The pragmatics and prosody of focused *some* in a corpus of spontaneous speech

DRAFT – Please do not cite without permission.

Anca Cherecheș

May 2014

1 Introduction

Traditionally, a number of different phenomena have been studied under the focus heading, phenomena that could be in nature information structural (the focus vs. background distinction, information novelty vs. givenness, question-answer congruence, explicit contrast marking, explicit corrections), semantic (association with focus in focus-sensitive quantification, exhaustivity effects) or pragmatic (implicit contrast which triggers implicatures). While it may be possible to give a uniform semantic-pragmatic account for all these phenomena, the details are still debated. On the surface, what ties these phenomena together is that English-speaking linguists (and probably naive informants) perceive them as prominent. Intuitively, this psychoacoustic notion of prominence is tied to acoustic features such as intonational events, intensity, duration and vowel quality.

But beyond this intuition that focus is linked to prominence, it is still not clear what the exact nature of this relationship is. Is the link grammatically mediated, wherein we would expect prominence to be a reliable, unfailing marker of focus, or is focus pragmatically inferred, and its acoustic markers a paralinguistic device akin to emphasis? Furthermore, if phonological prominence is the grammatical marker of focus in English, are different kinds of focus phenomena marked differently? This could weigh in on the question of whether focus is a uniform category of meaning: do the related phenomena have something substantial in common? On the other hand, we would also have to ask if differences in prominence automatically require us to assume different phonological categories.

There are two strands of research that address such questions. First, is focus always prosodically marked? In the case of focus in relation to information status, it is fairly well established that accessible, informative and/or unpredictable referents tend to be more prominent (Calhoun 2006: chap. 2; and references therein). For other types of focus phenomena, however, it is not as clear. Association with focus in focus-sensitive adverbs has been studied closely in the context of second-occurrence focus, where the second focus of a phonological phrase no longer sounds (as) prominent to the naked ear, thus calling into question whether focus is always marked by a certain level of prominence. Explicit contrast marking has rarely been investigated independently from the confound of information structure (focus vs. background). Less is known about how prominence is used to trigger implicatures.

Second, there are production studies that explicitly compare how different types of focus are marked, although they are mostly limited to the distinction between discourse novelty and contrastive focus, where contrastive is used in a broad sense to cover not only explicit contrast, but also corrections and answers to wh-questions. There are also a few corpus studies (Calhoun 2006: chap. 5.6.1; and references therein) which looked at various kinds of contrastive focus (only explicit, only contrastive themes, or various combinations of wh-answers, focus-sensitive adverb scope, corrections, explicit and implicit contrast) on its own or in comparison with discourse novelty. Findings suggest that contrastive foci are in general more likely to be marked by pitch accents (F0 peaks or valleys), by peak delay (F0 extrema aligning later in the stressed syllable), and by more extreme F0 values and stress correlates such as duration, intensity, and vowel quality.

This paper adds to the list of corpus studies on the realization of focus in English through intonation and aims to address the two questions expressed above. Unlike previous studies, however, I focus on explicit and implicit contrast marked on the determiner *some*. As a function word, *some* has not been annotated for contrastive focus in previous corpus studies of contrast. However, it is somewhat easier to annotate for implicit contrast (in other words, as triggering a scalar implicature), because in its unfocused state it is often significantly reduced, like most function words.

In terms of methodology, this study follows that of Howell (2012) in focusing on a single construction, *some* + noun (in particular, I looked at tokens of *some people* and *some money*). I restricted the dataset in this way because, like Howell 2012 and unlike previous corpus studies, the utterances are harvested from the Internet, a much larger pool of data than the average speech corpus. This allows us to collect a larger number of tokens for a single phrase, large enough to perform meaningful data analysis, while at the same time controlling for the segmental context to a certain extent. As a point of comparison, 384 tokens of *some people* were collected from the Internet, whereas 98 are available in the Switchboard corpus

(Godfrey, Holliman & McDaniel 1992).

In §2, I review the most common semantic and pragmatic phenomena associated with focus, as well as sketch out the theoretical approach that I adopt in this paper. I also discuss in detail the kinds of focus that feature in this dataset and point out how they can be given a uniform semantic-pragmatic account. In §3, I review basic facts about focus marking in English through intonation and other cues. I also introduce a phonological theory that can mediate between semantics/pragmatics and these acoustic cues. In §sec:analysis, I detail the data collection process, the pre-processing stage, and how the relevant measurements were collected. I then represent these acoustic features and their interactions in the two conditions, focus and lack of focus, through traditional plotting techniques as well as unsupervised machine learning algorithms. Finally, I train a classifier that can label new data as focused or unfocused and I report its accuracy on an unseen portion of the dataset. §5 concludes this study.

2 Focus

Focus is commonly construed as a device for structuring information (Calhoun 2006, Beaver & Clark 2008, Zimmermann & Onea 2011), highlighting certain constituents in order to optimize communication in some way. If we conceive of communication as sharing information about the world, modeled as adding/subtracting propositions from the common ground (Stalnaker 1978), then focus could affect common ground *management* or common ground *content* (Krifka 2008). In the following subsections, I describe the most common focus phenomena in relation to these two aspects of the common ground.

To model focus theoretically, I adopt the central claim of Alternative Semantics (Rooth 1985, 1992) paraphrased by Krifka (2008) as follows:

(1) **Definition (informal)**: Focus indicates the presence of alternatives that are relevant for the interpretation of linguistic expressions.

For example, take the sentence *Bernie met Bertie.* The meaning of the object NP is the individual in the model that the interpretation function picks out for the constant *Bertie.* Rooth calls this the ordinary semantic value of the utterance. Additionally, however, if the object is focused, a second level of meaning is calculated: the focus semantic value, which is the set of alternatives from which the denotation of *Bertie* is drawn, here the entire domain of individuals. In general, the alternatives of a focused expression are all semantic objects of the same type as that expression. For a complex expression like the sentence in (2), which contains both focused and unfocused constituents, we compute alternatives compositionally.

In our example, the verb composes through function application with each alternative of its focused object, producing alternative propositions that differ in the inner verbal argument.

(2) Bernie met BERTIE.

a. $[\![\text{Bernie met Bertie}_{\text{F}}]\!]^o = \text{Bernie met Bertie}$ **ordinary semantic value**

b. $[\![\text{Bernie met Bertie}_{\text{F}}]\!]^f = \{\text{Bernie met } x \mid x \in D_e\}$ **focus semantic value**
$= \{\text{Bernie met Bertie, Bernie met Ernie, Bernie met Ann, } \dots\}$

In (2) and in the examples below, I mark focus as the information structural category with a subscript F and its realization in English through phonetic prominence by typesetting the prominent constituent in small caps.

2.1 Focus and common ground content

Focus affects common ground content if it has truth-conditional effects. Such effects are observed with so-called focus-sensitive operators such as exclusive adverbs *only* and *just*, additives *also* and *even* and negation *not*, whose interpretation depends on the placement of focus (Kuroda 1965, Fischer 1968, Jackendoff 1972).[1]

(3) Focus-sensitive exclusive adverb *only*

a. Bert only gives PRESENTS to his children. (He doesn't give them food, money or affection. He's a terrible father.)
$\text{ONLY}(\text{PRESENTS})(\lambda x[\text{GIVES}(b, x, c)])$

b. Bert only gives presents to his CHILDREN. (He doesn't give presents to anyone else, including his own mother. He's a terrible son.)
$\text{ONLY}(\text{CHILDREN})(\lambda x[\text{GIVES}(b, p, x)])$

(4) Focus-sensitive negation

a. Bert doesn't give PRESENTS to his children. (He gives them shelter, food, and affection. He doesn't have money for anything else.)

b. Bert doesn't give presents to his CHILDREN. (He gives presents to his wife and his mother, but he doesn't want to spoil his children.)

The assertion *Bert only gives presents to his children* is ambiguous without prosodic information because either object could be an argument to *only*. However, phonological

[1] Although all of these expressions are focus-sensitive, it is actually not clear if focus affects the assertion of the sentence they appear in or some other level of meaning, such as presuppositions or conventional implicatures they are associated with (Kadmon 2001: §13.2).

prominence on either object disambiguates, as shown by the possible continuations in (3a) and (3b), and the corresponding proposition is added to the common ground.

Focus can also affect common ground content when the speaker highlights a constituent in order to trigger a conversational implicature (Rooth 1992, van Kuppevelt 1995, van Rooij & Schulz 2004, Zondervan 2010).[2]

(5) [*Context*: Bernie, Ernie and Bertie are talking about this week's linguistics colloquium talk. I stop by and ask how the talk went.]
Bernie: Well, *I* liked it.
a. ... but no one else did.
b. ... so the others must have too.
c. ... but I don't know about everyone else.

In (21), if the speaker makes the subject *I* the most prominent word in the sentence, he triggers an implicature, as illustrated by the reinforcement options in (21a-c), although what precisely the implicature is can depend on the context. (21a) illustrates a scalar implicature. As Rooth (1992) explains, the answer is weaker than (in the sense that it is entailed by) an alternative such as *Ernie, Bertie and I liked it.* By Gricean reasoning based on the Maxim of Quantity, the hearer might infer that if the stronger alternative was not said, it is because it does not hold, so no one else liked the talk (21a). However, imagine that everyone in the department knows that Bernie has high standards and is hard to please. In that case, Bernie's answer has the flavor of ***Even I*** *liked it* (21b). In this case, the implicature is that others, who have more mundane expectations, would certainly have appreciated the talk as well; basically a strengthened version of what is actually said. Typically, strengthening implicatures are believed to be introduced by the second Maxim of Quantity ("do not make your contribution more informative than is required"), in conjunction with Relation and the last two Manner maxims ("be brief" and "be orderly"), or what Horn (1984) calls the R-principle. Note that in either case, these implicatures would be less likely to be triggered if Bernie would not have stressed the subject, but would have answered with default prosody, making *liked* the most prominent word in the sentence.

An alternative implicature is also possible here. Suppose Bernie, Ernie and Bertie have not yet exchanged impressions about the talk. In that case, Bernie's answer could implicate that he simply does not know about the Ernie and Bertie, but as for himself, he enjoyed the talk (21c). This implicature follows from the second Maxim of Quality ("do not say that for which you lack evidence"). Of course, the hearer is unlikely to know to what extent the talk

[2] For the purpose of this classification of focus effects, I am assuming that implicatures get added to the common ground.

has already been discussed, so it is possible that this particular implicature is triggered by a distinct intonational event.[3]

2.2 Focus and common ground management

The notion of common ground management was introduced by Krifka (2008) to account for how interlocutors intend the common ground to develop, given their communicative goals and interests. For instance, questions do not add anything to the common ground, unless they have presuppositions that the hearer will accommodate, but they do direct the next conversational move, as by requesting that a piece of information be added to the common ground, such as in (6).

(6) Q: Who ate all the cheese at the reception?
A: BERNIE ate all the cheese.
A': #Bernie ATE all the cheese.
A": #Bernie ate all the CHEESE.

The answer to the question above is only uttered felicitously if the most prominent constituent is the information that the question demands; in this case, *Bernie* must bear the highest prominence. This restriction on prominence in answers is also known as question-answer congruence and is intuitively captured using alternative semantics for focus and Hamblin semantics for questions. Thus, we analyze the question as denoting the set of propositions which would constitute appropriate answers (Hamblin 1973): {*Bernie at the cheese, Ernie ate the cheese, Bertie ate the cheese, Cara ate the cheese, . . .*}. The focus semantic value of the appropriate answer works out to be a subset of the question denotation: {*Bernie at the cheese, Ernie ate the cheese, Bertie ate the cheese, Cara ate the cheese, . . .*}. So an answer is congruent with a question if its focal alternatives are the same as the alternatives denoted by the question.

A (possibly) related kind of focus is so-called 'presentational' or 'informational' focus, where the part of the sentence that is considered new or important is highlighted. In (7), there is no overt question to prompt Bernie's statements. But under the Question Under Discussion (QUD) framework (Roberts 1996, Büring 2003), discussion topics are modeled as answers to implicit questions that structure discourse. As a cooperative interlocutor, Ernie (in the examples below) would accommodate implicit questions to the effect of *What did you*

[3] For instance, perhaps the pitch accent on *I* is L*+H in this cases, as opposed to H* for the previous kinds of implicatures, to use Tone and Break Indices notation. Or perhaps the boundary tone is different for this uncertainty implicature: an H% rather than an L%.

do today? and *Have you ever baked anything?* respectively.[4] This approach to informational focus allows us to retain the insight that focus indicates the presence of alternatives by explaining what is the role of alternatives in structuring discourse in these cases: they allow the hearer to identify what the QUD is, and therefore what the discourse is about and where it is going.[5]

(7) a. *Out of the blue context.*
Bernie: So, I baked a CAKE today.
Ernie: Did you?

b. *Bernie:* Yeah, I'm a pretty bad baker, but I baked a CAKE once.
Ernie: See, now that's impressive.

Answers and informational focus accompany the addition of an element out of a set of alternatives to the common ground. Another common use for focus is in corrections, as in (8), where the element that is added to the common ground competes with its alternatives, at least one of which has usually been explicitly proposed and rejected in the discourse (see also Zimmermann & Onea 2011).

(8) It's not WEDNESDAY, it's already THURSDAY!

Finally, focus is used to highlight alternatives that might be added or accommodated into the common ground, but that contrast in some way. The textbook example of this use of focus is a symmetric contrast, as in (9a-b), where focus is marked on constituents in parallel syntactic structures and of the same semantic type. In (9a), where the contrast is at the sentential level, both subject and object are focus-marked, such that the subject/object of the first sentence is juxtaposed to the subject/object of the second sentence. In (9b), the contrast is at the sub-sentential (NP or DP) level and alternatives are constructed by substituting the noun modifiers with other objects of the same type.

(9) a. ERNIE baked the CAKE and BERNIE made the FROSTING.

b. An AMERICAN farmer was talking to a CANADIAN farmer ... (Rooth 1992)

[4] The QUDs are different in these two contexts even though the focus marking is the same (*cake* is prominent), because in (7a) the focus is interpreted on the whole VP, while in (7b) it is on the object alone. This brings us to the issue of focus scope ambiguity, or Focus Projection, which will be addressed in the next section.

[5] This interpretation of (7a-b) depends on the assumption that the most prominent constituent always indicates some kind of focus. This goes against some theories of accentuation, which instead assume that there exists a default prosody that is unaffected by focus, mostly resulting from phonological constraints on rhythmical patterns. While some current studies still adopt a similar position (e.g. Zubizarreta 1998), most researchers follow the Focus-to-Accent approach (Ladd 1980, Gussenhoven 1983, Selkirk 1984), which assumes that the location of sentential prominence is always meaningful in some sense (Ladd 2008: §6.1.2).

Contrast does not have to be symmetric, of course, and the alternatives do not have to be explicitly mentioned or entailed by the discourse. In (21), where focus triggered a pragmatic implicature, we could say that the speaker establishes an implicit contrast between himself and his alternatives and it is from this contrast that the scalar implicature arises.

Other types of contrast (broad vs. narrow, identificational, subset, verum, confirmative etc.) have also been proposed, but they are not as important for understanding the *some* data in this study, so I gloss over them here, but see Krifka 2008, Zimmermann & Onea 2011, Ladd 2008: §6, a.o.

2.3 Focus and the determiner *some*

The data used in this study consists of tokens of the determiner *some* in two phrases, *some people* and *some money*, with our focus being on the determiner itself, and not on the whole phrase or on the noun. So which of these focus effects do we find in the data collected for this study?

After I carried out the coarse-grained annotation (focused / unfocused) using both the audio and the transcript, I went back to the transcripts without audio information and classified the tokens for different types of focus. On a first pass, I did this without audio so as to focus on the semantic and pragmatic properties of the contexts, instead of potentially misleading information from the acoustic signal, since focus is not the only cause for prosodic prominence. On a second pass, I compared my fine-grained (no-audio) annotation with my coarse-grained (with-audio) annotation, to see if I had missed any cases of focus. This was an important step because implicit contrast, which gives rise to implicatures, is notoriously hard to spot from a written transcript alone.[6]

One possible focus effect on *some* that is somewhat difficult to take into account is discourse novelty. Could *some* be marked for discourse novelty at all? There are two scenarios we need to consider here. The first, that the determiner itself could be discourse new and thus prosodically marked; the second, that a phrase containing the determiner (e.g. the

Intuitively, the first scenario seems wrong. Indeed, information structural annotation standards mandate that only constituents denoting discourse referents (individuals, places, times, events, situations, and even propositions) are to be annotated for information status (e.g., see Dipper, Götze & Skopeteas 2007: 150), which practically excludes function words

[6] Riester & Baumann (2013: 235) bring up the difficulties of annotating implicit contrast in a top-down fashion (without access to the acoustic signal). Other studies with an information-structural annotation component have tended to focus on various types of explicit contrast (Böhmová, Hajič, Hajičová & Hladká 2003, Zhang, Hasegawa-Johnson & Levinson 2006) or to include implicit contrast in a catch-all “other” category (Calhoun 2006). Still other studies do not discuss annotation procedures in detail (Hedberg & Sosa 2007).

like *some*. Such annotation standards are generally based on theories of information status that factor in accessibility, informativity and predictability of discourse referents to predict how likely is a word of being prominent (e.g. Grosz, Joshi & Weinstein 1995, Bell, Brenier, Gregory & Girand 2009).

A more formal theory of givenness which is based on contextual entailment is Schwarzschild 1999. Schwarzschild's (informal) definition of givenness is reproduced in (10). The existential closure of *some* (11) only requires that there is a contextual antecedent which entails that there is some entity and some property which applies to that entity. This is trivially true even in an out-of-the blue context, since Schwarzschild allows for certain backgrounded information, including presumably information about the speech act, such as the fact that there exists a speaker. So according to Schwarzschild's definition, *some* is always given.

(10) Definition of GIVEN (informal version): (Schwarzschild 1999: ex. 25)
An utterance U counts as GIVEN iff it has a salient antecedent A and
a. if U is type e, then A and U corefer;
b. otherwise: modulo ∃–type shifting, A entails the Existential F-Closure of U.

(11) *Some* is given if it is entailed by an antecedent, modulo type-shifting:

$$\begin{aligned} [\![\textit{some}]\!] &= \lambda P.\lambda Q.\exists x.P(x) \wedge Q(x) \\ &= \exists P.\exists Q.\exists x.P(x) \wedge Q(x) \qquad \textbf{∃–type shifting} \end{aligned}$$

However, this does not mean *some* could not bear focus marking, since Schwarzschild does not equate givenness with lack of focus marking. It's the other way around: he defines a highly ranked constraint which equates lack of focus marking with givenness (12), but this constraint would not penalize a given element for being focus-marked. When could a given element bear intonational focus-marking for the discourse novelty? This kind of situation can arise because "old parts can be assembled in new ways" (Schwarzschild 1999: 160). Even if everything is given at the word level, at the phrase or sentence level we can still run into constituents which are not given. These constituents are focused and this focus is expressed somewhere inside the phrase.

(12) GIVENness: A constituent that is not focus-marked is given.
AVOIDF: Do not focus-mark.
FOCUS: A Foc-marked phrase contains an accent.
HEADARG: A head is less prominent than its internal argument.
(Schwarzschild 1999: 173)

Such an example is given in (13). First note that at the word level, everything in the answer is given: the pronoun co-refers with *Bernie*, *photographed* is entailed by the question, and, as argued above, *some* is always given. Finally, *people* is given if its existential closure, $\exists x, People(x)$ is entailed by the context. Bernie and the two interlocutors verify this condition, so we conclude that *people* is also given. However, the answer *He photographed some people* without any focus marking is not given at the sentential level. To see this, note that Schwarzschild defines givenness based on entailment relations between propositions. So questions need to be type-shifted to form propositions (Schwarzschild 1999: 157); in this case, informally, the question becomes $\exists x$[Bernie photographed x in the park yesterday], which does not entail *He photographed some people.* Therefore, an answer without focus marking incurs a violation of the GIVENness constraint

(13) Q: What did Bernie photograph in the park yesterday?
A: #He photographed some people.
A': He photographed [some PEOPLE]$_F$.
A": He photographed [SOME people]$_F$.

Compare this to an answer with focus on the QP. The existential F-closure[7] of such an answer is $\exists y$[Bernie photographed d], which is entailed by the existential closure of the question. This answer keeps the GIVENness constraint happy, and thus it is optimal.[8]

This example illustrates a scenario where every word is given, but a phrase still has to be focused. However, in this study we are interested in focus on *some*, not on the phrases that the determiner may be part of. Therefore, we need to know if *some* could sound focused (in other words, be prosodically focused) not because of semantic focus on itself (since it is always given), but because of focus on a larger phrase it is a part of. Could we expect answer A" in example (13), or do we predict A'? Schwarzschild's theory predicts that in this scenario, *people* should be more prosodically prominent than some (answer A') because of the HEADARG constraint (12), which captures head-argument asymmetries that had been previously observed with focus marking (e.g. Selkirk 1984).

To conclude this discussion of informational focus, we do not expect to see focus for discourse novelty on *some*, at least according to Schwarzschild's framework and standard corpus annotation practices for information structure.

Another focus effect which happens to be missing from our data is association with focus.

[7] Schwarzschild defines the existential F-closure of an utterance U as "the result of replacing F-marked phrases with variables and existentially closing the result, modulo existential type shifting" (Schwarzschild 1999: 150).

[8] One may wonder if an answer with focus on the entire VP (*He [photographed some* PEOPLE*]$_F$.*), or indeed on the entire sentence, wouldn't also satisfy the same constraints. It would, but Schwarzschild suggests that the AVOIDF constraint would prefer for the F-marker to cover as little material as possible (Schwarzschild 1999: 169).

Of course, there is no principled reason why this might be the case. We can easily construct examples where a focus-sensitive VP-level adverb associates with a focused *some people*, so we can only assume that this is an accidental gap.

Similarly, clear cases of focus due to question-answer congruence are missing, most likely because it is rarely the case in spontaneous conversation, even in radio interviews, which are a major component of our corpus, that exchanges consist of simple, direct questions and simple, direct answers. It might also be the case that wh-questions assume that the answer-giver can be and wants to be slightly more specific and overall more informative in her answer than the indefinite NP *some people* allows. For instance, take the exchange below. The baseball player clearly wants to remain vague and uncooperative, even after a direct question is asked. He uses *some* merely to assert the existence of a set of people that he has ill feelings towards, but will not identify this set any further.

(14) [*Context: Red Sox player Joshua Beckett is holding a press conference.*]
Beckett: Oh, I'm upset with myself for the lapses in judgment but, you know, there's–there's also some–some–some ill feelings towards **SOME people**.
Interviewer: In the clubhouse, Josh? Former teammates? . . . When you say people, err. . .
Beckett: There's–there's people.
[*Back in the studio, the radio show hosts are discussing this.*]
Host 1: You can't leave the door open like that, because we're all sitting back saying, well, is he pissed at an individual player? . . . Is he pissed at the people who left?

Although I did not come across association with focus or question-answer congruence in this corpus, I did find a significant number of focused *some* in explicit contrast constructions, usually in parallel pairs such as *some people / others*, *some people / I* or *some people / some people*. Most examples of explicit contrast use symmetric configurations (15), but this is not always the case. (16a) entails two contrasting propositions: 'some people complain' and 'the speaker does not complain,' which will be added to the common ground, so I consider this a case of explicit contrast, even though the two sentences that correspond to these propositions are not syntactically parallel. (16b) clearly entails 'Rider does not leave.' Appositives such as *like some people*, though not-at-issue content, are commonly treated as a proposal to update the common ground (Murray 2014: 4; and references therein), here with the proposition 'Some people leave.' Thus, here too we have two contrasting propositions in the common ground, although syntactically one proposition is expressed in the main clause and the second in an appositive.

(15) *Focused some in explicit contrast contexts, symmetric configurations*

a. When you bring in a guy like Chad. . . High profile guy, everybody knows him and **SOME people** love him, **SOME people** hate him, big tv star. . .

b. There hasn't been a lot of bitterness. I think it's emotional for **SOME people** and then there is anticipation and excitement for others.

(16) *Focused some in explicit contrast contexts, non-symmetric configurations*

a. I'm sure there are **SOME people** that complain about the officiating. I'm not going to be that guy today.

b. The thing about Rider is that he doesn't just leave just 'cause he wants to, like **SOME people**, he actually stays and does his job.

Comparatives are also commonly associated with focus in this corpus. In some cases we again have explicit contrast and a rather symmetric configuration (17a; see also Rooth 1992). The comparative construction in (17b) is different in that the clause 'some people are saying (that was devastating to x degree)' does not contrast directly with the main clause. However, *some people* in (17b) still reads as contrastive and sounds focused. So it still seems like a contrast is established, but not a symmetric contrast. The embedded clause instead would seem to contrast implicitly with a proposition to the effect of 'The speaker says that was not devastating (to x degree),' which is certainly entailed by the context and can be assumed to be in the common ground.

(17) *Focused some in contrast contexts, comparatives*

a. I'm not negative about the team like **SOME people** are.

b. I don't think that was as devastating as **SOME people** are saying.

In some cases the context strongly suggests a contrast, but does not strictly speaking entail it. For instance, (18) seems to implicate that the speaker is not fed up with this baseball player's antics like some people are, which would be a clear contrast, but this inference is cancelable, so it must be an implicature.

(18) So if you can get him to accept a deal like that, it's a steal for the Red Sox. I know **SOME people** are fed up with some of the antics, but the production is very good.

(18′) So if you can get him to accept a deal like that, it's a steal for the Red Sox. I know **SOME people** are fed up with some of the antics, *and in fact I am too,* but the production is very good.

I was able to spot a few cases of implicatures while doing the bottom-up, transcript-only annotation, particularly where the context suggested (but did not entail) a contrast between *some people* and the discussion participants (18) or another salient referent (19). In (19), for example, the topic of the discussion is Jeff Green and his heart condition, so all of a sudden bringing up other people in the last sentence seems irrelevant, giving rise to the pragmatic inference that the point of the last sentence is to communicate that Jeff Green **is** very fortunate.

(19) Jeff Green is going to miss the entire season. They found a heart condition called an aortic aneurysm in a physical last Friday. [. . .] We certainly metaphorically thank the Lord that Green was able to have a thorough and comprehensive physical and this WAS detected. [. . .] Because I remember when you came in here and told me about it. When it first happened. You know, it's . . . like you said, **SOME people** aren't that fortunate.

However, the implicature that we most expected to see, because it is so prevalent in discussions of scalar implicature, is *some, but not all* (Horn 2004). Of course, this implicature is entailed by the contrasts described above such as *some, but not me*, but the more specific contrast often seems more salient in context. Contexts which specifically call for the *some but not all* implicature have the flavor of (20a–b), where there is a contextually salient group of people and something is consequently predicated of a subset of those. In (20a), this group of people is explicitly mentioned in the preceding sentence, where the speaker asserts that they wanted someone fired. The following sentence strengthens the statement with the scalar additive *even*, and also adds *some*, which makes *some people* sound distinctly like a subset of the overall set of people that wanted this individual fired.

In (20b), the radio show guest seems to assume that people can get the link from his tweet, and the show host corrects this assumption, pointing out that a subset of radio listeners are not watching Twitter. This suggests a possible reason for why the *some, but not all* implicature did not seem as salient as others in our corpus: *people* might be too general of a restrictor to be so explicitly given as in (20a). In cases where it is not explicitly given, it might otherwise be hard to spot unless particularly obvious, as in the case of the implicit correction in (20b).

(20) *Subset focus with implicature: some, but not all*

a. They finally found that his system worked last year in the playoffs cause people wanted him fired. **SOME people**, as you had mentioned, wanted him fired even after he won a championship. Stupid, but this year it's much different.

b. *Guest:* I just tweeted the link right now. [. . .]
Radio show host: **SOME people** aren't watching your tweet right now, Chuck. We can promote that, that's fine.

Interestingly, after I consulted both the transcript-only annotation and the audio+transcript one, a few more implicatures jumped out. These had the flavor of *some, as opposed to none / many.* In (21a), the context is such that we can be quite confident as readers/listeners that the speaker believes that a limited, but non-zero number of people could not have ratted on Terry Francona. Could we instead interpret this as a case of implicit contrast, to the effect of *some people are easy to eliminate, some aren't*? Yes, and of course this does follow from the implicature *some, but not many*, but this does not seem to be the speaker's intention here, since he is not pursuing the question of who is suspicious and who isn't. The QUD seems more specific than this, perhaps something like "Who leaked the info?", which is developed into a strategy of inquiry that we can represent with two sub-questions: "Who had access to the info?" (answered by "There is a fairly limited circle of people …") and "Is anyone highly unlikely to have leaked the info?" (answered by "…it's fairly easily to kind of eliminate SOME people"). It does not seem maximally informative to deduce from this latter answer that some, but not all people are fairly easy to eliminate. After all, someone from this set of people who had access to the information must be responsible for the deed. Instead, it makes more sense to draw the pragmatic inference that some, but not many or some, as opposed to no one is fairly easy to eliminate.

(21) *Focused some with implicature: some, as opposed to none / many*

a. There is a fairly limited circle of people who would have had access to information in Bob Hohlers piece.[9] Pretty limited and it's–it's fairly easy to kind of eliminate **SOME people** like Terry Francona's wife. Sorry, I'm not buying that Terry Franconas wife dropped the dime to Hohler, or his kids, or …

b. *Host 1:* The programming department did not choose to have John Ryder on instead of the Red Sox Game last night. I–I mean **SOME people** might have wanted that.
Host 2: Not many, if any.
Host 1: But that was not the decision that was made. John Ryder was on last night because the Red Sox were rained out and I think for the Red Sox point of view, I think this was a good thing.

(21b) has a couple of potential interpretations, depending on whether the *some people* that the first host mentions are the programming department (in which case we have a subset interpretation: some but not others, or some but not all) or someone else, such as fans. On

[9] The commentator is referring to Sports reporter Bob Hohler's article in the Boston Globe, Oct. 12, 2011, titled *Inside the collapse of the 2011 Red Sox*, `http://www.bostonglobe.com/sports/2011/10/11/red-sox-unity-dedication-dissolved-during-epic-late-season-collapse/KL4IT0morzpzJR0TsO1LsI/story.html`.

the second interpretation, it seems likely that the implicature communicated by the focused *some* is *some, but not many*, as the second host explicitly adds in the very next utterance.

Some, as opposed to none / many inferences under focus are even clearer in the case of *some money* phrases, where of course we never get definite contrasts like *some money, but not others*, while *some, but not all (the) money* seems rare enough not to have come up in our corpus. There are only a couple of examples of prominent *some* followed by *money*, and in all cases it seems to mean *some, but not much* or *some, but (maybe) not a lot by contextually given standards*, as illustrated in the example below.

(22) *Context: Sports commentators are discussing a 2008 incident where then-Boston College football coach Jeff Jagodzinski ("Jags") interviewed for a different coach position, despite being warned that he would get fired if he did so.*
Commentator 1: No matter what Jags wanted out of this deal, going in with his eyes open, the Jets or the Seahawks or anything else, did he not handle this about as sloppy as he possibly could handle?
Commentator 2: Yeah, yeah, I mean I think it could have been handled... Obviously he wanted out. Obviously he decided, this is the way to get out. 'DeFilippo threatens to fire me,' he says, 'good! I get paid...' And we'll see how much he gets paid. He'll get **SOME money**.

2.4 Same focus, different alternatives

In the previous section, I identified focus on *some* with different semantic/pragmatic effects, mainly through analyzing the context of utterance, but also supplemented with the audio stream. Broadly speaking, I discussed two classes of phenomena: definite contrast and Horn-style scalar implicatures. Definite contrast (*some as opposed to others*, *some as opposed to me*) can be explicit or implicated and is often identifiable from context, without the audio. This kind of contrast is not available for mass nouns like *money*, but is common with *people.* Although it is not clear if it would still be as common with lower cardinality count nouns (e.g. *lawyers, doctors, Nobel Prize winners*) or nouns that do not include the speaker, it is notable in this corpus, but interestingly not in the literature on pragmatic inferences with *some.*

It is Horn-style scalar implicatures (*some, as opposed to many/much, all*) that dominate conversations about the pragmatics of *some.* From an annotation perspective, these are hard, if not impossible to identify without the audio signal, unless another speaker obligingly makes them explicit, as in (21b). This is unexpected because the standard theory of scalar implicature assumes that implicatures arise simply as a result of standard Gricean reasoning, coupled with the existence of Horn scales, but without the mediation of focus (Horn 2004: 10). Scalar implicatures are believed to be more robust to contextual variation than other kinds of inferences. Still, context is acknowledged to influence whether the inference arises, although little is known about such influences beyond the sentential level, where a lot of work has been done on scalar implicatures in downward and upward entailing environments, or with quantifiers, modals and factive verbs (Zondervan 2009: 94, Zondervan 2010: 14).

This project highlights the role of focus in triggering scalar implicatures, as well as definite contrast implicatures. This raises questions about the connection between focus and implicatures, as well as about the relation between the two kinds of implicatures and what determines which kind of implicature will arise. I will address such matters briefly in this section, but will not offer a complete solution.

Rooth (1992) formalizes the relation between focus and scalar implicatures by imposing a constraint on the scale of alternative assertions that is the object of Gricean reasoning and ultimately leads to a scalar implicature. For instance, in (23a) we want the scale of alternatives to be the set of propositions {*some people agree*, *many people agree*, *all people agree*}, ordered from weakest to strongest based on entailment, such that the strongest statement entails all the others (Horn 1972) and uttering a weaker proposition implicates the negation of the stronger ones. Rooth argues that focus provides information about these alternatives. Specifically, that the set of scalar alternatives is a subset of the focus semantic value of the sentence. This effectively provides a link between focus alternatives and scalar alternatives, without collapsing the two, and can be straightforwardly applied to Horn-style implicatures with *some*, as in (23a).

(23) a. [Some]$_F$ people agree (but not many / all). **Horn-style implicature**
b. [Some]$_F$ people agree (but not me). **definite contrast**

For the implicature in (23b) to work in a similar way, we conceptualize it in terms of a different kind of scale, which capitalizes on which individual can be called on as a witness. Assume, for instance, that there are three salient referents: the speaker, Ernie and Bertie. Then we can construct a partially ordered set such as (24). Asserting that *Some people agree* implicates that a stronger assertion is false. Depending on the context, the speaker could be implicating that some other referent doesn't agree (perhaps himself or perhaps one of the other salient referents), or some group of salient referents doesn't agree (perhaps the speaker and his two friends, Ernie and Bertie).

(24)
$$\left\{\begin{array}{c} \textit{some people agree} \\ \textit{some people including speaker agree},\ \textit{some people including Ernie agree}, \ldots \\ \textit{some people incl. speaker \& Ernie agree},\ \textit{some people incl. speaker \& Bertie agree}, \ldots \\ \textit{some people including speaker \& Ernie \& Bertie agree} \end{array}\right\}$$

This set of alternatives would be drawn from the focus semantic value of the sentence, which would include all possible combinations of individuals in the model that have the property that they are people: {*some people agree*, *some people including one referent agree*, *some people including two referents agree*, …}. Intuitively, a scale will be constructed if there are salient referents whose stance on the matter could be under discussion in the discourse. This would explain why definite contrasts are not found with mass nouns like *money*: in most circumstances, there are no salient subparts. However, if there were salient subparts, this approach would allow us to obtain a definite contrast interpretation. For instance, in (25) there are four stashes of money for the burglars to find. A response with focus on *some* could imply that not all or not a lot of the money was found (a Horn-style scalar implicature), but since the interlocutors both know where the money is, it could be more specific than that: the burglars did not find the money that was hidden in the more unusual places, such as in a waterproof case inside the toilet bowl (a definite contrast).

(25) *Context: you have money hidden in four places around your house. When you get home, you realize your house was broken into and a number of valuables were taken. You call your partner immediately.*
Partner: Did they find the money?
You: They found SOME money (but not the money in the toilet bowl).

I leave up to further investigation the question of how exactly definite contrast alternatives are built. It is possible, for instance, that there are two alternative constructors for *some*, one of which substitutes other quantifiers (or possibly other determiners) and the other which takes subsets of the restrictor. This solution recalls the literature on weak quantifiers, which include *some*, as well as *few* and *many* (Milsark 1974, 1977, Diesing 1992, a.o.). This literature distinguishes between two readings of such quantifiers: a cardinal, weak reading and a proportional, strong reading. The difference in meaning is more apparent with *few* and *many* than with *some*, so I use *many* to illustrate in (26). Consider (26) in a scenario where there are 7 children in the daycare and 6 of them are playing, then reading (26a) is false, while (26b) is true.

(26) Many children are playing.

a. 'The children, who are many in number, are playing.' *cardinal reading*

b. 'Many of the children are playing.' *proportional reading*

The two readings are distinct enough that they have different distributions[10] and, as described above, can have truth conditional effects (Partee 1989). The difference between them, as many hypothesize following Milsark, is whether the restrictor set is asserted or presupposed. In case of the cardinal reading, the restrictor set is asserted, and the determiner is believed not to have quantificational strength. It is this reading that is also associated with a reduced *some*, which many researchers, Milsark included, write out as *sm*.

This is not to say that the cardinal reading of *some* cannot be accented, however, just that the quantificational reading cannot be reduced. As an example of this, consider two environments which are believed to select for only one of the readings: the post-copular position in existential sentences (27a-b), which only takes weak determiners, and the subject position of individual-level predicates (28a-b), which only takes strong determiners.

(27) *Weak determiners in existential sentences*

a. There is/are **a/sm/a few/many/three** fly/flies in my soup.

b. *There is/are **the/every/all/most** fly/flies in my soup. (Diesing 1992: §3.2.2)

c. There are SOME flies in my soup.

(28) *Strong determiners with individual-level predicates*

a. **The/every/all/most** fly/flies is/are intelligent.

b. ***A/sm/few/mny/three** fly/flies is/are intelligent.

c. SOME flies are intelligent.

The reduced *sm* (the cardinal reading, which simply asserts the restrictor set), can appear in the existential sentence environment, but not as the subject of an individual-level predicate. The question is, can an accented *some* appear in both environments, and if so, what does it mean? Milsark observes that yes, the cardinal *some* in existential sentences can be accented, in which case it means *some as opposed to none or many* (Horn-style implicature). *Some* can also appear as the subject of individual-level predicates, but in this case, according to Milsark, it is not the same *some*: it has a strong, rather than a weak reading and it carries the inference *some, but not others* (definite contrast). It is effectively equivalent to the partitive *some of the*. This strong reading of the determiner, Milsark speculates, presupposes its restrictor set and is a true quantifier, while the weak reading is a non-quantificational determiner that gets existentially bound at some point in the derivation.

[10] Witness the definiteness effect, as discussed by e.g. Milsark (1974, 1977), Diesing (1992), Herburger (2000), a.o.

This corpus study confirms Milsark's observations on the different readings of stressed *some*, but takes the different possible readings of stressed *some* to be the result of focus on the determiner, which can give rise to different types of focus alternatives, and based on these, different kinds of scalar alternatives.

3 Prosodic prominence and focus marking

In the previous sections, as in most semantic-pragmatic work on focus, I simplified the phonetic reality of focus marking in a number of ways. I assumed that all cases of focus are marked identically with some form of stress, emphasis, or prosodic prominence, which I represented using small capitals. I also assumed that there is always only one focused constituent in a sentence, and it is always the most prominent element. Lastly, I assumed that focus as an information structural device can always be identified through its realization. However, all of these assumptions are problematic for reasons which will be discussed in this section.

3.1 Focus accentuation

Intonation events are probably the most prominent aspect of focus marking in English. Intuitively, it seems like focused elements are always accompanied by a pitch accent: a specific contour in the fundamental frequency (F0) of the speaker's voice, usually a rise or a fall towards a local extremum of fundamental frequency (Bolinger 1965, Jackendoff 1972, Pierrehumbert 1980).

One of the most influential frameworks for intonation is the autosegmental-metrical (AM) theory of intonational phonology, which, as the name suggests, developed out of autosegmental phonology and metrical phonology, mostly following Janet Pierrehumbert's dissertation (Pierrehumbert 1980), as well as other foundation works such as Liberman 1975, Bruce 1977. AM phonology argues that intonation is best described by a series of high or low tonal targets which are phonologically relevant, with transitions between them that are just a matter of implementation. For instance, (29) illustrates two tonal targets: a high target H*, aligned to the constituent in focus, and a low target L%,[11] which aligns to the right edge of the utterance.

```
(29)  BERNIE liked the talk.
      H*                 L%
```

This notation follows the conventions of the Tone and Break Index (ToBI) transcription system for intonation, based on work on the intonation of English and Japanese by Pierrehumbert 1980, Beckman & Pierrehumbert 1986, Pierrehumbert & Beckman 1988, developed into an annotation system by Silverman et al. 1992, Pitrelli, Beckman & Hirschberg 1994, Brugos, Shattuck-Hufnagel & Veilleux 2006.

[11] This example annotation is somewhat simplified in that there are two kinds of low boundary tones in ToBI: L–L% and H–L%.

Like AM phonology, ToBI distinguishes between two kinds of tonal targets: pitch accents, such as the H* above, and boundary tones, such as the L% in (29). Boundary tones are marked with the percentage sign. They are associated with the periphery of a prosodic phrase and, in English, may express illocutionary status (question, statement). Pitch accents are used, among other things, for marking focus, although the relationship is not one-to-one.

ToBI describes more complex tonal events by combining H and L tones. For example, contrastive topics like *Anna* in (30a) and *Manny* in (30b) are claimed to sound 'scooped' or 'fall-rise,' and are represented by a bitonal accent, L+H* (Liberman & Pierrehumbert 1984, Steedman 2000), where the star tells us which tone aligns with the stressed syllable of the word.

(30) a. *'Background-answer (BA) contour'*

What about Anna? Who did she come with?

ANNA came with MANNY.

L+H*   H*   L%

b. *'Answer-background (AB) contour'*

What about Manny? Who came with him?

ANNA came with MANNY.

H*   L+H*   L%

The ToBI inventory of pitch accents for English includes H*, L*, L+H* and L*+H. Additionally, there are a few pitch accents with a downstepped H target: a target which has a perceptually salient lowered pitch than a previous H. Downstepped tones are marked with an exclamation point: !H*, L+!H*, L!+!H*, H+!H*. This brings us up to a total of eight pitch accent types in ToBI (for English, at least). Interestingly, though, the distribution of these different accent types in hand-annotated corpora is very uneven. Taylor 2000 notes that a full 79% of the pitch accents in the Boston University Radio News corpus were H*, and another 15% were L+H*. Similar results are reported for other corpora (Calhoun 2006: 64). Furthermore, inter-annotator agreement is reasonably high for simple pitch accent identification (81-92%), but agreement on pitch accent type is relatively low (61-72%) (Calhoun 2006: 61, and references therein). This is despite the fact that none of these corpora included unrestricted, spontaneous speech (which is more difficult to annotate, so we would expect lower inter-annotator agreement), and annotators were allowed to inspect a pitch track, as well as listen to the audio and read the transcripts.

This bears upon one of the questions we started out with: are different kinds of focus marked differently? Some previous research has suggested that contrastive focus is marked by a bitonal L+H*, while discourse novelty is marked by an H* (Pierrehumbert & Hirschberg

1990). This is the distribution that we saw in (30). The information that is a direct answer to the question, such as *Manny* in (30a), is discourse-new and is believed to be marked by an H*. However, the other argument, *Anna*, is mentioned in the question, and in such a way that we get a contrastive interpretation, so *Anna* is believed to get a L+H* accent. But data on poor inter-annotator agreement calls this clear-cut distinction into question. Furthermore, corpus studies have not found contrastive focus to be exclusively associated with L+H* (Calhoun 2006).

As discussed in the previous section, we do not expect our data to include focus marking for discourse novelty on *some*, but we do see both explicit contrast, where the contrasted elements are directly mentioned or entailed by the context, and implicit contrast, where a contrast is implicated, often resulting in a scalar implicature. So the data lends itself to a similar question: are these two kinds of focus realized differently, perhaps with different pitch accents? Due to time constraints, I do not address this here, but this is also a topic of contention. Some authors claim that different pitch accents are used for "regular focus" (roughly, explicit contrast) than for restricted contrast (roughly, focus associated with scalar implicatures) (Pierrehumbert & Hirschberg 1990, Ladd 1980). Others argue that the intonational contour does not have to be different. Instead, some kinds of focus can be *perceived* as more prominent by virtue of the pragmatic context (Krahmer & Swerts 2001) or the prosodic context (viz. post-focal deaccenting, Wagner 1999), or could be marked through extra-grammatical means, such as general emphasis.

So far, we have considered different kinds of tonal targets as markers of focus, all of which consist of a local pitch extremum (usually a maximum). However, pitch accents are not only related to focus; their distribution also depends on structural criteria and may convey other forms of intonational meaning. In the next section, I present some of these structural criteria, which will inform my interpretation of the pitch data associated with *some people* and *some money.*

3.2 Structural constraints on pitch accents

Two concepts are critical for understanding the structural distribution of pitch accents: phonological phrasing and the nuclear pitch accent. Phrasing establishes domains that restrict the application of phonological rules or constraints. These domains are built up recursively from the segmental level, producing a hierarchy of nested prosodic constituents, most prominent of which are the syllable, the foot, the prosodic word, the phonological phrase and the intonational phrase (Selkirk 1984, Nespor & Vogel 1986, Beckman & Pierrehumbert

1986, Shattuck-Hufnagel & Turk 1996).[12] The higher levels of prosodic structure are relevant to intonation in various ways. One, which I have already mentioned in the previous section, is the presence of boundary tones, which, as the name suggests, align with the edges of prosodic phrases. Another is the presence of a nuclear pitch accent (also known as phrasal stress or primary accent): the most prominent and, in English, the last pitch accent of a prosodic phrase. All phrases (both phonological and intonational) must have at least one pitch accent; if it is the only pitch accent of the phrase, then it is by default the nuclear pitch accent.

Example (31), adapted from Shattuck-Hufnagel & Turk 1996: ex. 6, illustrates that phrasing depends on the speaker to a large extent: the sentence under consideration could be parsed into one (31a), two (31b) or three (31d) prosodic phrases. While there are some constraints, stemming for instance from the syntactic structure (31c), prosodic structure does not need to follow syntactic structure entirely (31b). However, each prosodic phrase must have at least one pitch accent. If there is room for only one pitch accent, it becomes the nuclear pitch accent, such as in the phrase containing only *George* in (31d). If there are multiple items which could be pitch accented, the pitch accent is last in the phrase (31a).

(31) *What happened?*
   a. (George and Mary gave BLOOD).
   b. (George and MARY) (gave BLOOD).
   c. *(GEORGE) (and Mary gave BLOOD).
   d. (GEORGE) (and MARY) (gave BLOOD). (Shattuck-Hufnagel & Turk 1996)

Focus interacts with both of these dimensions of prosodic structure. It tends to attract nuclear prominence (e.g., Calhoun 2006: §6.2.2), as illustrated in (32) using corrective focus in various positions. In this example, I used small capitals to indicate the location of focus / nuclear prominence and acute accents to indicate optional pitch accents. Note that in pre-nuclear position we can have (optional) pitch accents on all the lexical items (32a-b). These pitch accents may express paralinguistic features such as affect or they may be inserted for purely rhythmical purposes. In terms of information structural categories, however, they may not be meaningful in any way. In post-nuclear position, we generally see no pitch movement (32b-c). This latter phenomenon is also known as post-nuclear deaccentuation and will feature prominently in debates over examples that contain two foci in the same intonational phrase, to which I will return shortly.

(32) Bernie fed the cat.

[12] A number of other prosodic constituents have been proposed, some of which correspond roughly to the ones in this list and some which occupy an intermediate or a higher level.

a. No, Bérnie féd the FISH.

b. No, Bérnie WASHED the cat.

c. No, ERNIE fed the cat.

This distribution of pitch accents around the position nuclear prominence is relevant to our investigation, since pre-nuclear phrases have a much higher chance of being pitch accented than post-nuclear phrases. Granted, if a phrase like *some people* gets a pre-nuclear pitch accent, this will probably align to the noun *people.* But this should still have effects on the determiner. Since the speaker will reach an F0 peak (assuming a high tonal target) on the first syllable of the noun, the rise towards this peak should already have started along with the determiner. This suggest that we should control for effects of prosodic context such as position in relation to the nuclear pitch accent, as we would in an experiment or a corpus-study using a ToBI-annotated corpus. The data in this study is not ToBI-annotated, however, so I will take this into account as another source of variability.

Before moving on to the effects of focus on phrasing, consider what it means for the nuclear pitch accent to be in a "default" sentence-final position, as in (31a). In early accounts of pitch accenting, this was considered the "normal stress" that is specified by rule and has no meaning or function, but can be supplanted by "contrastive stress," which has interpretive effects (Newman 1946, Chomsky & Halle 1968 and their Nuclear Stress Rule, and more recently Cinque 1993, Zubizarreta 1998). This view has been supplanted by the Focus-to-Accent approach (Schmerling 1976, Ladd 1980, Gussenhoven 1983, Selkirk 1984). Proponents of this view dissociate pitch accenting, which necessarily applies to individual words, from the semantic/pragmatic notion of focus, which may apply to entire phrases. Consequently, they recognize that a pitch accent can correspond to focus on a larger phrase ("broad focus"), such as the entire sentence in (31a) or the VP in (31b,d), both of which are discourse new and thus focus-marked. In contrast, we can have "narrow focus," such as in (32), where the focus marker spans a single word (roughly, since the determiner in (32a) could be included), to which the nuclear pitch accent is also anchored.

This perspective brought into focus a question which became one of the central issues in intonational focus research: in the event of broad focus, which element of the focused phrase anchors the pitch accent? Or from a different perspective, how do we determine the scope of semantic/pragmatic focus from surface focus marking through intonation? This is known as the "focus projection" problem. One of the most influential solutions is Selkirk 1995, which was used as a starting point by Schwarzschild (1999) in his semantic theory of givenness and the phonology-semantics interface, discussed in section 2.3.[13] Selkirk's focus projection

[13] See, however, Büring 2006, Ladd 2008, Calhoun 2010 a.o. for phonological approaches to the scope of focus marking.

principles are based on syntactic structures; for instance, she proposes that focus-marking on an internal argument licenses focus-marking on the head, and focus-marking on the head licenses focus-marking on the phrase. These observations allowed us to conclude in section 2.3 that *some* would not be intonationally focus-marked because of informational focus on a larger phrase.

Finally, focus has been argued to not only affect the position (nuclear) pitch accents, but also prosodic phrasing. For instance, Steedman (2000) and Calhoun (2006) assume that clauses[14] are divided into a theme (roughly, the topic, that which is presupposed, given or accessible from context) and a rheme (roughly the new information, that which advances the discourse), a division which is usually reflected in the prosodic phrasing of the utterance. For instance, in the second sentence in (33), *Mona* and *Geoff* are themes and the VPs are rhemes, so each of these occupies distinct prosodic phrases. At the same time, *Mona*, *Geoff*, *Paris* and *Brussels* are contrastive, but this is considered a dimension of information structure that is distinct from the theme/rheme division. This approach offers a new solution to the focus projection problem by constraining projection to the prosodic phrase of the rheme (but see Ladd 2008: 220 for criticism).

(33) Mona and Geoff met at the train station. (Calhoun 2006: ex. 2.65)
(MONA) (was going to PARIS) (and GEOFF) (was going to BRUSSELS).

3.3 Focus and non-tonal prominence

The view that focus is primarily related to pitch accents comes under scrutiny from the phenomenon second-occurrence focus: a focus which is repeated and seems to no longer be intonationally marked in its new context. In (34), for instance, the first occurrence of *graduate students* bears a nuclear pitch accent as expected, but the second one comes after a prominent corrective focus on *Petr* and does not seem to be accompanied by a pitch accent. The issue has been widely discussed by semanticists because it poses problems to theories of association with focus according to which the interpretation of *only* depends on the placement of focus, which in the case of second-occurrence focus just does not seem to be realized at all (Rooth 1992, Partee 1999, Buring 2013).

(34) A: Eva only gave xerox copies to the GRADUATE STUDENTS.
B: No, PETR only gave xerox copies to the graduate students. (Partee 1991: ex.31)

Acoustic studies confirm the intuition that second-occurrence focus is not accompanied by pitch movements, but find other acoustic markers, such as increased vowel duration, intensity,

[14] In fact, the theme/rheme division could be at various levels, from utterance to individual words.

spectral tilt, and more peripheral vowels, so overall less vowel reduction than might be expected based on the prosodic context (Rooth 1996, Bartels 2004, Beaver, Clark, Flemming & Jaeger 2007, Howell 2011). For instance, Beaver et al. (2007) ran a production study with 20 speakers who read three-sentence discourses (pseudo-randomized, with fillers) that set up a second-occurrence focus associated with a focus-sensitive adverb (*only* or *always*). Discourses were constructed in pairs, such that every two discourses shared the last sentence, which differed only in the position of focus (e.g. *Even the state prosecutor only named [Sid]$_F$ in court today* vs. *Even the state prosecutor only named Sid in [court]$_F$ today*). The authors thus compared the pairs of NPs across utterances in different conditions (in the example above, focused *Sid* versus unfocused *Sid*), as well as pairs of NPs within utterances (e.g. *Sid* with *court*). They measured acoustic correlates of pitch (F0 maxima, F0 minima, F0 mean, F0 range), as well as word duration and word RMS intensity. Findings confirmed that F0 maximum and F0 mean were not significant predictors of second-occurrence focus, but duration showed significant ($p < 0.05$) focus effects. Moreover, the authors found marginal ($p < 0.1$) effects for RMS intensity, standardized F0 range and standardized F0 minimum.

Note, however, that these effects correspond to very small differences: across utterances, average differences were 6msec (duration), 0.31dB (intensity), 4.1Hz (F0 range), -3.4Hz (F0 minimum). The differences are so small, that they may not be perceivable; the just noticeable difference is believed to be 10–40msec (Lehiste 1980) and 1–4dB (Stevens 1998). Furthermore, pitch trackers are not 100% accurate. Even if all the parameters are set optimally, pitch trackers must deal with irregular phonation at various levels, from various voice qualities such as creakiness, phrase-final effects, dialectal and speaker characteristics, to disruptions of periodicity during frication and transitions from voiced to voiceless segments. Resulting pitch tracking errors that are easily detectable, such as halving or doubling, can then be cleaned up by a post-processing algorithm, but this might also affect legitimate sharp rises and falls, as well as smooth over small variations. Additionally, not all F0 variation translates into perceivable or linguistically significant properties of the speech signal. For instance, control of the vocal folds is affected by segmental gestures, such as during the closure and release portions of obstruents or with intrinsic F0 with high vowels (Gussenhoven 2004: §1.4). Beaver et al. recognize that the size of the effects is not compelling, and carry out a perception study using recordings from their production experiment. Subjects could discriminate between prominent and non-prominent productions, but accuracy was very low: 63% on average.

In a similar fashion, Howell (2011) carried out a production/perception experiment that controls for segmental context more carefully by using homophones and a specific rhythmic pattern in target sentences, and takes additional measurements such as spectral balance. The task of Howell's experiment was more sophisticated than a discrimination task: a forced-choice context retrieval task, it asked subjects to choose the context that would be most conducive to each production of the target sentence. Howell also used informed subjects: people with training in linguistics, including trained phoneticians. Interestingly, the production study found only found a significant effect on duration, but a subsequent study put this effect down to rhythm instead of semantic structure. However, it is very possible that the results are partly due to lack of statistical power: Howell used only three speakers for the first, and two for the second experiment. Unsuprisingly, accuracy rates in the forced-choice perception experiment are lower than we find in Beaver et al. 2007: an average of 57.5%, ranging from 45% to 68%, with six subjects.

Howell's results may be explained by a lack of statistical power, especially in the case of the production experiments. This is particularly poignant in light of research showing that different speakers mark prominence in different ways: some tend to use more extreme F0 targets, some primarily manipulate duration or vowel quality, and so on (Mo 2010: §8.3.1). It is possible that work such as Beaver et al. 2007 and Howell 2011, while uncovering only small effects, is pointing us in the right direction in taking emphasis away from pitch accents as primary markers of focus, and redirecting it towards prosodic prominence as a more general notion which may materialize as pitch movements, but also segmental properties traditionally associated with (phrasal or lexical) stress, such as duration, intensity, formant characteristics.

A more large-scale investigation of acoustic markers of focus points in the same direction. Howell (2012: §3) harvested utterances of the phrase *than I did* from the web and annotated them for focus on *I* or on *did* based purely on the context. Unlike with *some people* and *some money*, the matrix clause of *than I did* makes it clear if the subject or the verb will be focused: if the subject of the matrix clause co-refers with *I* in *than I did*, then focus will be on the verb (35b); otherwise, the subject of the matrix clause contrasts with the subject of *than I did* (35a). This data collection and annotation procedure created two datasets, containing 91 and 127 tokens of the phrase respectively, extracted from natural, conversational speech (podcasts and sports commentary), with almost as many speakers as there were tokens of the phrase. Additionally, Howell ran a production experiment wherein 26 participants were recorded reading 16 constructed sentences with *than I did* which varied along two dimensions: first or second occurrence focus, and declarative or interrogative contexts.

(35) a. He stayed longer than $[\text{I}]_{\text{F}}$ did. (Howell 2012: 64)

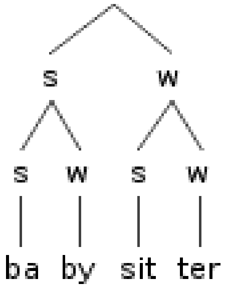
b. I should have liked that song a lot more than I $[\text{did}]_{\text{F}}$.

Howell trained a number of different models of the data using two kinds of machine algorithms (Support Vector Machines and Linear Discriminant Analysis) and various groups of acoustic features, picked either manually or using feature selection algorithms. The models were trained either on a web-harvested dataset, or on laboratory data, and were then tested for accuracy on new data, either web-harvested or lab-recorded. Best accuracy rates ranged from 83.5% for web-trained, lab-tested models to 92.9% for web-trained, web-tested models; these numbers are quite high, suggesting that focus is marked quite consistently even though in this case it is clear from the context, as illustrated in (35). No analysis of the errors was performed, so we do not know what might be possible causes of misclassified tokens. Interestingly, F0-related features turned out not to be good predictors of focus status; instead, non-intonational measures such as duration and formant measurements were robust predictors of focus. Howell does not discuss what kind of prosodic contexts the phrase was most likely to be found in; most importantly, we do not know how the phrase tends to be positioned relative to the nuclear pitch accent. But it is clear from these classification results that whatever role pitch accents play in focus identification, they are not the only or possibly even the most important / robust cue, at least for the kinds of overt, explicit contrasts that seem to form the bulk of Howell's *than I did* data.

Additionally, Howell ran two perception experiment: one where 40 subjects listened to a portion of the web-harvested tokens (without context) and chose the most prominent word, and a second one where another 41 listeners heard a portion of the lab-recorded tokens and performed the same task. Accuracy rates are on average reasonably high for human subjects, although somewhat lower than automatic classifiers, with 71.2% mean accuracy for lab-recorded second-occurrence focus data (higher than Beaver et al.'s 63% accuracy but lower than the 81% accuracy of Howell's top performing classifier) to 85.9% accuracy for web-harvested data (compared to 92.9% by the top performing classifier). Listeners differed greatly in their accuracy rates, however, ranging from 34.4% to 93.8% for lab-recorded second-occurrence focus. It could be that high accuracy listeners are able to extract more information from the acoustic signal: perhaps they are more sensitive to the prominence markers that happen to be more robust predictors for this data, or perhaps they are more flexible at recognizing different ways of marking prominence. Howell (2012: 128) also offers other potential explanations, pointing out that the task or the stimuli may be too different from speech production and understanding in natural conditions.

The experiments discussed in this section indicate a shift in perspective from intonational correlates of focus (pitch accents) to more complex bundles of cues related more generally to prosodic prominence. Non-focus related perceptual studies on prosodic prominence also consistently point towards prominence as a package of acoustic features that can vary from speaker to speaker, listener to listener, and context to context (Terken & Hermes 2000, Kochanski, Grabe, Coleman & Rosner 2005, Mo 2010). Supra-segmental phonology offers a theoretical approach to this new perspective: prominence (alongside prosodic phrasing) is a more fundamental concept in the Autosegmental-Metrical (AM) framework than intonation (Ladd 2008: §2). As in its precursor, metrical phonology (Liberman 1975, Liberman & Prince 1977), prominence in the AM framework is a structural property defined on binary prosodic trees that are built on top of syllables. In any prosodic subtree, one node is stronger than the other; the strongest syllable at various levels of prosodic structure then anchors various prosodic markers. The strongest syllable at the word level, for instance, is the syllable that is only dominated by strong nodes and is perceptually the bearer of lexical stress (such as *ba-* in the example below). At the phrase level, the strongest syllable anchors the nuclear pitch accent. Other strong syllables may anchor other pitch accents, subject to constraints mentioned above (the tune-text association).

(36) Metrical representation of a compound, from Ladd 2008: 56.



Under this view, then, pitch accents are simply cues to prominence, or more precisely to the prosodic structure of an utterance, alongside duration, intensity, the quality of the vowel and so on. Semantic focus tends to attract the highest prominence in a prosodic phrase and may do so by manipulating phrasing or strength relations between nodes in prosodic trees. However, focus (and information structural constraints in general) are not the only constraints on prosodic structure: there are also phonological (e.g. stress clash avoidance, rhythm), syntactic and paralinguistic factors influencing the suprasegmental organization of utterances.[15]

This paper carries out a corpus study on the relation between focus and prosodic structure which controls for only some of the other factors influencing supra-segmental structure. For

[15] This approach to the phonology of focus closely follows Calhoun's (2006) work on AM phonology using probabilistic modeling on corpus data.

instance, we control for some syntactic and phonological factors by selecting a specific target, the quantificational determiner *some*, in two fixed phrases: *some people* and *some money*, where the two following nouns have similar syllabic and rhythmic structure, and together give us a sizable collection of tokens of both focused and unfocused *some*'s. However, the strength of a corpus study lies precisely in the wide range of data that it presents for analysis, a topic which I develop in greater detail in the following section.

4 Corpus study

4.1 Data collection

Data for this study was from freely available audio content on the Internet, mostly from radio shows and podcasts hosted on websites which provide transcripts, allow keyword searches and return not only a link to the relevant audio content, but also the time index of the search terms. This allows researchers to cut the audio file, which can often represent several hours' worth of recording, to a more more manageable and useful size. The automated harvesting procedure is described in detail in Howell (2012) and in Rooth, Howell & Wagner (2013).

Following collection, the data was entered into a web-based annotation app, ezra (Lutz, Cadwallader & Rooth 2013), which we used to clean the data, identifying false hits, fixing the transcript where necessary and adjusting the time window of interest. Since most of the transcripts were produced by Automatic Speech Recognizer algorithms, they were not very accurate: about 20% of the search results did not actually contain the search terms. Many of these problems were due to the nature of the recordings, almost all of which were natural, spontaneous conversations and included speaker overlaps, a significant amount of background noise, poor quality recordings (sometimes from phone interviews or on-the-scene reporting at sporting events), mumbling, dialects and disfluency. Table 1 provides the exact number of data points at each step of the analysis.

The corrected transcripts were used to produce automatic segmentation of the data using the ProsodyLab-Aligner (Gorman, Howell & Wagner 2011), an HTK-based aligner which outputs Praat-style Textgrids. The Textgrids were checked manually and about 35% of tokens were discarded due to segmentation errors, especially from the *some people* context (see Table 1). Segmentation errors occurred for the same reasons as described above for automatic speech recognition: multiple speakers overlapping, background noises, mumbling and so on. The first of these problems could be avoided in the future if speaker diarization is applied to the dataset, so that the ASR algorithm and the aligner can be run on distinct channels, representing only the speech of a single speaker.

	Collected	Contain search terms	Alignment ok	Measurements ok
some people	448	358	204	197
some money	352	264	202	198

Table 1: Number of data points at each step of the data analysis.

The aligner also had problems with hesitations and repetitions when these were not represented accurately in the transcript. Additionally, the phrase *some money* was interpreted as always having two [m] segments, which sometimes resulted in the nasal being split into two segments based on some property of the signal, and other times it affected the accurate segmentation of surrounding phones. A similar problem affected unfocused *some*'s, which were often reduced to the surface form [sm]. In such cases, the aligner still erroneously tries to find a vowel segment, producing a misaligned output. These problems can be resolved by fixing the problematic transcripts and adding a new pronunciation for *some* in the aligner dictionary. However, for now, I simply discarded these tokens. This gets rid of a significant portion of the data, but it should not have a large affect on the analysis because the remaining sample is still representative of unfocused realizations: reduced *some* is still represented, as long at least 5msec or so of a vowel-like segment is present.

From the remaining data I extracted several measures of interest, which I describe in detail in the next section. In a few cases (3% of the data) the pitch trackers failed extract a usable F0 contour from the target word. These cases were also excluded from the analysis.

4.2 Annotation

I hand labeled the remainder of the data as focused or unfocused. For the annotation process, I had access to the sound file, the transcript, the waveform and spectrogram, and a rough pitch track, but I primarily based my decisions on the first two. I discussed different kinds of focus phenomena I came across in §2.3. Tokens of *some money* presented no annotation difficulties; most of them were clearly unfocused, although I did come across a few which were clearly accented and which gave rise to implicatures, as mentioned at the end of §2.3. Tokens of *some people* were more difficult overall, especially where the audio signal and the context could support an implicature, but it was unclear if the speaker intended it. In many such cases, I consulted native speakers to confirm my judgment, but there is still some amount of subjectivity in the annotation. Future work should quantify this uncertainty, for instance by assigning the annotation task to a team of annotators and calculating inter-annotator agreement.

4.3 Measures

The data analysis was conducted in Matlab using scripts written by Mats Rooth, Sam Tilsen and myself. I collected 11 measures, which I describe below.

duration - the duration of V1 (the *some* vowel) and V2 (the stressed vowel in *people/money*)

intensity - the root mean square amplitude of V1 and V2

formants - F1, F2 and F3 for V1 and V2. Formants were extracted using a Matlab script from Sam Tilsen, implementing a Linear Predictive Coding algorithm.

F0 level - minimum, maximum, range for *some* and for the first syllable of *people/money*, as well as the rise from the previous word to the maximum point in *some*. Two pitch trackers were used to obtain these measurements: the Praat pitch tracker (using autocorrelation; see Boersma 1993) and `fxrapt` from the `Voicebox` toolbox for Matlab (using normalized cross correlation; see Talkin 1995), but the Praat track was used for analysis because it performed optimally.

F0 extremum alignment - the alignment of the F0 peak/valley within *some* and *people/money*, represented as percentage of the duration of the coda or the stressed vowel respectively.

Duration measurements are distributed in 10msec bins due to the resolution of the automatic aligner.

Formants were difficult to extract in some cases for the *some* vowel. I considered anything outside the range [200,650] for F1 and [1000,1600] for F2 to be an extreme or unexpected value. I compared these to Praat formant readings and fixed those that seemed erroneously extreme by manipulating the formant tracker parameters in Matlab. Most of the problematic cases were due to some form of formant merger and could be fixed by changing the expected number of formants and choosing an optimal window duration.

Reliable F0 values were the most difficult to extract because of the quality of the recordings and the speech style. I initially experimented with `fxrapt` alone, which takes 21 customizable parameters, and a number of post-processing scripts written in Matlab by Sam Tilsen and McKee. To find the most reliable set of parameters, I selected a random test sample of 30 *some people* and 30 *some money* and plotted the pitch tracks for various combinations of the parameters that I judged to be most important: `vtranc, doubled,`

	Focused	Unfocused	Total
some people	143	54	197
some money	4	194	198
Total	147	248	395

Table 2: Distribution of focused and unfocused *some* in the dataset.

`freqwt, absnoise`.[16] Since all speakers were male, I restricted the pitch range to [75,375], which removed many (but not all) pitch halving errors.

Even with an optimal set of parameters, the resulting pitch tracks still contained a significant number of doubling and halving errors, most of which were cleaned up in a post-processing stage which identified extreme values, interpolated to replace them with a value more similar to neighboring frames, and smoothed the final measurements. The challenge in this stage was choosing a value for parameters regulating the largest F0 jump that should be allowed to survive in the pitch track, as the pitch tracks also contained a fair amount of legitimate sharp rises and falls.

After choosing a set of parameters for `fxrapt` and the contour post-processing algorithms, I also ran a script to collect pitch data from Praat, which I then imported into Matlab using a script by Sam Tilsen and Christina Bjorndahl. I plotted the Praat pitch track side by side with the `fxrapt` pitch track for all of the data and found that the Praat pitch tracker was more robust than the Matlab one in the majority of cases, with fewer doubling and halving errors[17] and more overall data points collected.

The data was normalized for machine learning analyses. Since I did not have speaker information or annotated utterance boundaries, I performed a z-transform across all speakers.

4.4 Analysis

Table 2 gives the break-down of the data in terms of focus labels. There are many more unfocused than focused tokens with *some money*, which is as expected (§2.4). There would be more unfocused *some people* if we had not discarded highly reduced *some*'s which gave rise to gross alignment errors.

[16] See the `fxrapt` documentation for an explanation of the parameters: `http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/doc/voicebox/fxrapt.html`.

[17] Jesus & Jackson (2008) compare the performance of eight open-source pitch tracking algorithms, including Praat and `fxrapt`, on a collection of (non-spontaneous) British English and Brazilian Portuguese utterances. They find the Praat autocorrelation algorithm to be the most accurate for F0 measurements. However, they do not provide any parameter settings for `fxrapt`, they do not mention any post-processing, and their data is non-spontaneous and most likely has less background noise and recording quality issues.

Figure 1 illustrates the distribution of segmental measurements on the [ʌ] vowel in *some* and (for duration and intensity) the stressed vowel in *people/money*. As expected from previous studies on non-intonational markers of focus (§3.3, the first boxplot shows a clear difference in the duration of [ʌ] based on the two conditions, where focused [ʌ] is on average much longer and the right tail of the distribution extends to values greater than 100msec. A smaller, but still difference can be seen in the V1 intensity boxplot: the distribution of focused [ʌ] significantly overlaps that of unfocused [ʌ], but it still centers around a noticeably higher average. On the other hand, the stressed vowel in the following noun looks just slightly lower in intensity in the focused condition, suggesting a shift in metrical prominence from the noun to the determiner. The duration of the vowel in the noun is, however, not noticeably different in the two conditions.

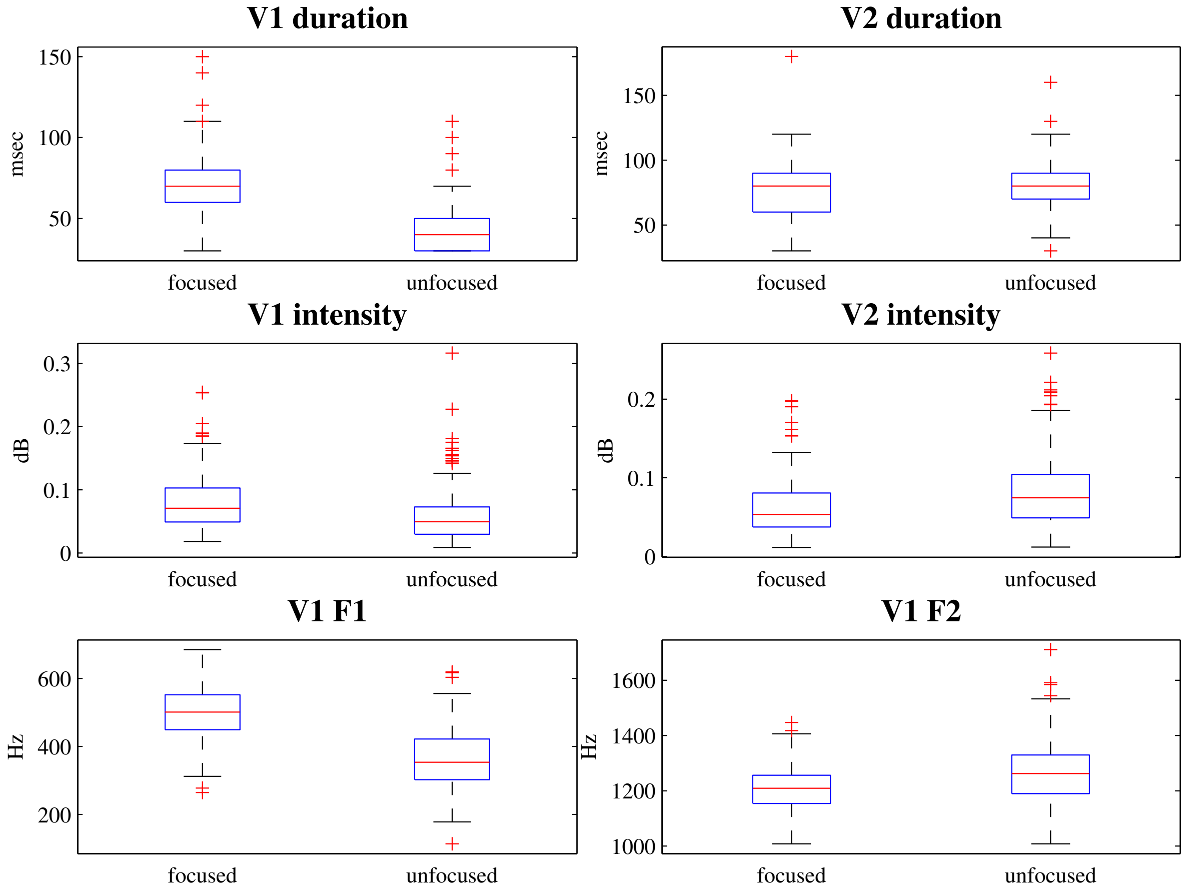


Figure 1: Raw segmental measurements by focus annotation for V1 (*some* vowel) and V2 (*people/money* vowel), across speakers.

Clear differences can also be observed for the vowel quality of the *some* vowel. First note that when *some* is not focused, the quality of the vowel is much more variable (the variance is much larger for both F1 and F2). This is consistent with a metrically less prominent unfocused *some* that is subject to target undershoot and is more variable. However, it could also be that we simply have more data for unfocused *some*, and the actual trend is not as strong as this plot suggests. Additionally, we observe a noticeably lower F1 and a higher F2 for unfocused [ʌ], suggesting a more centralized vowel, as (perhaps more clearly) illustrated in Figure 2. The vowel space depicted here again reveals some overlap in vowel quality, but more centralized and spread-out values for unfocused [ʌ].

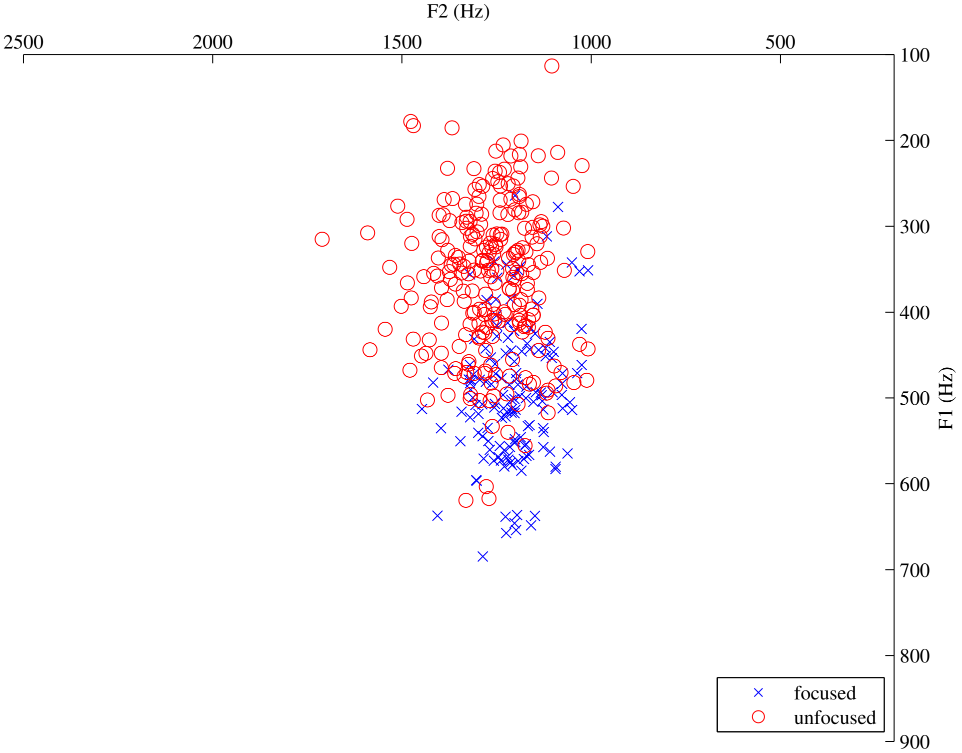


Figure 2: Raw formant measurements by focus annotation for V1 (*some* vowel), across speakers.

The *some money* context allows us to make another interesting comparison in terms of vowel quality: the stressed [ʌ] in *money* (in a variety of prosodic contexts: pitch accented, unaccented, pre-nuclear, post-nuclear etc.), the unfocused [ʌ] in *some* (unstressed, unaccented), and the focused [ʌ] in *some* (generally nuclear pitch accented). AM theory and Calhoun's (2006) probabilistic model of prosodic structure lead us to expect the noun to be more prominent overall, all things considered, because it is a lexical word. And this is what we observe: the *money* vowel is less centralized than the *some* money, despite the wide range of [ʌ] productions. However, focused [ʌ] in *some* is more similar to [ʌ] in the noun than to the bulk of *some* productions, suggesting that (if vowel quality is a correlate of phonological prominence), focus is a strong attractor for prominence.

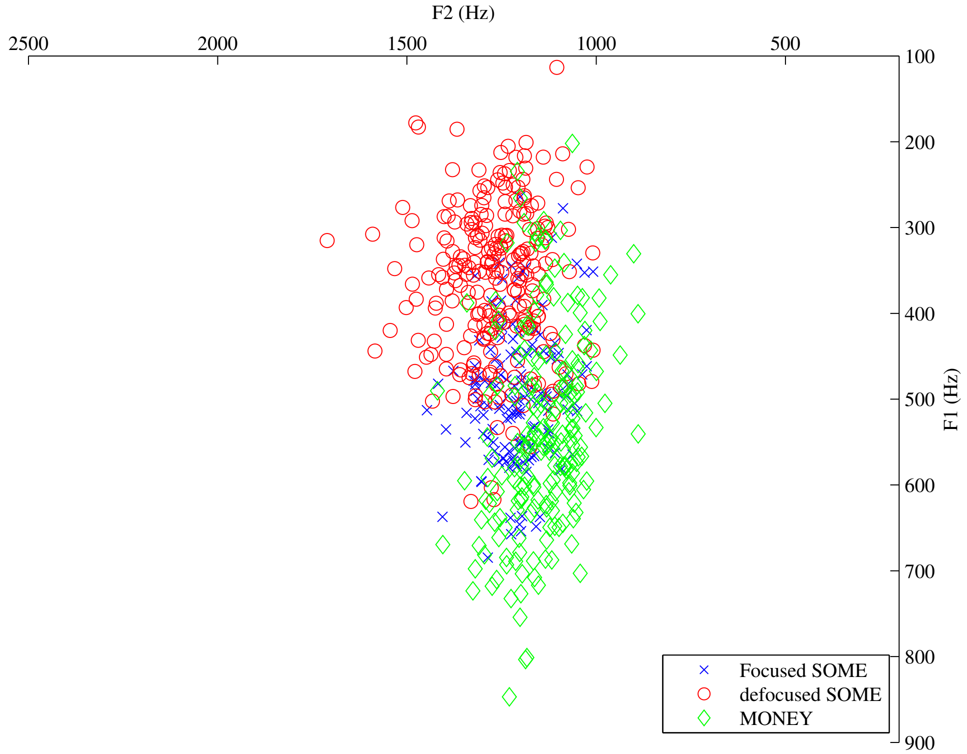


Figure 3: Raw formant measurements by focus annotation for V1 (*some* vowel) and V2 (*money* vowel), across speakers.

So far, we have seen that segmental measures seem strongly correlated with focus, a relationship which we assumed to be mediated by phonological prominence, as per the Autosegmental-Metrical framework. However, as discussed in §2, most focus studies highlight the role of pitch accents in signaling focus. Figure 4 represents the distribution of F0 measurements from the *some* coda, unnormalized, across all speakers. As expected, the first boxplot reveals higher F0 maxima for focused *some*s on average. Although unfocused *some* can have high F0 values, depending on context, speaker's characteristic range and extralinguistic goals (such as expressing affect), and so on, the right tail of the focus distribution is certainly thicker. On the other hand, the left tail seems quite comparable; in such cases, syntagmatic comparisons (within-utterance) would probably be more telling than paradigmatic comparisons (across-utterances). The second boxplot illustrates such a comparison: the ratio of V1 F0 maxima to V2 F0 maxima. This comparison confirms the trend in the first boxplot. As expected we see that a larger portion of the unfocused *some*'s have F0 maxima that are smaller than the F0 maxima of their following nouns. However, ratio values also extend in the other direction for unfocused *some*, which is however not surprising, given that basic declination can leave *some* with a higher pitch than its following noun in the absence of pitch accents. Declination makes it difficult to rely on F0 measurements alone as correlates of pitch, which is of course a psychoacoustic measure. The analysis would probably be more accurate if we could manually or automatically annotate the corpus with pitch accent information, or in some way separate the effects of downtrend from the effects of relative accentuation.

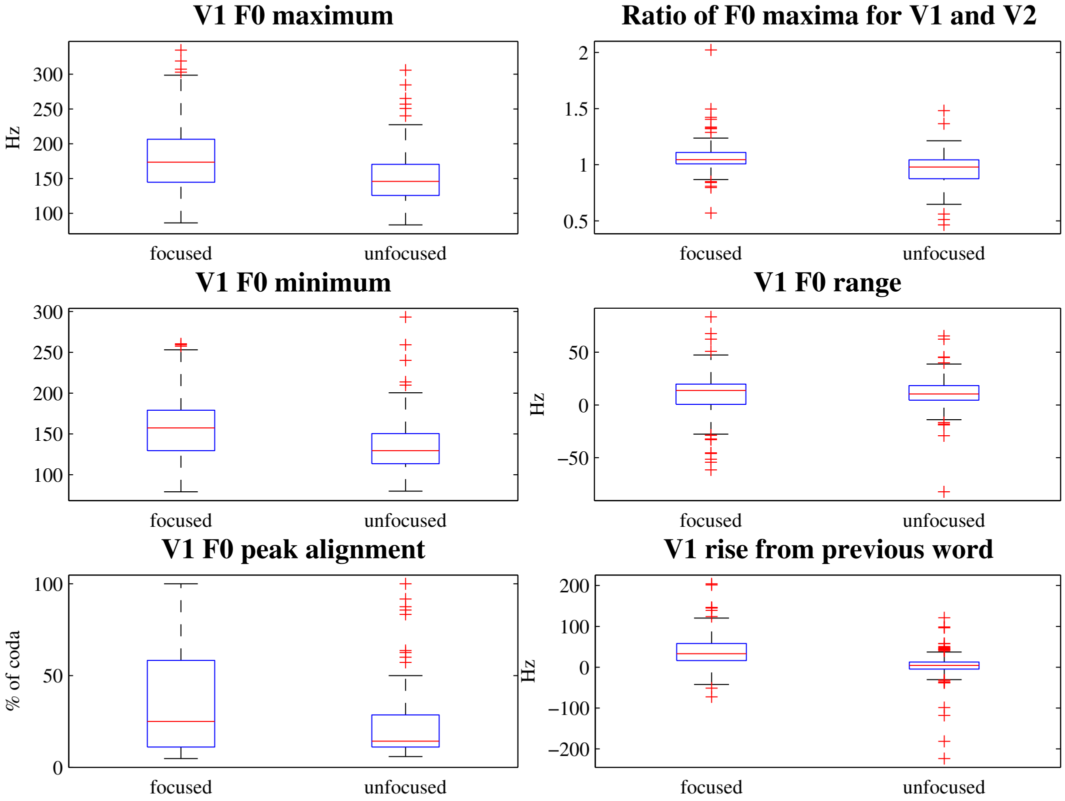


Figure 4: Raw intonational measurements by focus annotation for V1M (*some* coda), across speakers.

Another interesting intonational measure which seems to correlate with focus is the alignment of the F0 peak, measured as the distance from the onset of the *some* coda, normalized by coda duration. Thus, unfocused *some*'s will have their highest F0 measurements towards the beginning of the coda because of downtrend. A significant number of outliers align the peak with the end of the coda, probably because the speaker is preparing for a high F0 target due to a pitch accent on the following noun. On the other hand, the distribution of focused *some* alignment data is heavily skewed to the right and overall more variable in terms of F0 peak location. This suggests that peak alignment could be better cue to the presence of a pitch accent on *some* than simply the maximum F0 over that word.

Finally, I calculated the rise in F0 from the preceding word to the maximum F0 of *some*. As expected, it is much more common to see a rise in focused tokens, and a fall in unfocused tokens. But this measure is again difficult to interpret without any information on prosodic structure, since we have no way of knowing if *some* is starting a prosodic phrase, in which case it follows an F0 reset or perhaps even a different speaker, with a different F0 range.

Looking at each dimension of the data individually allows us to see that, despite the great number of factors that could be influencing prosodic structure in this dataset, focus still seems to come out as a strong predictor of prominence. To quantify this, I conducted a number of machine learning experiments in Matlab. Machine learning is more suitable for analyzing this dataset than traditional statistical techniques because it is more robust in the context of small and noisy datasets.

Unsupervised learning algorithms are commonly used for exploratory data analysis, since they do not need to be trained on gold standard labels. I use two of these in the sections below to answer questions such as how much structure there is in the data and how well different combinations of acoustic features can predict semantic/pragmatic focus labels. Supervised learning algorithms, on the other hand, need to be trained on a set of measures associated with the correct label that the algorithm must learn to predict. Based on this training set, the algorithm outputs a classifier which can be used to predict the label of any other data point based on the given measures.

4.5 Unsupervised learning: k-Means clustering

As their name suggests, clustering techniques in machine learning are used to identify groups of observations in a dataset. Thus, they can be used to look for patterns or for structure in the data. The k-means clustering algorithm takes in any number of measures for each data point (n) in the dataset and the number of groups (k) that the data should be partitioned into. It returns a group index for each data point, such that items within each group are maximally similar to each other and maximally different from items in another group.

The k-means learning algorithm treats each observation as an object in n-dimensional space, where n is the number of measures associated with each data point. It starts by taking k points at random in n-dimensional space that represent the starting centroids of the k groups.[18] It calculates the distance from each data point to these k centroids[19] and

[18] The number k is provided by the user or is determined through experimentation. The k starting centroids can be provided by the user, but are usually k data points chosen at random from the dataset

[19] The distance measure can be configured by the user. I used the default squared Euclidian distance.

All	V1 features	V1 no F0	V1 no F0min	V1 no F0min/max	V1 no F0rise/range
81%	83%	88%	79%	80%	81%
64-99	70-92	80-95	79-93	70-93	71-92

Table 3: Mean and range of accuracy figures of `kmeans` clustering for different combinations of features.

it associates each data point with the closest of the k centroids. This results in k groups. The algorithm calculates the means of each group and moves the centroid to the location of these means. The entire process is then repeated until the distance between group members and their centroid is minimized.

I used the `kmeans` function from Matlab's Statistics toolbox, version 8.2. I normalized the measure vectors by z-transforming the data across all speakers, since Euclidian distance is sensitive to scale changes across different types of measurements. I ran `kmeans` with $k = 2$ groups in order to find how well the algorithm can learn the distinction between focused and unfocused *some*'s with no access to user annotation, from the acoustic data alone, and for different kinds of acoustic measurements (features). I calculated the accuracy of the classification as percentage of tokens correctly classified, using my annotation as gold standard. Finally, I repeated the classification several times because the final distribution of clusters depends on the initial conditions (the randomly chosen centroids) and I took the mean accuracy over these repetitions as representative of the algorithm's overall success.

Table 3 gives the mean and range of accuracy figures for clusters built using different sets of acoustic features, with no access to focus annotation. We first note that classification accuracy is overall quite high; the numbers are in the same range as Howell's (2012) top performing classifier trained with manually-created focus labels. However, it is likely that *some* can be and is reduced much more when unfocused than Howell's focus items (*I* and *did*), since [sm] is still phonotactically licit. On the other hand, it is also not the case that all unfocused *some*'s were completely reduced to [sm], so the high accuracy rates are still surprising. The next interesting step would be to improve on the gold standard annotation by using a team of annotators. This would give us more information about how confident we can be in the focus labels and how much inherent ambiguity there is in the signal (versus classifier error).

In terms of the performance of different acoustic features as predictors of focus, note that segmental acoustic features (without F0 information; column 3) produce more accurate classifications (highest mean) and more robust classifications (highest range). On the other hand, all acoustic features taken together (both segmental and intonational, for *some* and the following noun), have produced the highest overall accuracy (99%), but also the lowest (65%). Thus, the predictive power of this set of features is less robust: it depends more on learning circumstances, such as initial conditions. This observation tends to generalize to other machine learning tasks: more features are not necessarily better. In this case, too many allow too much leeway in the kinds of patterns that the classifier can learn. In general, too many features can overfit the dataset under observation and thus not extend to unseen datasets, which is the whole point of the endeavor.

4.6 Unsupervised learning: principal component analysis

In §3.3 we observed that prominence is cued by a collection of acoustic markers, including pitch, duration, intensity, and vowel quality. However in §4.4 we could not graphically represent how all these features combined structure the data points into a focused and an unfocused group. We only looked at the effect of individual features using boxplots, and two features combined (F1 and F2) using a scatterplot. In the previous section, we used a clustering algorithm to reveal groups in the data, but we could only indirectly probe the effects of different combinations of features.

Principal component analysis (PCA) is a unsupervised machine learning algorithm which is especially useful for reducing the dimensionality of complex datasets, which allows us to better visualize and understand how each feature contributes to explaining the data. Like k-means classification, PCA considers each data point to be a point in n-dimensional space, where n is the number of features (here, acoustic measures). The goal of PCA is to find groups of features which are similar enough that we can collapse them into a single, complex, compound feature that structures the data in the same way. We can then plot the data points in terms of these new complex features and we can determine how each of the original measures contributes to these new complex measures. Often, the most important two or three measurements capture enough of the variation in the dataset that plotting them provides a fair representation of the data in two/three, rather than the original n dimensions.

I used the `pca` function from Matlab's Statistics toolbox, version 8.2. I normalized the measure vectors by z-transforming the data across all speakers. As with `kmeans`, I ran the algorithm on various combinations of features to try to capture as much of the variance of the data in a few components for better visualization.

In the first experiment, I used all 21 features described in §4.3. Table 5 shows the partial outcome of this experiment: the six most important principal components (PCs) and the percent of total variance explained by each of them. The `pca` function always sorts components in order of their explanatory power. We can then plot up to the first

Principal components	Percent variance explained
PC1	25.6%
PC2	14.7%
PC3	10.5%
PC4	8.1%
PC5	6.8%
PC6	6.3%
Total	72%

Table 4: Percent of variance explained by first 6 principal components calculated as combinations of 21 acoustic features.

three principal components in a biplot, which represents all the data points and the original features in relation to these new dimensions.

This biplot is shown in Figure 5. Since the first two components account for just slightly over 40% of the data, the biplot should not be considered a good representation of the data. Still, even with this caveat, we can see that the data is fairly well clustered such that most focused tokens are in quadrants one and four. The original 21 acoustic measures are represented in the labeled vectors, such that: A. the cosine of the angle between a vector and an axis indicates how much the feature contributes to the principal component, and B. the cosine of the angle between two vectors indicates how correlated the two features are, with highly correlated features pointing in the same direction.

Thus, we note that many of the original features were highly correlated. These are pairs such as F0 maxima and F0 minima on V1 (the *some* vowel) and the same for V2, the duration of V1 and its first two formants, the alignment of F0 peaks on V1 and the height of F0 rise from the previous word to the peak on V1 etc. However, there are also pairs that were not as predictable, such as the relation between segmental measures like F1 and duration on the one hand, and intonational measures like the height of the F0 rise and the alignment of the F0 peak. Such relationships can provide a basis for trimming the set of features even further in the hopes of capturing more of the variance of the data in the first few components for better visualization.

Some of the original features, such as the duration of V2 and the third formant of V1, do not contribute much analysis, at least for the first two PCs. The most important features for the first PC are, in order: F0 extrema for V1, V1 duration, size of V1 rise, and V1 first formant. Interestingly, the second formant of V2 also makes a significant positive contribution, slightly more significant than V1 intensity and V1 peak alignment. However,

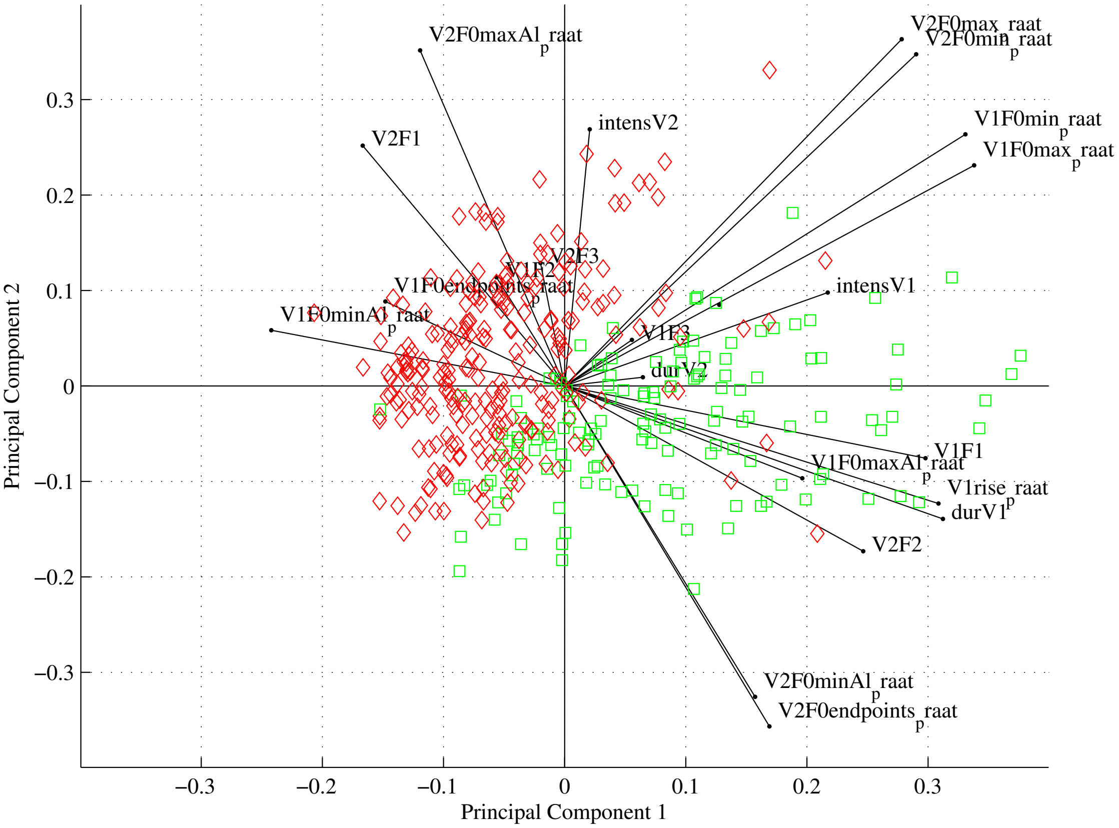


Figure 5: Biplot of first two principal components, based on 21 acoustic features. The original acoustic features are represented as vectors. The data points have been projected onto the PC planes. Green squares represent focused tokens and red diamonds unfocused tokens.

this is simply because most focused tokens come from the *some people* context, and the [i] in *people* has a higher F2 than the [ʌ] of *money.* Vectors pointing towards the negative side of the PC1 axis (the horizontal axis) make a negative contribution towards PC1. For instance, the higher the F0 range, the more likely a token is to be classified as unfocused, most likely because F0 continues to drop due to downtrend, whereas if *some* is pitch accented for focus, downtrend will temporarily be reversed.

The second PCA experiment used only V1 features, mostly segmental features and two F0 measures. Note that the top six principal components now explain almost all of the variance in the data, and the top two PCs explain 53.6%, a significantly larger portion than in the previous experiment, making the biplot a fair (though still not good) representation of the data. This is not surprising, given that the data has been significantly reduced, down to 7 sets of measurements from 21.

Principal components	Percent variance explained
PC1	36.3%
PC2	17.3%
PC3	13.8%
PC4	12%
PC5	8.4%
PC6	8.1%
Total	96%

Table 5: Percent of variance explained by first 6 principal components calculated as combinations of 7 acoustic features.

However, the biplot (Figure 6) still shows relatively good separation of the data, mostly based on duration, V1 first formant, intensity and F0 peak height/alignment on *some*. F2 contributes a small negative component, but is mostly important alongside F3 for the second component. It is unclear what kind of separation is created alongside the second dimension. Tokens with high F2 and F3 (and to some extent high F0 peaks) are distinguished from tokens without, but why this might the case is unclear. The distinction does not have to do with which noun follows *some*, so it remains a mystery for now.

To conclude, the PCA experiments carried out here suggest that segmental features, particularly duration, vowel height and intensity, are relatively robust predictors of focus, alongside at least one F0 measure: the height of the F0 peak.

4.7 Supervised learning: linear discriminant analysis

For the supervised learning, the dataset must be divided into a training set and a test set (I used a common 80-20 ratio). Based on the training set, the learning algorithm builds a classifier which learns the best combination of features that produces the desired labels. The classifier is then used on the test set and performance measures are calculated based on how the classifier's predictions compare with the gold standard. Feature engineering is just as important (if not more) in this kind of learning as in unsupervised learning. To arrive to the best combination of features, researchers reserve another portion of the training set for validation. The classifier that performs best on the validation set is selected as optimal, and its performance on the test set is reported as the final measure. The test set is thus reserved until the last moment to prevent the researcher from building a model which overfits the data and thus has inflated performance measures on the test set.

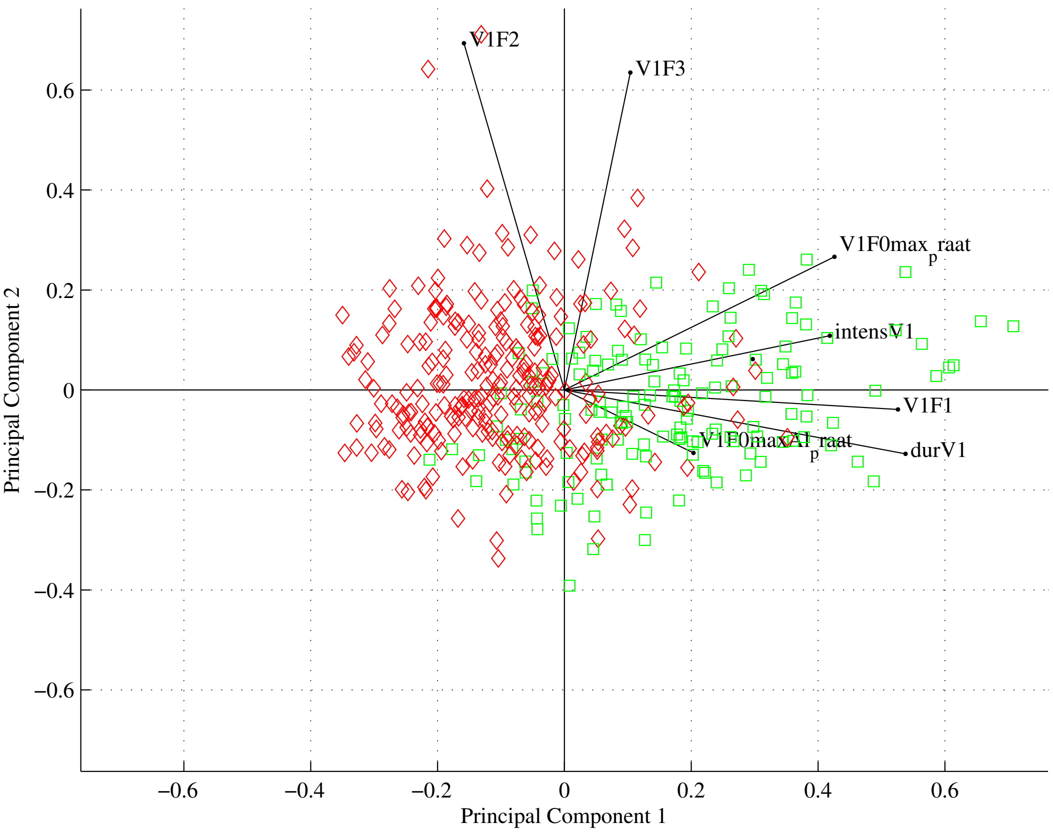


Figure 6: Biplot of first two principal components, based on 7 acoustic features. Green squares represent focused tokens and red diamonds unfocused tokens.

I used the Matlab function `cvpartition` to reserve 20% of the data for testing, and 20% of the remainder for validation. I then trained a Linear Discriminant Analysis classifier on the training set using as predictors the set of features from Figure 6 (V1 duration, intensity, F1, F2, F3, F0 peak height and F0 peak alignment). Linear Discriminant Analysis is somewhat similar to k-means clustering. It attempts to define a decision boundary in terms of the given features to divide the data into as many classes as there are labels. The decision boundary is set such that within-class distance is minimized and between-class distance is maximized, as with k-means clustering.

LDA is implemented by Matlab's `ClassificationDiscriminant.fit`, from the Statistics toolbox version 8.2, which I also use here. The algorithm returns a confusion matrix, which I reproduce in Table 6. Most of the classified tokens in the test set are represented in the main diagonal, which corresponds to correctly predicted labels. A total of 4 tokens were wrongly predicted as not being focused, and 2 respectively of being unfocused. This

	Predicted focus	Predicted no-focus
Actual focus	19	4
Actual no-focus	2	38

Table 6: Confusion matrix for LDA predictions based on 7 acoustic features.

corresponds to an accuracy rate of 99%. This is higher than the accuracy rate we saw with k-means clustering, but of course this is not surprising, since now we are telling the algorithm what kind of clusters we want. It is possible that the high accuracy rate is also due to the highly constrained nature of the dataset, but note also that while we have a limited number of tokens which are restricted to *some people* and *some money*, we also have high degrees of background noise, low quality recordings, and almost exclusively spontaneous speech from diverse speakers. Additionally, we had no information about prosodic structure or speaker identity, so we could not perform the most ideal kinds of normalization or create robust syntagmatic measures.

Future research could re-run these machine learning experiments on a larger dataset, perhaps after bringing in the many data points which were discarded due to segmentation issues, and expand to more instances of some in other contexts. The tentative conclusion we draw from this analysis is that, as expected, segmental measures of prominence are more robust predictors of focus than F0 measures, but information about F0 peak height and alignment can also provide important clues.

5 Conclusion

This paper has undertaken a corpus study of focused *some* in the context of QPs *some people* and *some money*. The study thus takes advantage of the enormous untapped resource of the Internet as source of natural, spontaneous speech, and deals with the inherent difficulties of such a noisy dataset by using machine learning techniques to explore this multi-dimensional data.

I started by reviewing the differences and similarities between the various pragmatic and semantic concepts that have been gathered under the label of 'focus' and I presented the kinds of focus that are represented in this corpus. This exercise serves multiple purposes. On the one hand, it allows us to better compare this study to previous studies in order to determine if we are comparing apples to apples. This is a valid concern given that the different kinds of focus could turn out to have slightly different acoustic signatures; for instance, we might expect more variability in the marking of implicit rather than explicit contrast. Secondly, we now have a partial fine grained annotation of the corpus, which allows us to extend the study in precisely the direction just described. Finally, while having a single (non-native) annotator could be a handicap of this study, analyzing how the context affected the annotator's judgment reveals strategies for maintaining consistency, and perceived levels of confidence about the annotation.

In §3, I discussed different acoustic markers of focus and I touched upon the phonological nature of the mapping between semantics/pragmatics and phonetics (through prosodic structure). I noted that recent studies have identified non-intonational cues such as duration, intensity and vowel quality, which mark loci of prosodic prominence even in the absence of pitch accents. These are important not only because of corner cases such as second occurrence focus. There is growing evidence that listeners must be able to combine these cues in order to prosodically parse an utterance, given that different speakers may rely on different (combinations of) cues. Furthermore, relating the prosodic structure to meaning is also a context-sensitive task, since there many factors that affect prosodic structure, some structural, some paralinguistic, and some actually meaningful.

In this respect, web-harvested data is interesting because a corpus is more representative of the large variety of influences on prosodic structure in a way that laboratory data is not, and this is what calls for a different type of quantitative analysis than the statistics that is employed in carefully controlled experiments. Spontaneous speech also has major advantages and disadvantages. On the one hand, it gives us the opportunity to study some types of meaning that are difficult to elicit, such as implicatures, and it samples a different range of the population than many lab experiments that recruit from the undergraduate population. On the other hand, it presents a challenge to data collection and analysis, and have to be carefully monitored and combed through to ensure accurate measurements. Additionally, some types of measurements are not reliable without further annotation.

For instance, this analysis suggests that F0 cues are not as important in this context as segmental cues such as duration and vowel quality. But it is possible that this is because we are forced to compare across speakers and across utterances. Even though all speakers were male, they still varied quite a bit in their base F0 level and in their working range. F0 peak height is also highly affected by phonetic effects such as declination. An unfocused *some people* could appear at the beginning of a prosodic phrase and have a higher F0 than a focused *some money* appearing at the end of a prosodic prosodic phrase. Some of these confounds could be mitigated by more thorough data pre-processing, including for instance speaker diarization, manual or automatic prosodic boundary and/or pitch accent annotation, and so on. However, we can also keep these conditions into account and interpret a weak showing from F0 in some of our machine learning models as stronger than it looks, since the trend remains intact and visible despite the great number of factors that affect F0.

This corpus confirmed intuitions from the semantic literature that unfocused *some* tends to be very reduced, perhaps even lacking in a vowel altogether. But while there were frequently two distinct distributions of [ʌ] that depended on focus condition for many acoustic features that we looked at, the distributions were certainly overlapping, so no classification could be done based on just one feature

The K-Means clustering algorithm produced surprisingly accurate clusters without access to any semantic/pragmatic labels. Additionally the LDA classifier had extremely good accuracy on the test set. This suggests that *some* is a particularly easy case for focus classification, perhaps because of the large expectation for *some* to reduce and to be less prominent than the noun it precedes.

The Principal Components Analysis allowed us to visualize the dataset for all the acoustic features we collected by reducing the dimensionality of the dataset. Interestingly, even the most minimal set of features that we tested did not produce principal components that could explain a large portion of the variance of the data in only the first two steps, so the biplot we produced was at most an ok representation of the data. This might suggest that there is a good amount of indeterminacy if we just look at the acoustic signal, and as an annotator I was influenced to a large extent by semantic and pragmatic context, or it might suggest that the current features could be improved upon for a simple clustering task.

References

Bartels, Christine. 2004. Acoustic correlates of 'second occurrence' focus: Toward an experimental investigation. In Hans Kamp & Barbara Hall Partee (eds.), *Context-dependence in the analysis of linguistic meaning*, Amsterdam; Boston: Elsevier.

Beaver, David I & Brady Z Clark. 2008. *Sense and sensitivity: how focus determines meaning.* Malden, MA: Blackwell Pub.

Beaver, David I., Brady Zack Clark, Edward Stanton Flemming & T. Florian Jaeger. 2007. When semantics meets phonetics: Acoustical studies of second-occurrence focus. *Language* 83(2). 245–276. doi:10.1353/lan.2007.0053.

Beckman, Mary E. & Janet B. Pierrehumbert. 1986. Intonational structure in Japanese and English. *Phonology Yearbook* 3. 255–309. `http://www.jstor.org/stable/4615401`.

Bell, Alan, Jason M. Brenier, Michelle Gregory & Cynthia Girand. 2009. Predictability effects on durations of content and function words in conversational english. *Journal of Memory and Language* 60(1). 92–111. doi:10.1016/j.jml.2008.06.003.

Boersma, Paul. 1993. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. In *Proceedings of the Institute of Phonetic Sciences*, vol. 17, 97–110. University of Amsterdam.

Böhmová, Alena, Jan Hajič, Eva Hajičová & Barbora Hladká. 2003. The Prague dependency treebank. In Anne Abeillé (ed.), *Treebanks: building and using parsed corpora*, 103–127. Dordrecht; Boston: Kluwer Academic Publishers. `http://link.springer.com/chapter/10.1007/978-94-010-0201-1_7`.

Bolinger, Dwight. 1965. *Forms of English; accent, morpheme, order.* Cambridge, Mass: Harvard University Press.

Bruce, Gösta. 1977. *Swedish word accents in sentence perspective*: Lund : LiberLäromedel/Gleerup Ph.D.

Brugos, Alejna, Stefanie Shattuck-Hufnagel & Nanette Veilleux. 2006. Transcribing prosodic structure of spoken utterances with ToBI. `www.tobihome.org`.

Büring, Daniel. 2003. On D-trees, beans, and B-accents. *Linguistics and Philosophy* 26(5). 511–545.

Buring, Daniel. 2013. A theory of second occurrence focus. *Language and Cognitive Processes* 1–15. doi:10.1080/01690965.2013.835433. `http://dx.doi.org/10.1080/01690965.2013.835433`.

Büring, Daniel, Danielring. 2006. Focus projection and default prominence. In Valéria Molnár & Susanne Winkler (eds.), *The Architecture of Focus*, 321–346. Berlin, New York: Mouton de Gruyter.

Calhoun, Sasha. 2006. *Information structure and the prosodic structure of English: a probabilistic relationship.*: University of Edinburgh Ph.D. `https://www.era.lib.ed.ac.uk/handle/1842/8120`.

Calhoun, Sasha. 2010. The centrality of metrical structure in signaling information structure: A probabilistic perspective. *Language* 86(1). 1–42. doi:10.1353/lan.0.0197.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English.* New York: Harper & Row.

Cinque, Guglielmo. 1993. A null theory of phrase and compound stress. *Linguistic Inquiry* 24(2). 239–297.

Diesing, Molly. 1992. *Indefinites.* Cambridge, Mass.: MIT Press.

Dipper, Stefanie, Michael Götze & Stavros Skopeteas (eds.). 2007. *Information structure in cross-linguistic corpora*, vol. 7 Interdisciplinary Studies on Information Structure. Potsdam: Universitätsverlag.

Fischer, Susan. 1968. On cleft sentences and contrastive stress. Ms.

Godfrey, J. J., E. C. Holliman & J. McDaniel. 1992. SWITCHBOARD: telephone speech corpus for research and development. In *ICASSP-92: 1992 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 1, 517–520. doi:10.1109/ICASSP.1992.225858.

Gorman, Kyle, Jonathan Howell & Michael Wagner. 2011. ProsodyLab-Aligner: a tool for forced alignment of laboratory speech. In *Proceedings of acoustics week in Canada*, 4–5. Quebec City.

Grosz, Barbara J., Aravind K. Joshi & Scott Weinstein. 1995. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics* 21(2). 203–225.

Gussenhoven, Carlos. 1983. Focus, mode and nucleus. *Journal of Linguistics* 19. 377.

Gussenhoven, Carlos. 2004. *The phonology of tone and intonation.* Cambridge; New York: Cambridge University Press.

Hamblin, Charles L. 1973. Questions in Montague English. *Foundations of Language* 10. 41–53.

Hedberg, Nancy & Juan Sosa. 2007. The prosody of topic and focus in spontaneous english dialogue. In Chungmin Lee, Matthew Kelly Gordon & Daniel Büring (eds.), *Topic and focus: cross-linguistic perspectives on meaning and intonation*, 101—120. Dordrecht, the Netherlands: Springer. `http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=196741`.

Herburger, Elena. 2000. *What counts: focus and quantification.* Cambridge, Mass.: MIT Press. `http://cognet.mit.edu/library/books/view?isbn=026258185X`.

Horn, Laurence. 2004. Implicature. In Laurence R Horn & Gregory L Ward (eds.), *The* *handbook of pragmatics*, Malden, MA: Blackwell Pub.

Horn, Laurence R. 1984. Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. In D. Schiffrin (ed.), *Meaning, Form and Use in Context: Linguistic Applications (GURT '84)*, 11–42.

Horn, Laurence Robert. 1972. *On the Semantic Properties of Logical Operators in English.* United States – California: University of California, Los Angeles Ph.D. `http://search.proquest.com.proxy.library.cornell.edu/docview/302676533/citation?accountid=10267`.

Howell, Jonathan. 2011. Second occurrence focus and the acoustics of prominence. In Folli Raffaella & Christiane Ulbrich (eds.), *Interfaces in linguistics: new research perspectives*, 278–298. Oxford; New York: Oxford University Press. `http://ecommons.library.cornell.edu.proxy.library.cornell.edu/handle/1813/28941`.

Howell, Jonathan. 2012. *Meaning and Prosody: On the Web, in the Lab, and from the Theorist's Armchair*: Cornell University PhD dissertation.

Jackendoff, Ray. 1972. *Semantic interpretation in generative grammar.* Cambridge, Mass.: MIT Press.

Jesus, Luis M. T. & Philip J. B. Jackson. 2008. Frication and voicing classification. In António Teixeira, Vera Lúcia Strube de Lima, Luís Caldas de Oliveira & Paulo Quaresma (eds.), *Computational Processing of the Portuguese Language* (Lecture Notes in Computer Science 5190), 11–20. Springer Berlin Heidelberg. `http://link.springer.com/chapter/10.1007/978-3-540-85980-2_2`.

Kadmon, Nirit. 2001. *Formal pragmatics: semantics, pragmatics, presupposition, and focus.* Malden: Blackwell.

Kochanski, Greg, Esther Grabe, John Coleman & Burton Rosner. 2005. Loudness predicts prominence: Fundamental frequency lends little. *The Journal of the Acoustical Society of America* 118(2). 1038–1054. `http://scitation.aip.org/content/asa/journal/jasa/118/2/10.1121/1.1923349`.

Krahmer, E. & M. Swerts. 2001. On the alleged existence of contrastive accents. *Speech Communication* 34(4).

Krifka, Manfred. 2008. Basic notions of information structure. *Acta Linguistica Hungarica: An International Journal of Linguistics* 55(3-4). 243–276. doi:10.1556/ALing.55.2008.3-4.2.

Kuroda, Shige-Yuki. 1965. *Generative grammatical studies in the Japanese language*: MIT PhD dissertation.

Ladd, D. Robert. 1980. *The structure of intonational meaning: evidence from English.* Bloomington: Indiana University Press.

Ladd, D. Robert. 2008. *Intonational phonology.* Cambridge; New York: Cambridge University Press.

Lehiste, Ilse. 1980. Phonetic manifestation of syntactic structure in english. *Annual Bulletin of the Research Institute of Logopaedics and Phoniatrics* (14). 1–27.

Liberman, Mark. 1975. *The intonational system of English*: Massachusetts Institute of Technology Thesis. http://dspace.mit.edu/handle/1721.1/27376.

Liberman, Mark & Janet Pierrehumbert. 1984. Intonational invariance under changes in pitch range and length. In Mark Aronoff & Richard Oehrle (eds.), *Language sound structure: studies in phonology*, 157—-233. Cambridge, MA: MIT Press.

Liberman, Mark & Alan Prince. 1977. On stress and linguistic rhythm. *Linguistic Inquiry* 8(2). 249–336.

Lutz, David, Parry Cadwallader & Mats Rooth. 2013. A web application for filtering and annotating web speech data. In *Web as Corpus 8*, 29—36. Lancaster, UK. https://sigwac.org.uk/raw-attachment/wiki/WAC8/wac8-proceedings.pdf.

Milsark, Gary L. 1974. *Existential sentences in English*: Massachusetts Institute of Technology PhD dissertation.

Milsark, Gary L. 1977. Toward an explanation of certain peculiarities of the existential construction in English. *Linguistic Analysis* 3. 1–29.

Mo, Yoonsook. 2010. *Prosody production and perception with conversational speech.* Urbana, IL.: University of Illinois Ph.D. http://hdl.handle.net/2142/18560.

Murray, Sarah E. 2014. Varieties of update. *Semantics and Pragmatics* 7. 1–53. doi: 10.3765/sp.7.2. http://semprag.org/article/view/2843.

Nespor, Marina & Irene Vogel. 1986. *Prosodic phonology.* Dordrecht, Holland; Riverton, N.J., U.S.A.: Foris.

Newman, Stanley S. 1946. On the stress system of english. *Word* 2(3). 171–187.

Partee, Barbara. 1989. Many quantifiers. In Joyce Powers & Kenneth de Jong (eds.), *ESCOL '88: Proceedings of the Fifth Eastern States Conference on Linguistics*, 383–402. Columbus: Ohio State University.

Partee, Barbara. 1991. Topic, focus and quantification. In Steve Moore & Adam Wyner (eds.), *Semantics and Linguistic Theory (SALT) 1*, 159–187.

Partee, Barbara. 1999. Focus, quantification and semantics-pragmatics issues. In Peter Bosch & Rob van der Sandt (eds.), *Focus: linguistic, cognitive, and computational perspectives*, 213–231. Cambridge; New York: Cambridge University Press.

Pierrehumbert, Janet. 1980. *The phonology and phonetics of English intonation*: MIT Ph.D.

Pierrehumbert, Janet & Mary Beckman. 1988. *Japanese tone structure.* Cambridge, Mass.: MIT Press.

Pierrehumbert, Janet & Julia Hirschberg. 1990. The meaning of intonational contours in the interpretation of discourse. In Philip Cohen, Jerry Morgan & Martha Pollack (eds.), *Intentions in Communication*, 271 – 311. Cambridge, Mass.: MIT Press. http://vp5qw4uf5x.search.serialssolutions.com/?ctx_ver=Z39.88-2004&ctx_enc=info%3Aofi%2Fenc%3AUTF-8&rfr_id=info:sid/summon.serialssolutions.com&rft_val_fmt=info:ofi/fmt:kev:mtx:book&rft.genre=book&rft.title=Intentions+in+communication&rft.date=1990-01-01&rft.isbn=9780262031509&rft.externalDBID=n%2Fa&rft.externalDocID=mdp.39015018351232&paramdict=en-US.

Pitrelli, John, Mary Beckman & Julia Hirschberg. 1994. Evaluation of prosodic transcription labeling reliability in the ToBI framework. In *ICSLP 1*, 123—-126.

Riester, Arndt & Stefan Baumann. 2013. Focus triggers and focus types from a corpus perspective. In *ICASSP-92: 1992 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4 2, 215–248. doi:10.5087/d&d.v4i2.2925. http://elanguage.net/journals/dad/article/view/2925.

Roberts, Craige. 1996. Information structure in discourse: Towards an integrated formal theory of pragmatics. *Working Papers in Linguistics* 49. 91–136.

Rooth, M. 1992. A theory of focus interpretation. *Natural language semantics* 1(1). 75–116. http://www.springerlink.com/index/k57211207j40p176.pdf.

Rooth, Mats. 1985. *Association with Focus*: GLSA, Dept. of Linguistics, University of Massachusetts, Amherst dissertation.

Rooth, Mats. 1996. On the interface principles for intonational focus. *Proceedings of SALT* 6(0). 202–226. http://elanguage.net/journals/salt/article/view/6.202.

Rooth, Mats, Jonathan Howell & Michael Wagner. 2013. Harvesting speech datasets for linguistic research on the web. white paper. http://hdl.handle.net.proxy.library.cornell.edu/1813/34477.

Schmerling, Susan F. 1976. *Aspects of English sentence stress.* University of Texas Press.

Schwarzschild, Roger. 1999. Givenness, AvoidF and other constraints on the placement of accent. *Natural Language Semantics* 7(2). 141–177. doi:10.1023/A:1008370902407.

Selkirk, Elisabeth. 1995. Sentence prosody: Intonation, stress, and phrasing. In John Goldsmith (ed.), *The Handbook of Phonological Theory*, 550–569. Cambridge, Mass.: Blackwell.

Selkirk, Elisabeth O. 1984. *Phonology and syntax: the relation between sound and structure.* Cambridge, Mass.: MIT Press.

Shattuck-Hufnagel, S. & A. E. Turk. 1996. A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research* 25(2). 193–247.

Silverman, Kim, Mary Beckman, John Pitrelli, Mari Ostendorf, Colin Wightman, Patti

Price, Janet Pierrehumbert & Julia Hirschberg. 1992. ToBI: a standard for labeling english prosody. In *ICSLP 2*, 867—-870.

Stalnaker, Robert C. 1978. Assertion. *Syntax and Semantics* 9. 315–332.

Steedman, Mark. 2000. Information structure and the syntax-phonology interface. *Linguistic Inquiry* 31(4). 649–689. `http://www.jstor.org/stable/4179127`.

Stevens, Kenneth N. 1998. *Acoustic phonetics.* Cambridge, Mass.: MIT Press. `http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=9234`.

Talkin, David. 1995. A robust algorithm for pitch tracking (RAPT). In W. B. Kleijn & K. K. Paliwal (eds.), *Speech coding and synthesis*, 495–518. Amsterdam; New York: Elsevier. `https://docs.google.com/viewer?url=http%3A%2F%2Fwww.ee.columbia.edu%2F~dpwe%2Fpapers%2FTalkin95-rapt.pdf`.

Taylor, Paul. 2000. Analysis and synthesis of intonation using the Tilt model. *The Journal of the Acoustical Society of America* 107(3). 1697–1714. doi:10.1121/1.428453.

Terken, Jacques & Dik Hermes. 2000. The perception of prosodic prominence. In M. Horne (ed.), *Prosody: Theory and experiment. Studies presented to Gösta Bruce*, 89–127. Dordrecht; Boston: Kluwer Academic Publishers. `http://link.springer.com/chapter/10.1007/978-94-015-9413-4_5`.

Wagner, P. 1999. The synthesis of german contrastive focus. In *Proceedings of ICPhS '99*, .

Zhang, Tong, Mark Hasegawa-Johnson & Stephen E. Levinson. 2006. Extraction of pragmatic and semantic salience from spontaneous spoken English. *Speech Communication* 48(3–4). 437–462. doi:10.1016/j.specom.2005.07.007. `http://www.sciencedirect.com/science/article/pii/S0167639305001743`.

Zimmermann, Malte & Edgar Onea. 2011. Focus marking and focus interpretation. *Lingua* 121(11). 1651–1670. doi:10.1016/j.lingua.2011.06.002.

Zondervan, Arjen. 2009. Experiments on QUD and focus as a contextual constraint on scalar implicature calculation. In Uli Sauerland & Kazuko Yatsushiro (eds.), *Semantics and pragmatics: From experiment to theory*, 94–110. New York: Palgrave Macmillan.

Zondervan, Arjen. 2010. *Scalar implicatures or focus: an experimental approach.* Utrecht: LOT dissertation.

Zubizarreta, Maria Luisa. 1998. *Prosody, focus, and word order.* Cambridge, Mass.: MIT Press. `http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=50391`.