Marta Wierzba

2022

A Paradigm for Testing Contrastive Topic Projection

A Paradigm for Testing Contrastive Topic Projection

Marta Wierzba

University of Potsdam

wierzba@uni-potsdam.de

1   Introduction

The relation between information-structural categories and their prosodic cues is not one-to-one. In German, the main phonological correlates of focus are prominence and a certain type of pitch accent, realized by virtue of increased intensity, pitch, and/or duration of a single syllable. What is interpreted as focused is however typically not a syllable or a single word, but often a larger unit, e.g., a DP. For example, in (1), focus on the subject phrase *die Katze* 'the cat' would be expressed by a nuclear pitch accent on the first syllable of the noun *Katze* 'cat', whereas focus on the object phrase *den Vogel* 'the bird' would be expressed by a nuclear pitch accent on the first syllable of *Vogel* 'bird'. In a strict sense, this already is a case of 'focus projection' (a term I am adopting from Höhle, 1982, and Selkirk, 1984; see also Chomsky, 1971; Jackendoff, 1972, for an early description of the phenomenon): a small unit is overtly marked for a certain property, and this property 'spreads' to a larger part of the structure (from a noun to a DP). If the word *Vogel* 'bird' carries a nuclear pitch accent in (1) (as indicated by the small caps), the focus interpretation can spread to an even larger than the object DP: the verb phrase or the whole clause can be interpreted as focused, i.e., with a nuclear accent on the object, the utterance is ambiguous with respect to focus.

(1)   *Die Katze   hat den* VOGEL *beobachtet.*
       the cat.nom has the bird.acc watched
       'The cat watched the BIRD.'

A similar observation can be made for contrastive topics (CTs): in (2), the noun *Vogel* carries the prosodic cue of this information-structural category (a fall-rise accent, represented by √), but intuitively, a larger constituent—at least the verb phrase—can be interpreted as the CT.

(2)   *Den* √*Vogel hat die Katze*   MORGENS   *beobachtet...*
       the   bird.acc has the cat.nom in.the.morning watched

CT projection is understudied in comparison to focus projection. There has been less theoretical and empirical work on the question in which way projection can proceed in this case: can the intuition that a fall-rise accent on the object licenses interpreting the whole VP as a CT be confirmed in controlled experiments? Can the interpretation even spread to the whole clause, similar to focus projection? Also, no standard paradigm has been established yet for studying this question. While focus projection can be investigated systematically using question-answer pairs, the semantics of CTs involve more elaborate discourse structures that are more challenging to control in an experimental setting.

In this paper, I first provide theoretical background about the semantics and prosody of foci and CTs (Section 2). In Section 3, I discuss a number of specific empirical challenges that arise with respect to the experimental investigation of CTs, and propose an experimental paradigm for perception studies that takes a first step towards resolving some of these issues: (i) by introducing an artificial scenario the paradigm makes it easier to construct the necessary

contrasts; (ii) by explicitly guiding the participants' attention to the prosodic cues it prevents the cues from being overlooked; (iii) and by using preposed embedded clauses it makes it possible to investigate contrast on various constituents without the need of syntactic movement. A first small-scale experiment in which this paradigm was employed is presented in Sections 4 and 5. The results suggest that in transitive clauses, there are parallels in the behavior of foci and CTs: for both categories, projection from the object to the whole clause is available to a greater extent than projection from the subject to the whole clause. The experiment helps to pinpoint some open issues which will need to be addressed in the further development of the paradigm. Some unexpected patterns were observed even in the focus conditions, which were intended to serve as a baseline: (i) projection from the object was not accepted by all participants, (ii) some participants accepted projection from both the subject and the object. I discuss potential sources of these unexpected results and how they could be addressed in further research.

2 Theoretical Background: Projection

2.1 Focus Projection: Semantics

The claim that example (1), repeated below in (3), is (like its English counterpart) ambiguous with respect to focus is supported by the observation that it can be a felicitous answer to all three questions in (3a-c). However, the information-structural interpretation is not completely unrestricted: for example, (3) would not be a felicitous answer to a subject question like *'Who or what watched the bird?'*.

(3)   *Die Katze hat den* VOGEL *beobachtet.* 'The cat watched the BIRD.'
    a.   Who or what did the cat watch?   (focus = object)
    b.   What did the cat do?   (focus = verb phrase)
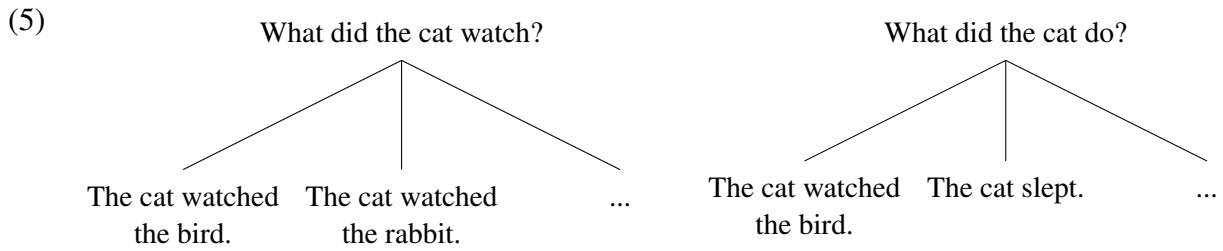   c.   What happened?   (focus = whole clause)

In contrast, (4) with a nuclear pitch accent on the subject is not a felicitous answer to the questions in (3a-c), but would instead be compatible with a subject question like *'Who or what watched the bird?'*.

(4)   *Die* KATZE *hat den Vogel   beobachtet.*
    the cat.nom has the bird.acc watched
    'The CAT watched the bird.'

The relation between focus and questions can be made more explicit in terms of alternative semantics (Rooth, 1981, 1992). The basic idea is that the interpretation of utterances containing focus-marked expressions involves alternatives, and these alternatives need to be related to the context in a certain way. For example, the expression *the cat* is marked as focused in (4). Uttering the sentence in this form is only felicitous in a context that makes propositions of the form *'X watched the bird'* relevant, where X are alternatives to *the cat*. This requirement is satisfied in a context like *'Who or what watched the bird?'*, but not in the contexts in (3a-c).

In Roberts's (1996; 1998) formal theory of discourse structure, the idea of correspondence between questions and focus marking can be represented with reference to discourse trees. In this model, focus marking can be described as indicating which expression is relevant to the current question under discussion (QUD).

Focus projection can then be analyzed as an utterance that is compatible with at least two different discourse structures, differing in the QUD, as shown in (5).

(5)

	What did the cat watch?			What did the cat do?	

The cat watched the bird.   The cat watched the rabbit.   ...   The cat watched the bird.   The cat slept.   ...

2.2 Focus Projection: Prosody

There have been various proposals how focus projection can be represented in models of the syntax-prosody-IS interface. One way to think about it is in terms of (information-structurally) neutral and marked prosody. Neutral prosody is systematically governed by syntax-prosody mapping rules in English and German: e.g., there is a preference for mapping lexical phrases (like NP, VP) to phonological phrases, whose head is typically realized with a pitch accent. This leads to a preference for accenting the object (a phrase) but not necessarily the verb (a head) in transitive sentences, resulting in pitch accents on the subject and the object. Usually, the rightmost accent, i.e., the one on the object, is most prominent at the level of the intonational phrase (ι). The formal representation of this idea in Pattern 1 of Figure 1 is based on Selkirk's (2011) model involving recursive phonological phrases (ϕs); see Truckenbrodt (1995) for a different implementation of the same core idea.

Pattern 1: sentence stress on the object	Pattern 2: sentence stress on the subject
(x $)_\iota$	(x $)_\iota$
(x) ((x) $)_\phi$ *prosody*	(x) ((x) $)_\phi$ *prosody*
[S]$_{NP}$ [V [O]$_{NP}$]$_{VP}$ *syntax*	[S]$_{NP}$ [V [O]$_{NP}$]$_{VP}$ *syntax*

Figure 1. Two different prosodic structures for a transitive sentence

This default, syntax-based pattern can be altered by information structure. Focused constituents need to be prominent within a certain domain (cf. Truckenbrodt's, 1995, FOCUS constraint); in the examples discussed in this paper, they need to be prominent at the level of the intonational phrase—more precisely, the focused part of the structure needs to contain the head of the intonational phrase. If the object is focused, the default pattern does not have to be altered. If, in contrast, the subject is focused, the default prosodic structure (Pattern 1 in Figure 1) does not satisfy the focus requirement, and the prosodic structure needs to be altered such that the subject is more prominent (Pattern 2 in Figure 1—a marked prosodic structure—would satisfy this requirement).

Under this view, focus projection to the level of the VP or the sentence can be modeled as a case in which the default pattern satisfies the prosodic requirement for such a larger constituent: when the VP or the whole sentence is focused, the focus contains the head of the intonational phrase by default, thus, no deviation from default prosody is required.

2.3 Contrastive Topic Projection: Semantics

For CTs, it has been sporadically noted (Jacobs, 1997: 96; Büring, 1997: 72–74; Constant, 2014: 42/105–106) that projection at least to the VP-level is also possible when the prosodic cue (fall-rise accent) occurs on the object. The example from Section 1 is repeated in (6).

(6)   *Den √Vogel hat die Katze*   MORGENS   *beobachtet...*
     the   bird.acc has the cat.nom in.the.morning watched

What exactly does projection mean in the case of CTs? Following Büring's (2003) analysis, unlike foci, CTs are not directly related to the question under discussion. Rather, CT marking points to a further (often implicit) related question. In (6), the speaker indicates that they are answering the question when the cat watched the bird (via focus marking on *morning*); but in addition, they either imply (via CT marking on *bird*) that a question about watching something else is relevant (object contrast, as in (7a)), or a question about another activity (VP contrast, (7b)). The latter is a case of CT projection to the level of the VP.

(7) a. *'When did the cat watch the bird, and when did it watch the rabbit?'* (CT = object)
Den √Vogel hat die Katze MORGENS beobachtet, und den √Hasen ABENDS.
~ 'As for the bird, the cat watched it in the morning, and the rabbit in the evening.'
   b. *'When did the cat watch the bird, and when did it sleep?'* (CT = VP)
Den √Vogel hat die Katze MORGENS beobachtet, und √geschlafen hat sie ABENDS.
~ 'As for watching the bird, the cat did that in the morning, and it slept in the evening.'

In Roberts's model of discourse structure, CT marking can be described as indicating which other questions besides the QUD are considered relevant, as shown in (8) and (9).

(8) *Contrastive topic: 'the bird'*
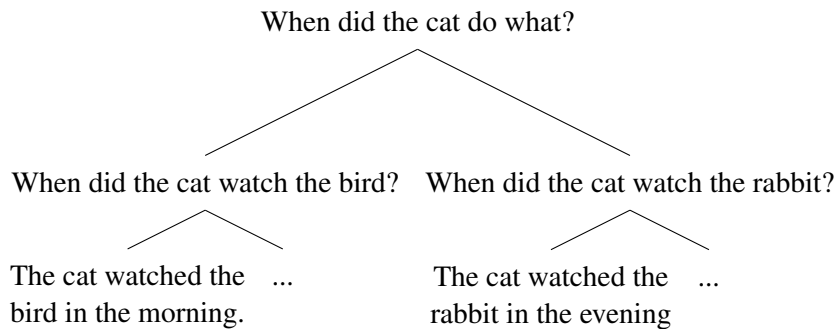
(9) *Contrastive topic: 'watch the bird'*
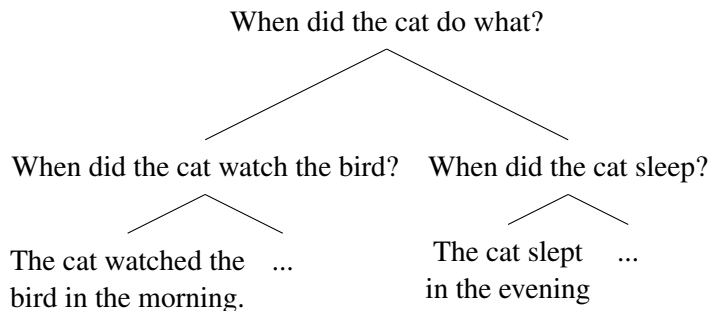
Foci and CTs thus differ in whether they are related directly to the QUD or to sister questions to the QUD, but have in common that both have to do with alternatives and with question-answer-correspondence. Based on this similarity, Wagner (2012) and Constant (2014) have argued that CTs *are* essentially foci; in their approaches, sentences containing both a focus and a CT are analyzed as cases of nested foci. Comparing the projection behavior of foci and CTs can inform us whether this approach is on the right track: if they are essentially the same category, a similar projection behavior would be expected.

2.4 Contrastive Topic Projection: Prosody

It has been proposed that focus and CT are subject to similar constraints with respect to prosodic prominence. According to Féry (2007), (contrastive) topics are subject to a constraint that requires them to have the highest prosodic prominence in their semantic domain, fully parallel

to the the constraint which requires focused phrases to have the highest prosodic prominence within their domain.

There is however a clear difference with respect to intonation: focused phrases typically carry a falling pitch accent and the following material is phonetically compressed at a low pitch register. CTs carry a fall-rise accent referred to as a 'root contour' Jacobs (1997), represented as $\sqrt{}$ here (and sometimes reduced to a simple rise). Together, a CT accent, a focus accent, and a high pitch plateau between them form a characteristic intonational contour referred to as 'bridge contour' or 'hat contour'.

Based on this, one possible hypothesis about the phonological realization of CT projection would be that the prosodic structure is identical to cases of focus projection and there is only a difference in intonation. For example, in a transitive sentence, the object would have the highest prosodic prominence at the level of the intonational phrase if the VP is focused. Analogously, if the VP is a CT, we would also expect the object to have the highest prominence, but instead of a falling accent, we would expect a fall-rise accent followed by a high plateau, as illustrated in (10). I will refer to this idea as the *prominence-based hypothesis*.[1]

VP interpretable as focus:

/⁻\ __ /⁻\ _____			*intonation*
(x)ₗ			
(x) ((x))_φ			*prosody*
[S]NP [[O]NP V]VP			*syntax*

VP interpretable as a contrastive topic:

/⁻\ __ √‾‾‾‾‾			*intonation*
(x)ₗ			
(x) ((x))_φ			*prosody*
[S]NP [[O]NP V]VP			*syntax*

Figure 2. Predicted patterns for a transitive German SOV clause under the assumption that foci and CT are associated with similar prominence constraints but different intonational contours.

An alternative hypothesis has been proposed in Wierzba (2013, 2017): based on the intuition that in a sentence like (10) a fall-rise accent on the direct object can mark the whole VP as a CT, whereas such an accent on the PP can only mark a part of the VP as a CT, I conjectured that the fall-rise might be left-aligned with the part of the sentence interpreted as the CT, i.e., that the fall-rise accent marks the left edge of the CT. In (10), this would mean that a fall-rise accent on the object as in (a) would indicate that the CT starts at that point and extends to the right, spanning the entire VP. A fall-rise accent on the PP as in (b) would indicate that the CT starts here and extends to the right, spanning only a part of the VP (the PP and the verb). Under this hypothesis, a larger domain spanned by the hat contour would correspond to a larger CT. I will refer to this idea as the *alignment-based hypothesis*.

(10) a. *Peter hat [das √Päckchen zur   Post   GEBRACHT]CT...*
        Peter has the   parcel   to.the post.office taken
        'Peter took the parcel to the post office...'
        (possible continuation: '...but he did not do the dishes.')   contrasted: DP + PP + V
    b. *Peter hat das Päckchen [zur √Post GEBRACHT]CT...*
        'Peter took the parcel to the post office...'
        (possible continuation: '...but he did not send it.')   contrasted: PP + V

The alignment-based and the prominence-based hypotheses make the same prediction for a transitive clause in which the VP is a CT (the pattern in Figure 2 would be expected), but there are cases in which their predictions differ. For example, if the whole clause is a CT,

---

[1] Typically, a CT occurs together with a focus within one sentence and they form the hat contour together. The illustration in Figure 2 shows the atypical, but possible case of a sentence containing only a CT (the nuclear pitch accent could fall within the following clause). In the experiment presented in Section 4, items like this will be used to avoid a confound that arises when CTs and foci co-occur within the same sentence.

the alignment-based hypothesis would predict that the fall-rise accent is on the subject: if the fall-rise marks the beginning of the CT, and the CT spans the whole clause, the fall-rise should be placed on the left-most element. Since the material following the fall-rise accent is typically non-prominent, this would entail a difference in prominence relations, with the subject being more prominent than the object. According to the prominence-based hypothesis, the fall-rise accent should fall on the object in this case, in parallel to focus projection. I will return to these prediction below in Section 4.3.

3   Previous Empirical Research on Contrastive Topic Projection

3.1   Previous Research

Whereas focus projection is well-studied from a theoretical and empirical perspective (see De Kuthy & Meurers, 2012, for an overview), there are fewer studies investigating how projection works in the case of CTs, and the results tend to be less clear-cut.

Initial experimental evidence in favor of CT projection is reported in Wierzba (2013). Sentences with three-place verbs like *stellen* 'put' were tested, containing a direct object as well as a directional PP argument. One of these arguments was fronted, for example:

(11) a. *Das √Päckchen hat Susi* NICHT *zur   Post   gebracht.*
          the parcel   has Susi not   to.the post.office taken
          'Susi did not take the parcel to the post office.'
          (more literally: 'The parcel Susi did not take to the post office.')
      b. *Zur   √Post   hat Susi das Paket* NICHT *gebracht.*
          to.the post.office has Susi the parcel not   taken

The stimuli were presented auditorily, with a fall-rise accent on the fronted constituent and a falling nuclear accent on the negation. Two different types of context were tested: the context was either compatible with the interpretation that only the fronted constituent was the CT, as in (12a-b), or compatible with the interpretation that the whole VP was the CT, as in (12c):

(12) a. Narrow object contrast: *'Do you know what Susi took to the post office?'*
          Das Päckchen hat sie nicht zur Post gebracht, aber mehr weiß ich darüber auch nicht.
          'Susi did not take the parcel to the post office, but I don't know more about that.'
      b. Narrow PP contrast: *'Do you know where Susi took the parcel?'*
          Zur Post hat sie das Päckchen nicht gebracht, aber mehr weiß ich darüber auch nicht.
          'Susi did not take the parcel to the post office, but I don't know more about that.'
      c. Broad VP contrast: *'Susi had several chores, right? Do you know which ones she completed?'*
          Das Päckchen hat sie nicht zur Post gebracht, aber wenigstens hat sie eingekauft.
          'Susi did not take the parcel to the post office, but at least she did the shopping.'

The results showed a significant interaction between the factors FRONTED XP and CONTEXT in the direction that both interpretations were similarly acceptable when the object was fronted and carrying the fall-rise accent, whereas only the narrow-contrast interpretation was acceptable when the PP was fronted.

In follow-up experiments reported in Wierzba (2017), this effect was not replicated, and much smaller differences were found between the tested conditions. That the results from Wierzba (2013) were not replicated in the later studies could have to do with acceptability ratings being especially susceptible to task effects. In Wierzba (2017: 5.9.2), there is a discussion of the results for the filler items of the follow-up experiments that points in this direction: ratings for items with clear violations of syntactic rules (e.g., extraposition of object DPs) were much more consistent than ratings for items with clear mismatches of prosody and context, as illustrated in

Figure 3. In the stimuli with a prosodic mismatch, a rating in the upper half of the scale (4 or higher on a 7-point scale) was chosen in a third of the trials.

Q-A mismatch (prosodic violation)   extraposed object (syntactic violation)
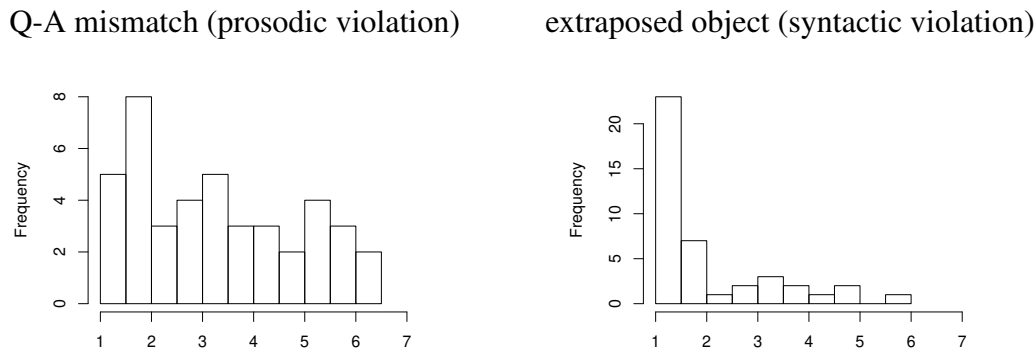
Figure 3. Histograms depicting the mean rating by subject for the control items with a prosodic question-answer mismatch / extraposition of an object DP (from Wierzba, 2017: 187)

In Wierzba (2017: 187), I suggested that "the effect of prosodic manipulations might be underestimated in the results because not all participants paid attention to them or considered them as relevant for the task" and proposed that this potential problems could be alleviated by more explicit instructions and allowing participants to listen to the stimuli several times. This suggestion will be taken up here.

Another potential source of the low replicability might be that in the case of CTs, it is more challenging to provide a context that unambiguously induces the intended information-structural reading. When studying focus, a question-answer-pair like "What did Susi do? – Susi took the parcel to the post office" or a correction like "Did Susi do the shopping? – No, Susi took the parcel to the post office" clearly favor interpreting the VP as focused. In the case of CTs, the context needs to make clear that the speaker is following a certain discourse strategy. In (12c), the context is intended to express that the speaker is addressing the superquestion ('Which chores did Susi complete?') by dividing it up into subquestions varying in the verb phrase, like "Did Susi take the parcel to the post office?", "Did Susi do the shopping?". In Wierzba (2013, 2017), the goal was to construct relatively natural, every-day dialogs without the need to provide an elaborate background story. In dialogs of this type, it is challenging to provide plausible motivation for why the speaker is comparing activities (which is necessary to study VP contrast), and it is even more difficult to construct a plausible context in which a speaker is comparing whole propositions (which would be necessary to study contrast at the clause level). If participants find the context implausible, this might lead them to accommodate different discourse structures than intended. An alternative approach would be to use a more artificial setting, which would require providing more detailed instructions, but which would in turn allow to spell out the question structure more explicitly and to construct any desired kind of contrast. This approach will be taken in this paper.

One further thing to note about previous research on CTs is that in German, there are also certain restrictions with respect to word order. A CT needs to precede the focus (in order to form a hat contour). In Wierzba (2013), sentences with fronted objects were used for this reason, which introduces further potential confounding factors (how much of the interpretative effect can be attributed prosody, and how much to the marked syntax?). In a part of the follow-up experiments in Wierzba (2017), embedded clauses without fronting were used (embedded clauses are verb-final in German, which is standardly assumed to be the basic word order). However, this comes with the problem that this highly restricts the options what can be the focus. If, for example, the VP is intended to be the CT in a transitive clause, and the focus needs to follow

the fall-rise accent, the focus accent necessarily falls on the verb or the auxiliary, with a verb or verum focus interpretation:

(13) *Ich glaube, dass Susi √das Päckchen* VERSCHICKT *hat.*
 I think that Susi the parcel sent has
 'I think that Susi has sent the parcel.'

This puts a further restriction on the type of context that can be used, and it additionally involves the complication that the focus is on an element (the verb) that is also included in the CT, creating a nesting of information-structural categories. Below, I will propose to adopt the idea of using embedded clauses, but adjust it in such a way that it allows to clearly separate CT and focus, and to flexibly investigate any type of contrast.

3.2 Challenges

To sum up, previous research suggests that studying CT projection raises some specific difficulties:

(i) Focus can be studied using question-answer pairs, but the semantic effects of CTs involve more complex discourse structures that are often partly implicit in dialogues. It is more difficult to set up a plausible unambiguous context, to exclude accommodation, and to make sure that the context is taken into account by participants.

(ii) Previous experiments indicate that intonational cues are often not taken into account or overlooked/deemed irrelevant.

(iii) In German, CTs have to precede the focus to form the hat contour. Therefore, most examples/studies of CT projection involve CT fronting. Syntactic movement might be a confounding factor; it is desirable to test projection in basic word order first.

3.3 Proposed Solutions

I propose an experimental paradigm for perception studies with the following features:

(i) In order to make it possible to contrast any pair of propositions without raising plausibility issues, a theater scenario with explicit scene descriptions is used.

(ii) Deliberately deviating from standard methodology, the purpose of the study is made transparent to the participants: in the instructions and in a training phase, the relevant prosodic cues and interpretative effects are pointed out based on uncontroversial examples with narrow contrast (where the CT is, e.g., a single adjective). Additionally, a forced-choice method (choosing between two realizations in a given setting) is used, making the prosodic contrasts especially salient.

(iii) Preposed embedded sentences with basic SOV order are used as the target clauses containing the CT; the focus is part of the following main clause. This makes it generally possible to form a hat contour without syntactic movement within the target clause, and it makes it possible to investigate projection to the whole clause, which is otherwise particularly difficult.

The details are illustrated in the following section based on a first experiment employing the proposed paradigm.

4 Experiment

In a first experiment, I tested whether focus/CT projection to the sentence level in transitive sentences is possible from the object and/or from the subject. The motivation to test this specific type of projection stems from the consideration that in this case, the predictions are relatively uncontroversial for foci—projection should be possible from the object and impossible from

the subject—, whereas the two hypotheses discussed above make different predictions for CTs. Thus, the focus items can serve as a baseline for interpreting the CT projection results and as a sanity check for the paradigm: any deviations from the expected results in the focus items should help to identify potential remaining methodological problems.

The materials and data are available at the Open Science Foundation under: https://osf.io/h742e/

4.1 Design and Materials

The experiment included an introduction in which participants were familiarized with the task and the type of context that would be used throughout the experiment. The instructions can be found in the appendix.

Every experimental trial began with an explicit scene description in a theater scenario, e.g.:

(14) You saw a theater play with a friend, including the following scenes:
Scene 1: A fisherman warns a sailor.
Scene 2: It storms.

For each of the 20 target items, a CT and a focus version was constructed. In the CT version, the following context contained two overt questions, establishing a contrast between the two scenes. Zero-place predicates were used for the second scene to exclude the possibility of construing a narrower contrast. Two different variants of the critical part of the sentence marked in boldface were presented auditorily, differing in the positioning of the fall-rise accent (subject/object).[2] The prosodic realizations are illustrated in Figure 4. The participants' task was to decide which of the versions sounded better in the given context. They also had the option to express that neither of the versions fit.

(15) Later you talk about the play. You say: "I remember that we saw that a fisherman warned a sailor and that it stormed. In which scenes did that happen again?" Your friend answers:
"**Dass der Angler den Segler gewarnt hat**, haben wir in der ersten Szene gesehen.
(*lit.* **That the fisherman warned the sailor**, we saw in the first scene). It rained later."

(16) a. dass der √ANGLER den Segler gewarnt hat...
b. dass der Angler den √SEGLER gewarnt hat...
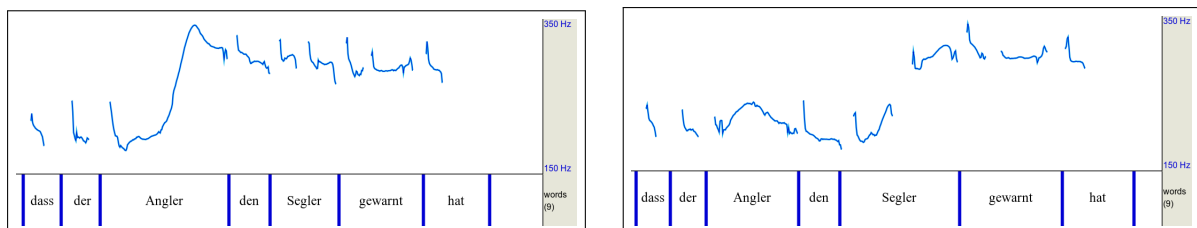
Figure 4. Examples of pitch contours: contrastive topic cue on the subject (left) vs. object (right)

To construct the corresponding focus version, a corrective context was used. This was done to make it especially clear what is focused through overt contrast; in broad information questions like "What happened?", a narrower contrast could potentially be accommodated by participants. A falling nuclear accent was realized on the subject or object. The prosodic realizations are illustrated in Figure 5 (generated by Praat, cf. Boersma & Weenink, 2020).

---

[2] The sentences were recorded by the author of the paper. The recordings only contained the highlighted, preposed embedded clause; i.e., they contained the fall-rise accent and ended with the high plateau. They did not contain the nuclear pitch accent, which would by assumption fall into the following main clause.

(17) [...] You say: "I think that in the first scene, we saw that it stormed." Your friend answers: "Nein, da haben wir gesehen, **dass der Angler den Segler gewarnt hat**. (*lit.* No, in that scene we saw **that the fisherman warned the sailor**." [...]

(18) a. dass der ANGLER den Segler gewarnt hat.
     b. dass der Angler den SEGLER gewarnt hat.
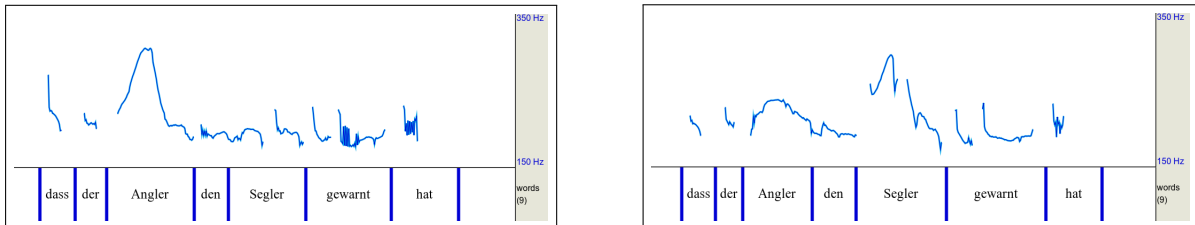
Figure 5. Examples of pitch contours: focus cue on the subject (left) vs. object (right)

Fourteen stimuli with narrow contrast, e.g., on an adjective as in (19), were used as control conditions. The expectation here is that the version with the relevant accent (falling nuclear accent in the case of foci; fall-rise accent in the case of CTs) on the adjective as in (20b) should always be selected.

(19) Scene 1: The old lioness is hunting
     Scene 2: The young lioness is hunting.

(20) a. ...die JUNGE Löwin... / ...die √junge Löwin...
     b. ...die junge LÖWIN... / ...die junge √Löwin...

Sixteen additional filler stimuli were included for exploratory purposes. They will be described in more detail in Section 5.

4.2 Participants and Procedure

Sixteen native speakers of German took part. They were recruited via Prolific (prolific.ac) and were paid for participation. The study was set up via L-Rex (Starschenko & Wierzba, 2019), a platform for web-based rating studies.

The CT/focus conditions were presented in separate experimental blocks to facilitate concentration on each of the two prosodic patterns. Half of the participants saw the focus conditions first, the other half saw the CT conditions first. Comprehensive instructions, including two examples, and a set of training items preceded each block, sensitizing participants for the relevant prosodic cues. Translated versions of the instructions can be found in the appendix. During the training phases, participants received feedback concerning the expected choice; i.e., when they did not choose the expected response, a message appeared, explaining what the expected response was and why. After that, participants had the possibility to change their answer and/or provide a comment. During the experimental phases, no feedback was provided. Each experimental phase (CT/focus) contained 10 critical items and 15 fillers. In sum, each participant saw 4 example items, 10 training items, 20 critical items (10 of them in the focus condition, and 10 of them in the CT condition; i.e., each participant saw each item in one of the conditions), and 30 filler items.

Every stimulus was presented on a separate screen, starting with a written text. The target sentence was marked by boldface font. The context was followed by two audio players, containing two different recorded versions of the target clause. Participants could choose to play/pause them freely and to listen to the files as many times as they wanted. Below, a question and three buttons were presented. The question was always the same: 'Which variant of the highlighted sentence fits better into this context?' Participants gave their answers by clicking on

one of the buttons. Optionally, they could enter a comment. They proceeded to the next screen by clicking on the 'Continue' button. The layout is illustrated schematically in Figure 6.

Figure 6. Illustration of the experiment's layout on the screen

4.3 Predictions

Under the prominence-based hypothesis, i.e., the idea that CT and focus projection work similarly with respect to prominence and only differs in the intonational contour (hypothesis 1), we would expect the same response pattern for both categories: the variant with the respective prosodic cue on the object (nuclear falling accent / fall-rise accent) should be preferred. In contrast, the alignment-based hypothesis, i.e. the idea that the hat-contour needs to be left-aligned with the CT (hypothesis 2), would predict a different prosodic realization in this case: when the whole clause is a CT, the fall-rise accent should be on the leftmost phrase—the subject. This predicts a main effect of category on the subject/object preference, and, in particular, a higher proportion of subject preference for CTs and a higher proportion of object preference for focus. It is also conceivable that CT projection to the clause-level is not possible at all or works in yet a different way, not represented by any of the auditory variants; in this case, more 'neither' answers for CTs would be expected. The predictions are illustrated schematically in Figure 7.
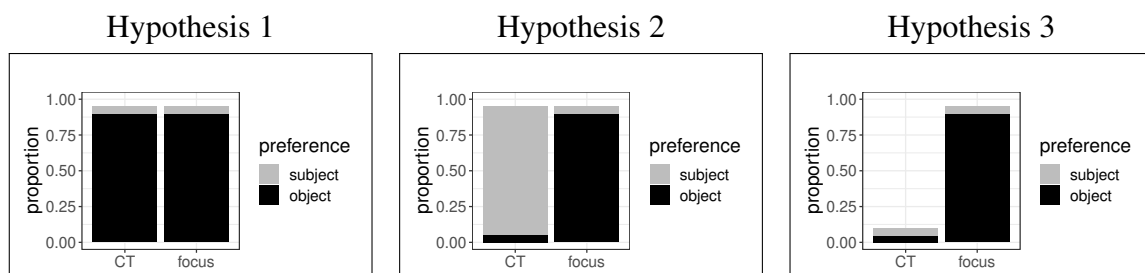

Figure 7. Predictions of the three hypotheses: 1. Prominence-based hypothesis: CTs and foci are subject to similar prosodic constraints; 2. CTs are left-aligned with the hat contour; 3. There is no CT projection or it works differently than anticipated.

4.4 Results

Three out of sixteen participants were excluded prior to analysis based on the exclusion criterion of less than 80% of expected responses in the narrow-contrast control conditions. The analysis is based on 130 data points per condition (10 per condition from each participant).

For analysis, the responses were encoded as two binary dependent variables: (i) *acceptance:* whether one of the two provided versions of the sentence was selected or whether the "neither" option was chosen (subject version or object version vs. "neither), and (ii) *preference:* which of the provided versions was chosen (within the data in which one of them was accepted: subject version vs. object version).

The response proportions are illustrated in Figure 8. With respect to *acceptance*, one of the two provided versions was chosen 59 out of 130 times (45%) in the CT items and 80 out of 130 times (62%) in the focus items. Within the trials in which one of the provided recordings was chosen, the *preference* was as follows: in the CT items, the recording with the fall-rise accent on the object was chosen 54/59 times (92%). In the focus items, the recording with the falling accent on the object was chosen 63/80 times (79%).
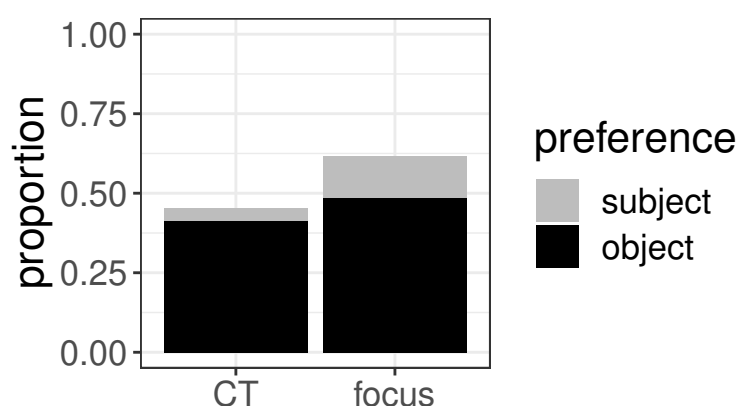
Figure 8. Results of the experiment (based on data from 13 participants): proportion of trials in which the recording with the prosodic cue on the subject/object was chosen; percent missing up to 100 in the plot correspond to trials in which the option 'neither' was chosen.

The R package *lme4* (R Core Team, 2013; Bates et al., 2015) was used for analysis. According to a logistic regression model[3] with *acceptance* as the dependent variable, there were a significant main effect of *category* (independent variable with two levels—CT, focus—, sum-coded): there were significantly more "neither" answers for CTs than for foci (significant main effect of *category*: z = −2.96, p = 0.003).

According to a logistic regression model[4] with *preference* as the dependent variable, there was a trend towards less projection from the subject for CTs than for foci within the subset of data in which one of the versions was accepted, but the main effect of *category* did not reach a significant level (z = −1.94, p = 0.052).

4.5 Discussion

First, it is encouraging to see that 13 out of 16 participants consistently (i.e., in more than 80% of the cases) chose the expected response in the controls; overall, the expected response was chosen in 85% of the control trials. Even though a direct comparison to Wierzba (2017) is not possible because a different method was used (acceptability ratings vs. forced-choice), this tentatively suggests that the methodology employed here yields relatively sharp and stable judgments.

The following methodological aspects were intended to ensure a high rate of expected responses in the control trials: participants were given the possibility to directly compare two versions and to listen to them several times in order to make it easier for them to perceive the prosodic differences, and explicit instructions and a training phase were added in order to make

---

[3] With random intercepts for participant and item; models with a random slope did not converge.

[4] With a random intercept for participant; models with a random intercept for items or with a random slope did not converge.

it clearer that prosody should be taken into account in the judgments. Two anonymous reviewers made the following suggestions how it could be determined empirically in future research whether the training phase had the desired effect: (i) to record and analyze the participants' behavior during the training phase in more detail (how often the feedback was shown, and how often the decision was changed)[5], (ii) to test a group of participants with a training phase, and another group without it, to check whether the training phase influenced the behavior. The second suggestions could also be applied to the other aspects of the paradigm (e.g., the option to listen to a stimulus several times) to determine their effectiveness.

As for the results of this first experiment, the observed patterns suggest that CT projection from the object to the whole clause is possible to a higher extent from the object than from the subject, just like focus projection. This is compatible with the prominence-based hypothesis (hypothesis 1) and incompatible with the alignment-based hypothesis (hypothesis 2). It is in line with the idea that the two categories are similar with respect to the mapping between semantics and prosody: for both, (default) prominence on the object is more compatible with a broad-contrast interpretation than (marked) prominence on the subject.

In addition, a significant difference was found between CT and focus with respect to the proportion of 'neither' responses. A difference in this direction was predicted by hypothesis 3 (no CT projection), but there are two caveats: first, the results suggest that CT projection is not completely impossible, and second, the contrast between focus and CT seems to be mainly due to a difference with respect to projection from the *subject*, which seems to be available to some extent in the case of focus (in addition to the expected projection from the object).

There are two open questions at this point. First, while the proportion of 'neither' responses is higher for CTs, it is also quite high for foci, which is unexpected—projection from the object is predicted to be possible. Second, it is unexpected that focus projection from the subject should be possible—what is behind this observation? In the following exploratory analyses, these open questions will be addressed again, and some tentative explanations for the observations will be suggested.

5 Exploratory Analyses

In addition to the confirmatory hypothesis tests reported in the previous section, several post-hoc analyses were conducted. In this section, I will have a look at the results patterns of individual participants, the comments that were provided, and exploratory fillers. I will discuss in which ways these additional analyses might shed some light on the open questions.

5.1 Results by Participant

Figure 9 shows the individual results the 13 participants who passed the inclusion criterion. The participants are numbered and presented in ascending order based on the proportion of 'object' responses in the focus condition.

Visual inspection of the first four plots shows two different patterns for those participants with the lowest availability of focus projection from the object. Two of these participants (#1–2) generally (for both CT and focus) show almost no acceptance of focus projection from either the subject or the object. As for the other two of these participants (#3–4), there seems to be a different reason for the low availability of projection from the object: they show a preference for projection from the subject in the focus case. This, however, does not carry over to CTs, which show roughly the same availability of projection from the object, but much less from the subject.

---

[5] Since in this experiment, only the participants' final decision was recorded, this suggestion could not be implemented yet with the data from the current experiment.
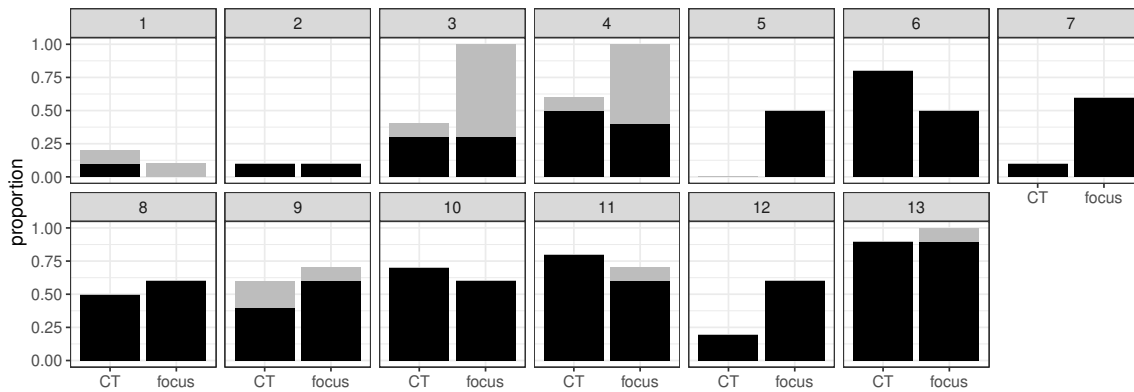
Figure 9. By-participant results, ordered based on proportion of 'object' responses for the focus version.

The remaining nine participants (#5-13) preferred projection from the object in at least 50% of the focus trials. The majority of these participants (#6, #8, #10, #11, #13) also show high availability of projection from the object (>50% of the trials) with CTs. Three of these participants (#5, #7, #12) show lower availability of projection with CTs, while one (#9) shows similar availability of projection, but a slightly higher preference for projection from the subject in the case of CTs.

Recall that two of the open questions discussed in Section 4.5 were that on average, there were more 'neither' answers for CT than for focus, and that there was higher availability of focus projection from the subject than expected. The by-participant analysis helps to see that both these effects are driven by a relatively small part of the participants: (i) some participants almost did not accept CT projection at all and almost always chose the 'neither' response (#1, #2, #5, #7, #12), and (ii) some participants accepted focus projection from the subject and object almost equally (#3, #4).

Observation (i) could mean that for some participants, CT projection is not available, and each phrase that is part of the CT need to be marked as such individually. Thus, these participants might prefer a version in which the subject and the object each carry a fall-rise accent. Alternatively, the observation is also compatible with the possibility that the participants would have accepted a version with a single fall-rise accent, but neither on the subject nor on the object; instead, they might have preferred the accent to fall on the verb. I will come back to these possibilities in the following sections. The comments and responses to the filler materials can help to tentatively distinguish between these possibilities.

As for observation (ii), it might be explainable by an unintended contrast between the subject and the object. Since the items contained two animate DPs, a part of the participants might have perceived them as contrasting with each other in the focus items, leading to the expectation that both should be equally prominent (multiple foci); since such an option was not provided, these participants might have alternated between choosing prominence on the subject/the object. Selecting DPs that are less likely to be perceived as contrasting with each other could potentially solve this issue in future experiments.

5.2 Comments

Participants had the option to add a comment to their response. Inspection of the comments provides some insight about the participants who did not meet the control criterion and whose data was excluded: one of these participants often commented that the two versions sounded the same or that the difference was perceived as marginal. This indicates that for some participants, the auditory differences employed in the experiment were not easily discernible perceptively even when close attention was paid.

The comments from the participants who did meet the criterion are informative with respect to the open question why the proportion of 'neither' responses was so high. Above, I conjectured that this might either be because participants would have expected several accents (on both the subject and the object), or because they would have preferred the main prosodic cue to fall on the verb. Several of the participants with a particularly high proportion of 'neither' responses left comments that point in the former direction, e.g. (translated from German):

(21) a. "The first version refers to the object (person), the second to the subject. The situation, whether it was raining or whether somebody was hugged, is not stressed."
      b. "I expected the 'greeting' to be stressed rather than the involved people."

All comments that make it explicit that an accent on the verb was expected stem from focus trials. For the CT trials, the comments were more ambiguous, e.g.:

(22) a. "The stress is not accurate."
      b. "Stress is only on one of the parts, not on the entire part."

This could mean that several accents were expected, but the comments are not as straight-forward as in the focus trials.

The preference for stress on the verb is unexpected in the focus trials—that projection from the object to the sentence level is possible is well-established in the theoretical and empirical literature. Thus, it is likely that there is a methodological issue at play. As suggested to me by Joseph De Veaugh-Geiss, using zero-place predicates might have increased the salience of the verb, and participants might have construed an unintended narrow contrast between the two verbs in spite of their different argument structures. This potential problem could be by solved using a different description for the second scene, e.g., a different transitive clause instead of a zero-place predicate (this was avoided here to prevent other undesired contrasts, e.g., between the individual arguments in the two sentences, but it seems like the attempt to avoid a confound here had the opposite effect).

5.3   Filler Item Results

The filler materials included a number of stimuli with different types of contrast. The primary goal of including them was to get a first impression as a basis for further experiments, but they are also informative with respect to the open questions of this experiment.

The first set consisted of eight items with contrast on the VP. As for focus, the expectation here would be that the relevant accent should fall on the object. For CTs, both the prominence-based hypothesis as well as the alignment-based hypothesis would also predict the cue to fall on the object, but this has not been reliably shown for CTs yet (see discussion above in 3.1). For this reason, these stimuli were not used as controls for potential exclusion of participants, but as exploratory materials.

(23) Scene 1: The horseman greets the passerby.
      Scene 2: The horseman dismounts.

(24) a. ...dass der Reiter den PASSANTEN gegrüßt hat... / ...dass der Reiter den √Passanten gegrüßt hat...
      b. ...dass der Reiter den Passanten GEGRÜSST hat... / ...dass der Reiter den Passanten √gegrüßt hat...

Another set of exploratory materials involved intransitive clauses, four with an unergative verb and four with an unaccusative verb; in both cases, the whole proposition was contrasted with a zero-place predicate. Based on experiments reported by Hoskins (1996) and Irwin (2011)

for English and Verhoeven & Kügler (2015) for German[6], the expectation for the focus cases here would be that the argument is more prominent in sentences with an unergative verb like (25), whereas the verb is more prominent in sentences with an unaccusative verb like (26). This has not been tested empirically for CTs yet. The prominence-based hypothesis would predict a similar pattern as for focus, whereas the alignment-hypothesis would predict preference for a fall-rise accent on the subject, irrespective of the verb type.

(25) a. ...dass der MANN geschrien hat... / ...dass der √Mann geschrien hat...
     b. ...dass der MANN geschrien hat... / ...dass der Mann √geschrien hat...

(26) a. ..dass der ARBEITER hingefallen ist... / ...dass der √Arbeiter hingefallen ist...
     b. ...dass der Arbeiter HINGEFALLEN ist... / ...dass der Arbeiter √hingefallen ist...

The results, shown in Figure 10, show that in the stimuli with VP contrast, one of the two versions was virtually always accepted in this type of item, and there was a preference for the nuclear pitch accent to fall on the object in the focus items. For CTs however, the two versions were chosen almost equally frequently.

transitive sentences with VP contrast   intransitive sentences with sentence-level contrast
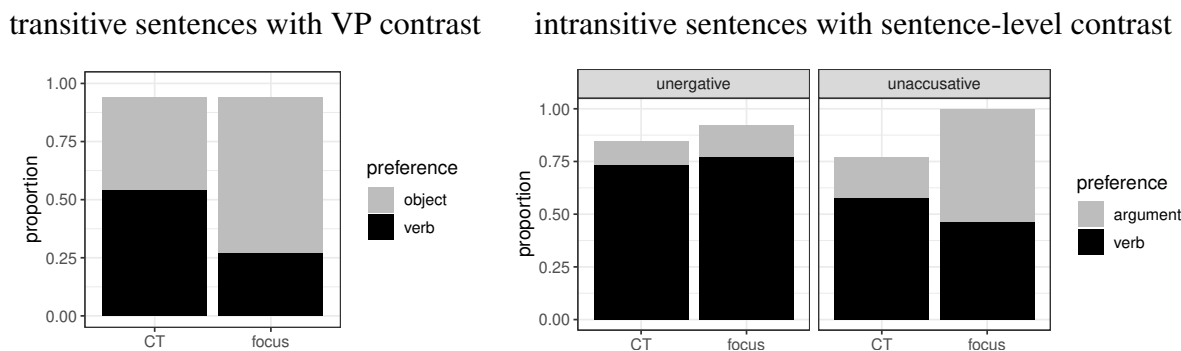
Figure 10. Results for the two groups of exploratory filler items (based on data from 13 participants).

An asymmetry between focus and CT is also visible in the results for the intransitive clauses. As expected, the focus items show a clear preference for the nuclear pitch accent to fall on the verb in the case unergatives; for unaccusatives, the version with the nuclear pitch accent on the argument was chosen roughly equally frequently. CTs show a similar pattern for unergatives; but for unaccusatives, the version with the prosodic cue on the argument was chosen less frequently than in the focus items.

These asymmetries might indicate that the similarity between foci and CTs might be limited to certain types of projection (e.g., they might behave similarly with respect to projection to the clause level in transitive clauses, but differ when it comes to projection to the VP level or to the clause level in intransitive clauses). None of the two hypotheses discussed above would predict the contrasts observed here.

The general trend in both types of fillers is that a fall-rise on the verb was chosen as the preferred version to a much larger extent than expected. Further research is required to establish why this was the case. Again, an unintended narrow verb contrast might have played a role, but this would not explain the observed difference between CT and focus. It is possible that the version with the fall-rise on the verb was preferred because it involved two accents (a prenuclear accent, and then a more prominent accent on the verb), while one of the elements was deaccented in the version with an early accent on the argument, because here, the verb was part of the high and tonally flat plateau of the hat contour. If this can be corroborated, the results of the main experiment should also be revisited—there, the version with an earlier onset of the CT contour

---

[6] See Hirsch & Wagner (2011) for discussion of further factors besides unergativity/unaccusativity that play a role in this kind of sentences.

(fall-rise accent on the subject) also involved deaccentuation of the object and the verb. It should be checked to what extent a preference against deaccentuation might have played a role in the results.

5.4 Discussion of the Exploratory Analyses

The post-hoc analyses provide some tentative insights with respect to the questions that were left open by the results of the main experiment. As for the question why there were so many 'neither' responses, the comments with respect to the focus items suggest that a version with prominence on the verb might have been preferred by some of the participants, potentially due to the unintentional emergence of a narrow verb contrast. Inspection of the filler materials suggests that also in the case of CTs, a version with the fall-rise accent on the verb might have been the missing preferred option and the reason behind the 'neither' responses. The filler materials show that in stimuli with contrast on a VP or an intransitive clause, accenting the verb was preferred to a high degree in the CT trials. It is conceivable that this would also have been the preferred realization if such a version had been provided in the main target items of the experiment. Investigating further what the reason behind this pattern is could potentially also help to explain the even higher proportion of 'neither' responses for CTs in the main experiment. The last open question concerned the unexpected availability of focus projection from the subject. The by-participant analysis showed that this was mainly driven by two participants. I tentatively suggested that this could also be due to an unintentional contrast, in this case between the subject and the object.

6 Conclusion

In this paper, I have taken a first step towards developing an experimental paradigm for systematic research on CT projection in German, tackling a number of methodological issues. Adopting the paradigm as a whole or some if its components comes with the following advantages: it makes it possible to investigate CTs of all categories, to study CTs in any syntactic position, and to directly compare CTs to foci.

The results of a first small-scale experiment employing the paradigm suggest that there are parallels between foci and CTs with respect to projection to the clause level: for both categories, projection from the object is more available than from the subject. This is compatible with the view that they are subject to similar constraints with respect to prosodic prominence. However, some of the exploratory materials that were included in the study indicate that there might be asymmetries between focus and CT projection in other cases that should be investigated further (concerning projection within the VP and in intransitive clauses).

The experiment also helps to identify ways in which the paradigm can be further improved. One of the main open issues is that the proportion of responses indicating that projection was neither possible from the subject nor the object was overall higher than expected, also in the focus cases. Furthermore, a by-participant inspection of the data showed that a subgroup of speakers accepted focus projection from the subject and the object. Both of these unexpected findings could have to do with unintended contrasts in the materials, suggesting that this is a potential confound that further studies should control for.

Acknowledgments

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1).

Boersma, P., & Weenink, D. (2020). *Praat: doing phonetics by computer [Computer program].* version 6.1.37.

Büring, D. (1997). *The Meaning of Topic and Focus. The 59th Street Bridge Accent.* London / New York: Routledge.

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

Chomsky, N. (1971). Deep structure, surface structure, and semantic interpretation. In D. Steinberg & L. Jakobovits (Eds.), *Semantics: An Interdisciplinary Reader in Philosophy, Linguistics, and Psychology* (pp. 183–216). Cambridge: Cambridge University Press.

Constant, N. (2014). *Contrastive Topic: Meanings and Realizations*. PhD thesis, UMass Amherst.

De Kuthy, K., & Meurers, W. D. (2012). Focus projection between theory and evidence. In B. Stolterfoht & S. Featherston (Eds.), *Empirical Approaches to Linguistic Theory (Studies in Generative Grammar 111)* (pp. 207–240). De Gruyter Mouton.

Féry, C. (2007). The prosody of topicalization. In K. Schwabe & S. Winkler (Eds.), *On Information Structure, Meaning and Form* (pp. 69–86). Amsterdam / Philadelphia: Benjamins.

Hirsch, A., & Wagner, M. (2011). Context and prominence in English intransitive sentences. poster presented at ETAP 2.

Höhle, T. (1982). Explikationen für 'normale Betonung' und 'normale Wortstellung'. In W. Abraham (Ed.), *Satzglieder im Deutschen: Vorschläge zur syntaktischen, semantischen und pragmatischen Fundierung (Studien zur deutschen Grammatik 15)* (pp. 75–153). Tübingen: Narr.

Hoskins, S. (1996). A phonetic study of focus in intransitive sentences. In *Proceedings of the ICSLP 96*.

Irwin, P. (2011). Intransitive sentences, argument structure, and the syntax-prosody interface. In M. B. Washburn, K. McKinney-Bock, E. Varis, A. Sawyer & B. Tomaszewicz (Eds.), *Proceedings of the 28th WCCFL* (pp. 275–284). Somerville, MA: Cascadilla Proceedings Project.

Jackendoff, R. (1972). *Semantic Interpretation in Generative Grammar*. Cambridge, MA: MIT Press.

Jacobs, J. (1997). I-Topikalisierung. *Linguistische Berichte*, *168*, 91–133.

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Roberts, C. (1996). Information structure in discourse: towards an integrated formal theory of pragmatics. In Yoon, J. H., & Kathol, A. (Eds.), *OSU Working Papers in Linguistics 49: Papers in Semantics*, (pp. 91–136).

Roberts, C. (1998). Information structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics & Pragmatics*, *5*, 1–69.

Rooth, M. (1981). *Association with focus*. PhD thesis, University of Massachusetts.

Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics*, *1*, 75–116.

Selkirk, E. (1984). *Phonology and syntax: the relation between sound and structure*. Cambridge, MA: MIT Press.

Selkirk, E. (2011). The syntax-phonology interface. In J. Goldsmith, J. Riggle & A. Yu (Eds.), *The Handbook of Phonological Theory, 2nd edition*. Oxford: Blackwell.

Truckenbrodt, H. (1995). *Phonological phrases: Their relation to syntax, focus, and prominence*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA.

Verhoeven, E., & Kügler, F. (2015). Accentual preferences and predictability: An acceptability study on split intransitivity in german. *Lingua*, *165*, 298–315.

Wagner, M. (2012). Focus and givenness: A unified approach. In I. Kučerová & A. Neeleman (Eds.), *Contrasts and Positions in Information Structure* (pp. 102–147). Cambridge: Cambridge University Press.

Wierzba, M. (2013). Subparts of contrastive topics and the syntax-information structure interface. In E. Boone, M. Kohlberger & M. Schulpen (Eds.), *Proceedings of ConSOLE XX*.

Wierzba, M. (2017). *Revisiting prosodic reconstruction: an interface-based approach to partial focus and topic fronting in German*. PhD thesis, University of Potsdam.

Appendix

Instructions: Introduction and Study Structure

In this study, you are going to listen to sentences and judge them.

The study is structured as follows:

- Part 1 (approx. 15-20 min)

    Explanation of the task (based on two examples)
    Practice phase (5 sentences)
    Experiment (25 sentences)

- Part 2 (approx. 15-20 min)

    Explanation of the task (based on two examples)
    Practice phase (5 sentences)
    Experiment (25 sentences)

Feel free to take a short break between the two parts. Please put on headphones now. Next, you will get a more detailed instructions for the first part of the study.

Instructions for the Focus Block

Listen to the first one of the two sentences that you can play at the bottom of this page. You can listen to them several times. Adjust the volume such that you can understand everything clearly.

The first sentence is *Der Ritter hat das Schloss erreicht* ('The knight reached the castle') – but with a special intonation. The sentence sounds as though the speaker was correcting something: that it was the knight (and not a different person) who reached the castle. The sentence would thus fit well into the following kind of context in which someone is being corrected:

Situation:
in the morning: The knight reaches the castle.
in the evening: The hiker reaches the castle.

Someone says: "I think that the following happened in the morning: The hiker reached the castle."
Someone else corrects them: "No, that's not right. In the morning, the following happened: **The knight reached the castle.**"

Listen to the second sentence now. Again, you can listen to it several times.

It is the same sentence: *Der Ritter hat das Schloss erreicht* ('The knight reached the castle') – but the intonation is a bit different. Something else is corrected here: that it was the castle (and not a different place) that was reached. This sentence would thus not fit as well into the above context. In this example, you could therefore judge the first version as a better fit for the context.

Please choose one of the answers and click on 'Continue'. You can leave the comment field empty here, we will explain later what it is for.

[next page]

In order to be able to make exact reference to concrete situations, we will use a theater scenario. Imagine that you saw a small theater play with a friend. Later you try to recall together what happened in the first scene. Every page of the questionnaire will thus look approximately like this:

This is what you saw:
Scene 1: *(In a picturesque landscape, a knight and a hiker are on their way.)* The knight reaches a castle.
Scene 2: The hiker reacher the castle.

Later you talk about it:
You say: "I think that in the first scene, we saw that the hiker reached the castle."
Your friend replies: "No, in that scene we saw that **the knight reached the castle**."

Listen to the two variants of the sentence and decide which one fits this context better in your opinion. You can leave the comment field empty again.

Instructions for the Contrastive Topic Block

Listen to the first one of the two sentences that you can play at the bottom of this page. You can listen to them several times. Adjust the volume such that you can understand everything clearly.

The first sentence is *Der Ritter hat das Schloss morgens erreicht...* ('The knight reached the castle in the morning....') – but with a special intonation. With this intonation, the sentence sounds incomplete, as if the speaker wanted to imply something specific: namely that somebody else reached the castle, too. The sentence would thus for example fit well into the following kind of context:

Situation:
in the morning: The knight reaches the castle.
in the evening: The hiker reaches the castle.

Someone says: "**The knight reached the castle in the morning.** The hiker reached the castle only in the evening."

Listen to the second sentence now. Again, you can listen to it several times.

It is the same sentence: *Der Ritter hat das Schloss morgens erreicht...* ('The knight reached the castle in the morning...') – but with a different intonation. Something else is implied here: that the knight reached something else besides the castle. This sentence would thus not fit as well into the above context. In this example, you could therefore judge the first version as a better fit for the context.

Please choose one of the answers and click on 'Continue'. You can leave the comment field empty here, we will explain later what it is for.

[next page]

In order to be able to make exact reference to concrete situations, we will use a theater scenario. Imagine that you saw a small theater play with a friend. Later you try to recall together in what order things happened. Every page of the questionnaire will thus look approximately like this:

This is what you saw:
Scene 1: *(In a picturesque landscape)* A knight is traveling on horseback. He reaches a castle.
Scene 2: A hiker is traveling on the same path. He also reaches the castle.

Later you talk about it: You say: "I remember that the knight and the hiker reached a castle. In which scenes did that happen again?"
Your friend replies: "**That the knight reached the castle, we saw in the first scene.** That the hiker reached the castle happened later."

(Later the recordings will be shorter – the part "...we saw in the first scene" will always remain the same and will not be part of the recordings anymore.)

Listen to the two variants of the sentence and decide which one fits this context better in your opinion. You can leave the comment field empty again.

Shorter Version of the Instructions (for second presented block)

(Note: Each participant saw a block with focus items and a block with contrastive topic items. The instruction for the second block were adjusted and shortened. This is illustrated here for the focus instructions.)

We will now explain based on examples what your task in the second part of the study will be.

Listen to the first variant of the two sentences that you can play below.

It is the same sentence again, about the knight and the castle – but with a different intonation. With this intonation, the sentence sounds as though the speaker was correcting something: that it was the knight (and not a different person) who reached the castle. The sentence would thus fit well into the following kind of context in which someone is being corrected (we are using the same kind of theater scenario as before):

Situation:
in the morning: The knight reaches the castle.
in the evening: The hiker reaches the castle.

Someone says: "I think that the following happened in the morning: The hiker reached the castle."
Someone else corrects them: "No, that's not right. In the morning, the following happened: **The knight reached the castle.**"

Listen to the second sentence now. It is the same sentence: *Der Ritter hat das Schloss erreicht* ('The knight reached the castle') – but the intonation is a bit different. Something else is corrected here: that it was the castle (and not a different place) that was reached. This sentence would thus not fit as well into the above context. In this example, you could therefore judge the first version as a better fit for the context.

Please choose one of the answers and click on 'Continue'.

Instructions for the Practice Phase

Next, there will be a practice phase. Here we will use examples for which we already have a certain expectation. The purpose of this phase is to ensure that the instructions are clear and to familiarize yourself with the interface of the experiment. You will still be able to listen to the recordings as many times as you want. As soon as you are ready, you pick the variant of the sentence that is a better fit for the given context in your opinion. If you think that both variants are fine, please nevertheless decide which one of the two is best. If you think that none of the variants fits, you can choose the option 'neither' (in some cases, this is exactly what we expect).

After you picked one of the answers, you will see feedback in which we tell you whether your response matches our expectations. You can then listen to the recordings again. You will have the possibility to change your response, e.g., if you misheard or misclicked. It his however also completely possible that you have a different intuition about the sentence than we expected – in that case, you can stay with your answer.

In the latter case, we would like to ask you to leave a short comment in the text field, if possible. This will help us to better understand why our expectations diverge in this case.

Example of a Practice Phase Trial

This is what you saw:
Scene 1: *(in a festively decorated hall)* A woman enters the stage and sings.
Scene 2: The woman dances.

Later you talk about it.
You say: "I remember that a woman sang and danced. In which scenes did that happen?"
Your friend replies: "**That the woman sang**, we saw in the first scene. That she danced happened later."

(Note: The first recording had a contrastive topic accent on "woman", the second recording on "sing".)

In case the participant chose the second recording, the following feedback was displayed:

"Your answer matches our expectations: the intonation in the second version implies that the woman did something else (besides singing). This fits into the context."

In case the participant chose the first recording or 'neither', the following feedback was displayed:

"Your answer does not match our expectations. Our expectation is that in this context, the second version fits better. The intonation in the second recording implies that the woman did something else (besides singing). This fits the context. The intonation in the first version implies that someone else (besides the woman) sang. This does not fully fit into this context.

You can now listen to the sentences again and change your response if you want. If you stick to your answer, it would be helpful for us, if you could enter a short comment into the text field."

Instructions for the Experiment Phase

You have now completed the practice phase. The following experiment phase will be about examples for which we do not know yet which version is perceived as a better fit for the context by native speakers. There will therefore be no feedback. There are no right or wrong responses – we are interested in your personal intuition about the sentences.

In this phase, comments are not required, but if you would like to add a remark to your response, you can use comment field.