor’ associated with ‘even’ in (7) and (8)) preceding the SOF.

(7)  a. Both Sid and his accomplices should have been named in this morning’s court session.
   b. But the defendant only named [Sid]F in court today.
   c. Even [the state prosecutor]F only named [Sid]SOF in court today.

(8)  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]F today.
   c. Even [the state prosecutor]F only named Sid [in court]SOF today.

Beaver et al. (2003, to appear) found that SOF expressions were realized phonetically different from unfocused expressions. Although duration, relative intensity, and relative energy were found to be the primary phonetic features of SOF, the data were compatible with optional pitch-marking on SOF.5 SOF expressions (e.g. ‘Sid’ in (7c)) were shown to be significantly longer, louder, and realized with more energy and higher pitch range than their unfocused counterparts (e.g. ‘Sid’ in (8c)).6

Interestingly, Beaver et al. (2003) observed that the two focus sensitive operators used by them in the study (‘only’ as above in (7) and (8), and ‘always’) seem to differ with regard to SOF-marking. While ‘only’ seems to trigger SOF-marking, expressions supposedly associated with an always-operator were not realized in a significantly different way from their unfocused counterparts. There are independent reasons to believe that not all operators traditionally considered focus sensitive operators are alike (Beaver & Clark 2003) and that ‘always’ may not be a focus sensitive operator in the sense relevant here. In what follows, I use terms like focus-marking (and its variants OFC- and SOF-marking) and association with focus only when referring to ‘real’ focus sensitive operators, such as, for time being, ‘only’ and ‘even’.

Bartels (1997) presents a pilot study that investigated the phonetic realization of SOF as compared to mere repetitions of a focused phrase. She recorded 6 native speakers of English speaking three short texts in a dialogue situation.7 In one type of discourse speaker B only repeated what speaker A said (to an imaginary audience who could not hear speaker A). In that scenario B would utter (10a) after A has uttered (9). In a second scenario speaker B actually reacted to speaker’s A utterances creating a sentence containing a SOF expression. This is exemplified by B uttering (10b) after A has uttered (9).

(9)  When I was in China, I lived only on [rice mush]F for a month.

---

5 Beaver et al. (2003) did not annotate their data for pitch accents and phrase boundaries (i.e. no ToBI-style annotation was available for their data). All analyses were performed on phonetic rather phonological properties. They found that some pitch measures (standardized range of f0 and standardized minimum f0) marginally correlated with second occurrence focus. I will come back to the role of pitch-marking for second occurrence focus in section 4 and 5.

6 Overall the intensity and pitch range effect was only marginally significant. There was also a marginally significant effect for minimum pitch, which, however, is shown to be due to an interaction effect between word position and focus assignment (SOF vs. unfocused).

7 This method is probably a more natural situation than the one used in Beaver et al.’s study because, even though participants were reading out texts they were interacting with each other.
(10) a. When you were in China, you lived only on [rice mush]$_{F}$ for a month.
   b. Gee, I am glad I wasn’t there. [I]$_{SOF}$ couldn’t live only on [rice mush]$_{SOF}$ for a month.

Unfortunately, Bartels’s design (as she admits herself) has several serious methodological drawbacks. Nevertheless, the results of her study are interesting for at least two reasons. First, even though tentative they support the observation by Beaver et al. (2003) that SOF is either not pitch-marked or at least not primarily marked by pitch. Second, Bartels shows that repeated focus phrases, (10a), and SOF expressions, (10b), share at least some phonetic properties. Both types of focus expressions don’t seem to be pitch marked, and are reduced in length and relative loudness compared to the first occurrence of the focus expression. The term ‘second occurrence focus’ in its most narrow use (the one employed here) refers to a special type of repeated focus expressions, namely those that follow an OFC expression (i.e. an ordinary non-repeated focus), which typically bear a strong pitch accent. Bartels’s observation is interesting since it suggests that, on the one hand, some differences between OFC-marking and SOF-marking are due to the fact SOF expressions are – by definition- repeated. On the other hand OFC (as above in (10a)) and SOF (as in (10b)) differ in that SOF – again, by definition- follows a nuclear accent (which is located on the OFC expression ‘I’). Thus some differences between SOF-marking and OFC-marking may be due to the post nuclear position of SOF expressions. I will come back to this possibility during the general discussion in section 5.

In sum, Beaver et al. (2003, to appear) and Bartels (1997) demonstrated that SOF is prosodically marked and that thus the argument from second occurrence focus is potentially void. However, they did not show whether hearers can perceive SOF marking. Only if SOF-marking is perceptible can it be used as a prosodic cue for processing focus assignment of sentence with SOF expressions. Given that the phonetic differences between SOF expressions and their unfocused counterparts were quite subtle (e.g. an average duration difference of 8ms, and a 2% difference in relative loudness for Beaver et al.’s experiment) it is by no means clear whether the phonetic correlates of SOF observed by Beaver et al. are strong enough to be a factor in language processing. In other words, while SOF-marking Hypothesis 1 is firmly grounded in empirical evidence, none of the studies on SOF has yet provided evidence for SOF-marking Hypothesis 2 (and, for that matter, neither for SOF-marking Hypothesis 3). SOF-marking Hypothesis 2 is addressed in the next section and SOF-marking Hypothesis 3 is addressed in section 4.

---

8 For example, note that (10b) does not present a classical case of second occurrence focus in which the second occurrence focus follows another expression that is in the scope of a focus-sensitive operator such as ‘even’ in example (7) and (8) of Beaver et al. (2003, to appear).
3 Second occurrence focus is perceptible

To address the pivotal assumption of the argument from second occurrence focus, that SOF is not marked by perceptible prosodic features, I conducted a forced-choice perception experiment. The results of the experiment support SOF-marking Hypothesis 2 (cf. section 1; repeated below) and thus argue against the argument of second occurrence focus.

<table>
<thead>
<tr>
<th>SOF-marking Hypothesis 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>(2) SOF-marking is perceptible (and used in processing of focus assignments).</td>
</tr>
</tbody>
</table>

Section 3.1 describes the methodology, materials, and the design of the experiment. Section 3.2 summarizes the predictions that follow from SOF-marking Hypothesis 2. Section 3.3 summarizes the results of the experiment. Finally, section 3.4 contains a discussion of the results and section 3.5 the summary of the study’s implications.

3.1 Description of experiment

Subjects: 14 native speaker subjects (10 female, 4 male; age range: 23-28) were recruited as follows. 10 linguistically naïve subjects (8 female, 2 male; age range: 19-26) in the study, all without prior experience in linguistic perception studies participated. 4 subjects (2 female, 2 male; age range: 24-28) were linguistically trained.

Stimuli: 40 minimal discourse pairs of stimuli were quasi-randomly selected from Beaver et al.’s (2003) experiment described in section 2 (the random selection was generated using the program Shuffle by Christophe Pallier 1997). For reasons discussed in section 2, only discourses with the focus sensitive operator ‘only’ were selected, yielding three different types of pairs (see appendix B). Discourses that contained stuttering, unusually long pauses, or other signs of ‘reading effects’ (recall that the production experiment elicited the stimuli in a reading task) were excluded from the experiment in a blind review process prior to the creation of the random selection. Of the selected discourses, only the final sentence was used. Both parts of a pair always were taken from the same discourse pair (recall that Beaver et al.’s 2003 study tested subjects on 14 different discourse pairs; cf. section 2) and had been uttered by the same speaker. In other words, the two parts of a pair only differed in their (second occurrence) focus assignment. This is illustrated by the examples given above in (7) and (8) and repeated below as (11) and (12). The sentences in (11c) and (12c) form a stimulus pair for the perception study. The preceding discourse from the Beaver et al.’s (2003) production study, given in (11a,b) and (12a,b) are not used in the perception experiment. Here SOF-marking is indicated by ‘[…]|SOF’ but subjects were presented acoustic stimuli and therefore had no access to any artificial annotation.

(11) 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]|SOF in court today.

(12) a. Both Sid and the defense attorneys 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]|SOF in court today.
(12) 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]_{f} today.
    c. Even [the state prosecutor]_{f} only named Sid [in court]_{SOF} today.

The list of stimuli pairs was balanced to meet the following conditions. Two pairs each from all 20 speakers of the production experiment were used. For each type of pair, there were an equal number of instances where the first NP was supposed to be SOF-marked in the first sentence of a pair and instances in which the second NP was supposed to be SOF-marked in the first sentence.\(^9\) Since, the two sentences in a pair only differed in their focus assignment, the second NP was SOF-marked if and only if the first NP was not SOF-marked.

Two versions of the experiment with inverse order were produced. 7 subjects took version 1 and 7 subjects took version 2 of the experiment. The following order constraints were observed when the list of stimuli was constructed. The same kind of internal order of a pair (the first sentence has a SOF-marked first NP or a SOF-marked second NP) was not allowed to occur more than three times in a row. The same kind of pair was not allowed to occur twice in a row.

**Presentation and procedure:** The acoustic stimuli were incorporated into an HTML page. Subjects first read through the instructions and then were allowed to ask the experimenter questions before they started the experiment. Subjects were asked to judge “[…] in which of the two sentences (A or B) the speaker wished to make the second word (which will be given in bold face [see Figure 1]) more prominent than the first.” Subjects were not allowed to make any changes once they reached a decision for a pair.

Each pair of acoustic stimuli was accompanied by two words that were displayed in the same row. This is exemplified below for two pairs (pair 4 and 5 out of 40):

![Figure 1](image)

**FIGURE 1** – Excerpt of the HTML page displayed to subjects during the experiment

10 pairs of stimuli were displayed at a time. After every 10\(^{th}\) pair of sentences subjects were asked to take a break of 30secs to 2 minutes, which they could time themselves, before they continued to the next page with 10 pairs of stimuli. Subjects were asked to proceed from pair to pair and always listen to the first stimulus in a pair first. After that they were allowed to listen to each of the stimuli as often as the wanted to. All subjects required between 15-25 minutes for the experiment and approximately 2 minutes to read the instructions.

To test whether subjects had actually understood the instructions, all subjects were asked to repeat in their own words what they had been asked to do after the experiment.

---

\(^9\) Beaver et al. (2003) also let subjects read every stimulus twice (see section 2). The stimuli used in this perception experiment were balanced for repetition. Both parts of a pair were from the same repetition.
3.2 Predictions

It follows from SOF-marking Hypothesis 2 that subjects should perform above chance in selecting from each pair the stimulus that was uttered with the intention of marking the second NP as SOF. This prediction is also based on the hypotheses that in most instances, the participants in Beaver et al.’s production experiment understood what they were asked to do and therefore produced natural stimuli10 and that the preceding discourses in the production study (i.e. (a) and (b) in e.g. (11) and (12) above) were unambiguous in which focus assignment they elicited. However, even if these two assumptions are not always met (for example, because participants in Beaver et al.’s production study did not always produce natural stimuli due to confounding factors like the artificial situation inevitably created in phonetics experiment), participants in perception study should still overall perform above chance.

3.3 Results

All subjects performed above chance (mean performance=63% correct answers; range= 52.5-77.5%). The one-sample t-test against an expected mean of 20 correct answers out of 40 reveals that subjects on average perform significantly above chance (t= 7.7; df= 13; p< 0.001) and that subjects performed ‘alike’ (i.e. there was no subject effect; $\chi^2 = 9.6$, df= 13; p> 0.7). The average effect size is small to medium ($\omega = 0.28$).

None of the between-subject factors (i.e. subject id, gender, version of the experiment, linguistic education) had a significant effect. Crucially for pooling the data from all subjects, the level of linguistics education did not matter either. Even though linguists performed slightly better than linguistically-naïve subjects (67% correct answers vs. 62%) this difference was not significant ($\chi^2 = 1.2$; df= 1; p> 0.25). Furthermore, all effects that hold for all subjects together also hold for linguistically-naïve subjects alone. There also was an item effect ($\chi^2= 163.9$, df= 39; p< 0.001). That is, subjects did not perform equally well on all sentence pairs. Interestingly, there were items that nobody answered as expected and items that everyone answered as expected by SOF-marking Hypothesis 2. A post-hoc test for the type of item (recall that there were three types of discourse pairs differing in their lexicalization) revealed no significant effect (although one type of discourse pair on average performed better than the other two types, 60 vs. 78%, this difference was not significant).

Finally, note that all subjects who were tested showed clear understanding of the task in the post-experimental interview although the precise strategy that subjects chose to solve the task seemed to differ from subject to subject. For a more advanced (and therefore theoretically more adequate) statistical modeling of the results in terms of logistic regression yielding essentially the same results as the analysis presented here, see Appendix C.

3.4 Discussion

The results show that the prediction made in section 3.2 is met. This provides strong supporting evidence for the SOF-marking Hypothesis. Even though the phonetic correlates of SOF were shown to be very small in Beaver et al. (2003, to appear; cf. section 2), subjects are

---

10 This assumption turns out to be not always met. I come back to this issue in section 3.4.
able to identify SOF-marked elements correctly in 63% of all cases in the experiment. This shows that subjects are able to identify SOF-marking without contextual clues (recall that the sentence pairs were spliced from their preceding discourse), but it also raises the question of why people did not perform even better (or in other words, why the effect size is relatively small). For one thing, the overall performance is decreased because of 4 items that were almost always judged contrary to the predictions of the SOF-marking Hypothesis (judged correctly by only 0-25% of all subjects), and 4 more items that performed below chance (correctly judged by 25-45% of all subjects). Note that this is neither what is expected according to the SOF-marking Hypothesis, nor what would be expected in the absence of prosodic cues to SOF (since subjects should perform at chance in the absence of phonetic cues, these answers should have been judged correctly by 50% of all subjects). It seems rather likely that, compared to naturally occurring speech the artificial experimental scenario of Beaver et al.’s experiment resulted in an above-average amount of intonationally ill-formed stimuli. An examination of the intonation well-formedness of a subset of Beaver et al.’s data confirmed this hypothesis. Prior to the perception experiment, I had already examined the intonation on the second sentences in the read discourses (see (11) and (12) above), which contained one OFC expression and no SOF expression. For a sentence of that type it is relatively easy to hear whether they are intonationally well-formed or not. While the majority of the stimuli were well-formed, the rate of intonationally ill-formed sentences was high enough to explain why some stimuli in the perception study were consistently judged counter to expectations. In several stimuli, the second sentence (e.g. (b) in (11) and (12) above) contained a strong pitch accent on the unfocused expression but not on the focused expression indicating that the subject had problems reading the stimuli (either because they were unnatural or because the subject was tired or not paying enough attention at that moment). The wrong placement of the pitch accent sometimes seemed to carry over to the third sentence containing the SOF expression. The decision to not a priori exclude those supposedly ill-formed stimuli was made because this may have seriously confounded the study. Although it remains a question for further research whether intonational ill-formedness was the reason for the five items’ performance, it seems plausible that the overall performance for exclusively well-formed stimuli would be higher than 63%. If one excludes the four worst items overall performance increases by 6% to 69%. If one also excludes the 4 items that performed below chance, the overall performance increases to 73%.

Whether or not one wants to exclude the five below-chance items, this still leaves open the question of what it means that SOF-marking was perceptible in 73%, 69%, or 63% of all cases respectively. Is SOF-marking optional? This would also account for the item effect

11 Although neither answer ‘A’ nor answer ‘B’ was chosen significantly more often overall, most of the items that everyone judged contrary to expectations were where the ‘correct’ answer was ‘A’. This could be due to some design confound that resulted in answer ‘B’ being preferred in the absence of phonetic cues. It is, however, hard to imagine that a bias of that type could be so strong that all subjects unexpectedly chose ‘B’ over ‘A’.

12 These kinds of problems are well-known from other reading studies and have been the target for critique by researchers who have pointed out that reading experiments of the type described here are suboptimal for eliciting naturally occurring speech. I agree with that assessment. Optimally, a comprehension task should have been included in Beaver et al.’s study to test whether for a given stimulus the subject had understood what he/she was reading. The perception study can be interpreted to also provide an additional filter to test for unnatural stimuli. The five experimental pairs judged counter to expectations by everyone probably contained at least one intonationally ill-formed sentence (see also section 4.4 for discussion of this point).
mentioned above. Alternatively, it could be the case that pretty much all items differ in how often subjects judged them correctly and that the overall mean performance of 63% is only due to a relatively small overall bias. That the latter is not the case can be seen in Figure 2. Only 8 out of 40 items (20%) performed clearly below chance, whereas 27 items (67.5%) were correctly judged clearly above chance.

![Figure 2](image.jpg)

**Figure 2** – Number of items for a given correct answer ratio (percentage of correct answers)

If we leave apparent outliers (items that performed below chance) aside for the moment, this still leaves 5 items (12.5%) that were judged at chance. This pattern is compatible with optional marking of SOF. Alternatively, the fact that a lot of items (13 items or 32.5%) were correctly judged by 55-75% of all subjects could be interpreted to mean that SOF-marking is gradient rather than optional. That is, speakers can decide to which degree they want to mark SOF in a given context. At this point, it is impossible to determine which of these hypotheses is the correct one.

Finally, note that one may object that the task subjects were asked to performed referred to intended prominence rather than directly to SOF-marking, and that it is therefore impossible to know from this experiment (a) whether an average listener is able to identify SOF-marking in a real life situation, and (b) if listeners are in principle able to do so, whether they actually use phonetic marking to infer SOF.

While (a) is a valid objection, it is important to keep in mind that language processing is a highly automated and (at least at the level relevant to the current discussion) subconscious task. Linguistics is primarily interested in the subconscious knowledge (cf. Jackendoff’s 2002:29 “f-knowledge”) people have about language. Eliciting judgments from subjects, on the other hand, is asking for a conscious act of accessing that knowledge, which is a drastically different task. It seems therefore fair to assume that the artificially experimental setting inevitably introduces confounds. Simplifying the task by explicitly pointing subjects to the difference in prominence in the stimuli was intended to make up for the weakening of the effect that may have been introduced due to the problem just mentioned.

In any case, the main purpose of the perception experiment was to establish whether SOF-marking is perceptible. In that sense, subjects were used as (very intelligent) minimal difference detectors. That they performed significantly above chance therefore means – if nothing else – that something in the phonetic realization of SOF is different from the phonetic realization of unfocused repeated expressions.
As for (b), the point of this paper isn't to show that pragmatic and discourse clues are not employed in determining focus assignments. Rather, the goal of the experiment was to show that, contrary to claims in the literature, phonetic clues are available to infer the focus assignments of a sentence even in the case of SOF. This goal has been achieved. Also, even though natural discourses are usually rich enough to provide unambiguous clues as to whether a constituent is in SOF, why shouldn't listeners simultaneously use all information available to them? As more and more research on language processing shows, the default in language processing seems to be massive simultaneous constraint satisfaction, where the constraints come from different levels of linguistics competence as well as from non-linguistic processes (such as eye-movement; e.g. Tanenhaus et al. 1995).

### 3.5 Conclusions

The perception experiment presented above provides strong support for a crucial part of SOF-marking Hypothesis 2, in so far as it shows that hearers in principle have access to prosodic cues when identifying second occurrence focus because SOF-marking is perceptible. Crucially, this makes the argument from second occurrence focus (e.g. Partee 1999) invalid as it stands.
4 Second occurrence focus-marking differs from non-repeated focus-marking

In this section, I take the results from the previous section and Beaver et al. (2003, 2004, to appear) on SOF-marking and compare them to the phonetic properties of OFC-marking (i.e. the marking of ordinary, non-repeated focus). The studies presented in this section provide strong support for SOF-marking Hypothesis 3.

<table>
<thead>
<tr>
<th>SOF-marking Hypothesis 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>(3) SOF-marking shares features with OFC-marking but is not identical to it.</td>
</tr>
</tbody>
</table>

While this section starts out by comparing OFC- and SOF-marking in a variety of different ways, ultimately, the following spell-outs of SOF-marking Hypothesis 3 are shown to be true for the data investigated (cf. section 4.5):

<table>
<thead>
<tr>
<th>Three specific spell-outs of SOF-marking Hypothesis 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>(3A) Duration, intensity, and energy are significant predictors (crucial measures) of SOF-perception.</td>
</tr>
<tr>
<td>(3B) Maximum, mean, and minimum $f_0$ are significant predictors (crucial measures) of SOF-perception.</td>
</tr>
<tr>
<td>(3C) Duration, intensity, and energy are stronger predictors of SOF-marking than any of the pitch measures. Particularly duration, intensity, and energy marking on SOF expression is not mediated by pitch-marking. 13</td>
</tr>
</tbody>
</table>

SOF-marking Hypothesis 3C raises the question as to whether pitch-marking is never employed in SOF-marking. I show that this is not the case. The primary difference between SOF- and OFC-marking is the absence of obligatory pitch-marking on SOF expressions. Based on the assumption that the two variants of SOF-marking do not differ in their meaning, this essentially means that there are two SOF-marking ‘allomorphs’. 14 The data examined in this paper suggest that one of the SOF-marking allomorphs is (one of) the OFC-marking 13 While we can test whether a given prosodic property correlates with SOF-marking and whether this correlation is mediated by any of the other measures taken, it is impossible to show that a given correlation is not an indirect effect (i.e. mediated by) another phonetic/phonological property of SOF-marking that was not measured. More concretely, this means that it is impossible to show that duration, intensity, and energy effects are not mediated by another unknown factor not investigated here (e.g. vowel quality and spectral balance/tint; cf. Huss 1978; Sluijter & van Heuven 1996a,b, 1997). It is, however, possible to address the question whether the correlation between PA and a given measure is mediated by another measure (in this case pitch) and whether one measure contributes more information than another measure. 14 Strictly speaking, the term ‘allomorph’ is only adequate if focus (or at least the type of focus considered here) is a morpheme. While I see no immediate problem with that assumption, nothing in the remainder of the paper depends on this assumption. I leave it open for future research whether there is a focus morpheme or even different focus morphemes (corresponding to different types of focus, such as contrastive focus, information focus, polarity focus, etc.). In any case, by using the term ‘allomorph’ here and below, I highlight the fact that the two variants of SOF-marking are identical in terms of their focus meaning. If the two variants aren’t real allomorphs, they are at the very least contextual variations in the realization of focus. I am in debt to Arnold Zwicky and Edward Flemming for making me aware of several of the above-mentioned issues.
allomorph(s). This insight together with SOF-marking Hypothesis 3A to 3C eventually leads to the following tentative refinements of SOF-marking Hypothesis 3 (cf. section 4.6 and 5.2):

<table>
<thead>
<tr>
<th><strong>SOF-marking Hypothesis 3’</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>(3’) SOF-marking employs (at least) two prosodic allomorphs. One of these allomorphs is also employed in OFC-marking.</td>
</tr>
<tr>
<td>(3’-1) The allomorph shared between OFC- and SOF-marking employs pitch-marking.</td>
</tr>
<tr>
<td>(3’-2) The second SOF-marking allomorph does not employ pitch-marking.</td>
</tr>
</tbody>
</table>

The results presented here therefore also bear on current theories of English intonation and of theories of focus-marking, since almost all of them primarily or even exclusively focus on pitch as the prosodic means of marking sentence level prominence. SOF-marking, I argue below, presents a case of sentence level prominence that is not due to pitch-marking.

It is because of the challenge that this claim presents to current theories of focus-marking and intonation that I go into a lot of detail about different types of pitch-marking below. Especially, the thorough investigation of different possible ways in which pitch could play a role in SOF-marking, although at times lengthy, is necessary. Some researchers (e.g. Katy Carlson, p.c.) have replied with skepticism to Beaver et al.’s (2003) claim that pitch-marking is not involved in SOF-marking at all, and, as will be shown below, their concern is well-motivated. The motivation for the detailed discussion in the remainder of this section thus lies in the possibility that pitch-marking is always, or if not that, at least, optionally part of SOF-marking. As already mentioned above, I will argue in the remainder of this section that both Beaver et al. (2003) and Katy Carlson (p.c.) were right. On the one hand SOF-marking can involve pitch-marking, on the other hand it doesn’t have to. I will also provide a tentative overview of what type of pitch-marking correlates with SOF.

Section 4.1 provides some background and the motivation for the four studies presented in the remainder of this section. Section 4.2 and 4.3 present two quantitative studies of the difference between focus- and SOF-marking. These two studies are based on a large data set and rely on automatic extraction of the relevant measures. Since especially pitch-tracking and automatic pitch extraction are known to be error-prone, section 4.4 presents two small qualitative studies that test the observations made in section 4.2 and 4.3 and elaborate on the status of pitch-marking as a factor in SOF-marking. Section 4.5 goes into some detail with regard to the relative weight of different prosodic correlates of SOF-marking. Finally, section 4.6 wraps up the results of the 5 studies presented in section 4 and concludes with a tentative refinement of SOF-marking Hypothesis 3 (i.e. SOF-marking Hypothesis 3’ given above).

### 4.1 Warm-up

In order to compare the phonetic properties of OFC and SOF, a similar data set for both types of foci is needed. Optimally, to prevent confounds due to word related factors the material should contain the two types of foci on the same words, and, to avoid confounds due to

---

15 The factors I refer to here include but are not limited to: frequency of occurrence, predictability in the discourse, phonetic length (all of which influence duration measurements), or the phonological structure of the word (which can influence the f0 contour over a word).
to phrasing, in the same syntactic position with approximately the same distance (roughly in words) from the beginning of the sentence boundary, or even better, the beginning of the utterance. Beaver et al.’s (2003) data set consists of discourses which contain three sentences each. In all cases, the second sentence contained a OFC on exactly the same word where the third sentence contained a SOF expression. Examples are given below (for another example see (7) and (8) in section 2). Whichever expression is in focus in sentence (b) is in SOF in sentence (c). Sentence (b) only differs from sentence (c) in the type of focus assigned to that expression, the lexicalization of the subject NP (‘nurse’ vs. ‘doctor’), whether the subject is in focus, and with regard to the position in the overall discourse (second vs. third sentence). In both (b) and (c), the target expression bears the same grammatical function, argument role, and is equally far away from the beginning of the sentence. Since in both (b) and (c) the same word is the target, there is also no difference in the phonological form (e.g. word length, phonological structure) or the lexical semantics (animacy, definiteness). In (13) the target word is given (it has been explicitly mentioned in the prior discourse) in both (b) and (c). In (14), the target word is given in (c), but not in (b).

(13) a. Both Pete and Edward are suffering from the flu.
    b. But the nurse only gave [Pete]\textsubscript{f} a pill today.
    c. Even [the doctor]\textsubscript{f} only gave [Pete]\textsubscript{SOF} a pill today.

(14) a. Pete really needed an injection to ease the pain.
    b. But the nurse only gave Pete [a pill]\textsubscript{f} today.
    c. Even [the doctor]\textsubscript{f} only gave Pete [a pill]\textsubscript{SOF} today.

Since, for the purpose at hand, the data set just described is the best one currently available, I compared OFC-marking and SOF-marking between the (b) and (c) versions of Beaver et al.’s (2003) experimental stimuli. As mentioned above, section 4.2 compares the effect of focus- and SOF-marking on paradigmatic differences in the phonetic realization of target words. By paradigmatic difference, I mean the difference on the same word in the same syntactic position of the phonetic realization depending on whether the word is in focus or unfocused (and mutatis mutandis in SOF or unfocused). Thus a paradigmatic comparison of SOF-marked vs. unfocused expressions compares, for example, ‘Pete’ in (13c) with ‘Pete’ in (14c). A paradigmatic comparison of focus-marked vs. unfocused expressions would include, for example, ‘Pete’ in (13b) with ‘Pete’ in (14b) or ‘a pill’ in (14b) with ‘a pill’ in (13b). One could also directly compare focus with second occurrence focus expressions, e.g. ‘Pete’ in (13b) with ‘Pete’ in (13c), but since all focus expressions are in second sentences and all SOF expressions are in third sentence, declination (and similar effects for duration and intensity) would confound that comparison. The indirect comparisons are therefore preferable.

A second study, presented in section 4.3, explores syntagmatic differences between focus- and SOF-marking. That is, the second study compares differences between the phonetic realizations of two adjacent words depending on the focus assignment, i.e. whether the first

---

16 Some stimuli contained additional minimal differences between the second and third sentences, such as the absence or presence of ‘would’ in “But mum only let the cat in the tent today” vs. “Even the kids would only let the cat in the tent today”. Since ‘would’ in this context cannot bear prominence, this should not matter.

17 I am grateful to David Beaver, Brady Clark, and Maria Wolters for giving me access to the data.
word is in focus and the second word is unfocused or vice versa (and mutatis mutandis for SOF and unfocused expressions). To illustrate this, consider the case of focus-marking. The relevant syntagmatic comparisons are between ‘Pete’ and ‘a pill’ in (13b), where the focus ‘Pete’ precedes the unfocused expression ‘a pill’, and ‘Pete’ and ‘pill’ in (14b), where the unfocused expression preceded the focus. Each syntagmatic difference by itself does not mean much. Imagine a syntagmatic difference in length. Since two absolutely different words are compared the resulting number (in this case the difference in length between the first and the second word) doesn’t convey much information. However, if one compares a minimal set of syntagmatic comparisons, such as ‘Pete’ vs. ‘a pill’ in (13b) compared to ‘Pete’ vs. ‘pill’ in (14b), the difference becomes meaningful. Imagine that in the same string of two words (here ‘Pete a pill’), the first word is longer than the second if the first word is focused but less longer, equally long, or even shorter if the second word is focused. This change in difference, which can be construed as lengthening or shortening compared to the expected length of the word given the prosodic context and the current speech rate, conveys information.

To further illustrate what I mean by paradigmatic and syntagmatic differences, consider that there seem to be three sensitive strategies that a listener may employ in order to detect phonological marking. A listener could have expectations about the local phonetic realization of an expression (paradigmatic) or rather about the form of the transition from one expression to the next (syntagmatic) or about both. In the paradigmatic comparison, whether an expression bears a certain function is then determined by comparing the observed phonetic realization with different marking prototypes corresponding to marking of different functions. In that model a hearer compares the observed marking to possible variants of prosodic marking on the same (type) of word in the same position. For a syntagmatic comparison on the other hand, the relevant unit of observation is the form of the transition from one expression to another (e.g. does the $f_0$ drop or rise? Does the speaker slow down?).

In the remainder of this section, several prosodic measures and their paradigmatic and syntagmatic differences on focus- or SOF-marked words will be discussed. Although other measures (or variants/standardizations) were tested for some of the differences discussed below, I only discuss the following measures in detail (since the other measures do not yield different results): duration, r.m.s. intensity, relativized energy, and four $f_0$ measures. Energy was relativized by dividing it by the overall energy on the discourse. The four $f_0$ measures are maximum, mean, and minimum $f_0$, and $f_0$-range. Since the perception of pitch changes below 1000Hz, which is the frequency range of interest here, is essentially linear, $f_0$ is an adequate approximation of perceived pitch. Nevertheless, to preserve the ontological distinction between $f_0$ and pitch, I will talk about $f_0$ measures to refer to the concrete phonetic properties but pitch-marking to refer to the strategy that results in those phonetic properties.

In addition to the absolute values, the $f_0$ measures and the intensity measure were also taken in a standardized form. $f_0$ measures were standardized by subtracting the overall mean $f_0$ of the whole utterance (all three sentences of a discourse, as e.g. (11) and (12) in section 3.1) and dividing the outcome by the standard deviation of the $f_0$ values of the whole utterance (range was only divided by the standard deviation). Intensity was standardized the same way.

---

18 It may be slightly confusing that I refer to this kind of comparison as syntagmatic since it also involves a paradigmatic comparison. In that sense, what I refer to as ‘paradigmatic comparison’ is a comparison that only involves paradigmatic differences, whereas what I refer to as ‘syntagmatic comparison’ involves both syntagmatic and paradigmatic differences.
The $f_0$ and intensity values were therefore essentially z-scores. If not mentioned otherwise, all measures reported below were taken as described in this paragraph. All statistical analyses reported below are performed on the standardized or relativized measures, even if the mean values and standard deviations of the absolute measures are given (with the exception of duration which was not standardized). The standardized measures are used for the statistical analyses since it is unclear what it would mean to be marked by an absolute loudness or $f_0$ increase (in most cases the results didn’t differ between standardized and non-standardized measures). The absolute values are given for the means since it is easier to interpret mean differences if given in the usual units (e.g. Hz for $f_0$, dB for loudness, etc.). All measures were also transformed to meet the normality assumptions of the employed statistical analyses.

4.2 Quantitative paradigmatic comparison

This section compares the paradigmatic effects of (non repeated) focus- and SOF-marking. The overall logic of comparing paradigmatic differences for focus-marked and SOF-marked expressions has been described in the previous section and is illustrated in Figure 3.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure3.png}
\caption{Comparison of paradigmatic differences for focus- vs. SOF-marked expressions}
\end{figure}

The OFC data was gathered from a subset of Beaver et al.’s (2003) data that was labelled for OFC as well as SOF. The SOF data is based on the complete data set. Table 1 summarized the effect of OFC-marking and SOF-marking as compared to repeated non-focused information in the same position and the same sentence. The first column on each side lists the mean differences for all phonetic measures taken on the target words. Energy differences are only given in percentage increases. The second column on each side gives the F-values and the level of significance. The left side of the table represents the comparison in the left box of Figure 3 and the right side of the table represents the comparison in the right box of Figure 3. The right side of Table 1 thus replicates the production data analysis of Beaver et al. (2004). Significant results for a measure mean that the paradigmatic differences for this measure are significantly different from 0 (if one always subtracts the focus from the unfocused variant of the same word). The F-tests are taken from a two 2-way analysis of variance with subjects and discourse pairs as random factors and information status (i.e. ‘focus vs. not focus’, and ‘SOF vs. not focus’ respectively) and word position as fixed factors. The analyses are based on the subjects and N given at the top of each side of the table. Significant differences are highlighted by shading.
### TABLE 1 – Paradigmatic marking on non-repeated vs. second occurrence focus

Since the measures given in Table 1 stem from slightly different analyses and the tests are based on different Ns, the results cannot be used to determine the precise nature of the relation between non-repeated and second occurrence focus. Nevertheless, Table 1 provides an approximation of the differences between the two kinds of marking. Non-repeat focus-marking and SOF-marking differ in at least two ways. First, while OFC-marking clearly employs pitch, SOF-marking does not do so in the same way. Second, also the non-pitch measures, i.e. duration, intensity, and energy, seem to be noticeably reduced on SOF expressions. The data in Table 1 also strongly suggest that is not paradigmatic differences in r.m.s. intensity that mark SOF. Even though there is a significant effect on intensity both for focus- and for SOF-marking, the lowest just noticeable difference for loudness under optimal conditions seems to be in the range of 0.3 to 1 dB, and for more complex sounds it ranges between 1 to 4 dB (Stevens 1998:section 4.2.2). The duration difference of 6ms for SOF-marked phrases on the other hand is in principle perceptible. To the best of my knowledge, not much is known about just noticeable energy differences, but it is known that energy serves as a good indicator for loudness of very short sound segments (Stevens 1998:section 4.2.2).

### 4.3 Quantitative syntagmatic comparison

In addition to the paradigmatic effects of focus- and SOF-marking just discussed, the two types of marking also have different effects on the syntagmatic differences between two words. This second type of effect is investigated next.

In order to determine whether the syntagmatic effects of focus and SOF-marking differ, I conducted a 2-way analysis of variance with subjects and discourse pairs as random factors and ‘type of focus’ (2 levels: focus sentence vs. SOF sentence; see above) and ‘focus assignment’ (2 levels: unfocused constituent precedes the constituent associated with the focus...
sensitive operator or vice versa) as fixed factors.\(^{19}\) For the current purpose, the interesting cases are those where a phonetic measure was significantly affected by focus-marking but not by SOF-marking (or, for that matter, vice versa). In other words, the interesting cases are those with an interaction effect of ‘focus assignment’ and ‘type of focus’ and optimally a main effect for ‘focus assignment’ (though insignificance of the main effect could be due to a significant interaction). In Figure 4, which schematically represents the idea of syntagmatic comparisons outlined in section 4.1, the 2-way analysis of variance corresponds to the bold arrow between the left and the right box.

\[\begin{array}{c}
\text{Pete’ in (13b)} & \text{‘a pill’ in (13b)} \\
\text{Paradigmatic difference between} & \text{syntagmatic differences} \\
\text{‘Pete’ in (14b)} & \text{‘a pill’ in (14b)} \\
\text{Average syntagmatic difference focus-marked vs. unfocused} \\
\text{‘Pete’ in (13c)} & \text{‘a pill’ in (13c)} \\
\text{Paradigmatic difference between} & \text{syntagmatic differences} \\
\text{‘Pete’ in (14c)} & \text{‘a pill’ in (14c)} \\
\text{Average syntagmatic difference SOF-marked vs. unfocused}
\end{array}\]

**FIGURE 4 -** Comparison of syntagmatic differences for focus- vs. SOF-marked expressions

The analysis (referred to as the ‘main analysis’ below) revealed a significant main effect of focus assignment on standardized maximum \(f_0\) \((F(1,7)= 5.2; p= 0.056)\), mean \(f_0\) \((F(1,9)= 7.0; p< 0.05)\), and duration \((F(1,10)= 4.7; p= 0.055)\), as well as a highly significant main effect on standardized intensity \((F(1,7)= 11.5; p< 0.01)\), and relative energy \((F(1,7)= 14.9; p< 0.01)\). Focus assignment and type of focus had significant interaction effects on standardized maximum \(f_0\) \((F(1,7)= 7.8; p< 0.05)\), mean \(f_0\) \((F(1,5)= 14.9; p< 0.05)\), and minimum \(f_0\) \((F(1,5)= 6.8; p< 0.05)\), as well as a highly significant effect on standardized intensity \((F(1,4)= 6.5; p< 0.1)\).

In sum, the analysis of variance shows that (a) standardized maximum and mean \(f_0\), duration, standardized intensity and relative energy are syntagmatic correlates of at least one type of focus-marking, and (b) syntagmatic focus- and SOF-marking differ with regard to the effect they have on standardized maximum, mean, and minimum \(f_0\), and standardized intensity. This raises the question of how the two types of focus-marking differ. The presence of a main effect and an interaction effect for a given measure can be due to two quite different situations. For a given measure which is significantly affected by both focus assignment and its interaction with the type of focus, this could either mean that the measure is involved in only one of the two types of focus-marking, or in both but in opposite ways. To find out which of

\(^{19}\) To exclude pitch tracking errors from the analysis a relative conservative outlier analysis was performed. All measures of a sentence were excluded from the analysis if any standardized pitch measure on any word of a sentence deviated more than 2 standard deviations from the overall mean of that standardized pitch measure. This way 19% of the data were excluded. Given this high rate of data loss, it is important to note that the lack of significance of syntagmatic second occurrence focus-marking on \(f_0\) measures shown here and below is not due to the radical exclusion of possible pitch tracking errors (separate analyses with less radical or no exclusion criteria confirmed this).
these situations is the one that holds for the above-mentioned measures, I conducted two analyses of variance, one on only the focus sentences and one only on the SOF sentences respectively (this is similar to investigating simple main effects). As in the analysis just presented, subjects and discourse pairs were included as random factors, but this time ‘focus assignment’ was the only fixed factor (since each type of sentence was analyzed separately). The results are discussed in the following paragraphs. In the two tables that will be presented the upper half corresponds to the comparisons in the left box of Figure 4 and the lower half of the tables corresponds to the comparisons in the right box.

Table 2 summarized the mean differences for four $f_0$ measures (maximum, mean, and minimum $f_0$, and $f_0$-range) under four different conditions. The first two rows refer to focus sentences (e.g. the (b) sentences in (13) and (14) above). There are two versions of focus sentences, which only differ in their focus assignment. In the first version the focused NP (i.e. the NP associated with the focus sensitive operator) precedes the unfocused NP (referred to by ‘un’ in the table). This version is represented in the first row of Table 2. In the second version the focused NP follows the unfocused NP. This version is represented in the second row of Table 2. The fourth and fifth rows of the table refer to SOF sentences (e.g. the (c) sentences in (13) and (14) above). The third and sixth rows give the F-values for significant effects. Significance for a measure means that the difference of that measure on the two words in the sentences differed significantly depending on the focus assignment. A ‘*’ indicates that the differences between the two focus assignments is significant at $p<0.05$, ‘**’ stands for highly significant at $p<0.01$, and ‘+’ for marginally significant at $p<0.1$. Significant differences are highlighted by shading.

<table>
<thead>
<tr>
<th>Focus/SOF assignment</th>
<th>N</th>
<th>Maximum $f_0$ difference in Hz</th>
<th>Mean $f_0$ difference in Hz</th>
<th>Minimum $f_0$ difference in Hz</th>
<th>$f_0$-range difference in Hz</th>
</tr>
</thead>
<tbody>
<tr>
<td>focus-un</td>
<td>107</td>
<td>Mean</td>
<td>18.7</td>
<td>18.1</td>
<td>13.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Std. Error</td>
<td>3.4</td>
<td>2.7</td>
<td>3.1</td>
</tr>
<tr>
<td>un-focus</td>
<td>113</td>
<td>Mean</td>
<td>-4.3</td>
<td>-4.4</td>
<td>-2.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Std. Error</td>
<td>2.9</td>
<td>2.7</td>
<td>3.5</td>
</tr>
<tr>
<td>SOF-un</td>
<td>113</td>
<td>Mean</td>
<td>6.2</td>
<td>6.4</td>
<td>6.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Std. Error</td>
<td>3.9</td>
<td>2.9</td>
<td>2.7</td>
</tr>
<tr>
<td>un-SOF</td>
<td>116</td>
<td>Mean</td>
<td>7.3</td>
<td>5.0</td>
<td>7.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Std. Error</td>
<td>2.8</td>
<td>2.1</td>
<td>3.3</td>
</tr>
</tbody>
</table>

TABLE 2 – Syntagmatic mean differences of $f_0$ values for focus vs. unfocused and SOF vs. unfocused.

The data in Table 2 confirms again that focus is pitch-marked and that SOF-marking clearly differs from focus-marking with regard to pitch. Despite the fact that the overall variation (represented by the standard error of the mean) is approximately the same for the focus and the SOF sentences, the mean differences are almost completely absent for SOF sentences. For both focus- and SOF-marking $f_0$-range does not seem to carry much information. Note that the asymmetry between the two focus assignments on the focus-
sentences (the first and the second row in Table 2) is due to declination and therefore expected. If the focus precedes the unfocused element the unfocused element is likely to be de-accented, which results in a large positive $f_0$ difference. If the focused element follows the unfocused expression, the unfocused expression can bear a pitch accent. The nuclear accent on the focused expression only has to be about as high as or slightly higher than the maximum $f_0$ on the unfocused expression in order to be perceptually salient. This is reflected in negative differences with relative low absolute values in the second row of Table 2. The same pattern is not observed for SOF vs. unfocused expressions. $f_0$ differences seem to be primarily determined by the position since they are about the same for both SOF assignments (see row three and four in Table 2). The $f_0$ differences in SOF sentences seem to be primarily determined by declination.

Statistically, this is reflected by the fact that ‘focus assignment’ has a significant effect on standardized maximum, mean, and minimum $f_0$ measures in focus sentences but not in SOF sentences. This explains the interaction effect on those three measures found above for ‘focus assignment’ and ‘type of focus’. It also follows that the main effects on the three measures found in the main analysis for ‘focus assignment’ only holds for focus sentences. SOF-marking does not employ statistically significant syntagmatic pitch-marking.

In sum, with regard to syntagmatic $f_0$ differences, focus- and SOF-marking do not only differ quantitatively, but also qualitatively. While focus-marking (in the data considered here) was strong enough to counter declination, resulting in positive differences for the ‘focus-unfocused’ order and negative differences for ‘unfocused-focused’ order, $f_0$ differences for SOF sentences were primarily determined by declination and therefore positive and approximately of the same size for both the ‘SOF-unfocused’ and the ‘unfocused-SOF’ order. This was supported by the statistical analysis. Note, however, that it doesn’t follow from this that pitch is never part of SOF-marking. It only follows that pitch-marking is not consistent enough or does not occur often enough to reach statistic significance. In fact, I will show below that optional pitch-marking does play a role in SOF-marking.

Next, I present a syntagmatic comparison of focus- and SOF-marking with regard to duration, loudness, energy and relative energy in Table 3. The structure of Table 3 is parallel to Table 2. Intensity is given as ratios rather than differences because this is known to be more relevant for perceived loudness. Significant results are highlighted by shading.

<table>
<thead>
<tr>
<th>Focus/SOF assignment</th>
<th>N</th>
<th>Duration difference in ms</th>
<th>Intensity ratio in dB</th>
<th>Relative energy Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>focus-un</td>
<td>107 Mean</td>
<td>15.7</td>
<td>1.1649</td>
<td>.01100884</td>
</tr>
<tr>
<td></td>
<td>Std. Error</td>
<td>7.0</td>
<td>.1623</td>
<td>.00185331</td>
</tr>
<tr>
<td>un-focus</td>
<td>113 Mean</td>
<td>-6.5</td>
<td>.0888</td>
<td>-.00238121</td>
</tr>
<tr>
<td></td>
<td>Std. Error</td>
<td>7.0</td>
<td>.1287</td>
<td>.00192141</td>
</tr>
<tr>
<td>Significance</td>
<td>$^F(1,8)= 4.6$</td>
<td><strong>$F(1,6)= 16.3$</strong></td>
<td><em>$F(1,6)= 11.8$</em></td>
<td></td>
</tr>
<tr>
<td>SOF-un</td>
<td>137 Mean</td>
<td>10.1</td>
<td>.8519</td>
<td>.00494595</td>
</tr>
<tr>
<td></td>
<td>Std. Error</td>
<td>5.7</td>
<td>.1354</td>
<td>.00084303</td>
</tr>
<tr>
<td>un-SOF</td>
<td>142 Mean</td>
<td>-8.1</td>
<td>.6354</td>
<td>.00267638</td>
</tr>
<tr>
<td></td>
<td>Std. Error</td>
<td>6.2</td>
<td>.1272</td>
<td>.00076972</td>
</tr>
<tr>
<td>Significance</td>
<td>$^F(1,9)= 3.7$</td>
<td>n.s.</td>
<td><em>$F(1,8)= 3.8$</em></td>
<td></td>
</tr>
</tbody>
</table>

**TABLE 3** – Syntagmatic mean differences of duration, loudness, and energy for focus vs. un-focused.
Crucially, Table 2 shows that SOF-marking has syntagmatic correlates. This strengthens the result that pitch-marking does not seem to be significantly correlated with SOF (a null result by itself doesn’t really show anything but the presence of an effect on another variable lends support for the validity of the null result). Most noticeably, the syntagmatic duration difference is essentially the same for focus sentences (15.7 ms vs. -6.5 ms) and SOF sentences (10.1 ms vs. -8.1 ms). This means that the difference in length between the same two words (which in the sample on average were between 300-400ms long) depending on whether the first one is associated with a focus sensitive operator or the second one differs by 22.2 ms or 18.2 ms in non repeated focus sentences and SOF sentences respectively. The comparison for intensity on the other hand reveals similarities to syntagmatic \( f_0 \) marking. While focus-marking is clearly correlated with syntagmatic intensity differences, intensity differences are almost completely reduced for SOF-marking. On the first view this is also the case for energy differences. Note, however, that the standard error is noticeably reduced for SOF sentences, which makes the smaller overall difference correlated with SOF-marking more significant.

The statistical analysis confirms this. While standardized intensity patterns with the standardized pitch measures in that it is significantly affected by focus assignment only in focus sentences but not in SOF sentences, duration and relative energy are significantly affected in both types of sentences. This is in line with the main analysis presented at the beginning of this section. The main analysis did only find a main effect of focus assignment on duration and relative energy but not interaction with the type of focus. For intensity on the other hand, the main analysis revealed a main and an interaction effect – just as for standardized maximum, mean, and minimum \( f_0 \). Thus all four measure for which the main analysis revealed an interaction effect of focus assignment and type of focus seem to be due to the fact that focus-marking but not SOF-marking employs those measures (rather than the two types of focus-marking employing the measures in a different way).

Summing up, the quantitative comparison of focus- and SOF-marking has revealed that SOF-marking does not obligatorily include pitch-marking, neither paradigmatically, nor syntagmatically. Duration and energy marking, on the other hand, seem to be involved in both focus- and SOF-marking. The role of intensity remains unclear. Although there is a significant effect of SOF-marking on paradigmatic intensity differences, the difference in loudness seems too low to be perceptible. One way to interpret this result is that the loudness increase correlated with SOF-marking is not evenly distributed over the whole spectrum, but that what changes is rather the spectral tilt/balance (cf. Sluijter & van Heuven 1996). The small average increase may thus be due to a larger increase of energy on some frequencies which may very well be perceptible. Since the spectral balance was not investigated in the current study, I leave this question open for future research.

---

While Sluijter & van Heuven (1996; also Sluijter et al. 1996) show that, unlike intensity (which is a correlate of accent) spectral tilt is a highly reliable predictor of *lexical stress* in Dutch, Campbell & Beckman (1997) present results that spectral tilt *and* intensity are correlates of accent rather than stress in English. These results are in conflict with Sluijter & van Heuven (1997) who present results arguing that in English, too, spectral tilt (along with other glottal parameters, duration and vowel quality) is a correlate of stress rather than accent.
4.4 Qualitative comparison

The data presented in the two previous sections suggest that pitch-marking may not play a role in SOF-marking at all. However, listening to some of the data produced by participants in Beaver et al.’s (2003) study, SOF expressions seem to be sometimes pitch-marked (also Edward Flemming, Katy Carlson, p.c.). To investigate the possibility that SOF is optionally marked by pitch, I compared focus- and SOF-marking qualitatively. Using Boersma & Weenink’s (2004) Praat 2.1.28 program for phonetic analyses, I examined the pitch contours of a randomly chosen subset of the data (both the second and the third sentences of each chosen discourse).

The results are mixed. While the data clearly show that OFC was marked by mostly high pitch accents (H* accents in ToBI terminology; cf. Silverman et al. 1992), the data yielded less clear answers with regard to accents on unfocused expressions and with regard to SOF expressions. To a first approximation, the relation between pitch-marking of focus and SOF in the data set can be described by two cases. Either both the focus expression in the focus sentence (e.g. the (b) sentences in (13) and (14) above) and the SOF expression in the SOF sentence (e.g. the (c) sentences in (13) and (14) above) were marked by a pitch accent, or the focus in the focus sentence was pitch-marked but the SOF in the SOF sentence did not receive a pitch accent.

Figure 5 and Figure 6 illustrate the case where the target expression (here ‘Pete’) was marked by a pitch accent both when in focus and when in SOF. Although the pitch movement (also approximated by the difference in maximum f0 on ‘Pete’ vs. ‘pill’) is reduced in the SOF sentence (54 vs. 37 Hz difference in maximum f0), ‘Pete’ bears an H* accent in both sentences (this can be confirmed both audibly and visually). As a ballpark figure it may be useful to keep in mind that, under optimal conditions, the just noticeable difference in frequencies is about 1-3 Hz for the f0-range considered here (Ladefoged 1996:78; Stevens 1998:section 4.2.3; t’Hart et al. 1990:29f.). Here and below the HiF0s are given in the figures for the two NPs that are in the scope of a focus sensitive operator (here ‘only’).

21 For all pitch analyses shown below, the auto-correlation algorithm of Praat 2.1.28 was used. The maximum and minimum pitch parameters differed from speaker to speaker. Default settings were used for all other parameters.

22 There also were two further categories, which, although interesting for other reasons, I do not discuss here. Sometimes subjects produced what I would call an ill-formed pitch contour. That is, in the (b) sentence there was a clear pitch accent on the unfocused word. In all cases I looked at this error was carried over into the second occurrence focus sentence. In other cases, the (b) sentence was well-formed but the (c) sentence contained a clear H* on the unfocused expression. Although it is a questions for further research whether these are actually speech errors (probably due to the artificiality of the experimental situation and the complexity of the reading task), I ignore those cases for now. Example of more pitch tracks including some probably ill-formed ones are given in Appendix D.
Figure 5 and Figure 6 illustrate the same for a target word (here ‘chimps’) in the second position. In this case the pitch movement on the SOF expression is even more reduced in comparison to the pitch movement on the focus expression (105 Hz vs. 28 Hz). Nevertheless, ‘chimps’ seems to be pitch-marked in both sentences as also shown by the de-accenting of the material following it.
If all discourses in Beaver et al.'s (2003) experiment had been produced like the ones shown in Figure 5 to Figure 8, the insignificance of $f_0$ measures for SOF-marking would be surprising (see also Beaver et al. 2004, to appear). However, as mentioned above, not all SOF expressions were marked by a pitch accent. Figure 9 and Figure 10 serve to exemplify cases, where sentence (b) shows a clear pitch accent on the target word (here 'court'), but the SOF expression in sentence (c) is unaccented. All expressions following the nuclear accent on 'state prosecutor' in (c) are not accented and the maximum $f_0$ on the SOF expressions is only 2 Hz higher than the one on the unfocused expression in the same sentence. Figure 11 gives another example of a SOF expression without pitch-marking (this time the first position, i.e. ‘Pete’ is in SOF). The apparent pitch movement on ‘Pete’ and ‘pill’ in Figure 11 is due to interference of the /pʰ/ in the onset and not part of the perceived pitch (cf. Ladefoged 2003:87).
So far, the data presented in this section seem to suggest that SOF can be optionally marked by pitch accents. When SOF is pitch-marked, the pitch accent is typically an H* (at least in all examples that I examined) just like the typical pitch accent on focused items (e.g. Figure 5 to Figure 8). On the other hand, SOF expressions seem to be frequently realized without pitch-marking (e.g. Figure 10 and Figure 11). Even though the analyses presented in section 4.2 and 4.3 (see also the next section) argue that duration, energy, and maybe intensity were not only sufficient but even more reliable markers of SOF than any pitch measure, it could be the case that prominence on SOF expressions is only perceptible in the presence of a clear pitch accent. The results of the perception experiment in section 3 would then only hold for pitch-marked SOF expressions, which arguably are not the most common case. This in turn would cast doubt on the conclusions of the experiment that SOF-marking Hypothesis 2 is correct and the argument from second occurrence focus invalid.

To address this possibility, I examined the seven best items from the perception experiment (i.e. the items that most participants had answered as predicted by the SOF-marking hypothesis). The goal of this study was to determine whether subjects only performed clearly above chance if the SOF expression was pitch-marked. The study also investigates the possibility that SOF is marked by low pitch accents (i.e. L*) as suggested by e.g. Katy Carlson (p.c.). Consistent marking of SOF by L* accents should have shown up as a statistically significant effect on minimum f0 in the quantitative studies presented above. Nevertheless, it could be that participants in the perception study only got the right answer if there was an L* accent on the right word. The current study shows that this is not the case.

Recall that participants in the perception experiment were asked to identify the element in a pair in which the second target word was more prominent than the first. For the seven best items, more than 85% of all subjects chose the right item (there were 12 such items, but only

---

23 Even though focus is typically marked by an H* pitch accent this is not always the case (Cohan 2000; Hedberg 2003; Hedberg & Sosa 2001). So far not enough data are available to determine whether SOF can also be marked by other type of pitch accents than H*.

24 Consider a world in which SOF is marked by low or high pitch accents which are used equally often. The f0 measures taken (maximum, mean, minimum, and range) would therefore not reach significance. Now consider that in this world, all measures that seem to strongly correlate with SOF-marking are mediated by pitch-marking (e.g. there is a duration increase on a phrase whenever that phrase is marked by pitch accents, regardless of whether it is a low or high pitch accent). The correlation study presented below in section 4.5 addresses this possibility by also looking at partial correlations. However, since qualitative studies are sometimes easier to interpret, I have included such a study.
Thus for these seven items, the prominence on the second target word in that item must be very clearly perceptible. The question is which prosodic properties made the second target word so prominent. In the current study I focus exclusively on the question whether it was only due to clear pitch-marking (i.e. pitch accents) that these seven items were judged correctly by almost everyone.

Table 4 summarized the hand checked maximum (MAP) and minimum \( f_0 \) (MIP), and the \( f_0 \) contour for both words in both sentences that had to be compared in the experiment. The column ‘Contour’ contains information about whether the \( f_0 \) was falling (\( ð\)), rising (\( Ù\)), or flat (\( Ö\)) on the word, as well as the overall pitch movement in Hz (given in parentheses). For one item, no \( f_0 \) values could be obtained on the second word of sentence B because the speaker’s voice was too creaky. The first four rows in Table 4 are items for which participants always (correctly) chose answer ‘A’. The remaining three rows represent items for which answer ‘B’ was always (correctly) chosen. This is indicated by shading in the table. For ease of reference the item ID is given in the first column.25

| Item ID | Sentence A | | | Sentence B | | |
|---------|------------|------------|--------|------------|------------|
|         | word1      | word2      |        | word1      | word2      |
|         | MAP        | MIP        | Contour | MAP        | MIP        | Contour |
| 3       | 204        | 190        | (14)\( ð\) | 194        | 180        | (14)\( ð\) |
| 12      | 204        | 171        | (33)\( Ù\) | 213        | 44         | (169)\( ð\) |
| 20      | 182        | 167        | (15)\( ð\) | 180        | 167        | (13)\( ð\) |
| 25      | 176        | 172        | (4)\( ð\) | 181        | 175        | (6)\( ð\) |
| 6       | 265        | 203        | (62)\( Ù\) | 193        | 177        | (15)\( ð\) |
| 21      | 111        | 99         | (12)\( ð\) | 102        | 100        | (2)\( ð\) |
| 29      | 213        | 103        | (110)\( Ù\) | 227        | 119        | (108)\( Ù\) |

TABLE 4 – Maximum, minimum \( f_0 \), and \( f_0 \) contour on word1 and word2 in sentence A and B for pairs that were judged correctly by all participants in the perception experiment

Two important pieces of information can be extracted from Table 4. First, it is not apparent how one can explain that all items given in Table 4 were judged correctly almost all the time by only referring to pitch differences between the words in a sentence or difference in pitch between the sentences. On the one hand there are items where the chosen sentence had more pitch movement on the second word (e.g. item 12) and/or more pitch movement on the unfocused item in the sentence not chosen (e.g. item 12, 20). On the other hand there are items with more pitch movement and higher pitch on the unfocused word in the chosen sentence (e.g. item 3). Finally, there are items that contain almost no pitch movement on any of the target words (e.g. item 21, 25). Although, as expected, pitch-marking is a good predictor of SOF-marking for the data at hand if present, Table 4 shows that the prominence on a SOF-marked word could be clearly perceived by subjects even in the absence of pitch-marking. Second, Table 4 also shows that it cannot be the case that SOF is consistently marked by L* accents. Although a clear low pitch accent (indicated by a fall in \( f_0 \) in Table 4) on the second

---

25 By ‘hand-checked’ I mean that in addition to the automatic pitch analysis, I measured the duration of pitch periods by hand and compared the resulting \( f_0 \) values with the automatically extracted ones to see whether there were any inconsistencies.
word is a good predictor, only two out of the seven sentences have an L* accent on the second word of the chosen sentence.

To sum up what has been shown so far, SOF-marking optionally employs pitch-marking (and those cases look quite similar to OFC-marking though smaller in magnitude). Unlike focus-marking, however, SOF-marking is often realized without a pitch accent. Crucially, even in the absence of a pitch accent, SOF-marking is perceptible. Since several non-pitch factors show paradigmatic and syntagmatic correlations with SOF-marking, this raises the questions of how well those factors predict SOF-marking and whether, overall, pitch or non-pitch measures contribute more information to SOF-marking. These two questions are addressed in the next section.

4.5 Regression study

To determine which combination of the stimuli’s phonetic properties determined which stimulus in a pair the participants of the perception study chose, several regression models were compared. Specifically, I focus on specific versions of the SOF-marking Hypothesis 3. SOF-marking Hypothesis 3A is merely stating what was observed in the previous three sections. SOF-marking Hypothesis 3B states the suspicion that, although not significant in the overall data set, pitch measures were significant predictors of subjects’ behavior in the perception experiment. This hypothesis is based on the well-known fact that pitch in English is a strong marker of prominence and on the fact that the results of the previous section has shown that pitch-marking was a good predictor if present. Hypothesis 3C is a stronger version of Hypothesis 3A.

<table>
<thead>
<tr>
<th>Three specific spell-outs of SOF-marking Hypothesis 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>(3A) Duration, intensity, and energy are significant predictors (crucial measures) of SOF-perception.</td>
</tr>
<tr>
<td>(3B) Maximum, mean, and minimum $f_0$ are significant predictors (crucial measures) of SOF-perception.</td>
</tr>
</tbody>
</table>
| (3C) Duration, intensity, and energy are stronger predictors of SOF-marking than any of the pitch measures. Particularly duration, intensity, and energy marking on SOF expression is not mediated by pitch-marking.  

The measures considered here are duration, intensity (standardized and absolute), energy (relative and absolute), maximum, mean, and minimum $f_0$, and $f_0$-range. Pitch-related measures (i.e. standardized and non-standardized maximum, mean, and minimum $f_0$, and $f_0$-range) of the stimuli used in the perception study were hand checked because of the

---

26 While we can test whether a given prosodic property correlates with SOF-marking and whether this correlation is mediated by any of the other measures taken, it is impossible to show that a given correlation is not an indirect effect (i.e. mediated by) another phonetic/phonological property of SOF-marking that was not measured. More concretely, this means that it is impossible to show that duration, intensity, and energy effects are not mediated by another unknown factor not investigated here (e.g. vowel quality and spectral balance/tilt; cf. Huss 1978; Sluijter & van Heuven 1996a,b, 1997). It is, however, possible to address the question whether the correlation between PA and a given measure is mediated by another measure (in this case pitch) and whether one measure contributes more information than another measure.
possibility that pitch-doubling or pitch-halving had gone unnoticed in the automatic pitch analysis used in Beaver et al. (2003).\textsuperscript{27} This way a large portion (24\%) of the stimuli underwent corrections for at least one pitch-related measure of at least one of the two words in the stimuli. Crucially, the hand corrected data yielded a qualitatively different and more refined analysis with regard to the influence of pitch on the perception of SOF.

Based on the measures on each target word (i.e. each potentially SOF-marked word), measures of (a) the syntagmatic difference between the two target words in a sentence, $dw(measure)$, and (b) the paradigmatic differences between the same target word in the two sentences of a pair, $ds(measure)$, were calculated:

$$
dw(item, measure, S) = \text{measure on word1 in sentence } S - \text{measure on word2 in sentence } S
$$

$$
ds(item, measure, W) = \text{measure on word } W \text{ in sentence A} - \text{measure on word } W \text{ in sentence B}
$$

*Item* refers to the ID number of the item pair (1 to 40; cf. section 3.1) that the measurement is taken on. *Word1* is the word that linearly precedes *word2*. *Sentence A* and *B* refer to the two sentences in a pair (recall that they were always labeled ‘A’ and ‘B’). The relations between the distance measures are further illustrated by the following figure:

<table>
<thead>
<tr>
<th>Word1, sentence A</th>
<th>$dw(i, m, A)$</th>
<th>Word2, sentence A</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$ds(i, m, w1)$</td>
<td>$ds(i, m, w2)$</td>
</tr>
<tr>
<td>Word1, sentence B</td>
<td>$dw(i, m, B)$</td>
<td>Word2, sentence B</td>
</tr>
</tbody>
</table>

**FIGURE 12** – Distance measures used in the regression study

The $ds$ distance measures describe the relation between sentence A vs. sentence B rather than the relation between the sentence representing the correct/expected answer and the sentence representing the incorrect/unexpected answer. This makes sense, since this study aims to understand what influenced the decisions of participants in the experiment (i.e. why they chose answer ‘A’ over answer ‘B’) even if their decision was not the one expected by the SOF-marking Hypothesis. As mentioned in 3.4, the artificiality of the reading task in Beaver et al.’s experiment resulted in a substantial amount of intonationally ill-formed stimuli, which were nevertheless included in the perception experiment. The motivation for this study is to investigate what determined subjects’ perception of prominence and it therefore does not matter that in some cases a supposedly unfocused expression (i.e. an expression elicited to be unfocused) was perceived as more prominent than a focused expression because the stimulus was intonationally ill-formed.

\textsuperscript{27} Not all pitch values were hand checked. All ‘suspicious’ values (i.e. values that were more than 2 standard deviations above or below the speaker’s mean pitch) were checked and a subset of non-suspicious data points was also checked. Since the pitch-tracker seemed to work correctly for the non-suspicious cases, not all of them were examined for pitch-doubling or halving.
To illustrate the meaning of \(dw\) and \(ds\), consider the measure duration. If \(dw(item, duration, S)\) is positive for a given sentence \(S\) word1 is longer than word2, and vice versa if \(dw\) is negative. If \(ds(item, duration, W)\) is positive the word \(W\) was longer in sentence A than in B. Intuitively, the two measures \(dw\) and \(ds\) or any combination thereof correspond to parts of strategies that subjects could theoretically have employed in deciding which stimulus they should choose.

In addition to the above-mentioned distance measures, an overall distance measure \(d\) was calculate from the values of all four target words in a pair:

\[
d(item, measure) = \text{ds}(item, measure, 2) - \text{ds}(item, measure, 1) = \text{dw}(item, measure, B) - \text{dw}(item, measure, A) = \text{measure on word2 in sentence A} - \text{measure on word2 in sentence B} - \text{measure on word1 in sentence A} + \text{measure on word1 in sentence B}
\]

In the following significance is given for \(p< 0.05\), two-tailed. Effects referred to as highly significant are at \(p< 0.01\). PA refers to the percentage of ‘A’ answers. All analyses were performed using the SPSS 11.5 for Windows statistical package.

First, the overall difference \(d\) was examined for all measures taken. In what follows, the correlation coefficient \(r\) and the \(r^2\) of the model in which the measure is the only predictor given. \(R^2\) described how much of the overall variation of PA is accounted for by the model (an \(r^2\) of 0.1 corresponds to 10% of the variation). Duration correlated significantly with PA (\(r= 0.4; r^2= 0.16\)) and standardized r.m.s. intensity (\(r= 0.64; r^2= 0.41\)) and relative energy (\(r= 0.66; r^2= 0.44\)) were all highly significantly correlated with PA. Thus SOF-marking Hypothesis 3A was met. Of the non-standardized \(f_0\) measures only mean \(f_0\) had a significant correlation with PA (\(r= 0.44; r^2= 0.19\)). Standardized \(f_0\) measures with the exception of \(f_0\)-range, however, were correlated significantly (maximum \(f_0\): \(r= 0.36; r^2= 0.13\); minimum \(f_0\): \(r= 0.33; r^2= 0.11\)) or even highly significantly (mean \(f_0\): \(r= 0.46; r^2= 0.21\)). Thus SOF-marking Hypothesis 3B was met as well.

These results raise the question as to whether the stronger SOF-marking Hypothesis 3C is met as well. In order to answer this question, three types of analyses were conducted. First, I compared how much information a model based only on pitch provides about PA vs. how much information a model based on duration and intensity provides about PA. Since a model with all four standardized \(f_0\) measures would violate the requirement of independence (because of the collinearity of the four factors), I removed the weakest pitch factor, maximum \(f_0\), and compared the resulting significant model (\(F(3, 34)= 3.6; \text{adjusted } r^2= 0.17\)) with a model that contained only intensity and duration as predictors (\(F(2,35)= 15; \text{adjusted } r^2= 0.43\)). The latter model accounts for twice as much of the variation (\(r^2= 0.17\) vs. \(r^2= 0.43\)).

Second, I tested whether duration, intensity, and energy would still be strongly correlated with PA after all \(f_0\) measures are controlled for. All partial correlations were calculated (for the overall distance \(d\) as well as for the four distance measures). In all cases duration, intensity,

\[28\] Only the standardized \(f_0\) measures are used in the remainder of the analysis since they clearly contained more information. A comparison between the correlation results for the uncorrected \(f_0\) measures from Beaver et al.’s study and the hand-corrected \(f_0\) measures used here showed that none of the three significant correlations would have been detected on the uncorrected data. This stresses the importance of detailed checks of outputs returned by automated pitch extraction programs.
and energy were equally or even more correlated with PA after all \( f_0 \) measures were controlled for. This result suggests that subjects would have performed equally well or better without access to any pitch information.

Third and finally, I used a stepwise multiple linear regression to automatically derive a hierarchical model with the optimal subset of predictors for PA. The stepwise method begins with the strongest predictor, adds it to the model, and then checks whether adding the second strongest predictor to the model would significantly improve the model. At each step, the algorithm also checks whether all predictors in the model are still significant. Although this method does not necessarily arrive at the best possible model, it usually does an adequate job in creating a parsimonious model that predicts most of the variation in the data set. If fed with the d of standardized maximum, mean, minimum \( f_0 \), \( f_0 \)-range, duration, standardized r.m.s. intensity, and relative energy, the stepwise algorithm suggests a model based on only relative energy (\( F(1,36)=28.4; \text{adjusted } r^2=0.43 \)). That energy and not the highly correlated duration and/or intensity measures was chosen is not surprising since the model is more parsimonious. It is however striking that adding any of the \( f_0 \) measures did not result in a significant improvement in the model's fit. \( D(\text{energy}) \) accounts for 44% of all variation of PA. After removing energy from the input variables, the same method results in a model with only duration and intensity as factors (\( F(2,35)=15.0; \text{adjusted } r^2=0.43 \)). Thus the \( f_0 \) measures contain no significant amount of information that is not already contained in the duration and intensity (or energy) measure. This is partly due to the fact that some of the \( f_0 \) measures are correlated with intensity (and therefore with energy). Both standardized maximum and mean \( f_0 \) are highly significantly correlated with r.m.s. intensity (and energy). Nevertheless, this means that intensity and duration (or energy) are better predictors than any \( f_0 \) measure. This is intriguing in light of the well-known fact that \( f_0 \) measures are better predictors than duration and intensity for OFC (Liberman & Pierrehumbert 1984; Terken & Hirschberg 1994). Even though pitch is more important for SOF-marking than suggested in Beaver et al. (2003), it is less important than duration and intensity (or energy). Thus SOF-marking Hypothesis 3C is met as well.\(^{29}\)

Summing up, this section has provides further support for SOF-marking Hypothesis 3 by showing that duration, intensity, and energy, rather than pitch-measures seem to be the determining factors in perceiving prominence on SOF expressions. A partial correlation analysis also showed that duration, intensity, and energy would have accounted for as much or more of the overall variation of PA if there had been no pitch-marking at all. While OFC has been shown to be most reliably indicated by pitch-marking rather than duration and intensity (Liberman & Pierrehumbert 1984; Terken & Hirschberg 1994), duration, intensity, and energy are the most reliable predictors of prominence on SOF expressions. Rather than examining only the production data (as done in Beaver et al.’s 2003 study), this section examined which prosodic features correlated with the actual decisions of participants in the perception study. This is an important difference given that there is reason to believe that Beaver et al.’s (2003) production study contained a considerable (though acceptable) amount of prosodically ill-formed utterances (see section 3.4). Since the current regression study investigates which element in a pair participants chose rather than which ones they should have chosen, it is

\(^{29}\) It can also be shown that the \( f_0 \) measures are not mediating the duration and intensity effects. Several plausible mediation analyses were performed and in all cases the results were non-significant.
relatively unaffected by most kinds of ill-formed utterances (i.e. those that still employ SOF-marking but on the wrong target word).

4.6 Conclusions

The results presented throughout this section show that SOF-marking is optionally achieved by pitch-marking, but is perceptible even in the absence of pitch-marking. The regression study in section 4.5 argues that, of the factors investigated, duration, intensity, and energy (rather than $f_0$) differences seemed to most reliably predict subjects’ choice in the perception experiment. These results support the hypothesis that the apparent lack of pitch accents on SOF expressions has misled proponents of the argument from second occurrence focus (e.g. Partee 1999) to conclude that SOF expression are not prosodically marked at all (for a similar though slightly different argument, see Bartels 1997; Beaver et al. 2004). The potential lack of pitch accents on SOF expressions has some interesting consequences that go well beyond the argument from second occurrence focus.

The observation that duration (and energy) seems to be a more consistent correlate of SOF than pitch also presents a problem for all theories implicitly or explicitly assuming that focus is always marked by a (nuclear) pitch accent (e.g. Erteschik-Shir 1997, 1999; Halliday 1967; Jackendoff 1972; Kadmon 2001; Selkirk 1995) and to theories of intonation focusing primarily on the role of pitch-marking (e.g. Pierrhumbert 1980; Silverman et al. 1992). The consequences for those theories are discussed in the next section.

Pulling together everything discussed in this section, I suggest a refinement to SOF-marking Hypothesis 3. To capture that, on the one hand, there are cases of SOF-marking that seem to show all properties of OFC-marking including pitch-marking (the observed reduction of prosodic features even in those cases is not surprising given that all instances of SOF discussed here were in discourse final position) and that, on the other hand, SOF doesn’t have to be pitch-marked, I suggest that there are two SOF-marking allomorphs, and the one which employs pitch-marking is also used in OFC-marking:

<table>
<thead>
<tr>
<th>SOF-marking Hypothesis 3’</th>
</tr>
</thead>
<tbody>
<tr>
<td>(3’) SOF-marking employs (at least) two prosodic allomorphs. One of these allomorphs is also employed in OFC-marking.</td>
</tr>
<tr>
<td>(3’-1) The allomorph shared between OFC- and SOF-marking employs pitch-marking.</td>
</tr>
<tr>
<td>(3’-2) The second SOF-marking allomorph does not employ pitch-marking.</td>
</tr>
</tbody>
</table>

Even though this refinement has to be tentative at this point, I take it to be valuable since it provides a clear frame for further investigation and it enables me to make more concise predictions about the distribution of different types of focus-marking (cf. section 5.2 where I come back to this point).

30 For one thing, more detailed studies of SOF- and OFC-marking as well as other types of focus-marking are necessary before the proposed revision can be taken for certain. Second, it is arguably the case that there are several ways to mark focus by pitch accents (e.g. $H^*$, $L^*$, and $L+H^*$; cf. Hedberg & Sosa 2001:21; see also footnote 23 and 31) which – to be precise – correspond to several allomorphs of focus-marking (if the meaning of focus is always the same) or even allomorphs of several different type of focus morphemes (if there are several types of focus and each of them is marked by different types of pitch accents; cf. footnote 14).
General discussion and conclusions

At the beginning of this paper, I have stated three hypotheses about second occurrence focus (SOF), throughout the paper referred to as the SOF-marking Hypotheses. The three SOF-marking Hypotheses are repeated below. SOF-marking Hypothesis 1 has been shown to hold by Beaver et al. (2003, to appear; see also Bartels 1997; Rooth 1996 for related studies). SOF-marking Hypotheses 2 and 3 have been confirmed in the current paper. SOF-marking Hypothesis 2 has been confirmed by the perception experiment in section 3 and SOF-marking Hypothesis 3 has been confirmed (and refined) by several studies in section 4.

The Second Occurrence Focus-marking Hypothesis (SOF-marking Hypothesis)

1. SOF is prosodically marked.
2. SOF-marking is perceptible (and used in processing of focus assignments).
3. SOF-marking shares features with OFC-marking but is not identical to it.

In the following three subsections, I discuss the consequences drawn from the studies presented above on the argument of second occurrence focus (section 5.2), current theories of focus (section 5.2), and, finally, current theories of intonation (section 5.3). Finally, section 5.4 provides a brief overview of possibilities for future research.

5.1 Consequences for the argument from second occurrence focus

The results of the perception experiment in section 3 provide clear evidence for SOF-marking Hypothesis 2, that is, SOF-marking is in principle perceptible. Even though these results do not show that hearers use phonetic information to infer SOF assignment, they arguably have access to prosodic information correlated with SOF.

Together with Beaver et al’s (2003, to appear) work and Rooth’s (1996) study, both of which showed the validity of SOF-marking Hypothesis 1, section 3 provides a strong empirical basis against the argument from second occurrence focus (as stated in the introduction; cf. Partee 1999) and in favour of semantic theories of focus (e.g. Jacobs 1983; Krifka 1992; Rooth 1985; von Stechow 1985/1989). The only way to maintain the argument from second occurrence focus would be to show that in online processing of focus assignments hearers do not make use of phonetic information to infer SOF. Since this would come down to not making use of in principle available information, the burden of proof has shifted to proponents of the argument from second occurrence focus.

5.2 Consequences for theories of focus-marking

projection (e.g. Selkirk 1984, 1995; Gussenhoven 1984, 1999, forthcoming) focus is predominantly assumed to be directly correlated with pitch-marking, primarily with high pitch accents. I will use Ladd’s (1996) term ‘focus-to-accent’ (henceforth FTA) approaches to refer to any kind of theory that attempts to predict pitch accent placement based on focus assignment or, more importantly here, focus assignment from pitch accent placement (e.g. Selkirk 1995).

While it is uncontroversial that OFC-marking in English is realized by a “conspicuous pitch change in or near the lexically stressed syllable of the word”, i.e. pitch-marking (Terken & Hirschberg 1994:126; see also Pierrehumbert 1980; t’Hart; Collier & Cohen 1990)31, SOF-marking as discussed in this paper, presents a problem for FTA approaches. In the next paragraph, I outline why the data in this paper presents a challenge to FTA approaches.

SOF is a type of focus. That is, in terms of its meaning impact, SOF is undistinguishable from ordinary focus (OFC). Thus FTA approaches should be able to capture the distribution of prominence due to SOF. FTA approaches (as defined above) predict pitch accent placement rather than prominence placement. However, section 4 provides clear evidence that SOF-marking, unlike OFC-marking, does not require a pitch accent. While SOF-marking optionally can employ pitch-marking, the most consistent prosodic correlates of SOF are syntagmatic and paradigmatic duration and energy differences, and paradigmatic intensity differences. Importantly, the qualitative analysis of SOF-marking (as compared to OFC-marking; cf. section 4.4) revealed that SOF prominence was perceptible whether pitch-marked or not.

In section 4.6, I proposed to capture the optional employment of pitch in SOF-marking by positing two SOF-marking allomorphs. One SOF-marking allomorph employ pitch accents and is – for all that we know so far – identical to (one of) the OFC-marking morpheme(s). The other SOF-marking allomorph does not employ pitch-marking. This proposal based on the data discussed in section 4 is captured in the refinement of SOF-marking Hypothesis 3 repeated below (cf. section 4.6).

**SOF-marking Hypothesis 3’**

(3’) SOF-marking employs (at least) two prosodic allomorphs. One of these allomorphs is also employed in OFC-marking.

(3’-1) The allomorph shared between OFC- and SOF-marking employs pitch-marking.

(3’-2) The second SOF-marking allomorph does not employ pitch-marking.

FTA approaches have no means of inferring SOF if the second SOF-marking allomorph is used (cf. SOF-marking Hypothesis 3’-2) because they only deal with the placement of pitch accents. Such accounts thus fail to account for the nature of mapping between focus (of whatever type, including SOF) and the prosodic realization of a sentence. If one does not want to give up the generalization that SOF is a type of focus, theories of the relation between

---

31 It is, however, neither the case that all pitch-marked occurrences of focus are marked with a high pitch (Hedberg & Sosa 2001:21) or even any pitch accent (Ladd 1996:223f.), nor is it the case that a high pitch always marks focus (Cohan 2000; Hedberg 2002:142, 2003; Hedberg & Sosa 2001; Hocket 1998). Here I do not discuss this lack of direct correspondence between accents and focus assignment although it might be taken as further support for sentence level prominence independent of pitch accents, if it turns out that non-accented foci are prosodically marked by other means (similarly to SOF).
focus and accent, such as current FTA approaches (e.g. Erteschik-Shir 1997, 1999; Gussenhoven 1991; Selkirk 1984, 1995; Steedman 2002), have to be revised to make reference to a more general level of sentence level prominence (recall that the perception experiment in section 3, arguably, shows that SOF prominence can be perceived at the sentence level, or at least at the level of a complete prosodic phrase). Any theory that doesn’t reduce sentence level prominence to pitch-marking, such as, for example, theories that treat phrasal stress as a phenomenon independent of pitch accent placement (e.g. Hayes 1995; Ladd 1996; Liberman & Prince 1977), on the other hand, are compatible with the data presented above.

In the next section, I discuss such theories of intonation. As for theories of focus-marking, it follows from the existence of a non-pitch employing SOF-marking allomorph that theories of focus-marking have to be based on theories of intonation that treat pitch accent placement and sentence level/phrase level stress assignment as two independent phenomena.

5.3 Consequences for theories of intonation

If sentence/phrase level prominence can be perceptible even in the absence of pitch-marking, it is likely that this information will be used in the processing of prosodic information. Much of recent research on sentence level prominence has focused on the role of pitch accents, their distribution and their type (e.g. Bolinger 1958, 1964:285, 1972a:22; Pierrehumbert & Hirschberg 1990; Beckman & Ayers 1994; Beckman & Hirschberg 1994; Silverman et al. 1992). As argued for the case of SOF-marking in the previous section, this is problematic.

In this section I assume a framework like Hayes (1995; see above for further references) that provides a notion of secondary prominence and a clear division of prominence and pitch accent. I outline an analysis that provides tentative answers to the following questions. How could the distribution of the two hypothesized SOF-marking allomorphs be accounted for given a theory of intonation that is in principle compatible with the data? Why do the two allomorphs of SOF-marking differ in the way they differ?

I follow Hayes (1995) and Pierrehumbert (1980) in assuming that perceptual prominence and accent placement are independent phenomena. I also assume that the relation between these two phenomena is that, in English, primary prominence is usually (maybe always) marked by a pitch accent. In this view the nuclear stress rule says something like ‘within a prosodic phrase no element of greater or equal prominence can follow the location of the nuclear stress’. Since the final pitch accent in a phrase is often perceived as the most prominent one, it follows that final pitch accent of a phrase should typically lie on the element that is determined to carry the nuclear stress (i.e. the most prominent element in the phrase). It also follows that secondary prominence can be encoded by pitch accents if the secondary prominence precedes the primary prominence but not if it follows it.

Now, recall that SOF (as considered in the current paper) by definition follows a nuclear accent. In (15) and (16), repeated from (13) and (14), the SOF expressions ‘Sid’ and ‘a pill’ follow an OFC expression ‘the doctor’, which associates with the even-operator:

(15) a. Both Pete and Edward are suffering from the flu.
    b. But the nurse only gave [Pete]₁ a pill today.
    c. Even [the doctor]₁ only gave [Sid]SOF a pill today.
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 Sid [a pill] today

The OFC expression (‘the doctor’) in (c) always received a high nuclear pitch accent (see, for example, Figure 8, Figure 10, and Figure 11 in section 4.4). For the SOF expression following that accent this means that it (a) either falls into a new intermediate phrase, or (b) the remainder of the sentence - including the SOF expression - falls in the same phrase as the nuclear accent on ‘the doctor’. The two possible phrasings are illustrated below by parenthesis:

(17) a. (Even [the doctor] only gave Sid [a pill] today).

b. (Even [the doctor] only gave Sid [a pill] today).

In the case of (a), the new phrase – by definition (cf. Pierrehumbert & Hirschberg 1990) – has to contain a nuclear accent. I propose that if the sentence containing the SOF expression is phrased as illustrated in (a) above, SOF is marked by a pitch accent, just like OFC. In other words, for phrasing (a), the SOF-marking allomorph that employs pitch-marking is chosen.

In case of (b), the SOF expression – like everything else following ‘the doctor’ is predicted to be realized without a pitch accent (i.e. is de-accented) because, as mentioned above, the nuclear pitch accent is the ultimate accent in a phrase. In an environment like (b), the SOF-marking allomorph also used for OFC-marking is not available (because it employs pitch-marking) and thus only the other SOF-marking allomorph can be used if the sentence is phrased like (b). There is another possible interpretation of the lack of a pitch accent on the SOF expression. Even with the phrasing in (b), speakers could use the pitch-employing allomorph but reduce the pitch accent enough to make the accent on the SOF expression secondary in prominence. Since this alternative doesn’t make different predictions as to the distribution of the two SOF-marking allomorphs, I do not explore this possibility further here.

Note that I have presented an account in which the lack of a pitch accent on the SOF-expression is due to the post-nuclear position of SOF (by definition). One other phenomenon (or rather set of phenomena) that is commonly associated with the lack of accents is de-accenting of given/repeated/predictable material (e.g. Terken & Hirschberg 1994; cf. Ladd 1996:166 for more references). In (15c) and (16c), as in all SOF examples considered here, everything following ‘the doctor’ is given information (Prince 1981) and structurally identical to the preceding sentence (i.e. the same words bear the same grammatical functions and are in the same syntactic position), which has been shown to result in de-accenting (Terken & Hirschberg 1994).

One way to interpret this is that it is the speaker’s intention to actually mark the given material in (15c) and (16c) as given which causes it to be realized in the post-nuclear position (rather than in a new phrase). In this view, givenness is the motivation for the post-nuclear positioning of the repeated material including the SOF expression. There is, however, an alternative interpretation. In many examples (e.g. (15b)) the OFC expression in the second sentence of the discourse was also given material. Nevertheless, they received a clear pitch

---

32 Terken & Hirschberg (1994) have shown that givenness is significantly more likely to result in de-accenting if the given phrase is repeated in the same syntactic (and linear) position as its ‘antecedent’. The lack of
accent. It may thus be the case that givenness/repetition/predictability reduced the strength of a pitch accent (which is compatible with the fact pitch-marking on SOF-expression - if present – is reduced), but it doesn’t prevent pitch-marking, whereas being in a post-nuclear position prevents pitch-marking. This view is highly compatible with the following two facts. First, the material in the SOF sentences (e.g. (15c) and (16c)) is always given but only sometimes realized without pitch. Second, all prosodic marking (including duration and energy) seems to be reduced on SOF expressions. With the current data, it is impossible to find out which of the two alternative explanations for the role of givenness is correct (or whether both of them are correct). In my view, the second proposal has a slight conceptual advantage since it also is in line with the fact that non-given information (that is not predictable) is usually not reduced: to increase perceptual clarity (i.e. to increase the chance to be understood; cf. Bell et al. 2003; Fowler & Housum 1987; Gregory et al. 1999; Jurafsky et al. 1998).

Summing up the previous paragraphs, the two phrasing options (a) and (b) jointly account for optional pitch-marking on SOF-expressions. So far, however, this leaves unexplained why duration (and energy) were the primary predictors of prominence on SOF expression. The strength of the correlation between SOF-marking and duration (and energy) strongly suggests that duration and energy are always part of SOF-marking whereas pitch is only optionally part of SOF-marking. Interestingly, for all prosodic properties investigated here, it holds that whenever a feature is involved in SOF-marking it is also involved in OFC-marking (cf. section 4). In other words, the OFC-marking allomorph (which as mentioned above is one of the SOF-marking allomorphs; henceforth OFC-SOF allomorph) seems to incorporate a superset of the features of the non-pitch employing SOF-marking allomorph (henceforth SOF allomorph).

---

33 A systematic comparison of repeated focus in pre-nuclear or nuclear positions (i.e. repeated focus which is not SOF) with SOF would help to decide between the two alternative accounts to the role of givenness proposed above. Bartels's (1997; cf. section 2) pilot study – although conducted for a different purpose provides a first glance at this comparison. Bartels compared OFC with repeated focus (Bartels's “echo NP”) and SOF respectively. The results were mixed (cf. Bartels 1997:18). On the one hand, if the focus is on a lexical NP all prosodic features employed in SOF-marking (Bartels measured maximum $f_0$, amplitude, and duration, as well as the relative pitch excursion on the test word,) are significantly more reduced than in prosodic marking of repeated focus. This argues that givenness generally reduced the prosodic marking (rather than suppressing only pitch marking). On the other hand, if the focused item is a pronoun, only pitch is significantly more reduced in SOF-marking than in marking of repeated focus, whereas duration and amplitude are equally reduced. This observation supports a view in which it is post-nuclear position of SOF causes pitch accents to be suppressed and the reduction of other prosodic features is due to givenness. Clearly, more research is needed to understand the role of givenness as a determining factor in the distribution of the two SOF-marking allomorphs.

34 It is well-known that pitch accent marking in English systematically correlates with other phonetic features. Syllable final lengthening of expressions marked by a nuclear pitch accent has been observed by, among others, Beckman (1986), Beckman & Edwards (1994), Kohler (1991), and Shattuck-Hufnagel et al. (1994). That focused phrases are longer than otherwise identical unfocused phrases has been observed by, among other, Cooper et al. (1985), Eady et al. (1986), Fry (1955, 1958), Hockey (1998), Lehiste (1970), Sluiter and van Heuven (1996, 1997), and Terken & Hirschberg (1994:126). OFC-marking is also known to correlate with a local peak in amplitude/loudness or increased energy (e.g. Beckman 1986; Beckman & Edwards 1994; Fry 1955, 1958; Hockey 1998; Kohler 1991; Lehiste 1970; Shattuck-Hufnagel et al. 1994). Van Son & Pols (1992) have shown that vowel quality differs between pitch-marked and non-pitch-marked syllables. Sluiter & van Heuven (1996, 1997) argue that some of these features are correlated with stress and others with accent.

---

35  To de-accent the phrasal boundary in examples like (15b) could thus be due to the fact that, in (15a), ‘Pete’ is in the subject position and in (15b) ‘Pete’ is an object.
This raises the question of why the two allomorphs differ in the way they differ – i.e. is it merely arbitrary that duration and energy marking is shared between the two allomorphs but pitch (and to some extent intensity) is not? I propose that – far from being arbitrary – the SOF allomorph is a ‘de-pitched’ version of the OFC-SOF allomorph (a.k.a. ‘de-accented’ if accent is used to exclusively refer to pitch accents rather than Beckman’s 1986 ‘stress accents’). This immediately explains the distribution of the two allomorphs among the two phrasing contexts (17a) and (17b) as described above. Furthermore, there are independent reasons in favour of the proposed analysis.

Sluijter & van Heuven (1997) show that, in American English, intensity and pitch seem to be correlates of sentence accent rather than lexical stress, but duration, vowel quality, and glottal parameters (e.g. spectral tilt and glottal leakage) are correlates of stress rather than accent (see also Huss 1978; Sluijter & van Heuven 1996; Sluijter et al. 1996; but see Campbell & Beckman 1997 for conflicting results with regard to spectral tilt). Further support for the claim that lengthening correlates with stress rather than accent comes from Campbell (1993) who shows that duration, if properly quantified, serves as a good indicator of stress even in the absence of pitch. Under normal conditions pitch accents are realized on the syllable that also receives the primary lexical stress. It follows naturally from this that duration, energy, and to some extent intensity (due to spectral tilt) are apparently correlated with all instances of both OFC- and SOF-marking. The lack of pitch- and to some extent intensity-marking in the SOF allomorph is also accounted for. As proposed above, the SOF-allomorph is then nothing else but a de-pitched version of the OFC-SOF allomorph and the distribution of the two allomorphs is naturally accounted for (see above). The proposed account thus receives independent supported from the above-mentioned research on accent vs. stress. Addendum (3’-3) to SOF-marking Hypothesis 3’ captures the suggested analysis.

<table>
<thead>
<tr>
<th>SOF-marking Hypothesis 3’</th>
</tr>
</thead>
<tbody>
<tr>
<td>(3’) SOF-marking employs (at least) two prosodic allomorphs. One of these allomorphs is also employed in OFC-marking.</td>
</tr>
<tr>
<td>(3’-1) The allomorph shared between OFC- and SOF-marking (OFC-SOF allomorph) employs pitch-marking.</td>
</tr>
<tr>
<td>(3’-2) The second SOF-marking allomorph (SOF allomorph) doesn’t employ pitch-marking.</td>
</tr>
<tr>
<td>(3’-3) The SOF allomorph is a ‘de-pitched’ version of the OFC-SOF allomorph.</td>
</tr>
</tbody>
</table>

To give an intermediate summary, I have provided answers to the two questions posited above:

Q: Why do the two allomorphs of SOF-marking differ in the way they differ?
A: Because the SOF allomorph is a de-pitched version of the OFC-SOF allomorph.

Q: How could the distribution of the two hypothesized SOF-marking allomorphs be accounted for given a theory of intonation that is in principle compatible with the data?
A: Whenever something (e.g. a nuclear accent together with the phasing in (17b)) prevents the pitch-employing OFC-SOF allomorph from being realized, the SOF allomorph is chosen instead.
In the analysis proposed here, SOF-marking is perceptible in the absence of pitch-marking because of the features that mark lexical stress. However, this does not and cannot mean that SOF-marking is nothing else but lexical stress. First, Huss (1978) has shown that lexical stress differences (as in minimal pairs such as ‘import’ vs. ‘impórt’) are marked in the post-nuclear domain but not perceptually distinguishable. This argues against equating the prominence on SOF expressions with the non-reduction of lexical stress because SOF prominence is perceptible (see section 3).

To conclude, it follows that the SOF allomorph marks stress above the word level in the absence of a pitch accent. Any account, such as the one discussed at the beginning of this section, that disconnects pitch accent placement from sentence level prominence and has a notion of secondary prominence (e.g. Halliday 1967; Hayes 1995; Ladd 1996; Liberman & Prince 1977; Pierrehumbert 1980; Vanderslice & Ladefoged 1972) is compatible with this result, whereas any theory of intonation that conflate pitch accent placement and prominence (e.g. Erteschik-Shir 1997, 1999; Gussenhoven 1991; Selkirk 1984, 1995; Steedman 2002) is incompatible with the results presented here.

5.4 Future research

The results presented in this paper motivate more investigations of post-nuclear prominence to determine how phrasal stress and pitch accent marking interact to create sentence level prominence.

Another interesting question that remains open is whether the SOF allomorph is also used in other cases where prominence is marked in a domain in which pitch-marking is not an option. One possibility is that the so called SOF allomorph is simply phrasal stress in the absence of a pitch accent. Non-accent based stress is then in stark contrast to accent-based stress since there are arguably many different accent morphemes (i.e. H*, L*, etc.).

Finally, one fundamental issue that this paper hasn’t touched on at all is processing. Little work has been done on the online processing of focus assignments and the role of prosodic cues in processing these assignments. The current paper argues, that, in determining the role of prosodic cues in online processing, it will be necessary to distinguish between pitch-based and non-pitch-based prosodic marking. Questions that will have to be answered include: To which extent are listeners influenced by non-pitch marking? How do different ways to create prosodic prominence interact and how do they differ cross-linguistically (since arguably, there are some universal aspects involved – i.e. if prominence has the purpose to create increased attention and understandability/clarity, there will be only so many ways to do this given the physiology of our ears and brain).
References


BEAVER, DAVID, BRADY CLARK, EDWARD FLEMMING, AND MARIA WOLTERS. 2003. Debunking the argument from Second Occurrence Focus. Presented at the LSA 2004 Meeting, Atlanta, GA.


BOLINGER, DWIGHT. 1972. Accent is predictable (if you are a mind reader). *Language* 48, 633-44.

BOLINGER, DWIGHT. 1981. Two kinds of vowels, two kinds of rhythm. IULC.


ERTESCHIK-SHIR, NOMI. 1997.

FAUL, F. & ERDFELDER, E. 1992. GPOWER 2.0: *A priori, ad-hoc, and compromise power analyses for MS DOS* [computer program]. Bonn, Germany: Bonn University, Department of Psychology.


To the best of my knowledge there has been no systematic investigation of the distribution of ‘only’ in spoken and written English. Thus, it is unclear whether the ambiguous configuration ‘only + VP’ is actually used at all and, first of all, whether it exists in spoken English. If it did not exist all work on second occurrence focus in this configuration would null and void. No experimental work can address this question since, even though subjects may be able to read and pronounce examples with ‘only’ left-attached to VP, it does not follow that this potentially ambiguous context exists in natural speech.

I determined the distribution of ‘only’ in the relevant constructions (e.g. ‘only + VP’) in large corpora of English. Two syntactically annotated corpora of written English (Wall Street Journal, release 3; Brown Corpus, release 3) totalling approximately 80,000 sentences and one syntactically annotated corpus of spoken English (Switchboard Corpus, release 3) totalling approximately 50,000 sentences were searched for occurrences of ‘only + VP’ and ‘VP(only)’, where ‘VP’ stands for a verb phrase with one NP and one PP daughter or with two NP daughters, and ‘VP(only)’ stands for a verb phrase in which ‘only’ immediately precedes one of VP’s daughters. The former pattern is the potential ambiguous one. The corpus search revealed that potentially ambiguous ‘only’ was used significantly more often in speech than in

---

35 The two tgrep2 search patterns were: ‘VP< (/VB,*/ , only) < (NP $ (/PP|NP/))’ and ‘VP< (NP $ (/PP|NP/ , only ))’ respectively.
written language - both in comparison to overall number of occurrences of ‘only’ (n=2524; \( \chi^2(1)= 28.7; p< 0.001 \)) and (marginally) in comparison to the number of occurrences of ‘only’ preceding any kind of VP (n=455; \( \chi^2(1)= 2.96; p< 0.09 \)). Table 5 summarizes the distribution of ‘only’ in spoken and written English.

<table>
<thead>
<tr>
<th>TYPE</th>
<th>overall</th>
<th>preceding a ditransitive VP</th>
<th>preceding a VP</th>
<th>within a ditr. VP</th>
</tr>
</thead>
<tbody>
<tr>
<td>CORPUS spoken</td>
<td>722</td>
<td>35</td>
<td>201</td>
<td>1</td>
</tr>
<tr>
<td>written</td>
<td>1747</td>
<td>20</td>
<td>199</td>
<td>9</td>
</tr>
<tr>
<td>Total</td>
<td>2469</td>
<td>55</td>
<td>400</td>
<td>10</td>
</tr>
</tbody>
</table>

**TABLE 5 –** Occurrences of ‘only’ in corpora of spoken and written English.

To sum up, potentially ambiguous ‘only’ exists in naturally occurring spoken English (approximately 5% of all occurrences of ‘only’ in the Switchboard corpus were potentially ambiguous). While potentially ambiguous uses of ‘only’ seem to be preferred to unambiguous uses for ditransitive VPs in both spoken and written English, spoken English, unsurprisingly, has a significantly higher ratio (approximately 16 times higher) of potentially ambiguous uses to unambiguous uses in ditransitive VPs (35 times more potentially ambiguous uses) compared to written English (2.2 times more potentially ambiguous uses).

Undoubtedly, various interpretations for this difference can be entertained that do not refer to the availability of prosodic cues (e.g. stylistic guidelines for written language), but one possible interpretation is that ambiguous ‘only’ is avoided in written English because of the lack of disambiguating prosodic cues.

**Appendix B – Stimuli used in the perception experiment**

Three types of different pairs were used in the perception experiment. The overall 40 pairs split up as follows. 16 pairs each of the types given (18) and (19) and 8 pairs of the type given in (20). Less examples of type 15 were chosen because even though it is a quasi-minimal pair, there is a small difference between the two parts (a, b), ‘would’ is present in the (a) version but not in the (b) version.

(18) a. Even [the state prosecutor]F only named [Sid]SOF in court today.
   b. Even [the state prosecutor]F only named Sid [in court]SOF today.

(19) a. Even [the doctor]F only gave [Pete]SOF a pill today.
   b. Even [the doctor]F only gave Pete [a pill]SOF today.

(20) a. Even [the kids]F would only let [the cat]SOF in the tent today.
   b. Even [the kids]F only let the cat [in the tent]SOF today.

For the time being, a copy of the experiment can be viewed online at [http://www.stanford.edu/~tiflo/?experiments/SOF-PERC](http://www.stanford.edu/~tiflo/?experiments/SOF-PERC) [sic].
Appendix C – Modelling categorical perception with logistic regression

I performed a series of logistic regression analyses to determine whether any predictors besides EXPECTED ANSWER are necessary to account for the data. A model without any variables is correct in 56.2% of all cases in the sample. Adding the predictor EXPECTED ANSWER significantly improves the model ($\chi^2=38.52; p<0.001$) and leads to correct predictions in 62.5% of all cases. Next, I tested whether adding GENDER, SUBJECT, and/or VERSION, or any interaction thereof would improve the model significantly.

![FIGURE 13 - Model after entering EXPECTED ANSWER into the regression](image)

This was not the case (using the forward-likelihood stepwise method none of those variables was entered into the regression). Adding ITEM, however, does improve the model significantly ($\chi^2=170.613; p<0.001$) and the change in -2 log likelihood is dramatic.

![FIGURE 14 - Model after entering ITEM into the regression](image)

The resulting model makes correct predictions in 73% of all cases in the sample. This raises the question whether the item effect is driving the effect of EXPECTED ANSWER, especially since the coefficient of EXPECTED ANSWER is reduced to non-significance in the new model.
Variables in the Equation

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>S.E.</th>
<th>Wald</th>
<th>df</th>
<th>Sig.</th>
<th>Exp(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1(a)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EXPECTED(1)</td>
<td>2.128</td>
<td>1037.779</td>
<td>.000</td>
<td>1</td>
<td>.998</td>
<td>8.399</td>
</tr>
<tr>
<td>ITEM</td>
<td></td>
<td>73.451</td>
<td>38</td>
<td>.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-2.262</td>
<td>1037.778</td>
<td>.000</td>
<td>1</td>
<td>.998</td>
<td>.104</td>
</tr>
</tbody>
</table>

**FIGURE 15** – Significance of factors in the resulting model

Note, however, that the latter fact is not surprising since ITEM contains all the information of EXPECTED ANSWER and more. Nevertheless, the fact that ITEM is a significant predictor and its relative strength (Change in -2 Log Likelihood for adding ITEM= 170.613 > 38.52= -2 Log Likelihood for adding EXPECTED ANSWER) demands a closer examination of the ITEM effect.

**Appendix D – Pitch tracks**

This appendix provides more pitch tracks of focus- ((b)-type sentences; cf. section 4) and SOF-marking ((c)-type sentences). In particular, examples of apparently ill-formed pitch contours referred to in footnote 12 are also given.

Figure 16 is an example of a focus sentence, (b), with a lack of clear pitch-marking on the focused item. Figure 17 is an example of a focus sentence, (b) with not only a lack of pitch marking on the focused item but also a clear pitch accent on the unfocused item.

**FIGURE 16** – Pitch track of a (b) sentence with a focus on ‘sweets’ (Speaker: je, List: D, Item: 1)

**FIGURE 17** – Pitch track of a (c) sentence with a SOF on ‘Pete’ (Speaker: je, List: D, Item: 7)
Figure 18 and Figure 19 are further examples of SOF sentences, (c), with a lack of clear pitch-marking on the SOF item:

**FIGURE 18** – Pitch track of a (c) sentence with a SOF on ‘sweets’ (Speaker: je, List: D, Item: 1)

**FIGURE 19** – Pitch track of a (c) sentence with a SOF on ‘cats’ (Speaker: je, List: D, Item: 22)

T. Florian Jaeger,
Linguistics Department,
Stanford University,
Margaret Jacks Hall, Bldg. 460,
Stanford, CA 94305-2150
USA
Email: tiflo@stanford.edu
Url: www.stanford.edu/~tiflo/