

# Converging evidence for the relevance of alternative sets: Data from NPs with focus sensitive particles in German

KATHARINA SPALEK\*  
*Humboldt-Universität zu Berlin*

AND

AMIR ZELDES\*  
*Georgetown University*

## ABSTRACT

Recent psycholinguistic studies on the reality of alternative sets in processing focus NPs have shown that focus particles like ‘only’ play a special role in activating the mental representation of alternatives to focused nouns. In this paper we present a new corpus study which provides converging evidence to support psycholinguistic findings and suggests that alternatives preceded by a focus particle are not only more activated in experimental contexts, but are also more likely to be discussed in the subsequent context. To this end we develop and evaluate inter-annotator agreement on two novel annotation tasks in naturally occurring German corpus data: recognition of nominal alternatives in general without any context, and recognition of alternatives in the context of sentence pairs. We show that while annotators agree poorly on the first, they agree strongly on the second. We also develop a concept of ‘alternative density’, the number of alternatives realized in a sentence following a target NP, and present a mixed effects model showing a very significant rise in density after the presence of German *nur* ‘only’ independently of other factors.

KEYWORDS: *Alternative semantics, information structure, annotation, focus particles, alternative sets, converging evidence, corpus linguistics*

## Introduction

Empirical studies on the processing of information structure often characterize *focus* as the “information which is new and unrecoverable from preceding discourse” (Cutler & Fodor, 1979:41), and even recent psycho- and neurolinguistic experiments on focus make use of this characterization (e.g., Ganushchak, Konopka, & Chen, 2014; van Leeuwen, Lamers, Petersson, Gussenhoven, Rietveld, Poser, & Hagoort, 2014). However, in the theoretical literature, it has become increasingly common to define focus with respect to a set of alternatives. Krifka (2008:247), for example, gives the following definition for focus, which goes back to Rooth’s (1985, 1992) alternative semantics and Roberts’ (1995) domain restriction in context: “Focus indicates the presence of alternatives that are relevant for the interpretation of linguistic expressions”.

According to Rooth, every sentence has an ordinary semantic value and a focus semantic value (here and in what follows we will avoid semantic formulae in favor of everyday language to present the main arguments). The focus semantic value is “a set of alternatives from which the ordinary semantic value is drawn, or a set of propositions which potentially contrast with the ordinary semantic value” (Rooth 1992:2), whereby the ordinary semantic value is also part of the alternative set. Initially, which members can become part of the alternative set is only constrained by the requirement that they be of the same semantic type as the focus semantic value, or, as Rooth puts it, “the set of propositions obtainable from the ordinary semantic value by making a substitution in the position corresponding to the focused phrase” (ibid.). Rooth does not assume that the entire alternative set is always considered. Rather, a subset, C, is determined by factors like context, recency, relevance, frequency, and other cognitive factors. To sum up, focus alternatives must be of the same semantic type as the focused constituent, they must be able to be substituted for the focused constituent and their number is constrained by pragmatic and cognitive considerations.

Interestingly, even though Cutler and Fodor (1979) do not refer to an alternative set in their characterization of focus, they still use alternatives when discussing two examples (capital letters indicate the primary sentence accent):

(1) *The man on the CORner was wearing the blue hat.*

(2) *The man on the corner was wearing the BLUE hat.*

“Thus, the new information in (1) is that it was the man on the corner, NOT SOME OTHER MAN, who wore the blue hat, whereas in (2) the new information is that the hat was blue and NOT SOME OTHER COLOR” (p. 49, emphasis ours).

There is an ongoing debate about whether two different types of narrow focus should be assumed, that is, new-information focus and contrastive focus (see Krifka 2008:259). We are looking at examples of both types of focus in the present study but we will not be overly concerned with this theoretical controversy because, as Krifka (2008) shows, the above definition works for both types of focus. In practice, our data seems to consist overwhelmingly of new information focus cases (for a rough estimate of the presence of contrastive cases in our data and some discussion, see the Results section).

Focus can be expressed in different ways. In German, the language of interest here, focus can be expressed by accenting or by syntactic means such as fronting or clefts, and focus sensitive particle like *only* or *even* may additionally associate with the focused element. As we will discuss in more detail below, in the case of focus particles the association with focus is qualitatively different compared to foci marked by intonation or syntax without an associated particle. In a recent study, Spalek, Gotzner & Wartenburger (2014) demonstrated that listeners remember a set of focus alternatives better if a focused element was preceded by a focus particle than if it was not. There are at least two reasons why such a memory benefit could arise. One is that

it is an automatic consequence of processing in the language system: simply the way focus works. A second reason is discourse-functional: a response to frequent patterns in the discussion of alternatives. In the latter case, remembering alternatives might be advantageous for processing the upcoming discourse. This would be possible, for example, if alternatives to an element preceded by a focus particle are more likely to be talked about in subsequent sentences than alternatives to an element not preceded by a focus particle. We would therefore like to explore whether it is in fact the case that alternatives are more frequently mentioned in utterances after a focus particle is used in naturally occurring corpus data, outside of the laboratory setting. If not, this raises doubts that such a habituation effect could be learned, but if so, then habituation would have to be taken seriously as a possible explanation. Regardless of the explanation we suggest, finding evidence of higher alternative frequencies in the presence of focus particles would support the reality of alternatives as a cognitive factor, one which can be manipulated not only in controlled conditions but also in spontaneous language use. Some essential prerequisites to carrying out our study are a discussion of what constitutes alternatives, their identification in language data, and the reliability with which human annotators can carry out that identification.

In order to set the stage, we will first provide a brief description of the syntax and semantics of focus particles, followed by psycholinguistic evidence for the cognitive reality of alternative sets. We will finish the introductory section by discussing the psycholinguistic study on memory for discourse representations by Spalek et al. (2014), which this study converges with from a corpus perspective. This will be followed by the empirical part of our study, a corpus investigation of the impact of the German focus-sensitive particle *nur* ‘only’ on the occurrence of alternatives in discourse.

## **Background**

### *Syntax and semantics of focus particles*

Focus particles are syntactic elements with a relatively free word order (König, 1991). However, a focus particle’s position in a sentence is not arbitrary but depends on where the sentence focus is placed. Focus particles then associate with the specific constituent on which focus is placed.

As far as the meaning of focus particles is concerned, it is assumed that “the focus of a particle relates the value of the focus expression to a set of alternatives.” (König, 1991: 32, see also Krifka, 2008). Sentences with focus particles imply the corresponding sentence without the focus particle and quantify over a set of alternatives to the focused constituent. Note that here and throughout we will speak of focused *constituents* and their alternatives, as this is also relevant to determining the extent of the linguistic material associated with the particle, but it is understood that alternatives refer more strictly not to constituents, but to their denotations. Depending on their lexical meaning, focus particles can exclude (e.g., *only*) or include (e.g., *also*,

*even*) the alternatives for the constituents that they associate with (cf. König, 1991: 33).

The theoretical account that makes the assumption of an alternative set most explicit is Rooth's Alternative Semantics (Rooth, 1985; 1992). Moving away from the specific meaning contribution of focus particles for a moment, Rooth assumes, as outlined above, that each phrase has two semantic values, its ordinary semantic value  $[[\alpha]]^0$  and its focus semantic value  $[[\alpha]]^f$ . The focus semantic value is "the set of propositions obtainable from the ordinary semantic value by making a substitution in the position corresponding to the focused phrase" (Rooth, 1992: 76). Thus, bare focus, as expressed by intonation or word order, already implies the existence of a set of alternatives. However, the relation between focus particles and alternatives is stronger, or, as Beaver and Clark (2008) put it, the association of focus particles with the alternative set is "conventionalized" because this dependence is lexically encoded in the meaning of the focus particles. For example, in the case of *only*, the focus particle of interest here, what happens to the alternatives affects the truth conditions of the utterance: The utterance is true if the proposition is true for the focused constituent but not for any of its alternatives.

In the past few years, a number of psycholinguistic studies have attempted to test whether the alternative set is merely a theoretical construct or whether it also has cognitive reality, in the sense of exhibiting reliably reproducible empirical effects. It is to these studies that we will turn next.

#### *The cognitive reality of the alternative set*

Recent studies in psycholinguistics have investigated the notion of alternative sets, trying to answer the question whether alternatives to a focused constituent become activated in a listener's mind during discourse processing. A number of studies have used the cross-modal primed lexical decision task. In a lexical decision task, participants are presented with a letter string and have to decide whether or not this is a real word in the language tested. The time it takes them to perform this task is taken to reflect lexical access. If a word's representation has been activated by a prime, lexical access is faster than if no such activation has occurred. Braun & Tagliapietra (2010) used this paradigm to investigate the hypothesis that focus (in their study: contrastive focus expressed by contrastive accent) activates a set of focus alternatives (but not associatively related items). They presented Dutch native speakers with auditory sentences like *De sculpturen stonden in het museum* ("The sculptures stood in the museum") with either neutral or contrastive accent on "museum". Directly afterwards, participants saw a letter string on the screen and had to perform a lexical decision. In the critical condition, the word was either contrastively related (*archieff* "archive" vs. "museum") or unrelated (*noordoosten* "north east") (Experiment 1a) or associatively related (*kunst* "art" vs. "museum") or unrelated (*noordoosten*) (Experiment 1b). The authors chose items in the following way: For the non-contrastive associates, they relied on a pre-study in which they had asked participants to report the first word that came to mind upon reading the target word – excluding

words that fulfilled the criteria for contrastive associates; for contrastive associates, they chose words that were semantically related and could be replaced for one another in the given sentence context. Braun & Tagliapietra observed general facilitation for related items in Experiment 1b (that is, *kunst* was reacted to faster following *museum*, independent of the accent on *museum*). However, and critically, in Experiment 1a, they observed an interaction between relatedness of prime and target and the accent on the prime: Contrastive associates were reacted to faster than the control word if they appeared after a contrastively accented prime but not if they appeared after a prime with neutral accent. The authors conclude that this shows that contrastive accenting activates alternatives to the focused constituent, speeding up recognition of these alternatives in a lexical decision task. With a similar paradigm (cross-modal priming with lexical decision), evidence for the existence of a set of focus alternatives has also been reported by Byram-Washburn (2013), Gotzner, Wartenburger & Spalek (submitted), Husband & Ferreira (submitted), and Norris, Cutler, McQueen & Butterfield (2006).

In a memory study, Fraundorf, Watson & Benjamin (2010) investigated the effect of contrastive focus on participants' memory. Participants listened to brief stories of the type illustrated in (3) and (4).

(3) *Both the British and the French biologists had been searching Malaysia and Indonesia for the endangered monkeys.*

(4) *Finally, the French spotted one of the monkeys in Malaysia and planted a radio tag on it.*

The stories introduced two sets of alternatives, here: British vs. French and Malaysia vs. Indonesia. In the critical sentence (4), both sets were reduced to one item (here: French, Malaysia). One of these was produced with a contrastive accent (L+H\*, according to the ToBI system) while the other was produced with an H\* accent. Although there is an ongoing discussion as to what exactly these two types of accents encode and whether they are distinct categories or rather different values on a continuum (cf., e.g., Pierrehumbert & Hirschberg, 1990; Selkirk, 2002 vs. Ladd & Schepman, 2003), it is generally agreed that L+H\* induces a stronger contrastive interpretation than H\* (cf. Watson, Tanenhaus & Gunlogson, 2008; Weber, Braun & Crocker, 2006; Ito & Speer, 2008).

After having listened to 48 stories of the above type, participants were tested. They were given the stories in written form and with the critical words replaced by underscores, as in (5):

(5) *Both the British and the French biologists had been searching Malaysia and Indonesia for the endangered monkeys. Finally, the \_\_\_\_\_ spotted one of the monkeys in \_\_\_\_\_ and planted a radio tag on it.*

Participants had to fill in the blanks. It turned out that their memory was better for words that had been contrastively accented than for words that had carried the more neutral accent. In a subsequent set of experiments using identical materials, Fraundorf and colleagues also investigated recognition memory. In a first session, participants listened to the story. A day later, they returned to the lab and were given sentences of the form “The British scientists spotted the endangered monkey and tagged it” and had to indicate whether this was correct or incorrect. Contrastive accenting during the presentation phase (i.e., on the first day) improved participants’ discrimination ability between correct items and contrast items (i.e., their performance on the second day). The authors concluded that contrastive accenting allowed participants to encode both what happened (the ordinary semantic value of the proposition) and what did not happen (its explicitly introduced alternative) with higher accuracy.

So far, we have seen that contrastive focus, realized by means of pitch accent, evokes a set of alternatives and that these alternatives are better remembered than the same items produced without contrastive accenting. In the following section, we will turn to the additional function of focus sensitive particles like “only”, “also”, and “even”.

#### *The effect of focus particles on discourse processing*

Gotzner, Wartenburger & Spalek (submitted) investigated the question of whether alternatives are activated more strongly when listeners process focus in combination with focus sensitive particles compared to a condition where focus was realized only by intonational means. The authors carried out both a probe recognition study and a lexical decision study, but we will only introduce the probe recognition study here. Probe recognition is a paradigm intended to look at short-term memory effects. In the particular study described here, participants listened to short dialogues and immediately afterwards, they were presented with a written word. Upon presentation of the written word, they had to decide whether this word had appeared in the dialogue they had just heard. The time to reply *yes* or *no* was measured. Many psycholinguists assume that probe recognition is particularly well suited for tapping into the construction of a discourse model (e.g., Gernsbacher & Jescheniak, 1995; Glenberg, Meyer & Lindem, 1985).

The stimuli consisted of short dialogues in which a first speaker introduced a set of elements and formulated an assumption, repeating two of the elements. This assumption was then corrected by the second speaker who used the third alternative in focus. An example is given in (6) (note that actual the stimuli in the experiments were presented in German).

- (6) *Speaker 1: In the fruit bowl, there are peaches, cherries, and bananas. I bet Carsten has eaten cherries and bananas.*  
*Speaker 2: No, he only/ \_\_\_ ate peaches.*

On top of ‘only’ versus no particle, a third condition with less acceptable ‘even’ instead of ‘only’ was also tested, but is not relevant for the current discussion.

The dialogue was followed by a probe word presented visually on the computer. Participants had to decide whether the probe word had appeared in the dialogue. Probe words were either mentioned alternatives (i.e. one of the items that was not focused in the third sentence, e.g. *cherries* or *bananas*), an unmentioned alternative (i.e. an item from the same taxonomic category as the alternatives which has not been mentioned in the dialogue, e.g., *apples*) or an unrelated item that had no semantic or associative relation to the focused element (e.g. *clubs*).

The reason for including unmentioned alternatives was twofold. We wanted to prevent participants from using semantic category information as a decision criterion in the probe recognition task (i.e. we did not want them to base their decision only on the fact that a cherry is a piece of fruit), but we also wanted to investigate whether participants consider more than just mentioned items when representing focus alternatives in their mental model. We observed that it took participants longer to reject unmentioned alternatives than to reject the unrelated control words which suggests that they considered the unmentioned alternatives as part of the alternative set – at least temporarily. It took them longest to accept the mentioned alternatives. The latter finding seems counter-intuitive at first, but it is due to the paradigm used. Remember that in probe recognition, participants have to decide whether they have encountered the probe in the previous linguistic material. Given that the effect of focus is to highlight the presence of alternatives, we assume that during the initial stages of building the mental model, a large number of alternatives, both mentioned and unmentioned, are active. Therefore, it takes some effort (and hence, time) on the part of the participant to decide whether the probe has actually been mentioned or whether it just feels like it has been mentioned because it is semantically related to the focused element.

Additionally, and this is the focus of our research interest, there was a main effect for particles. In the presence of particles like ‘only’, the rejection of unmentioned alternatives and the acceptance of mentioned alternatives took longer than in their absence. No effect of focus particles on the rejection of unrelated items was observed. As outlined above, we proposed that the findings can be explained as a competition effect of alternatives and the focused element during the construction of a mental model and that the activation of alternatives is even stronger in the presence of a focus particle than in its absence.

While the study by Gotzner et al. (submitted) had investigated the impact of focus particles on the immediate comprehension process, in particular through the establishment of an alternative set in the mental model of a discourse, Spalek et al. (2014) investigated whether using a focus particle improved listeners’ memory for alternatives. For doing so, we used the so-called delayed recall paradigm. In its general form, participants are given some material to remember and are later presented with a ‘cue’ upon which they have to recall the previously presented information. In our study, we presented participants with the same items as in Gotzner et al. (submitted) (see example (6)). After participants had listened to ten such items, a

recall phase (the ‘cue’) followed during which they were asked “what was in the fruit bowl?”. In response to this question, they would have to say “cherries, bananas, peaches”. We used a number of filler items such that participants could not simply follow a strategy of memorizing the three items mentioned in the first sentence but rather had to memorize all the information present in the dialogue. The time between presentation and recall was about four minutes.

Participants recalled the focused element more often than the alternatives. However, while their memory for the focused element (here: *peaches*) was not affected by the presence or absence of a focus particle, they recalled the two alternatives (here: *cherries*, *bananas*) significantly more often if a focus particle had been used in the critical sentence than if no focus particle had been used. Somewhat surprisingly, it did not matter whether the particle was inclusive or exclusive. Rather, the mere presence of a focus particle seemed to cause improved memory for alternatives. To make the observed pattern very clear – in those cases, in which participants had heard “only peaches” or “even peaches” in the last sentence, they recalled more often that *bananas* and *cherries* had been mentioned in the first and second sentences than when they heard “peaches” in the last sentence. The improvement in recall was modest (from 71% to 77% in Experiment 1 and from 60% to 64% in a second experiment with a different item structure), but statistically significant.

We interpreted the findings in the following way: Because of their conventional association with focus (Beaver & Clark, 2008), focus particles activate alternatives. That is, in terms of so-called spreading activation models, which are often assumed in psycholinguistics, focus alternatives are activated by means of their relationship to the focused element, and the strength of this coactivation increases if a focus particle is present (though this relationship is also modulated by context, which restricts the available interpretations, cf. Roberts 1995). Elements that are more strongly activated during processing will be called into the listener’s mental model, or at least be more vividly represented than otherwise. Thus, the memory benefit might be a completely “automatic” effect caused by activation spread in the language processing system. However, it could also be the case that the improved memory is not only an automatic by-product, but also discourse functional. It would be advantageous for a listener to remember items if there is an increased probability that these items will reappear in the discourse.

In an as yet unpublished continuation to our data collection, we repeated the delayed recall task from Spalek et al. (2014) but presented all items before asking participants to answer the recall questions. This experiment differed in two important ways from the first one: First, memory load was much higher, because now participants had to remember 50 items before receiving a recall cue; second, the time between presentation and recall was prolonged (the delay between presentation and recall was now half an hour rather than 4 minutes). The results differed in two ways: First (and not surprisingly), participants recalled fewer items than in the first experiment; second, we no longer observed a positive effect of the presence of a focus particle on recall of the alternatives. The only effect that persisted was a main effect

of focus: The focused element was recalled more often than the alternatives. This data pattern suggests that there is a short-term memory advantage for focus alternatives while long-term, there is an advantage for the focused element only. Intuitively, this makes sense – the aim of successful communication should be a representation of the proposition a speaker has uttered – not alternative propositions. More precisely, the alternatives may be relevant to the communication at a certain point in the discourse, but can then be discarded. This can also be seen as ‘popping out’ the highest level of focus alternatives, still in short term memory, from a stack once they are no longer relevant, similarly to the stack of questions that forms the Question Under Discussion in Roberts’ (2012) terms; though of course there is no strict ‘last in first out’ processing for alternatives (we thank Paul Portner and an anonymous reviewer for commenting on this issue).

However, even if communication is primarily concerned with elements actually occurring in propositions, and not their alternatives, it might be the case that in utterances directly following the focused element, speakers are more likely to refer back to an alternative if they had used a focus particle than if they hadn’t. In that case it would also be useful for the listener to retain the alternatives in memory – at least for a little while.

In her dissertation, Kim (2012) reports a number of experiments concerned with the effect of focus particles on discourse processing, in particular the question which expectation a listener will have for the following discourse based on the use of a particular focus particle. The method of choice in these experiments is the visual world paradigm: Participants listen to brief discourses and are presented with a display of four different objects in the four quadrants of a computer screen. By monitoring participants’ eye movements, Kim was able to find out which referents were expected in the upcoming discourse (and how quickly after a focus particle had been processed). Kim reported that using the focus particle *only* strongly biased participants towards expecting a recently mentioned discourse referent after *only*. She found a preference for discourses to be continued with previously mentioned material in all cases, but this bias was much stronger in the case of *only*. Note that while this is a closely related question, it is not the question we are pursuing in the present paper. Kim wanted to know what listeners expected after having processed *only*, given a previous discourse. We want to know what language users produce in the upcoming discourse, given that they or their interlocutors have just produced a focused element preceded by *only*.

While Kim (2012) looked at expectations of the listener, Kaiser (2010) investigated the likelihood that a speaker will continue a given discourse by making reference to alternatives. She used dialogues with corrective focus, that is, dialogues in which an alternative was rejected by a second speaker. Participants were requested to provide a sensible continuation sentence to the presented discourse, imagining themselves to be speaker B. See (7) for an example of her stimuli:

(7) *Speaker A: The maid scolded the bride.*

*Speaker B: No, that’s wrong. It was the [secretary]F that scolded her.*

Kaiser (2010) manipulated both the syntactic construction (cleft constructions vs. canonical SVO sentences) and the syntactic function of the focused element (subject vs. object). Kaiser was interested in a number of different questions; the one most relevant to the present study was which element of the discourse speakers tended to refer back to if they used a full noun phrase in their sentence continuation. She observed that for those conditions in which the focused element was the sentence subject, the continuations most frequently referred to the alternative of the focused entity, for example “the maid wouldn’t have the audacity to scold the bride”. Kaiser concluded that a sentence subject that is corrected still maintains some significance for the upcoming discourse and is therefore more likely to be mentioned again.

Note that the studies by Gotzner et al. (submitted), Kaiser (2010) and Spalek et al. (2014) all used contrastive focus by means of a correction. It is possible that in discourses of the type “A not B”, the rejected alternative remains particularly prominent because a) there must have been a reason why it was mentioned in the first place and b) a need is felt to justify its rejection, thus speakers are likely to come back to it (this is evident in Kaiser’s results). In order to generalize the results from the delayed recall study, in Spalek et al. we replicated the findings with items that had a narrative structure rather than a dialogue and which involved no corrective focus. An example is given in (8), translated into English from the original German:

(8) *Carsten reaches for a basket full of peaches, cherries, and bananas. He wondered what he would like to eat. He only/ even/ \_\_\_ took out the peaches.*

We replicated the finding that alternatives (here: *cherries* and *bananas*) were recalled more often if the focused element had been preceded by a focus particle (‘only/ even the peaches’) than if it had not (‘the peaches’). Thus, a strong contrast is not necessary to keep the alternatives prominent. Still, the question in how far the improved availability of alternatives in the presence of *only* generalizes to all sorts of “real-life” language contexts was one of the motivations for the present study.

### *The present study*

Findings from our own previous work (Spalek et al., 2014) as well as the literature review above suggest that (a) hearers have improved memory for focus alternatives if the focused element was preceded by a focus particle and that (b) this might be advantageous because speakers have a tendency to refer back to an alternative to a focused element in subject position (Kaiser, 2010). Note that Kaiser’s study does not directly address the question we are asking, which concerns the particular effect of conventional association with focus (Beaver & Clark, 2008).

However, findings from experiments often suffer from the short-comings that experimenters only investigate particular structures in a controlled, but less natural laboratory context. In order to study the effects of focus particles on alternative production “in the wild”, we developed an annotation task and carried out a corpus

study, testing the hypothesis that alternatives in naturally occurring data will be referred to more often later in the discourse if the focused element was preceded by the focus particle “only” than if it was not.

## **Annotating corpus data**

### *Identifying alternatives*

In order to look at the effect of focus particles on the likelihood of alternatives in discourse, we must first address the question of identifying what qualifies as an alternative in the context of an annotation task. Generally speaking, alternatives are usually defined as belonging to the same taxonomic category as the focused element (e.g. Braun & Tagliapietra, 2010; Spalek et al., 2014). Although the examples from the experimental literature presented in the previous section may lead us to believe that some alternatives are always viable even without context (e.g. *apples : oranges, child : adult*), moving from the laboratory context to unconstrained corpus data is sure to test this belief, since it means handling a much larger lexical variety.

Is it actually possible to identify a list of prominent alternatives for each noun in a language, independent of context? In the next section, we attempt to find a number of such alternatives, that is, a set of referents that frequently replace one another, independent of context. To anticipate the results, it turned out that this was not possible, or rather that annotators did not agree about which nouns represent alternative pairs. Hence, in a second attempt we tried to determine whether annotators can agree in identifying alternatives given a certain context. As we shall see in the subsequent section, this turns out to be possible, and relevant with regard to the behaviour of focus particles in naturally occurring data.

### *General Alternatives*

To get an idea of what some likely alternative pairs in corpus data might be with no regard to semantic context, we decided to look for overt expressions of alternative semantics. König (1991: 35) states that “where the alternatives under consideration are overtly given, the expressions denoting the alternatives are frequently the focus of another particle”. In his example, an alternative expression is marked using the focalizing *not only ... but also*: “One expects a guide not only to know the terrain, but also to choose good roads and perhaps even to find a few short-cuts.” (ibid.). We therefore extracted from the deWaC Web corpus (1.63 billion tokens of German Web data, see Baroni et al. 2009) pairs of nominal heads in three German constructions that are expected to embed viable alternatives in a similar way:

- *entweder X oder Y*                    ‘either X or Y’
- *sowohl X als auch Y*                ‘both X and Y’
- *nicht nur X sondern auch Y*       ‘not only X but also Y’

In all cases, X and Y were allowed to be simple NPs in the same sentence with optional determiners (articles or possessives) and attributive adjectives, as identified by automatic part-of-speech tagging sequences (using the TreeTagger, Schmid, 1994 and the STTS tagset for German, see Schiller et al., 1999; see also Baroni et al., 2009 on the preparation of the corpus). In order to simplify grouping together similar items, we group together NPs by their head nouns, though as we shall see below, in some cases this leads to problems. Table 1 gives some head lemmas for selected results.

Table 1. Automatically extracted alternative pairs (excerpt).

| X head                       | Y head                    | frequency |
|------------------------------|---------------------------|-----------|
| <i>Recht</i> ‘right’         | <i>Pflicht</i> ‘duty’     | 183       |
| <i>Krokodil</i> ‘crocodile’  | <i>Ding</i> ‘thing’       | 119       |
| <i>Feinstaub</i> ‘fine dust’ | <i>Problem</i> ‘problem’  | 94        |
| <i>Kind</i> ‘child’          | <i>Erwachsene</i> ‘adult’ | 91        |
| <i>Mann</i> ‘man’            | <i>Frau</i> ‘woman’       | 68        |
| <i>Frau</i> ‘woman’          | <i>Mann</i> ‘man’         | 59        |
| <i>Erwachsene</i> ‘adult’    | <i>Kind</i> ‘child’       | 26        |

The items in the table illustrate some of the difficulties with a contextless approach, which, as we will suggest below, is not quite viable. On the one hand, some common oppositions like ‘rights’ and ‘duties’ are captured, which appears promising. On the other hand, some contexts are essential to understanding alternatives, or at least the entire NP must be considered, especially when the qualifier ‘other’ is used: ‘fine dust’ and ‘problem’ are not ordinarily perceived as alternatives, but in all cases the NP internal context was the adjective ‘other problems’, making it clear that ‘fine dust’ is only an alternative in its aspect as being a subtype of problem. In other cases, the noun semantics are so underspecified that much more context is required to establish the alternative. For ‘crocodiles’ versus other ‘things’, the context is established by the following text in the corpus, which happens to be repeated and quoted over a hundred times in the Web data (a problem of non-exact duplicate document recognition):

- (9) *Nicht nur ausgestopfte **Krokodile** können bei der Einreise nach Deutschland Ärger am Zoll verursachen, sondern auch weitaus unauffälligere **Dinge***  
 ‘Not only stuffed **crocodiles** can cause problems with customs when entering Germany, but also much more inconspicuous **things**’

In this context it is clear how crocodiles are contrasted with other things, but ‘things’ in general are not expected to be a salient alternative in the mind of hearers when crocodiles are mentioned. However even considering the entire NP in cases like these is not very useful: without the special context of smuggling things through customs, ‘other things’ are not a conventional general alternative to ‘stuffed crocodiles’.

It should also be noted that the relationship between X and Y may be more or less symmetrical, even in cases where the alternative relationship is uncontested. The

quantities for ‘woman’ vs. ‘man’ suggest that both orders of X and Y are similarly expected in language, but for ‘child’ vs. ‘adult’, children are much more likely to occur first in our data. This is likely to be non-coincidental: the comparison to adults is usually discussed when children are the topic of discourse, at least for the writers in this corpus, who are by and large expected to be adults themselves.

While these problems are substantial, it is possible that human judgment can agree as to which of the pairs suggested above are valid alternatives independent of context (*women* vs. *men*: yes; *crocodiles* vs. *things*: no). In order to evaluate this proposition, we conducted an annotation experiment in which three annotators evaluated the alternatives detected in the search above, for nouns occupying the position X and having at least three different alternatives, each of which occurred at least five times. This was decided in order to ensure we were working on more conventionalized alternatives, in so far as these can be shown to exist, rather than ad hoc oppositions, and that the nouns were not fixed expressions, which only admit very few alternatives despite being frequent, or the result of duplicate sentences (e.g. the case of ‘crocodile’). These criteria resulted in 109 nouns with 648 alternatives (average of 5.94 alternatives per noun) being evaluated, some examples of which are given in Table 2.

Table 2. Automatically suggested alternatives for human evaluation.

| node word (X)                    | alternative suggestions (Ys) | total frequency |
|----------------------------------|------------------------------|-----------------|
| <i>Mensch</i> ‘person’           | 45                           | 1162            |
| <i>Möglichkeit</i> ‘possibility’ | 15                           | 725             |
| <i>Kind</i> ‘child’              | 13                           | 666             |
| <i>Recht</i> ‘right’             | 8                            | 560             |
| <i>Frage</i> ‘question’          | 11                           | 542             |
| <i>Entwicklung</i> ‘development’ | 8                            | 522             |
| <i>Problem</i> ‘problem’         | 11                           | 514             |
| <i>Information</i> ‘information’ | 7                            | 492             |
| <i>Interesse</i> ‘interest’      | 11                           | 475             |
| <i>Qualität</i> ‘quality’        | 13                           | 447             |
| ...                              | ...                          | ...             |

Although it can be assumed that all or nearly all of the alternative suggestions were valid alternatives in context (on account of the highly selective constructions used to retrieve them), annotators unanimously agreed to reject 443 / 648 suggestions as not qualifying for general alternative status. However, they only unanimously agreed on including 31 cases as representing context-independent alternatives of the type ‘man’ : ‘woman’. This means that the total rate of unanimous agreement is 73.14%, but that positive alternatives were difficult to identify unequivocally. Figure 1 gives the breakdown of votes for the evaluation (from 0 votes, not an alternative, to 3 votes, unanimously an alternative). Solid areas are unanimous votes, while dashed areas represent alternatives identified by only one or two of the three annotators.

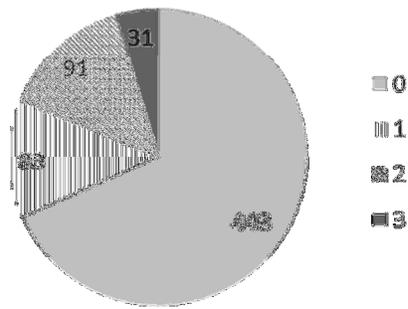


Figure 1. Votes for word pairs forming an alternative out of context.

The distribution of the disagreements within possible alternatives (dashed areas) was fairly unpredictable amongst the annotators, leading to a kappa value of  $\approx .402$ , which can be considered rather low (cf. Artstein & Poesio, 2008). This is partly due to the fact that kappa takes into account the probability of chance agreement, which is rather high in this case (binary decision with overwhelmingly negative cases – most noun pairs are not alternatives out of context). For comparison, part-of-speech tagging often achieves values of .9, while more difficult tasks, such as annotation of information structure, range between .6-.8 for information status (e.g. nominal phrases being given, accessible or new) and .4-.6 for focus annotation (see Lüdeling et al., to appear for an overview).

These results indicate that while there is substantial agreement between annotators on prototypical cases, many alternatives are not identified in the absence of context, and that there is room for variation in at least a quarter of cases. It therefore appears unlikely that we will be able to produce concise lists of alternatives to specific nouns without consulting context, in order to evaluate large amounts of data with and without focus particles automatically. In the next section we therefore move on to consider the identification of alternatives in context and the formulation of a testable hypothesis to answer the question of focus particles and their relationship to the likelihood of alternatives.

### *Alternatives in context*

Although the treatment of alternatives in context brings a substantial challenge to the task of alternative identification (we cannot expect to find a definitive list of alternatives to each word), there are good theoretical and experimental reasons to prefer it. Allowing a broad range of alternatives in a specific discourse situation corresponds to the permissive view of focus alternatives suggested by Rooth (1992). König (1991: 35) also emphasizes the role of the specific discourse situation by stating that “the selection of alternatives is highly context-dependent”. A psycholinguistic model grounded in the notion of spreading-activation would be hard to reconcile with a non-context sensitive approach. Although spreading activation as a concept is indiscriminate in the first instance, context interacts with activation based

on the stimulus in question (cf. Abdel Rahman & Melinger, 2009) – conceivably by changing resting level activation or altering connection weights. Empirically in the context of psycholinguistic studies on alternatives, Gotzner (2014) was able to show the relevance of context in the evaluation of alternatives unrelated to the focused item. She reanalysed the reactions to the unrelated probes in the probe recognition study from Gotzner et al. (submitted). She subdivided the unrelated items into those which could replace the focused element in the target utterance and those which could not. For example, in a sentence like “Matthias has bought [trousers]<sub>F</sub>”, the unrelated probe *lychees* can be substituted in the position corresponding to the focused phrase. By contrast, in the case of a sentence like “Carl has caught [flies]<sub>F</sub>”, *sofas* (the unrelated probe) cannot sensibly be substituted for the focused phrase. In the original analysis, it took longer to reject unmentioned alternatives (i.e., another piece of clothing for the target *trousers*, e.g., *shirts*) than unrelated probes. We had argued that the unrelated probes could quickly be discarded because they did not ‘feel as if they might have been mentioned’. However, in Gotzner’s (2014) novel analysis, it turned out that the unrelated probes actually formed two distinct groups: Contextually plausible substitutions (e.g., *lychees* in the case of Matthias buying [trousers]<sub>F</sub>) patterned along with unmentioned alternatives, that is, words belonging to the same hyperonym as the focused element. By contrast, contextually implausible substitutions (e.g., *sofas* in the case of Carl catching [flies]<sub>F</sub>) did not pattern along with the unmentioned alternatives, instead, they were quickly rejected. Thus, items which were sensible substitutions of the focused phrase were not easy to reject in the probe recognition task, because apparently, participants perceived them as good alternatives and had difficulty in deciding whether or not they were part of the mental model. The context-dependency of alternative sets makes the question all the more crucial: are we able to identify the occurrence of alternatives in natural contexts?

In order to design an annotation experiment for alternatives to complement the laboratory findings in Spalek et al. (2014), several factors had to be considered. Firstly, the choice of node nouns needed to be constrained, ideally in a way closely matching the laboratory data. We therefore took nouns from the same classes used in the experimental contexts, but extracted spontaneous corpus examples from deWaC for their usage with and without focus particles. We selected two intuitively prototypical members for each class, as summarized in Table 3, and looked for occurrences of the relevant lemmas. Lemmas are used to group together inflected forms of the same lexicon entry, thus the item *Schuh* encompasses both singular *Schuh* ‘shoe’ and plural *Schuhe* ‘shoes’, as well as inflected forms like *Schuhen* ‘shoes (dative)’. Note that the word *Arzt* ‘doctor’ in German is only used for physicians, and is not ambiguous with the title ‘doctor’.

Table 3. Categorized node words for annotations of alternatives in context

| category    | node words                                    |
|-------------|---|
| clothing    | <i>Hose</i> ‘pants’, <i>Schuh</i> ‘shoe’      |
| consumables | <i>Wasser</i> ‘water’, <i>Obst</i> ‘fruit’    |
| professions | <i>Lehrer</i> ‘teacher’, <i>Arzt</i> ‘doctor’ |
| body parts  | <i>Herz</i> ‘heart’, <i>Kopf</i> ‘head’       |
| animals     | <i>Hund</i> ‘dog’, <i>Katze</i> ‘cat’         |

The nouns have varying degrees of potential polysemy and represent a broad spectrum of syntactic and morphological types (genders, count/mass, inalienable possessions etc.), but since we are interested in the behavior of data in a naturalistic setting, we view this as largely desirable (though see below on some properties of specific nouns).

A second important consideration was what to consider as context. Since we are interested in the likelihood that alternatives will occur in the immediately following discourse, we chose to look at pairs of adjacent sentences in which the first sentence contains the node word, with or without the word *nur* ‘only’ modifying the node’s phrase. A study of larger discourse contexts is left as a possible point for further research. Sentences are segmented automatically in deWaC based on simple punctuation (periods, exclamation and question marks). For each node noun with and without *nur* ‘only’, 20 random pairs were extracted, resulting in 40 pairs for each of the ten nouns, or 800 sentences (400 pairs) to be annotated in total. Pairs were rejected from the sample if they were: a. a verbatim repetition of another pair; b. if *nur* did not associate with the node noun; or c. if any of the ‘sentences’ did not constitute a normal predication, e.g. when it corresponded to a nonsensical alphanumeric string, but also in the case of headings, e.g.:

- (10) [*Im Aufsichtslabor liegt eine Liste geeigneter Ärzte vor.*]<sub>S1</sub> [*Pflicht zur Ersten Hilfe*]<sub>S2</sub>  
 ‘[There’s a list of appropriate doctors in the supervision lab.] [Obligation for first aid]’

In many cases it appears that the second unit is the beginning of a new discourse unit, quite possibly written by a different author at a different time, which led to the exclusion of such pairs. We evaluated agreement on pair exclusion according to the criteria, again using three annotators, and received 83% exact agreement, with two annotators rejecting only 1 case against the opinion of the third for principled reasons (this case involved what was likely a heading that occurred between the two propositions containing the node word and alternative candidate). The remaining disagreements were caused by unclear instructions in the event that the second ‘sentence’ contained multiple propositions not properly separated by punctuation. It was decided to allow the latter by considering only context from the first such proposition as a possible locus for alternatives.

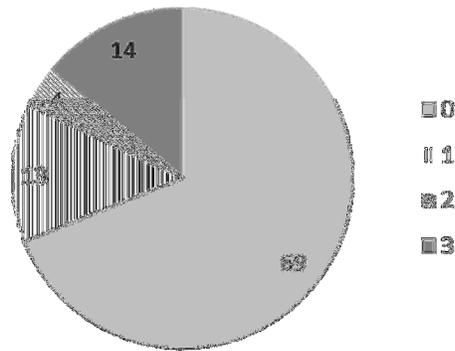


Figure 2. Agreement on the validity of sentence pairs in terms of votes to exclude a pair from three separate annotators.

For sentence pairs judged to be admissible, the three annotators were instructed to intuitively judge where, in the context of the relevant sentences, alternatives to the node word from S1 could be found in S2. To test this intuition, S1 could be supplemented introspectively by adding ‘as opposed to [ALTERNATIVE]’ (e.g. ‘oranges’ contrast with ‘apples’ if we can add in context: ‘John likes oranges’ + ‘as opposed to apples’). Alternatives were tagged at the position of the head noun of the phrase containing the alternative (though see below on the possible inclusion of pronouns).

100 random sentence pairs were used to evaluate annotators’ ability to agree on the identification of alternatives. The three annotators identified 19, 20 and 26 alternatives respectively, agreeing completely on the presence of the alternatives (0 or 3 votes for each alternative in a sentence pair) in 80% of cases. Disagreement was not sporadic, with mostly single annotators not marking an alternative found by two others, and a kappa value of  $\kappa=.86$  was reached. This can be considered quite high, especially for a novel annotation task.

Some of the disagreements seem to result from narrow vs. broad alternative readings, while others could be resolved by refining the guidelines. For example, two out of three annotators identified the alternative candidates marked in bold in (11-12) (some of the context has been omitted to save space), but only one annotator identified (13):

- (11) [*An der Haltestelle (...) riss einer der Täter dem Mädchen das Basecap vom **Kopf** (...)*]<sub>S1</sub> [*(...) einer kam aber zurück und versuchte dem Opfer ins **Gesicht** zu treten*]<sub>S2</sub>  
 ‘[At the station (...) one of the culprits ripped the baseball cap from the girl’s **head** (...)] [(...) but one of them came back and tried to kick the victim in the **face**]’

- (12) [(...) *habe einfach mal **Obst** gekauft (...)*]<sub>S1</sub> [*Auch neue **Gerichte** für die Küche machen echt spass.*]<sub>S2</sub>  
 ‘[ (...)I just bought **fruit** for once (...)] [New **dishes** for the kitchen are really fun too.]’
- (13) [*Nur meine **Katzen** fressen Kuhragout aus Dosen.*]<sub>S1</sub> [***Ich** werd niemals dick und rund weil ich mich gut ernähr.*]<sub>S2</sub>  
 ‘[Only my **cats** eat cow ragout out of cans.] [I will never become fat and round because I nourish myself well.]’

In (11), it’s clear that the baseball cap would not have been ripped from the girl’s face (unless possibly she was wearing it to cover her face); instead, an opposition of head to face is interpretable in the somewhat wider context of body parts that the culprits could do violence to, and the two are semantically and physically related. Similarly in (12) new dishes cannot be bought in the same manner as fruit (narrow reading in the context of the predicate ‘buy’), but buying fruit may well be an alternative to cooking a new dish. It therefore appears that for some cases the extent of the phrase associated with the particle matters, and an annotation of alternatives may have to refer to a more complex, hierarchical analysis of the sentences. In other cases, like (13), the problem is more straight-forward: it is clear that eating ragout from a can oneself is an alternative to cats doing so, but it may not be clear to an annotator if a pronoun such as ‘I’, while not being a lexical noun, is a candidate for an alternative. Arguably for annotation in context it should be, which can easily be added to the guidelines specifically.

Our reaching a relatively high kappa value and the fairly consistent annotation results led us to use the annotation of alternatives in context as the basis for the evaluation of the relationship between alternatives and the presence of a focus particle.

## Results for alternatives and the Focus Particle *nur* ‘only’

### *Alternative density in the data*

In addition to the triply annotated 100 sentence pairs above, the remaining 300 sentence pairs were divided into three random subsets for individual annotation by each of the annotators. The sentences contained an equal amount of examples with and without *nur* for each node noun, though 76 pairs were rejected as invalid as described above. With these removed, the dataset for the evaluation contains 324 sentence pairs, of which 161 were with *nur* and 163 without, but these were no longer divided equally between the test items. In order to receive a balanced design, we therefore added additional sentence pairs from the corpus until we had 400 valid pairs divided equally across the different conditions.

We also inspected the data manually to measure the proportion of contrastive foci in the data and estimate that only 1.25% of cases can be found (i.e. 5 sentence pairs;

we thank an anonymous reviewer for the suggestion to extract this information), which exhibit a clearly contrastive structure, though we note that focus annotation is known to be unreliable (see Lüdeling et al., to appear, for an overview). The low ratio itself is not surprising given other corpus data for German, for example only 2.7% as annotated in the Potsdam Commentary Corpus (Stede 2004), which contains argumentative texts such as newspaper editorials. Although it is very possible that contrastivity interacts with frequency of alternatives, the amount of data is too small to draw any conclusions in this regard.

Within these 400 pairs we proceeded to count the occurrences of alternatives, or in cases where annotators recognized different amounts of alternatives in the multiply annotated pairs we took the majority vote (there were no cases of a 3-way split in counts). This resulted in the data in Table 4.

*Table 4.* Alternatives in sentence pairs with and without *nur* ‘only’.

| condition          | pairs | alternatives |
|--------------------|-------|--------------|
| with <i>nur</i>    | 200   | 80           |
| without <i>nur</i> | 200   | 45           |

The results indicate that, in the observed data, the density of alternatives in the *nur* pairs is judged to be .4, as opposed to .225 in pairs without *nur*, a highly significant difference ( $p=.0002448$ ,  $\chi^2(1,N=400)=13.4516$ ). Note that this is not the proportion of pairs containing alternatives, but rather the expected amount of alternatives per sentence pair, which we will call ‘alternative density’. We consider this to be a more accurate measure of the influence of the focus particle on the presence of alternatives than a binary indication of whether or not alternatives were observed. In other words, the presence of *nur* is significantly connected to an observed increase of over 77% in the amount of alternatives per sentence pair.

#### *A mixed effects model*

The observed results above do not take into account a variety of factors that may covary with the presence of alternatives as identified in this study: for example, there could be influences from the identity of the annotator (more or less conservative annotation practice), the class that the word belongs too, the word itself, or the length of the sentences in the pair (the latter is to be expected, since there are more chances to realize an alternative in a longer sentence). We therefore constructed a multifactorial model of the data using generalized linear mixed-effects using the `glmer()` function in the R library `lme4`.

In the model selection process we considered the factors ‘word class’, ‘sentence length’ (for S1 and S2), and the occurrence of the focus particle to be fixed effects, and modeled the identity of the annotator and the specific word being examined as random effects, since the model should remain valid for other annotators and words (though we recognize classes of words such as professions versus clothes might be systematically differently inclined to promote discussion of alternatives). Table 5

presents the results for the best model, which is discussed below. Since there were no notable correlations between the fixed effects, we have left correlations out of the model description.

Table 5. Generalized linear mixed-effects model (family: poisson (log)) for the occurrence of alternatives in sentence pairs.

| <b>Model quality:</b>                              |                    |                   |                 |                    |     |
|--|--------------------|-------------------|-----------------|--------------------|-----|
| <b>AIC</b>   | <b>BIC</b>         | <b>logLik</b>     | <b>deviance</b> |                    |     |
| 536.2521   | 576.1667           | -258.126          | 516.2521        |                    |     |
| <b>Random effects:</b>                             |                    |                   |                 |                    |     |
| <b>Groups</b>                                      | <b>Name</b>        | <b>Variance</b>   | <b>Std.Dev.</b> |                    |     |
| <i>word</i>  | <i>(Intercept)</i> | 1.69E-12          | 1.30E-06        |                    |     |
| <i>annotator</i>                                   | <i>(Intercept)</i> | 2.12E-12          | 1.46E-06        |                    |     |
| Number of obs: 400, groups: word, 10; annotator, 4 |                    |                   |                 |                    |     |
| <b>Fixed effects:</b>                              |                    |                   |                 |                    |     |
|  | <b>Estimate</b>    | <b>Std. Error</b> | <b>z score</b>  | <b>Pr(&gt; z )</b> |     |
| <i>(Intercept)</i>                                 | -2.68855           | 0.365979          | -7.346          | 2.04E-13           | *** |
| <i>nurY</i>  | 0.544404           | 0.188024          | 2.895           | 0.00379            | **  |
| <i>classbodypart</i>                               | -0.18344           | 0.358179          | -0.512          | 0.60855            |     |
| <i>classclothes</i>                                | -0.19561           | 0.365892          | -0.535          | 0.59291            |     |
| <i>classconsumable</i>                             | 0.92296            | 0.28589           | 3.228           | 0.00124            | **  |
| <i>classprofession</i>                             | 0.578334           | 0.28568           | 2.024           | 0.04293            | *   |
| <i>lenS1</i>                                       | 0.000228           | 0.008158          | 0.028           | 0.97769            |     |
| <i>lenS2</i>                                       | 0.042894           | 0.006333          | 6.773           | 1.26E-11           | *** |

Because the number of alternatives is not normally distributed and has a lower bound of 0 (there cannot be fewer than 0 alternatives, and this is the mode value), we chose the Poisson distribution family, which closely matches the form of the data, as illustrated in Figure 3.

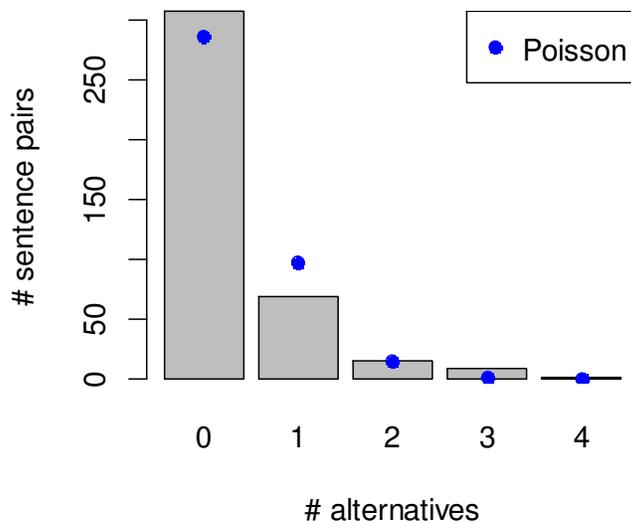


Figure 3. Random Poisson distribution superimposed on the observed alternative distribution for 400 elements with identical distribution means.

As Table 5 shows, the random effects account for a very small amount of the variance. This indicates that annotator bias and word specific behavior above and beyond word class are rather inconsequential, though it should be noted that the groups for each random effect have very few members. An evaluation of more words in each class is still needed to explore within class variation.

The fixed effects show that the most significant factor is the length of S2 ( $p=1.26e-11$ ), with a large positive, length-dependent coefficient. This is to be expected as discussed above and would only be surprising if it applied more or less to cases with or without *nur*. We therefore tested the model against a model also including an interaction of S2 length and focus particle occurrence; this model was not significantly better than the simpler model with no interaction. The length of S1, as expected, did not turn out to have a significant effect.

The second most significant effect is that of the class ‘consumable’ ( $p=.00124$ ), which is also linked to the fourth significant effect for the class ‘profession’ ( $p=.04293$ ), which taken together indicate that the classes are not equally likely to promote the occurrence of alternatives. Intuitively, the annotators also felt that discussions about doctors and teachers as well as what people eat and drink immediately lent themselves to the discussion of alternatives (especially patients and children as contrasted to the former, and alternative nutrition choices for the latter). This effect too is independent of the condition on the appearance of *nur*, and incorporating an interaction into the model did not improve it significantly.

Finally, the factor ‘nurY’ (occurrence of the particle *nur*) was significant in promoting alternatives ( $p=.00379$ ), an effect that was not eliminated by considering the multifactorial model. The effect of *nur* can be seen across word classes, as illustrated in Figure 4 (some jitter has been introduced into the alternative count to make the amount of points visible).

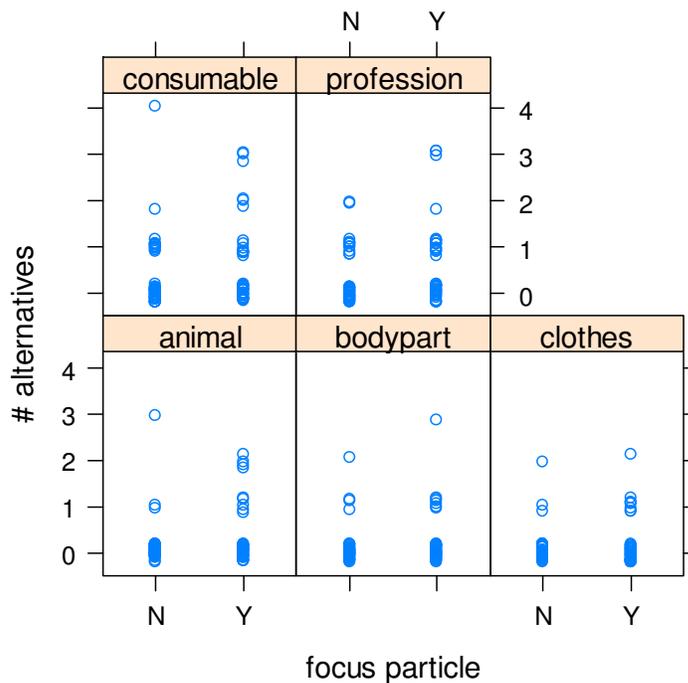


Figure 4. Number of alternatives under the conditions focus particle: Y/N divided by word class groups.

As the figure shows, in every group alternatives tend to cluster higher on the right side (particle=Y) than on the left (N).

## Discussion

The results of our study raise two important points for modelling alternatives. On the one hand, the evaluation of the annotation task for general alternatives suggests that talking about which sets of nouns function as alternatives in a mental model (either for a single test subject or for a language ‘at large’ based on multiple raters) is difficult and may involve too much subjective judgment. While we are able to come up with some alternatives that produce oppositions in an experimental context, it seems doubtful that human annotators will be able to reach a high level of agreement for arbitrary items, even for those that appear most frequently with overt alternative marking in corpus data. On the other hand, our results converge directly with psycholinguistic experiments on the reality of alternative sets and the effect of the focus particle, suggesting for example that whatever is viewed as an alternative to focus information in context is not only ‘redundantly’ activated in the mental model as evidenced by memory and reaction time effects, but also substantially more likely to be mentioned in the immediate subsequent context. In other words, the activation found in the laboratory setting is quite likely related to speaker’s subsequent behaviour after use of focus particles in the naturalistic corpus data setting. The

validity of these findings gains added credence from the fact that annotators could agree rather well about which nouns were to be seen as alternatives, and the fact that the effect was robust in a multifactorial evaluation of the data.

The finding that language users are more likely to refer back to an alternative if the focus particle ‘only’ has been used converges with the results of the production study by Kaiser (2010) and extends them to unconstrained language use as observed in a large, randomized dataset of texts crawled from the World Wide Web. Kaiser looked at cases in which an alternative was discarded using an explicit correction (“No, that’s wrong ...”), whereas we investigated the impact of the focus particle ‘only’ which asserts that none of the alternatives to the focus constituent are true. In both cases, Kaiser’s conclusion that discarded alternatives still maintain some significance for the upcoming discourse can be upheld.

From the psycholinguistic perspective, it is not surprising that naturalistic, contextual alternatives are likely to be mentioned in subsequent discourse. Kim (2012) shows evidence that participants tend to expect discourse continuations (in her case: with the focused element, not alternatives) to share a conceptual category with recently mentioned discourse content, and as discussed above, alternatives typically come from the same conceptual category. Usually, such categories are operationalized in psycholinguistic research as taxonomic categories, that is, cohyponyms of a given hyperonym (e.g., apples, pears, and cherries for FRUIT, or shirts, trousers, and jackets for CLOTHING). However, in her Experiment 5, Kim investigated whether focus alternatives could also stem from “ad-hoc categories” that are relevant for a particular context (cf. Barsalou, 1983) such as pens, olive oil, and shoe laces for “things I need to buy”. Just as Gotzner (2014), Kim (2012) showed that alternatives need not necessarily come from the same taxonomic category but rather from such ad-hoc categories, and this type of alternative is part of the data that contextualized alternative annotation can capture.

Above we have also suggested alternative density as a more refined metric than binomial alternative counts (present vs. not present). However the other possibility, modelling whether the focus particle effect is also found in a binomial family model representing only presence vs. absence of alternatives, should also be considered (we thank an anonymous reviewer for pointing this out). Indeed, the model remains significant with precisely the same factors at the same alpha thresholds for a binomial response, except that the noun class ‘profession’ rises in significance from alpha .05 (\*) to .005 (\*\*) (p-value .00316). Although both the binomial and the Poisson family models are tenable, we feel that the binomial model is less desirable, since it discards information which seems relevant from a cognitive perspective. One could argue that discussing multiple alternatives indicates more activation in networks that contrast with the node word than a single alternative, all other things being equal.

Measuring a scalar alternative density would normally be complicated by an interaction with sentence length in natural data (longer sentences can present more alternatives), but the regression model in the previous section allows us to separate this effect and make the density measure more useful. Alternative density gives us additional information that was not obtainable in the context of previous experiments,

since respondents are neither guided nor constrained in the form of continuation we find in the corpus data, whereas the experimental contexts generally targeted a specific response with a non-variable number of alternatives. Based on the significant effect of the focus particle for alternative density, our results suggest that a particle like ‘only’ not only raises the prominence or probability of any one alternative, but also correlates with a rise in the number of subsequent alternative mentions. The corpus data also provide evidence that the context need not be contrastive in order for this rise in the number of alternative mentions to happen, as the data contained overwhelmingly new information foci.

Finally, the findings support our hypothesis that it might actually be advantageous to maintain a stronger short-term representation of alternatives if a focus particle has been used: The memory benefit observed for alternatives in an utterance including a focus particle compared to an utterance not including a focus particle (Spalek et al., 2014), is potentially helpful in processing upcoming discourse because a likely referent will be more salient in the mental model, facilitating the comprehension process for the listener. For many purposes it is also impossible to interpret what ‘only X’ means accurately without considering the alternatives that are being excluded. Whether alternatives are more likely after focus particles because of the structure of the cognitive architecture, or whether a habituation effect is involved that is caused by experiencing alternatives more frequently after focalized NPs is something of a chicken-and-egg problem – a context in which we speak of ‘only X’ is immediately also a context in which alternatives to X can be discussed. But since alternatives can always be discussed, the rise in alternative density following the focus particle is a non-trivial and even surprising finding. By considering alternative density, which shows not only a qualitative (alternatives used or not) but also a quantitative effect (more alternatives), we see that focus particles can signal to the listener that alternatives are likely to be discussed next, even though the null hypothesis should be that alternatives may be discussed with or without the particle.

We believe that further studies on the presence of a habituation component are needed, for example by comparing experimental effects in probes that are more or less expected to appear with alternatives based on corpus data (does adding a particle always exhibit an effect? Are some nouns more ‘alternative happy’, and if so which and why? Can we predict this in an experimental setting based on training data from a corpus?). Both in corpus data and in the laboratory, we expect the issues to be hard to disentangle, but we believe a converging evidence approach is likely to be stronger than applying any one method in isolation.

## **Acknowledgments**

Both authors contributed equally to this paper. This work was supported by the German Research Foundation (DFG) as part of Collaborative Research Centre 632 on Information Structure (projects C7 and D1). We would also like to thank Linda Giesel for her help with the annotation experiments, as well as Laura Michaelis, Paul Portner and two anonymous reviewers for valuable comments on previous versions of this paper.

## REFERENCES

- Abdel Rahman, R. & Melinger, A. (2009). Semantic context effects in language production: A swinging lexical network proposal and a review. *Language and Cognitive Processes* 24(5), 713–734.
- Artstein, R. & Poesio, M. (2008). Inter-coder agreement for computational linguistics. *Computational Linguistics* 34(4), 556–596.
- Baroni, M., Bernardini, S., Ferraresi, A. & Zanchetta, E. (2009). The WaCky wide web: A collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation* 43(3), 209–226.
- Barsalou, L. W. (1983). Ad hoc categories. *Memory and Cognition* 11(3), 211–227.
- Beaver, D. I. & Clark, B. Z. (2008). *Sense and Sensitivity: How Focus Determines Meaning*. Oxford: Wiley-Blackwell.
- Braun, B. & Tagliapietra, L. (2010). The role of contrastive intonation contours in the retrieval of contextual alternatives. *Language and Cognitive Processes* 25(7/8/9), 1024–1043.
- Byram-Washburn, M. (2013). *Narrowing the Focus: Experimental Studies on Exhaustivity and Contrast*. PhD Thesis, University of Southern California. Available at: [http://www-scf.usc.edu/~byram/on\\_the\\_nature\\_of\\_the\\_focus\\_set\\_of\\_alternatives.pdf](http://www-scf.usc.edu/~byram/on_the_nature_of_the_focus_set_of_alternatives.pdf).
- Cutler, A. & Fodor, J. A. (1979). Semantic focus and sentence comprehension. *Cognition* 7, 49-59.
- Fraundorf, S. H., Watson, D. G. & Benjamin, A. S. (2010). Recognition memory reveals just how CONTRASTIVE contrastive accenting really is. *Journal of Memory and Language* 63(3), 367–386.
- Ganushchak, L. Y., Konopka, A., E., & Chen, Y. (2014). What the eyes say about planning of focused referents during sentence formulation: a cross-linguistic investigation. *Frontiers in Psychology* 5, doi: 10.3389/fpsyg.2014.01124.
- Gernsbacher, M. A. & Jescheniak, J. D. (1995). Cataphoric devices in spoken discourse. *Cognitive Psychology* 29, 24-58.
- Glenberg, A., Meyer, K. & Lindem, K. (1987). Mental models contribute to foregrounding in text comprehension. *Journal of Memory and Language* 26, 69-83.
- Gotzner, N. (2014). What's included in the set of alternatives? Psycholinguistic evidence for a permissive view. In *19th Annual Conference Sinn und Bedeutung*. Göttingen, Germany.
- Gotzner, N., Wartenburger, I. & Spalek, K. (submitted). *The impact of focus particles on the recognition and rejection of contrastive alternatives*.
- Husband, E. M. & Ferreira, F. (submitted). *The role of selection in the comprehension of focus alternatives*.
- Ito, K. & Speer, S. R. (2008). Anticipatory effects of intonation: Eye movements during instructed visual search. *Journal of Memory and Language* 58(2), 541–573.
- Kaiser, E. (2010). Investigating the consequences of focus on the production and comprehension of referring expressions. *International Review of Pragmatics* 2, 266–297.

- Kim, C. (2012). *Generating Alternatives: Interpreting Focus in Discourse*. PhD Thesis, University of Rochester.
- König, E. (1991). *The Meaning of Focus Particles: A Comparative Perspective*. London: Routledge.
- Krifka, M. (2008). Basic notions of information structure. *Acta Linguistica Hungarica* 55, 243–276.
- Ladd, D. R. & Schepman, A. (2003). "Sagging transitions" between high pitch accents in English: Experimental evidence. *Journal of Phonetics* 31, 81–112.
- Lüdeling, A., Ritz, J., Stede, M. & Zeldes, A. (to appear). Corpus linguistics and information structure research. In C. Féry & S. Ichihara (eds.), *The Oxford Handbook of Information Structure*. Oxford: Oxford University Press.
- Norris, D., Cutler, A., McQueen, J. M. & Butterfield, S. (2006). Phonological and conceptual activation in speech comprehension. *Cognitive Psychology* 53(2), 146–193.
- Pierrehumbert, J. B. & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. Cohen, J. Morgan & M. Pollack (eds.), *Intentions in Communication* (pp. 271–311). Cambridge MA: MIT Press.
- Roberts, C. (1995). Domain restriction in dynamic semantics. In E. Bach, E. Jelinek, A. Kratzer & B. H. Partee (eds.) *Quantification in Natural Languages* (pp. 661–700). Dordrecht: Kluwer.
- Roberts, C. 2012. Information structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics* 5(6), 1–69.
- Rooth, M. (1985). *Association with Focus*. PhD Thesis, University of Massachusetts.
- Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics* 1, 75–116.
- Schiller, A., Teufel, S., Stöckert, C. & Thielen, C. (1999). *Guidelines für das Tagging deutscher Textcorpora mit STTS*. Universität Stuttgart, Institut für maschinelle Sprachverarbeitung and Universität Tübingen, Seminar für Sprachwissenschaft.
- Schmid, H. (1994). Probabilistic part-of-speech tagging using decision trees. In *Proceedings of the Conference on New Methods in Language Processing* (pp. 44–49). Manchester, UK. Available at: <http://www.ims.uni-stuttgart.de/ftp/pub/corpora/tree-tagger1.pdf>.
- Selkirk, E. (2002). Contrastive FOCUS vs. Presentational focus: Prosodic evidence from the right node raising in English. In B. Bel & I. Marlin (eds.), *Speech Prosody 2002: Proceedings of the First International Speech Prosody Conference* (pp. 643–646). Université de Provence: Laboratoire Parole et Langage.
- Spalek, K., Gotzner, N. & Wartenburger, I. (2014). Not only the apples: Focus sensitive particles improve memory for information-structural alternatives. *Journal of Memory and Language* 70, 68–84.
- van Leeuwen, T. M., Lamers, M. J. A., Petersson, K., M., Gussenhoven, C., Rietveld, T., Poser, B. & Hagoort, P. (2014). Phonological markers of information structure: An fMRI study. *Neuropsychologia*, 58, 64-74.
- Watson, D. G., Tanenhaus, M. K. & Gunlogson, C. A. (2008). Interpreting pitch accents in online comprehension: H\* vs. L+H\*. *Cognitive Science* 32, 1232–1244.

Weber, A., Braun, B. & Crocker, M. W. (2006). Finding referents in time: Eye-tracking evidence for the role of contrastive accents. *Language and Speech* 49(3), 367–392.