

# Second occurrence focus: it's not size that matters\*

David Beaver

DRAFT, September 22, 2008

## 1 Introduction

This paper tackles some puzzling effects involving focus, using a combination of insights from Watson *et al.* (2008) and Beaver and Clark (2008). In particular, I develop an account of *Second Occurrence Focus* and related phenomena, and argue that the account produces superior predictions to that of Büring (ms). The new proposal centers on the following principles:

- (1) **Prominence Principle:** If one expression is more *communicatively significant* than another expression, then the first should be more *surface prominent* than the second.<sup>1</sup>
- (2) **Communicative significance:** The communicative significance of one expression relative to another may be affected independently by different factors which affect the communicative significance of either expression. These factors concern (i) how predictable expressions are for hearers, and (ii) how important they are as regards the speaker's goals.

---

\*This paper relates to ongoing joint work with Florian Jaeger, who I must thank for many comments and suggestions. I am also grateful to Edgar Onea for discussion.

<sup>1</sup>The *Prominence Principle* is related to empirically validated principles of constant acoustic redundancy (Aylett 2000; Aylett and Turk 2004; Van Son and Pols 2002; Van Son and Van Santen 2005) and constant information at the level of strings (Genzel and Charniak 2002; Levy and Jaeger 2007). One respect in which the Prominence Principle differs from these forbears is that the notion of communicative significance is more general than measures of information that are standardly used. These standard measures are based on the probability of some surface form occurring (i.e. predictability), but do not attempt to factor in the relative importance of different expressions relative to the communicative goals of the interlocutors.

The idea that predictability and importance are independent measures affecting acoustic prominence is taken from Watson *et al.* (2008). They showed not only that both informational measures contribute to prominence, but that they may have different acoustic correlates. The specific results they report on these acoustic correlates do not transfer directly to the issues that will be studied in this paper. The reason is that their results tell us only how utterances which are interesting differ from those which are unpredictable, but do not tell us anything about how speakers produce utterances which simultaneously combine separate items with various levels of interest and predictability.<sup>2</sup> I thus ignore the specific results they obtained on acoustic realization, and simply adopt their insight that both importance and predictability might be expected to contribute to surface prominence in some way. And here I should perhaps note that I twist their proposal in yet another way. Whereas the notion of importance they use appears to be intended as a general psychological index, having nothing to do with language *per se*, I will base calculations of importance rather more narrowly on the linguistic notion of *discourse function*.<sup>3</sup>

Here are some minimal assumptions I will make about how importance and predictability contribute to a single measure of communicative significance:

- (3) **Significance of unpredictable information:** If a constituent is unpredictable, e.g. because it is discourse new, then that unpredictability provides communicative significance for the constituent.
- (4) **Significance of answers:** Providing the information needed to answer the current question under discussion (and, to a lesser extent, any question under discussion) is important to the discourse participants. So if a constituent answers or partially answers the

---

<sup>2</sup>The main data I will be considering involves utterances which include one item which is both interesting and unpredictable, and another which is interesting but predictable. Though one could extend the model of Watson *et al.* (2008) to make predictions in this type of case, e.g. by assuming that acoustic prominence effects of predictability and importance were additive, it is not immediately apparent what the best approach would be. I take it that the single experiment they reported does not yet settle the question of how acoustic correlates of importance and predictability differ in general.

<sup>3</sup>One natural way of combining predictability with importance would be to view importance as utility, so that communicative significance would become expected utility. One could then study whether e.g. principles of constant information discussed by e.g. Hale, Levy, and Jaeger, were in fact better thought of in terms of constant expected utility. I will not pursue this ambitious but speculative line of research in the current paper.

current question, then the importance of that discourse function gives communicative significance to the constituent.

- (5) **Significance of contrast:** A contrastive expression fulfills an important discourse function (marking the difference between otherwise parallel units), and this gives communicative significance to that constituent.

One additional principle, drawn from Beaver and Clark (2008), will play an important role. It concerns the meaning and function of exclusives such as *only*, which occur in many of the crucial examples of focus that I will discuss, and allows us to relate the use of exclusives to their discourse function, and hence to communicative significance:

- (6) **Discourse function of exclusives:** Exclusives are grammatically constrained to associate with the Current Question, i.e. a possibly implicit question answered by the smallest constituent containing the exclusive and its complement.

What I term *surface prominence* is a catch-all that may potentially be affected by a large number of factors, including acoustic properties such as pitch movement, intensity, duration, and vowel quality. I will only consider acoustic prominence in this paper, and not consider the psycho-acoustic question of why certain physical properties of the sound signal should lead to one expression being perceived as more prominent than another. I take it that the following principles of surface prominence are not in and of themselves controversial: they are largely similar to those of Büring (ms) and Beaver *et al.* (2007). A well-known development of the idea that prominence and accenting should be studied as an extension of a general system of stress is that of Truckenbrodt (1995, 1999); for a recent detailed study see Calhoun (2006).

- (7) **Surface prominent:** Prominence is established partly as a result of phonological stress assignment. Other ways to mark prominence are diverse, and include word choice, word order, and use of intensive modifiers (*totally, really, fucking, etc.*), but also effects of gesture such as eyebrow movement and hand movement.
- (8) **Stress:** Sentential stress in English is realized through a combination of acoustic factors, including pitch excursions in proximity to syllables bearing primary word-stress, duration, intensity, spectral tilt, and vowel quality.

- (9) **Nuclear stress:** The most prominent element is the one bearing nuclear stress within an intonational phrase. This is realized in English through a combination of acoustic factors notably including major pitch excursion, followed by a reduction in pitch range for following expressions within the phrase.
- (10) **Non-nuclear stress:** A second tier of prominence is realized through non-nuclear stress in pre-nuclear or post-nuclear positions. In all such positions, increased duration and intensity are among the acoustic factors indicative of stress. In pre-nuclear positions, pitch accents are also commonly used to mark stress.

## 2 Basic data

I will now consider basic properties of the intonational marking of information structure, and how they are explained in the model proposed above.

First, I consider information status, in the sense of Prince (1981). There is a tendency for NPs which are discourse new to be more prominent than those which are discourse old, (or *given*). This obviously follows from the fact that things which have been previously mentioned are more easily predictable than things which have not been previously mentioned, and thus have greater communicative significance.<sup>4</sup>

And let us note right away that the facts are not simple. In a recent studies of corpora of spontaneous speech hand-annotated for newness of NPs, it was found that while there was a statistical correlation between information status and accent, this correlation is not strong relative to other factors (Brenier *et al.* 2006; Nenkova *et al.* 2007). In fact, and simplifying slightly, the best automatic classifiers in an accent prediction task (predicting which words are accented based on the raw text) do not benefit significantly from having access to annotators' judgments as to whether NPs were discourse new or discourse old.

Of course, corpus results on the weakness of the relation between accent and information status merely extend the already widely recognized observation that at least sometimes given NPs bear accents, perhaps most obviously in the case of an accented anaphoric pronoun, which can even bear nuclear

---

<sup>4</sup>Prince's influential work on Givenness has a correlate in the Natural Language Processing community, stemming from Rosenfeld's (1996) influential work on maximum entropy models. Rosenfeld showed that previous occurrence of an expression was a very significant predictor of whether that expression would occur again. All else being equal, this tendency favors the reoccurrence of given words over the introduction of new words.

stress. In fact, the main phenomenon I will consider in this paper, second occurrence focus, involves at least some degree of prominence on given NPs. So far, I have explained (trivially) why new information tends to be more prominent than old information, but I clearly have not explained the facts in anything like their full complexity: I have not explained why sometimes new NPs are less than maximally prominent, or why sometimes old NPs are at least somewhat prominent.

Now consider answers. Let us term a constituent that provides the core piece of information that answers or partially answers a question under discussion an *answering constituent*. It is well known that such constituents tend to be prominent. Indeed, semanticists analyzing intonation often present constructed data suggesting, at least by omission in their annotations, that the answering constituent is the only intonationally prominent constituent in the utterance. The contrast between these artificial data and naturally occurring speech is quickly seen when one inspects speech corpora. In the Switchboard Corpus (Godfrey *et al.* 1992) and the WBUR corpus (Ostendorf *et al.* 1994), utterances of more than a few words tend to have multiple constituents which are prominent to some extent, i.e. more prominent than the least prominent constituent in the sentence. In portions of these corpora that have been annotated using ToBI conventions, there are typically multiple words in each sentence which are analyzed as carrying an accent. Nonetheless, I speculate that there is something right about the semanticists' simplified picture (one answering constituent, one so-called *focus*). Specifically, I suggest that the answering constituent typically bears nuclear stress, at least when the immediately prior utterance is the question being answered.

I have already stated as a principle that constituents which provide answers to questions are communicatively significant, so obviously the model correctly predicts that answering constituents will have some degree of prominence. I suggest that the reason it is natural for answering constituents to bear nuclear stress is that typically answering constituents are the most communicatively significant expressions in the sentences that contain them.

And in turn, they tend to be the most communicatively significant expressions because they always have one trait that makes them communicatively significant, and they typically have another such trait. The trait that they always have, by definition, is the important discourse function of providing information that is sought by a participant. The additional trait that they typically have is that of being to some degree unpredictable, since otherwise the question probably would not have been under discussion. Indeed, answering constituents are often completely new and highly unpredictable.

A rather trivial example of an answering constituent that should be expected to bear nuclear stress is *Fred* in (12a), which, uttered as a reply to (11), is both an answering constituent and unpredictable. All other material in the same clause as *Fred* in (12a) is highly predictable, since it is directly repeated.<sup>5</sup> As a result, the word *is* is reduced to a leaner, and the word *here* is not expected to bear any major prominence marking, unless the speaker intends contrast, to which I will turn shortly.<sup>6</sup>

(11) Who's here?

- (12) a. Well, Fred's here.  
b. You are!  
c. Well, Fred's in the lounge.

The answer in (12b) is more subtle, in that the information conveyed explicitly by the utterance is completely known to the addressee, and neither word in the utterance could be considered new in any interesting sense. None the less, the word *you* should be expected to bear the nuclear accent, and this is predicted by the model, since *you* is an answering constituent.<sup>7</sup>

Answer (12c) illustrates why I say above that answering constituents typically bear nuclear stress, rather than that they always bear nuclear

---

<sup>5</sup>In contexts like that set up by (11), the single word constituent answer *Fred* would often be preferred to the full clause given in (12). I hypothesize that speakers choose a constituent answer precisely when other words that might be added to make a full sentence answer would not be of communicative significance. However, a theory of constituent answers, or of the presumably related phenomenon of ellipsis, is beyond the scope of the current paper.

<sup>6</sup>Related to the contraction of *is* in the above constructed examples, note that Frank and Jaeger (2008) show that the verbs *be* and *have* are more likely to be contracted when they are more predictable. Such work demonstrates why a general theory of predictability is to be preferred to a simplistic division between given and new information, since occurrences of weak (i.e. low semantic information) verbs like *be* and *have* could not easily be separated into those which carry new information and those which do not.

<sup>7</sup>Florian Jaeger (p.c.) points out that the copula *are* in (12b) cannot be reduced to a leaner to produce the answer *You're!* This suggests that the account of prominence presented here must be supplemented with constraints specific to the grammar of English in order to explain why certain constituents cannot be reduced. Further crucial constraints on reduction of communicatively insignificant expressions result from processing considerations. Thus Fox Tree and Clark (1997) and related publications show that even low information expressions are produced with heavy stress when the speaker encounters processing difficulty with following material. Specifically, they show that it is primarily processing considerations which determine whether the definite article *the* is produced with little stress and a schwa, or is lengthened, accented, and produced with the canonical vowel [i], thus *thee*.

stress. The example would typically be produced with some prominence for both *Fred* and *lounge*. There are a range of possibilities for how this might be done, but the point I want to make is that there need not be a phrase break between *Fred* and *lounge*, even when both carry pitch accents. In that case, the utterance of *Fred* would provide an example of an answering constituent which does not carry nuclear stress, since nuclear stress would fall on the word *lounge*.

As already hinted at, the above examples can be used to illustrate effects of contrast. In (12a), stress on *here* might convey *as opposed to somewhere else*, though the speaker would need to clarify further if the reply is to be a pragmatically acceptable response to the question under discussion. Perhaps more obviously, stress on *lounge* in (12c) may be used to convey *as opposed to here*, if the speakers are not in a location with a lounge, or to convey *as opposed to some other part of the current location*, if there is a lounge in the immediate vicinity. I suggest that if *lounge* is given equal or greater prominence than *Fred*, and given that both words are, by assumption, discourse new, then *lounge* must in fact mark contrast, for otherwise it would have strictly lower communicational significance than *Fred*. This would certainly be the case if there is no phrasing between the *Fred* and *lounge*.

At this point, I must make an important admission. The model I describe in this paper is very incomplete, since I offer no theory of phrasing. The best known analysis of English intonation, the ToBI model descending from Pierrehumbert's dissertation (Pierrehumbert 1980), involves two levels of phrasing marked by various boundary tones. And certainly, (12c) would naturally be produced with at least some phrasing between *Fred* and *lounge*, as well as after *well*. If both *Fred* and *lounge* are given high tone accents, but there is a large drop in pitch between them, that would be evidence of phrasing.<sup>8</sup> And this phrasing could potentially involve a minor break (separation into what are termed *intermediate* phrases), or a major break (thus two full *intonational* phrases). In either case, the question of what counts as nuclear stress becomes vexed. It would make sense to say that the notion of nuclear stress has to be restricted to the intonational (or intermediate) phrase, and is not, as I will implicitly assume at various times in this paper in order to simplify, a notion associated with sentential clauses.

I will not consider phrasing in any detail. What is important for the cur-

---

<sup>8</sup>The drop in pitch could also be explained by *lounge* bearing a low/high hybrid accent, though this would result in a different shaped contour than two high tone accents with a boundary in between.

rent paper is that in principle a theory relating various intonational patterns to prominence effects could be developed, and that given such a theory, communicational significance would determine intonation. Thus while I accept that the notion of nuclear stress I adopt is inadequate, or at least insufficiently developed, for our purposes it suffices to note that if *Fred* and *lounge* are separated by a phrase break, then the two words must have at least similar prominence, and therefore similar communicative significance. And this in turn tells us that if *Fred* and *lounge* are both new, and if *Fred* fulfills an important discourse function (as an answering constituent), then *lounge* too must perform an important discourse function. Most obviously, it can be contrastive.

Note here that the initial *well* in the above examples occurs in a separate intonational phrase from the rest of the sentence, as indicated orthographically by a comma. It must bear at least some accent, and significant prominence in the utterance as a whole. This prominence would seem odd in a model that was based on predictability only, and did not factor in importance. The word *well* is relatively predictable, since it is a frequent word which often appears in answers to questions. Although we will not pursue the idea further in this paper, it would be natural in the current framework to postulate that *well* performs a relatively important discourse function, perhaps that of resetting the hearer's expectations. This function would bestow upon *well* communicative significance, and hence prominence, despite its predictability.

The last phenomenon to be considered in the *basic data* category is so-called *association with focus*, as seen in examples like (13). The central observation is that if *cheesecake* bears nuclear stress and nothing else is prominent, then the utterance conveys that Mary ate nothing other than cheesecake. But if *party* bears nuclear stress and nothing else is prominent, the utterance conveys that Mary ate cheesecake on no other occasion. In each case, I term the item that bears nuclear stress, i.e. the item with which prior scholars would have said that *only* associates, the *semantic focus (of only)*.

(13) Mary only ate cheesecake at the party.

The *Discourse function of exclusives* principle I introduced earlier takes a cue from Beaver and Clark (2008), which itself builds on Rooth (1992) and Roberts (1996). The principle implies that the semantic variability of (13) is best seen not as direct association with focus but as association



with a question.<sup>9</sup> Since, by the principle, *only* in (13) is grammatically constrained to associate with the current question, we can calculate the meaning once we have identified the current question. The current question can be identified once we know what the answering constituent is. But since the answering constituent is communicatively significant, it is prominent. So, working backwards, if *cheesecake* bears nuclear stress, we can tell that it is the answering constituent. After some further work following the account of questions and focus developed in Beaver and Clark (2008:33–40), which itself builds on Roberts (1996), we can then tell that the current question is: *What did Mary eat at the party?*

Following Beaver and Clark (2008) once more, we can approximate the meaning of *only* as: the strongest true answer to the current question is the one which *only* modifies. In the case at hand, we predict the meaning: the strongest true answer to the question of what Mary ate at the party is the proposition that she ate cheesecake at the party. And since the latter proposition is the strongest true answer, it follows that she ate nothing else, as desired. Equivalent reasoning predicts that if *party* rather than *cheesecake* carries nuclear stress then Mary ate cheesecake nowhere except at the party. Presto! We have a story about focus sensitivity that explains how changes in prominence are tied to changes in meaning. This will form a crucial component of the account of second occurrence focus to which I now turn.

### 3 Second occurrence focus

The phenomenon of second occurrence focus involves words which are both repeated and the semantic focus of a focus sensitive particle such as *only*. I divide up examples of second occurrence focus using a three way taxonomy, depending upon whether the examples involve vegetables, rice, or, in the most sophisticated paradigm to date, crêpes.

#### 3.1 Second occurrence vegetables

The phenomenon of second occurrence focus was originally taken to be of interest because second occurrence foci are typically unaccented, so that second occurrence focus was seen as a case of unaccented focus. Thus *vegetables* in B's reply in (14), adapted from Partee (1999:p.215) is an example

---

<sup>9</sup>As noted by Beaver and Clark (2008), once we have adopted the view that focus is grammatically constrained by the current question, a theory in which *only* associates with the current question may be predictively indistinguishable from a theory in which *only* associates directly with focus. Here I take the former position.

of a second occurrence focus: a repeated word which is the semantic focus of an exclusive, and which need not carry a pitch accent.

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

When examples of this sort first entered the literature, semanticists were not much concerned with correlates of acoustic prominence other than accent. So it was further assumed, without direct evidence, that second occurrence foci were examples of foci which bore no phonological reflex of focus at all. This was taken to imply that the semantic and pragmatic effects of focus could be divorced from phonology. In turn this led some, e.g. Roberts (1996), to a strong conclusion: that association with focus is not to be analyzed as being tightly constrained by the grammar, but is instead mediated pragmatically.

As it happens, further experimental results of Rooth (1996), Beaver *et al.* (2007), and Ishihara and Féry (2006), have shown that second occurrence foci do bear correlates of acoustic prominence. In particular, Beaver *et al.* found that second occurrence foci have significantly increased duration and energy. We further showed that these correlates, though not large effects by any means, are perceptible by hearers. Given that second occurrence foci are perceptibly more prominent than otherwise similar non-foci, the phenomenon of second occurrence focus should not be seen as a case of semantic focus with no phonological reflex. Thus the phenomenon can no longer play the pivotal role that some hoped for in terms of locating the boundary between semantics and pragmatics. But second occurrence focus remains a puzzling phenomenon both for semanticists and intonational phonologists. Let us consider how the pattern of acoustic prominence in vegetable-type examples can be explained in the model advocated in the current paper, before moving on to *rice* and *crêpes*.

The explanation is simple. Consider B's utterance in (14). The word *Paul* is unpredictable, because it is new in the immediate discourse context. But furthermore, *Paul* is important, because it is contrastive: *Deux points!*<sup>10</sup>

---

<sup>10</sup>The word *Paul* is not only new and contrastive, but also the focus of a focus sensitive operator, the scalar additive *even*, so perhaps that should be *trois points*. But counting something both as contrastive and as the focus of an additive particle may be a case of double counting, since such foci are arguably always contrastive. This issue does not affect the analysis, and anyway it is important to realize that ultimately measures of communicative significance will depend on a more sophisticated operation than totting up

By comparison, the occurrence of *vegetables* in B’s utterance is discourse old and highly predictable, the VP that contains it being string identical to one in the previous utterance. The only thing *vegetables* has going for it in the communicative significance stakes is that it is an answering constituent for the current question, the question of what Mary eats. True, we already had an answer to that question, but now we are learning something new about the answer, namely what the significance would be of Paul knowing that it was the strongest true answer. So *Mary* is at least a little bit important: *un point!* Since *Paul* is the most communicatively significant word in the utterance, it can be realized with nuclear stress, and a pitch accent to boot. But that means that *Mary*, although it deserves some acoustic prominence, will fall in the post-nuclear tail, and as such should not be produced with a significant pitch movement. The result is that *Mary* bears some signs of acoustic prominence, including slightly increased duration and energy, but no pitch accent. And this is just what is observed in the controlled experiments reported by Beaver *et al.* (2007).

### 3.2 Second occurrence rice

In (14), B’s utterance of *vegetables* is not only repeated and a focus, it is also a repeat of a focus, since when it occurred previously it was also the focus of a focus sensitive operator (a previous occurrence of *only*, in fact). It is natural to wonder whether second occurrence focus effects depend crucially on whether the second occurrence focus was a focus in its previous occurrence. And this takes us from *vegetable* examples to *rice* examples. In the following, italics indicate that contrastive intonation is intended:

(15) People who *grow* rice generally only *eat* rice. (Rooth 1992:109)

The second occurrence of *rice* in (15) is the intended semantic focus of *only*, and it is repeated, but in its first occurrence it is not a semantic focus of any overt operator.<sup>11</sup> And the central observation is that...not much

---

points, even if they are French points. Of course, it is also notable that while linguists often treat given/new as a qualitative binary distinction, measures of unpredictability in computational linguistics are invariably probabilities, and measured using real numbers. We would expect that a full mathematical development of the current informal model would follow this latter approach. Thus our totting up of points should only be seen as an intuitive approximation to what should be operations on much finer grained measures of predictability and importance.

<sup>11</sup>It is also notable that the first and second occurrences of the *rice* are in the same sentence in (15), while first and second occurrences of *vegetables* occur in separate utterances in (14). We will not study the significance of this difference here.

has changed. That is, the utterance can still be naturally produced with no pitch accent on the second occurrence focus, in this case, *rice*.<sup>12</sup>

The explanation of second occurrence *rice* examples should now be obvious, since it is not essentially different from the explanation for the second occurrence *vegetable* examples. The word *eat* is both important and unpredictable. The second occurrence of *rice* is important, because it is the semantic focus of *only* and hence, by our theoretical assumptions, must be an answering constituent, but it is predictable.<sup>13</sup> Therefore *eat* is more communicatively significant than the second occurrence of *rice*, and it is natural for *eat* to carry nuclear stress, and *rice* to carry no pitch accent, as observed.<sup>14</sup>

Note that it remains an open empirical question whether in examples like (15), when produced with no phrasing within the VP *only eat rice*, the second occurrence focus *rice* bears any of the hallmarks of acoustic prominence. I know of no experimental study. But at least some suggestive evidence comes from considering pronominal variants of the original version. First note that in general it is possible to replace *rice* with the weak pronoun *it* in the phrase *eat rice*, as in (16). However, when we try to make the substitution in (15), the resulting sentence, as in (17), perhaps slightly odd, seems to imply that rice growers do nothing but eat the rice they have grown, e.g. as opposed to selling it or giving it away.

(16) People who *grow* rice generally *eat* it several times a week.

---

<sup>12</sup>In the main text we discuss utterances of (15) in which *eat* and the second occurrence of *rice* occur in the same intermediate phrase. However, it is also possible to produce (15) with a phrase break after *eat*. This would put the second occurrence of *rice* in a separate phrase, and since every phrase must contain at least one expression with a pitch accent, it follows that the second occurrence of *rice* would be pitch accented. To study such effects, it would be better to use heavier constituents than *eat* and *rice*, and to use words with pre-final stress, so that boundary tones could more easily be separated from pitch accents.

<sup>13</sup>The model I assume requires the postulation of an implicit current question for which *rice* is the answering constituent in (15). That question would presumably be: *What do people who grow rice eat?* To date none of the various question based analyses of discourse, and here I include Carlson (1985), Roberts (1996), and Büring (2003), as well as Beaver and Clark (2008), have provided a predictive and general model of how implicit questions are calculated. This is, perhaps, the greatest single shortcoming of these models.

<sup>14</sup>One complication is that (15) is naturally produced with a phrase break before the word *generally*. This would be explained in the current model by the fact that *grow* and *eat* have approximately equal communicative significance, both being new and contrastive. Putting each into its own intermediate phrase, or even each into its own intonational phrase, provides one way of making both of them the most prominent items within their containing phrases, and thus manifesting their similar communicative significance. But I must leave a full discussion of this issue for a hypothetical point in the future when I have a theory of phrasing.

(17) People who *grow* rice generally only *eat* it.

The word *it* does not make a good semantic focus for *only*, and I take it that the reason is that the weak pronoun *it* cannot normally be acoustically prominent.<sup>15</sup> And this, in turn, I tentatively take as indirect evidence that *rice* might carry at least some degree of sentential stress in (15).

### 3.3 Second occurrence crêpes

There is no way to intone the string in (18a,b) such that it successfully answers the question of what John ate in Paris and nowhere else. Thus Buring (ms) describes this answer as *ineffable*, although he is quite clear that he only means by this that the string in (18a,b), the answer form most straightforwardly obtained from the given question form, cannot represent the desired meaning. He recognizes that other forms might successfully answer the question: example (19) has such a form. The issue that Buring considers, though, is why the string form in (18a,b) cannot be used to answer the question.

(18) What did John only eat in PARIS?

- a. #John only ate crêpes in PARIS.
- b. #John only ate CRÊpes in Paris. (Buring ms, cited there as Schwarzschild p.c.)

(19) What John only ate in PARIS was CRÊpes.

We will turn to Buring's analysis shortly. First, let us see how the model proposed here explains the data. The first thing to note here is that the crêpes sentence seems very similar to the rice sentence. Like the rice sentence, the crêpes sentence involves two items that are of communicational significance, in this case *crêpes* and *Paris*. And like the rice sentence, the crêpes sentence involves one item that is both important and unpredictable, i.e. *crêpes*, preceding an item which is important but predictable, i.e. *Paris*. So the only difference between rice examples and crêpes examples, at least in the idiolects of a small group of professional semanticists, is that the former are good, and the latter are bad.

Or so it would seem. But in fact there is a further crucial difference: in rice-type examples, the most communicatively significant word is important

---

<sup>15</sup>See the discussions of Krifka (1997) and Rooth (1996), who note the awkwardness of pronominalizing the second occurrence foci in standard *vegetable*-type examples.

because it is contrastive, but in crêpes-type examples, the most communicatively significant word is important because it is the answering constituent for the current question. This immediately provides an explanation for why crêpes-type examples are infelicitous, and it turns out to that for the most part the explanation runs at a level removed from intonation *per se*. The problem with these sentences would seem to involve the presence of two separate answering constituents, and thus two separate questions, only one of which can be the current question for the smallest constituent containing *only* and its complement.

Suppose that *crêpes* is an answering constituent and *Paris* is not. Then the focus sensitive particle *only* associates with the question *what did John eat in Paris?* This yields the meaning: *the strongest true answer to the question of what John ate in Paris is that he ate crêpes in Paris*, which implies that John ate nothing but crêpes in Paris. But this meaning is not appropriate for a sentence intended as an answer to the question at hand, the question of what John eats nowhere but in Paris.

Now let us suppose that *Paris* is an answering constituent and *crêpes* is not. In that case, we will get the desired semantics, entailing that John eats *crêpes* nowhere but Paris. But in this case by definition of what an answering constituent is, the utterance cannot be an answer to the question of what John eats nowhere but in Paris. For in that case, *contra* to our assumptions, *crêpes* would be an answering constituent.

This leaves us with one more option worth considering. What if both *crêpes* and *Paris* are answering constituents? This is a complex case. The *discourse function of exclusives* principle requires that there is a unique current question for the smallest constituent containing *only*. There are two ways to realize this, and we have to break down the option under consideration into two further sub-options. The first sub-option is to assume a syntactic structure in which the complement of *only* is just *eat crêpes*, and *in Paris* is attached at a higher level. The structure would be: *[ [only [eat crêpes] ] in Paris]*. But in that case, *only* would incorrectly associate with a question for which *crêpes* was the answering constituent, such as: *What did John eat in Paris?* The second sub-option is to assume that *only* in effect associates with a complex question answered by the combination of both *crêpes* and *Paris*, by analogy with standard examples of association with multiple foci considered in the literature.<sup>16</sup> In that case, we would have an exhaustive answer to the question: *what did John eat where?* This

---

<sup>16</sup>I am thinking of cases like: *I only introduced BILL to SUE*, implying that there's no other pair of people such that the speaker introduced one to the other.

is a fine question, but it is not the question being asked, again producing a mismatch. So neither of the two sub-options work.

In sum, whatever we take to be the answering constituent in a *crêpes* example, there is a mismatch between the reply and the question that was asked. And it is this mismatch that is responsible for the oddity of *crêpes* examples.

Further evidence that this analysis is on the right track comes from considering slight variants on the original *rice* and *crêpes* examples. The alchemical trick we must somehow manage is to turn *rice* into *crêpes*, and *crêpes* into *rice*. Fortunately, Büring has already performed the first trick for us, creating a minimal pair for Rooth's original rice example in which the word *eat* is the answering constituent for an explicit question instead of being a contrastive element:

(20) What do you do only with rice? # I only EAT rice. (Büring ms)

Example (20) provides us with strong evidence that the problem with the original *crêpes* examples is just as I argued above: there is no way of assigning answering constituent status so as to make the reply semantically and pragmatically appropriate for the current question. For as soon as *eat* in the perfectly acceptable *rice* sentence is converted from being contrastive to being an answering constituent, we get infelicity. That is, we apparently have a *crêpes*-type sentence instead of a *rice*-type sentence.

To check the proposal thoroughly, we should also consider the reverse transmutation, from *crêpes* to *rice*. It is tricky to form a really good minimal pair with the original *crêpes* sentence, but here is an attempt:

(21) Daniel, rather than eating them in his hometown of Essen, only eats *bratwurst sausages* in Paris. And Roger only eats *crêpes* in Paris.

In (21), there is no explicit question under discussion, and we can understand the second sentence in such a way that the word *crêpes* is not an answering constituent for any implicit question. Rather, we can take *crêpes* as contrasting with *bratwurst sausages*. And, magically, all the original *crêpe* oddity just melts away. Example (21) is packed with a peculiar mixture of German meat and French pancakes, and with not a grain of rice in sight. But it could be no more felicitous even if it lay on a bed of the finest jasmine scented basmati from the foothills of the Himalayas, white and moist as virgin snow, yet steaming in the hot Punjab sun.

## 4 Comparison with Büring’s model

The model developed by Büring (ms) involves assumptions about intonational phonology and prominence which, although tidily expressed, are reasonably well in line with other work in the area such as Beaver *et al.* (2007) or the account advanced in the current paper. The main differentiating feature of Büring’s model is his innovative adaption, from the work of Jacobs (1991), of the notion of a *domain* of focus.

According to Büring, the interpretation of each focus takes place with respect to some larger syntactic domain. In the case of a direct answer to a question, or of cases of contrast that Büring considers, that domain is typically the entire sentence containing the focused expression. But in the case of the semantic focus of a focus sensitive operator like *only*, the domain of the focus may be smaller, provided that it is a constituent containing both the operator and the focus.

Büring’s crucial theoretical development, put in terms of the current paper, is his hypothesis that acoustic prominence of a focus depends on the size of its domain. Specifically, if one focus has a domain which strictly contains the domain of another focus, then the first must receive stronger stress. Thus size of focus domain plays the same role in Büring’s account as communicative prominence in the current account.

The model Büring develops accounts for standard vegetables-type second occurrence focus examples straightforwardly. In a case like (14), repeated below, the occurrence of *Paul* in B’s utterance is taken to involve what Büring calls a *free* focus, in addition to the focus bound by the operator *even*. A free focus is taken to always have a domain that includes the entire sentence. On the other hand the repeated focus *vegetables* has as its domain just the VP *only eats vegetables*. Thus *Paul* must bear nuclear stress, and *vegetables* ends up in the post-nuclear tail. This entails the desirable result that *vegetables* carries some hallmarks of stress, but no accent, just as in the model advanced in the current paper.

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

*Crêpes*-type examples provide perhaps the most striking predictive success for Büring’s model. In an utterance of *John only ate crêpes in Paris*, the word *crêpes* should have a domain with sentential scope, since it is the answer to a question. On the other hand, *Paris* should have a smaller domain, namely the VP *only ate crêpes in Paris*. This implies that *crêpes*



must be more prominent than *Paris*. But Büring adapts from Truckenbrodt (1995) a principle termed *FocusProminence*, which says that for a focus sensitive operator like *only* to associate with an expression, there must be some domain containing both the operator and the expression such that the expression is the most prominent item in that domain.<sup>17</sup> In the case at hand, any domain which contains *Paris* also contains *crêpes*, and we have already established that the latter is more prominent. Therefore, *only* cannot associate with *Paris*, and must associate with *crêpes*, thus resulting in an unintended meaning, and infelicity.

The success of Büring’s model with *crêpes* examples is unfortunately also its downfall. The problem is that Rooth’s rice examples involve exactly the same structural configuration as the *crêpes* examples. The one relevant difference, as I have already pointed out, is that *crêpes* is an answering constituent while *eat* in the rice example is contrastive. But this is no help for Büring’s model, for differences in discourse function do not enter directly into that model. The only parameter that Büring’s model has to work with is differences in domain size.

To understand the problem, note that for Büring to explain the prominence of *eat* in the rice example, he must assume that it has the largest domain. Therefore the only way for him to account for association of *only* with *rice* would be to drop his assumption that *only* must associate with the most prominent expression in its scope. But if he drops that assumption, then he loses his explanation of ineffability effects in the *crêpes* examples.

## 5 Last words

There seems to be no straightforward way for Büring to adapt his explanation so as to predict both ineffability of *crêpes* type examples and acceptability of the rice type examples, at least without adopting wholesale something like the analysis proposed in the current paper. I suggest, not without bias of course, that while making that adaption it might be just as well to drop

---

<sup>17</sup>This principle of Büring’s, *FocusProminence*, closely resembles a principle argued for by Beaver *et al.* (2007), who conclude: “The focus of a focus-sensitive operator receives the strongest phrasal stress in the scope of the operator, usually realized as a nuclear accent.” However, the (implicit) assumption of Beaver *et al.* is that the scope, in this sense, is a fixed syntactic unit, e.g. for VP modifying *only*, the scope would be the VP complement. On the other hand, Büring’s domains may potentially vary utterance by utterance, e.g. sometimes just including VP-*only* and a verb sitting next to it, and sometimes including *only* and the whole VP complement. Note that Rooth’s rice example is problematic not only Büring’s version of the principle, but also Truckenbrodt’s, and Beaver *et al.*’s.

the notion of a *focus domain* altogether, and replace it with *communicative significance*.

A model of communicative significance offers the possibility of a uniform explanation of a wide range of puzzling cases involving focus, and also a way to approach the marking of given and new information, as a special case of predictability. It may even help bridge the curious gulf that lies between the type of predictability-based model of accent common in the computational literature, and the accounts of *discourse importance* found in work by semanticists.

I end with two morals that can serve as a summary of the conclusions. First, and with specific reference to Büring: it's not size that matters. And second: major stress is only needed when dealing with stuff that is both important and unpredictable.

## References

- AYLETT, M. 2000. Stochastic Suprasegmentals: Relationships Between Redundancy, Prosodic Structure and Care of Articulation in Spontaneous Speech. In Sixth International Conference on Spoken Language Processing. ISCA.
- , and A. TURK. 2004. The Smooth Signal Redundancy Hypothesis: A Functional Explanation for Relationships between Redundancy, Prosodic Prominence, and Duration in Spontaneous Speech. *Language and Speech* 47.31–56.
- BEAVER, DAVID, BRADY CLARK, EDWARD FLEMMING, T. FLORIAN JAEGER, and MARIA WOLTERS. 2007. When semantics meets phonetics: Acoustical studies of second occurrence focus. *Language* 83.
- BEAVER, D.I., and B.Z. CLARK. 2008. *Sense and Sensitivity: How Focus Determines Meaning*. Blackwell.
- BRENIER, JASON, ANI NENKOVA, ANUBHA KOTHARI, LAURA WHITTON, DAVID BEAVER, and DAN JURAFSKY. 2006. The (non)utility of linguistic features for predicting prominence in spontaneous speech'. In Proceedings of the IEEE/ACL 2006 Workshop on Spoken Language Technology.
- BÜRING, DANIEL. 2003. On D-Trees, Beans, and B-Accents. *Linguistics and Philosophy* 26.511–545.

- BÜRING, DANIEL, ms. Been there, marked that a tentative theory of second occurrence focus.
- CALHOUN, SASHA. 2006. Information structure and the prosodic structure of English: a probabilistic relationship. University of Edinburgh dissertation.
- CARLSON, LAURI W. 1985. *Dialogue Games: An Approach to Discourse Analysis*. Kluwer Academic Publishers.
- FOX TREE, J.E., and H.H. CLARK. 1997. Pronouncing the *as thee* to signal problems in speaking. *Cognition* 62.151–167.
- FRANK, A., and T. FLORIAN JAEGER. 2008. Speaking rationally: Uniform information density as an optimal strategy for language production. In *The 30th Annual Meeting of the Cognitive Science Society (CogSci08)*.
- GENZEL, D., and E. CHARNIAK. 2002. Entropy rate constancy in text. In *Proceedings of ACL–2002, Philadelphia*.
- GODFREY, J., E. HOLLIMAN, and J. MCDANIEL. 1992. SWITCHBOARD: Telephone speech corpus for research and development. In *Proceedings of ICASSP-92*, 517–520.
- ISHIHARA, SHINICHIRO, and CAROLINE FÉRY. 2006. Phonetic Correlates of Second Occurrence Focus. In *Proceedings of NELS 36, Amherst, MA, GLSA*.
- JACOBS, JOACHIM. 1991. Focus Ambiguities. *Journal of Semantics* 8.1–36.
- KRIFKA, MANFRED. 1997. Focus and/or Context: A second look at second occurrence expressions. In *Context-dependence in the Analysis of Linguistic Meaning (Proceedings of the workshops in Prague and Bad Teinach)*, ed. by Hans Kamp and Barbara Partee, 253–276, University of Stuttgart. Institut fuer maschinelle Sprachverarbeitung.
- LEVY, ROGER, and T. FLORIAN JAEGER. 2007. Speakers optimize information density through syntactic reduction. In *Advances in neural information processing systems (NIPS) 19*, ed. by B. Schölkopf, J. Platt, and T. Hoffman, 849–856. Cambridge, MA: MIT Press.
- NENKOVA, ANI, JASON BRENIER, ANUBHA KOTHARI, DANIEL JURAFSKY, SASHA CALHOUN, DAVID BEAVER, and LAURA WHITTON. 2007. To memorize or to predict: Prominence labeling in conversational speech. In *Proceedings of NAACL-HLT 2007*.

- OSTENDORF, M., P. PRICE, and S. SHATTUCK-HUFNAGEL. 1994. The Boston University radio news corpus. Technical Report ECE-95-001, Boston University.
- PARTEE, BARBARA H. 1999. Focus, Quantification, and Semantics-Pragmatics Issues. In *Focus: Linguistic, Cognitive, and Computational Perspectives*, ed. by Peter Bosch and Rob van der Sandt, 213–231. Cambridge University Press.
- PIERREHUMBERT, JANET B. 1980. *The Phonology and Phonetics of English Intonation*. Cambridge, MA: MIT dissertation.
- PRINCE, ELLEN. 1981. Toward a taxonomy of given-new information. In *Radical pragmatics*, ed. by Peter Cole, 223–256. New York, NY: Academic Press.
- ROBERTS, CRAIGE. 1996. *Information Structure in Discourse: Towards an Integrated Formal Theory of Pragmatics*. OSU Working Papers in Linguistics 49. Papers in Semantics.
- ROOTH, MATS. 1992. A Theory of Focus Interpretation. *Natural Language Semantics* 1.75–116.
- . 1996. On the interface principles for intonational focus. In *SALT VI*, ed. by Teresa Galloway and Justin Spence, 202–226, Ithaca, NY: Cornell University.
- ROSENFELD, RONI. 1996. A maximum entropy approach to adaptive statistical language modeling. *Computer Speech and Language* 10.187–228.
- TRUCKENBRODT, HUBERT. 1995. *Phonological phrases—their relation to syntax, focus, and prominence*. Massachusetts Institute of Technology dissertation.
- . 1999. On the relation between syntactic phrases and phonological phrases. *Linguistic Inquiry* 30.219–255.
- VAN SON, R.J.J.H., and L.C.W. POLS. 2002. Evidence for efficiency in vowel production. In *Proceedings of ICSLP2002*, volume 1, Denver, USA.
- VAN SON, R.J.J.H., and JAN P.H. VAN SANTEN. 2005. Duration and spectral balance of intervocalic consonants: a case for efficient communication. *Speech Communication* 47.100–123.

WATSON, D.G., J.E. ARNOLD, and M.K. TANENHAUS. 2008. Tic Tac TOE: Effects of predictability and importance on acoustic prominence in language production. *Cognition* 106.1548–1557.