An experimental investigation of (non-)exhaustivity in *es*-clefts[1]

Joseph P. DE VEAUGH-GEISS — *Universität Potsdam*
Swantje TÖNNIS — *Universität Graz*
Edgar ONEA — *Universität Graz*
Malte ZIMMERMANN — *Universität Potsdam*

Abstract. We present an empirical study on exhaustivity inferences in German *es*-clefts compared to definite descriptions (pseudoclefts with an identity statement), exclusives, and focus constructions. Our study uses a novel mouse-driven picture-verification task in which the incremental updating of the context allows one to determine at which point participants take exhaustivity into consideration. Our results are compatible with a parallel analysis of clefts and definite pseudoclefts (see Percus 1997; Büring and Križ 2013; cf. DeVeaugh-Geiss et al. 2015), in so far as both structures are indeed interpreted on a par. In striking contrast to these analyses, however, we found that clefts do not systematically receive exhaustive interpretations, nor do definite pseudoclefts. We conclude that exhaustivity is not conventionally coded in either clefts nor definite pseudoclefts but rather arises through the anaphoric existence presupposition in both constructions together with a number-based implicature.

Keywords: *es*-clefts, definite pseudoclefts, exhaustivity, number-based pragmatic implicature, experimental study

1. Introduction

It has long been discussed in the literature that focus-background *it*-clefts (in German, *es*-clefts) give rise to both an existence as well as an exhaustivity inference, illustrated in (1). (For existence see Horn 1981; Delin 1992; Dryer 1996; Rooth 1999; Abrusán 2016; and for exhaustivity Halvorsen 1976, 1978; Atlas and Levinson 1981; Percus 1997; É. Kiss 1998; Velleman et al. 2012; Büring and Križ 2013, among others.)

(1) Es ist MAX, der einen Cocktail gemischt hat.
it is M. who a cocktail mixed has
'*It is* MAX *who mixed a cocktail.*' (it-*cleft*)
Max mixed a cocktail. (*canonical meaning*)
→ Someone mixed a cocktail. (*existence presupposition*)
→ No one except Max mixed a cocktail. (*exhaustivity inference*)

While there is general consensus that the existence inference is a presuppositon (although see the remarks in Büring and Križ 2013: Sec. 6 for arguments to the contrary), there is a long-standing debate on the semantic-pragmatic source of exhaustivity in clefts. On the one hand,

[1]We would like to thank the Priority Program XPrag.de funded by DFG SPP 1727, without which this research would not be possible. We would also like to thank (in alphabetical order) David Beaver, Judith Degen, Emilie Destruel, Agata Renans, and Maribel Romero for the insightful and critical comments at various stages in this research, as well as the participants at Sinn und Bedeutung 21 and the XPrag.de annual meeting for their feedback. Anna-Christina Boell, Julia Bauermann, and Mareike Philipp were essential in preparing and running the experiments at the Universität Potsdam and Universität Göttingen. All errors are our own.

there are accounts which take exhaustivity to be conventionally-coded and context-independent, which we will refer to as the semantic accounts. In this camp are: (a) the Definite-Semantic analyses that claim *it*-clefts share the presuppositional semantics of definite descriptions (Percus, 1997; Büring and Križ, 2013); and (b) the Inquiry-Terminating account, which takes clefts to encode the same discourse-semantic operators as exclusives, differing only in what is at-issue and what is not-at-issue (Velleman et al., 2012). On the other hand there are the non-conventionally-coded and context-dependent exhaustivity analyses, which we refer to as the pragmatic accounts, that take the exhaustive inference to be a generalized conversational implicature (Horn, 1981, 2014) or a focus-triggered scalar implicature (DeVeaugh-Geiss et al., 2015).[2]

One of the crucial ways in which the two camps differ is in the predictions regarding the robustness and systematicity of the exhaustive inference. With *robustness* we are referring to the strength of exhaustivity in terms of its obligatoriness and cancellability; by contrast, *systematicity* refers to the regularity of the exhaustive inference within and across both experimental settings and participants. In short, a robust inference is context-independent and non-negotiable, and a systematic inference will show up using various experimental methods and not just for a subset of the participants. Semantic analyses predict cleft exhaustivity will be robust and systematic; pragmatic analyses generally make the opposite prediction.

Another crucial way in which the theories differ is in their predictions in terms of *parallelism* with other exhaustivity effects, in particular for our study here, definite pseudoclefts. The definite semantic accounts of Percus 1997 and Büring and Križ 2013 make the clearest of all predictions in this respect, namely that clefts and definite pseudoclefts will pattern alike. The presuppositional account in Velleman et al. 2012, however, does not make any particular predictions regarding parallel behavior of clefts and definites. The pragmatic accounts, by contrast, predict clefts and definite pseudoclefts to elicit different response patterns, but plain focus and clefts may show comparable behavior.[3]

The goal of the two experimental studies presented here is to evaluate cleft exhaustivity along these three parameters, that is, robustness, systematicity, and parallelism. This paper will proceed as follows: In Section 2 we provide the theoretical and empirical background to cleft exhaustivity, and in Section 3 we describe the methods, design, and results of two joint experiments conducted on German *es*-clefts. Anticipating the results a bit, we find that clefts and definite pseudoclefts elicit nearly perfectly-parallel behavior, although neither showed robust nor systematic response patterns, results which no existing theory in the literature predict. In Section 4 we present two proposals which take advantage of the anaphoricity of clefts and definite pseudoclefts in order to account for the (non-)exhaustivity found in both constructions. Section 5 concludes.

[2]See, however, Pollard and Yasavul 2016 for the argument that clefts do not encode exhaustivity *per se*.

[3]However, see DeVeaugh-Geiss et al. 2015, in which clefts are predicted to show stronger exhaustivity effects than plain focus—even in subject position—given unambiguous focus-marking in the former but not in the latter. See also Horn 1981 for the claim that cleft exhaustivity is strengthened by the speaker going "out of her way" by using the cleft construction.

2. Background

Here we provide an overview of the theoretical and experimental work on cleft exhaustivity in the literature. We first take a closer look at the semantic approaches, which we divide into two groups: (a) the Definite-Semantic approaches, in particular Percus 1997 and Büring and Križ 2013, and (b) the Inquiry-Terminating (IT) approach, namely Velleman et al. 2012. Following this we discuss the pragmatic accounts in Horn 1981, 2014, and DeVeaugh-Geiss et al. 2015.

Theoretical approaches Starting with the semantic approaches, Percus (1997) proposed that *it*-clefts of the form *It is X who P* and definite descriptions of the form *The one who P is X* are parallel in their underlying syntax, with the former being derived from the latter syntactically through a process of extraposition (see Percus 1997 for details). Accordingly, the semantics for both constructions are claimed to be identical, with exhaustivity in clefts being equal to the uniqueness/maximality presupposition of the definite. Büring and Križ (2013), by contrast, make no claims regarding the syntax of the two constructions; rather, they employ algebraic semantics (Link, 1983) to model the exhaustivity effect for clefts as a homogeneity presupposition, which they argue accounts for the uniformity presupposition of definite descriptions as well. In this account, a cleft as in (1)—and the definite description counterpart—presupposes that Max is not one of a plurality of cocktail-mixers; i.e., either Max was the only cocktail-mixer, or he did not mix a cocktail at all. Since in both of these analyses cleft exhaustivity is conventionally-coded and context-independent, they make a clear and direct prediction that exhaustivity effects will be robust and systematic, with parallel behavior predicted for their definite description counterparts.

For the IT-construction approach, Velleman et al. (2012) make no particular claim about the parallelism between clefts and definites, instead taking clefts to share the same discourse-semantics with exclusives, namely by encoding the operators MIN and MAX. In particular, exhaustivity is encoded with MAX, which necessitates that no answer to the Current Question be above a certain upper bound. Thus, the answer terminates an inquiry by marking it as maximal. In exclusives MAX is at-issue, whereas in clefts it is not-at-issue. Similar to the semantic definite accounts, the IT-construction account would predict robust and systematic exhaustivity effects, given that the MAX operator is coded in the cleft-structure, but says nothing about parallel behavior between clefts and definites. Nevertheless, one problem with this approach—as pointed out by Pollard and Yasavul (2016)—is that clefts do not always provide a maximal answer, as illustrated in the following:

(2) Q: Who took the last cookie?
A: It was John or Mary, but I don't remember which.

Turning now to the pragmatic approaches, since exhaustivity in clefts does not appear to be as obligatory as semantic accounts might predict, Horn (1981) argued that exhaustivity in clefts is a (generalized conversational) implicature. This analysis is supported in part by data such as the following, in which—unlike the existence presupposition—the exhaustivity inference does not project out of the embedding environment, illustrated in (3).

(3) It isn't Max who mixed a cocktail.
→ Someone mixed a cocktail. (*existence*)
↛ No one except Max mixed a cocktail. (*exhaustivity*)

Furthermore, cleft exhaustivity appears to be easily violable, illustrated in (4).

(4) "Yes, it is bread we fight for—but we fight for roses too!"
(in a poem by James Oppenheim in 1911,[4] cited by Horn 2014)

Horn (1981, 2014) proposes that the exhaustivity inference in clefts has parallels to exhaustivity inferences in other focus constructions, which is argued to be strengthened for clefts by the fact that the speaker has chosen the more marked structure over the canonical word order. This additional strengthening is necessary to account for differences in acceptability for cancellation of plain focus and clefts, illustrated in the following contrast.

(5) a. Max has three children; indeed, he has four.
b. #It was a pizza that Mary ate; indeed, it was a pizza and a calzone.
(examples (18c)–(18d), pg. 133, Horn, 1981)

Although DeVeaugh-Geiss et al. (2015) similarly argue exhaustivity is focus-derived, they take a different pragmatic approach from Horn 1981. In their analysis, clefts are a structural device for marking focus unambiguously, and exhaustivity is a focus-triggered scalar implicature derived from Grice's Maxim of Quantity: that is, "[t]he focal alternatives build scales, and the quantity maxim conversationally implicates the exclusion of alternatives higher on the scale" (DeVeaugh-Geiss et al., 2015: 387). The weaker exhaustivity for plain focus is argued to be due to focus projection, since projection to a higher constituent makes the focal alternatives ambiguous. By contrast, projection out of the cleft pivot appears not to be possible, and thus *it*-clefts provide optimal environments for further pragmatic enrichment. These pragmatic accounts both predict that exhaustivity in clefts is context-dependent and neither robust nor systematic; furthermore, there are no predicted parallels to definite descriptions. A summary of the semantic and pragmatic approaches to cleft exhaustivity and their predictions is found in Table 1. It is worth noting that there are no accounts which predict exhaustivity to be *–robust/systematic* and *+parallel* to definites.

	±robust/systematic	*±parallel definite descriptions*
(A) Semantic Definite	+	+
(B) Semantic IT-Construction	+	+/–
(C) Pragmatic Accounts	–	–
?	–	+

Table 1: Predictions of three theoretical approaches to cleft exhaustivity.

[4] https://web.archive.org/web/20160216133611/https://en.wikipedia.org/wiki/Bread_and_Roses#Words

Experimental work While the theoretical literature tends to bias toward semantic accounts of cleft exhaustivity, experimental work rather supports a pragmatic account, since cleft exhaustivity has been shown to be less robust and less systematic than would be predicted for an inference which is conventionally-coded and context-independent. Prior studies have typically taken exhaustivity in exclusives as a baseline comparison for cleft exhaustivity. One such example is Onea and Beaver (2009), who tested exhaustivity effects in Hungarian preverbal focus using the "*Yes, but...*" test, and the subsequent studies on clefts by Destruel et al. (2015). In this paradigm, participants are presented with question-answer pairs in which, given a visual stimulus with or without an exhaustivity violation, they had to choose between three possible continuations corresponding to different exhaustivity strengths: (i) strictly exhaustive (*No. ...*), (ii) slightly exhaustive (*Yes, but...*), and (iii) non-exhaustive (*Yes, and...*). While exclusives elicited a majority of strictly exhaustive responses (i) in contexts depicting an exhaustivity violation, clefts in English and French elicited a majority of slightly exhaustive (ii) and non-exhaustive (iii) responses. Destruel et al. (2015) argue that the *Yes, but ...* paradigm is a test of the at-issue status of an inference, and given their results, that exhaustivity in clefts is not-at-issue while in exclusives it is at-issue.

However, not-at-issue inferences can have either a semantic or a pragmatic source (Horn, 2014). One diagnostic for the source of an inference traditionally found in the literature is to test whether or not an inference is defeasible: whereas semantic inferences are truth-conditional and cannot be cancelled in unembedded environments,[5] pragmatic inferences are non-truth-conditional and thus typically more easily cancellable, as illustrated in (5a). The cancellation diagnostic is the basis for the experimental studies in Saur 2013 and DeVeaugh-Geiss et al. 2015 on German *es*-clefts. In both studies, the exhaustivity inference from a cleft was followed by a violation (e.g., *Es ist X, der P. Außerdem, Y P.* 'It was X that P. Furthermore, Y P,' Saur 2013). In comparison to exclusives, clefts showed weaker exhaustivity effects when the exhaustive inference was violated: that is, participants more readily accepted cleft sentences despite the violation, which was not the case for sentences with *nur* 'only.' Based on the relative acceptability found in both studies, exhaustivity is argued not to be conventionally coded in clefts, but rather derived as a pragmatic implicature.

Although most studies on exhaustivity compare clefts directly to exclusives, not all have. One such study is Byram-Washburn et al. 2013, in which the acceptability of clefts was tested using written dialogues across various conditions. They find that conditions with violations of contrastiveness led to much lower acceptability ratings than conditions with violations of exhaustivity (see also Destruel and Velleman 2014 for experimental studies on contrastiveness inferences in clefts). Based on these findings, they conclude that exhaustivity is not conventionally-coded but a conversational implicature (in line with Horn 1981). However, it is not clear that the target stimuli were in fact violating exhaustivity, since there is a possible domain narrowing given the temporal adverb *yesterday* found in the target but not in the context. A direct comparison to exclusives could have, in fact, served as a useful control.

[5]Compare to embedded environments, in which semantic inferences, i.e., assertions and presuppositions, are (potentially) interpreted under the embedding operator, and thus cancellable; for presupposition cancellation, see, e.g., Karttunen 1971; Stalnaker 1974; Chierchia and McConnell-Ginet 1996; Abbott 2006; Abrusán 2016, among others.

To sum up, the experimental literature has found that the at-issue status of cleft exhaustivity is different from that of exclusives, with exhaustivity in clefts being not-at-issue and that of exclusives at-issue (in line with previous claims in the theoretical literature, e.g., Horn 1981). Moreover, the results from a growing number of studies are more compatible with a pragmatic approach than a hardwired semantic approach to cleft exhaustivity, given that exhaustivity violations are generally found to be more acceptable for clefts than might be expected for a context-independent and truth-conditional inference. Cancellation of exhaustivity is nonetheless not as acceptable as might be expected for a pragmatic analysis (DeVeaugh-Geiss et al., 2015), and no study to date—as far as the authors are aware of—have directly compared cleft exhaustivity to the maximality inference of definite descriptions, a notable gap between the theoretical and experimental literature.

3. Methods & results

We conducted two joint-studies on German *es*-clefts in order to test the robustness, systematicity, and parallelism of cleft exhaustivity when compared to exclusives, focus, and definite pseudoclefts using a mouse-guided incremental information paradigm written in Python (GNU/Linux v.3.4.2; Windows v.3.3.5) with the PyGame module (v.1.9.2a0, LGPL, Shinners 2011). Participants were presented with contextual information one box at a time in order to measure which truth-value judgment was made ('true' or 'false') as well as at which point there was enough information for participants to make their judgment. The design of both experiments was the same and will be presented together here.

At the beginning of the experiment participants were introduced to four roommates, Jens, Max, Tom, and Ben, who they were told undertook various activities together. Participants were informed that only these four roommates would appear throughout the entire experiment. After a brief introduction to the experiment including three practice trials, the experiment started. At the beginning of each trial, participants heard the audio stimuli in their headphones, after which they uncovered up to four boxes revealing the various activities of the four roommates. Their task was to uncover only as many boxes as necessary to make a judgment whether the sentence they heard was true or false. An example of the four target auditory stimuli is presented in (6)–(9). Pitch accent was always on the subject, illustrated by capital letters in the example stimulus.

(6) EXCLUSIVE
Nur MAX hat einen Cocktail gemischt.
only M. has a cocktail mixed.
'*Only Max mixed a cocktail.*'

(7) PLAIN FOCUS
MAX hat einen Cocktail gemischt.
M. has a cocktail mixed
'*Max mixed a cocktail.*'

(8) DEFINITE PSEUDOCLEFT
Derjenige, der einen Cocktail gemischt hat, ist MAX.
the.one who a cocktail mixed has is M.
'*The one who mixed a cocktail is Max.*'

(9) CLEFT
Es ist MAX, der einen Cocktail gemischt hat.
it is M. who a cocktail mixed has
'*It is Max that mixed a cocktail.*'

Note that in the definite pseudocleft sentences, the complex definite forms *derjenige*, *diejenige*, and *dasjenige* are compounds of the singular determiner elements *der-* 'the.MASC,' *die-* 'the.FEM,' or *das-* 'the.NEUT' plus *-jenige*, the latter derived etymologically from the demonstrative marker *jene/jener/jenes* meaning 'that one (over there).' We will argue that this inherent deictic or context-anaphoric meaning component will prove to be crucial for the interpretation of definite pseudoclefts in our experiments. For all stimuli in the definite pseudocleft condition, the complex-definite subject was singular and masculine, and it displayed singular nominative marking and gender agreement with the masculine proper name in predicative position.

The boxes were manipulated to present information in such a way as to be informative about the source and status of the exhaustivity inference. That is, whereas Box 1 was always irrelevant to the purposes of the experiments, the two experiments differed in a critical way at Box 2.

- Experiment I tested the at-issue status of exhaustivity by verifying in Box 2 the canonical meaning or prejacent of the target sentences. For example, for sentences (6)–(9) Max says: '*I mixed a cocktail*.' If exhaustivity is at-issue, we predict that after revealing the second box participants must continue uncovering Box 3 and Box 4 in order to check that exhaustivity holds.

- Experiment II tested the semantic-pragmatic source by falsifying exhaustivity in Box 2. For example, for sentences (6)–(9) someone other than Max, e.g., Ben, says: '*I mixed a cocktail*.' We predict that if exhaustivity is semantic (asserted or presupposed), falsifying exhaustivity should be enough to judge the target sentence as false; by contrast, if it is pragmatic the exhaustivity implicature is defeasible and thus participants can continue to uncover Box 3 and Box 4 in order to check the as-yet-unverified prejacent.

Although participants were free to choose which box to uncover next, we programmed the experiment such that the order was pre-determined. Moreover, once inside a box the cursor could not exit the box for at least 2000ms. This was done to keep participants from unnecessarily uncovering too many boxes, such as, e.g., by automatically mousing over all four boxes and then making a judgment. Further details of the individual experiments will be presented below.

Our experiments exhibit the following important design features that will allow for a controlled and systematic study of exhaustivity inferences in clefts and definite pseudoclefts.

1. The experiments explicitly control for at-issue semantic exhaustivity triggered by exclusive particles, on the one hand, and for bona fide pragmatic exhaustivity, as triggered by instances of in situ prosodic focus in auditory stimuli.
2. The experiments explicitly control for domain restriction in order to rule out any attempts at explaining exhaustivity violations away in terms of a subsequent enlargement of the quantificational domain.
3. The experiments involve no scalar ordering of alternatives in order to rule out attempts at explaining exhaustivity effects away by recourse to ordering on a contextually-supplied scale.

In both experiments there were two dependent variables: *Early Judgment* and *Late Judgment*. The first dependent variable was measured at Box 2 and had three values: *richtig* 'correct,' *falsch* 'false,' or 'continue' by uncovering Box 3. The second dependent variable was the final evaluation of the stimuli once all relevant information at the third or fourth box had been revealed. Clearly there was data for the *Late Judgment* only when participants chose to continue in the *Early Judgment* measure.

There was a 1:1 target-to-filler ratio. Filler items consisted of sentences with universal quantifiers (*jeder* 'everyone'), expletive constructions (*es ist klar, dass* ... 'it is clear that ...'), plural conjunctions (e.g., *Ben und Max haben* ... 'Ben and Max have ...'), and scalar constructions (*weniger als drei Leute* 'fewer than three people'). There were 32 target items and 32 filler items for a total of 64 sentences.

For Experiment I, we tested 32 German native speakers (24 female, 8 male), all students in Potsdam and Berlin, Germany (average age: 25.6).[6] For Experiment II we tested 32 German native speakers (20 female, 12 male), most of them students in Göttingen, Germany (average age: 27.8). The experiments took part in a laboratory environment, and participants were compensated for their time. We now turn to the individual experiments and their results.

Factorial design of Experiment I Experiment I involved a 4*2 factorial design, the two factors being *Sentence Type* (EXCLUSIVE, FOCUS, DEFINITE PSEUDOCLEFT, CLEFT) and *Exhaustivity* (±EXHAUSTIVE). Recall that the early judgment was at Box 2, which verified the canonical meaning or prejacent, whereas the factor *Exhaustivity* was measured as a late judgment in Box 3 and Box 4. For the condition +EXHAUSTIVE, exhaustivity holds. For instance, given a target sentence as in (6)–(9), in Box 2 Max says he mixed a cocktail, and in the other boxes none of the remaining roommates had the property of mixing a cocktail, as illustrated below. For the condition –EXHAUSTIVE, by contrast, one of the roommates in Box 3 or Box 4 also has the relevant property. For example, in Box 3 Jens says 'I mixed a cocktail,' which, for the sake of space, is not illustrated here.

BOX 1: Tom '*I fetched a straw.*'
BOX 2: **Max** '***I mixed a cocktail.***'
BOX 3: Jens '*I opened a bottle.*'
BOX 4: Ben '*I provided a schnaps.*'

[6]There were 33 participants in Experiment I, but one participant was removed since judgments were erratic on the controls.

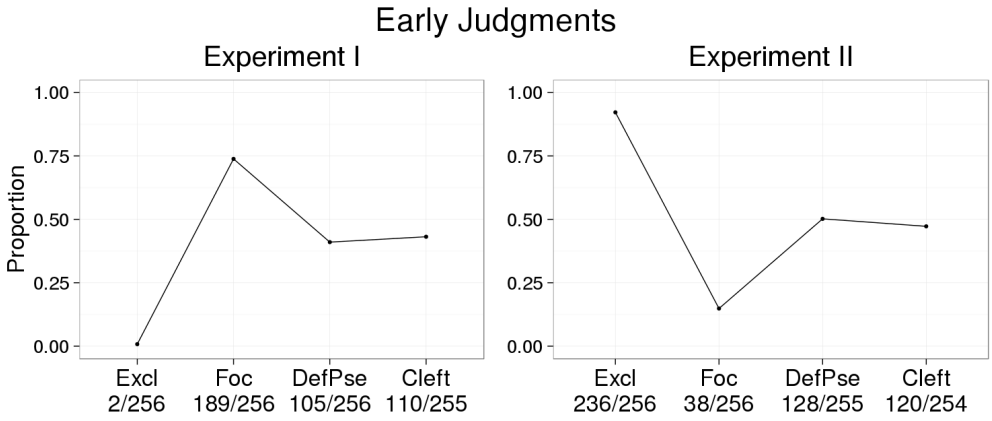

Figure 1: Proportion of early judgments (judgment = 1, continue = 0) for Experiment I (left) and Experiment II (right).

Experiment I results For data preparation of the target items, 1 response out of the 1024 potential responses at Box 1 was treated as an error and removed. All data at Box 2 were coded with 1 for judgment made and 0 for continue. Note that in Experiment I when a participant made a judgment at Box 2, the sentence was always judged 'correct'; that is, no participant selected 'false' at this point. Exclusives elicited a judgment at Box 2 only 1% of the time (2/256 responses): most participants chose to continue uncovering Boxes 3 and 4, as expected. By contrast, clefts and definite descriptions elicited a 'correct' judgment 43% (110/255 responses and 41% of the time (105/256 responses), respectively, and focus 74% of the time (189/256 responses). See the left graph in Figure 1 for the proportion of judgments made per sentence type in Experiment I. Note that when participants continued and exhaustivity was violated in Box 3 or Box 4, the sentence was consistently judged 'false,' suggesting that the experiment was indeed sensitive to exhaustivity.

We conducted a generalized linear mixed effects model (binomial family) in R with the dependent variable being the early response at Box 2 and the independent variable the sentence type (i.e., *glmer(Judgment.Box2 ~ SentType + (1|Item) + (1|Participant), family = binomial, data = ExpI)*).[7] Contrast coding was non-orthogonal: clefts were the baseline comparison for each of the other sentence types. There was no significant difference found between clefts and definite descriptions (SE = 0.286, p = 0.392); by contrast, there was a highly significant difference found between clefts and exclusives, and clefts and focus ($p < 0.001$ in all cases).

Factorial design of Experiment II Experiment II again had a 4*2 factorial design, as in Experiment I, the two factors being *Sentence Type* (EXCLUSIVE, FOCUS, DEFINITE PSEUDO-CLEFT, CLEFT) and *Canonical* (±CANONICAL). Recall that the early response was at Box 2, which falsified the exhaustive inference. The factor *Canonical* was measured as a late response in Box 3 and Box 4. The condition +CANONICAL means the canonical meaning or prejacent holds. For instance, for the sentences in (6)–(9) Max says in Box 3 that he in fact mixed a

[7] We ignore the late response in our computation, because in this early stage of evaluation it plays no role.

cocktail, as illustrated below. For the condition –CANONICAL, by contrast, Max in Box 3 or Box 4 says he did something other than mix a cocktail. For example, in Box 3 Max says 'I provided a schnaps,' which, for the sake of space, is not illustrated here.

BOX 1: Jens	'*I opened a bottle.*'	BOX 3: Max	'*I mixed a cocktail.*'
BOX 2: **Ben**	'***I mixed a cocktail.***'	BOX 4: Tom	'*I fetched a straw.*'

Experiment II results For data preparation of the target items, there were 3/1024 'correct' judgments for the falsifier at Box 2, which were treated as errors. The remaining data at Box 2 were coded with 1 for judgment made and 0 for continue. As predicted, exclusives elicited 'false' judgments 92% of the time (236/256 responses): most participants chose not to continue uncovering contextual information, even though the prejacent had not yet been verified. By contrast, focus elicited judgments only 15% of the time (38/256 responses): that is, most participants continued to uncover the remaining boxes to see if the canonical meaning held. Definite descriptions elicited 'false' judgments 50% of the time (128/255 responses), and clefts were similar in eliciting judgments 47% of the time (120/254 responses). Note that for participants who continued uncovering and found that the prejacent did not hold, they consistently judged the sentence as 'false'; furthermore, when they found that the prejacent was verified, they consistently judged the sentence as 'true.' See the right graph in Figure 1 (on page 355) for the proportion of judgments made per sentence type for Experiment II.

Again, we conducted a generalized linear mixed effects model in R with the dependent variable being the early response at Box 2 and the independent variable the sentence type (i.e., *glmer(Judgment.Box2 ~ SentType + (1|Item) + (1|Participant), family = binomial, data = ExpII)*). Contrast coding was non-orthogonal: clefts were the baseline comparison for each of the other sentence types. Again, there was no significant difference found between clefts and definite descriptions (SE = 0.2223, $p = 0.403$). By contrast, there was a highly significant difference found between clefts and exclusives, and clefts and focus ($p < 0.001$ in all cases).

Post hoc analysis In both experiments we measured whether and at which point participants made a truth-value judgment given the incremental evidence provided. Crucially, we were interested in participant response behavior at Box 2, which differed between the two experiments: in Experiment I the exhaustive inference was verified at Box 2, whereas in Experiment II the exhaustive inference was falsified at Box 2. The questions associated with the early evaluation variable are as follows. In Experiment I we established whether verifying at Box 2 that the canonical meaning holds was sufficient to make a judgment, or whether the exhaustivity inference was taken into consideration by further uncovering Boxes 3 and 4. That is, if a participant judged the sentence they heard as 'true' upon revealing the information at Box 2, exhaustivity did not matter (enough) to justify further investigation. By contrast, if a participant continued this means that exhaustivity was significant enough to warrant checking the upcoming information. In this case we predict participants will judge the sentence as 'true' in the +EXHAUSTIVE late response condition and 'false' in the –EXHAUSTIVE late response condition. This is precisely what we found, confirming that participants understood the logic of the experiment.

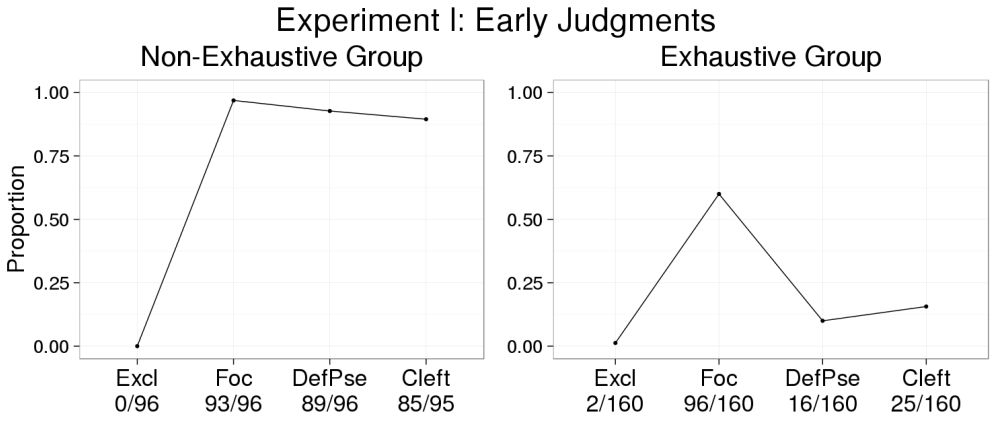

Figure 2: Proportion of early judgments (judgment = 1, continue = 0) for Experiment I for non-exhaustive group (left) and exhaustive group (right).

When analyzing participant behavior individually we found two main groups for definite pseudoclefts and clefts: either participants treated definite pseudoclefts and clefts as exhaustively as exclusives (Exp. I: 20 participants; Exp. II: 16 participants) or as non-exhaustively as focus (Exp. I: 12 participants; Exp. II: 16 participants). These categories were calculated like so: In Experiment I, if they chose 'continue' for definite pseudoclefts 5/8 or more times they fell into the exhaustive interpretation group, since exhaustivity was significant enough to warrant further uncovering a majority of the time; otherwise, they were in the non-exhaustive interpretation group (i.e., 4/8 or less times they chose 'correct' when the canonical meaning of the sentence was verified without checking the remaining boxes). In Experiment II, if they made a 'false' judgment for definite pseudoclefts 5/8 or more times they again fell into the exhaustive interpretation group, since falsifying exhaustivity at Box 2 was enough to judge the auditory stimuli as false; otherwise, they were in the non-exhaustive interpretation group (i.e., 4/8 or less times they chose 'continue' to check that the canonical meaning holds despite exhaustivity having been falsified).

Again, the post hoc analysis, illustrated in Figure 2 for Experiment I (on this page) and Figure 3 for Experiment II (on page 358), shows quite clearly that definite pseudoclefts were indeed interpreted in full parallel to clefts in both the participant groups as well as in the experiments, albeit in two different ways, either entirely exhaustively or entirely non-exhaustively. It is therefore reasonable to assume that the source of the exhaustivity inference is identical, or at least very similar, in both clefts and definite pseudoclefts.

4. Analysis

The experimental data are not in line with any of the major theories of cleft exhaustivity. As pointed out in connection with Table 1 on page 350, neither semantic nor pragmatic analyses would predict exhaustivity in clefts to be non-robust/unsystematic and parallel to definite pseudoclefts at the same time. The post-hoc analysis in particular indicated that our results are incompatible with semantic analyses of cleft exhaustivity (the effect being non-robust and unsystematic across speakers and experiments). Finally, it seems unlikely that the existence of

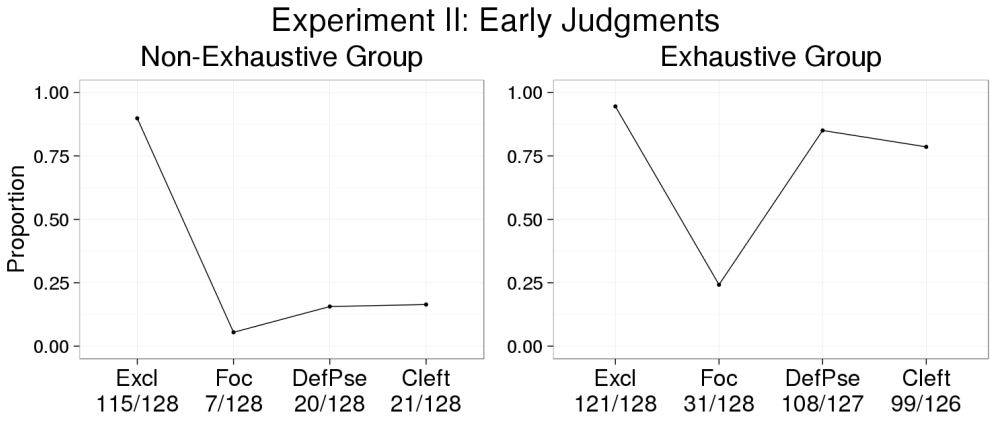

Figure 3: Proportion of early judgments (judgment = 1, continue = 0) for Experiment II for non-exhaustive group (left) and exhaustive group (right).

(unambiguously) marked focal alternatives is responsible for triggering the pragmatic exhaustivity implicature, given that there is weak exhaustivity in the plain focus condtions. In light of all this, we propose an alternative pragmatic analysis of the exhaustivity effect in clefts and definite pseudoclefts. In particular, we argue that our participants systematically differed in their assumptions regarding the potential linguistic contexts required for licensing the critical audio stimulus, in particular its existence presupposition.

Anaphoric presupposition of clefts It is standardly assumed in the literature that clefts have anaphoric potential (Prince 1978; Horn 1981; Soames 1989; Delin 1992; Hedberg 2000; and many others). They introduce as part of their constructional meaning a presupposition that marks the information conveyed by the cleft as known-fact (Prince, 1978) or, simply, as anaphoric (Delin, 1992). The anaphoric potential of clefts can be formally expressed in the form of an existence presupposition, following van der Sandt 1989 and Rooth 1996.

Turning next to the experimental setting of our experiments, there is no linguistic context against which to evaluate the audio stimulus. As a result, the existential presupposition of the cleft condition must be accommodated. This amounts to saying that the hearer will integrate into her discourse model some discourse referent with the relevant property described by the cleft relative that she takes the experimental speaker to (anaphorically) refer to. Crucially, we do not adopt claims in Szabolcsi 1994 (on pre-verbal focus in Hungarian) and Percus 1997 (on English *it*-clefts) that the existential presupposition of cleft sentences comes with an obligatory maximality effect. Instead we propose that part of what the experiment participants did was to reason about the anaphoric antecedent of the existential presupposition. On this proposal, the exhaustive group would take the presupposed discourse referent x to refer to an accommodated maximal discourse referent with property P denoted by the cleft relative.

One way of constructing a suitable discourse referent x would consist in assuming the denotation of an implicit question, as provided in Pollard and Yasavul 2016. That is, participants could have assumed that the cleft addresses the issue "who P?," thus resolving the existence

presupposition to a maximal discourse referent x with property P. Linking this with an identificational at-issue semantics for clefts, namely x = Pivot, the result will be that the maximal individual x with property P equals the pivot, which comes down to an exhaustivity claim. By contrast, the non-exhaustive group would accommodate a non-maximal discourse referent, as suggested by Pollard and Yasavul 2016 for indefinite antecedents. For this group the interpretation is then simply that there is some x with property P, and x = Pivot, which does not trigger an exhaustivity inference.

The foregoing assumptions would, in principle, suffice in order to explain our experimental findings in terms of differences in the way in which the experimental subjects accommodated the anaphoric existence presupposition of clefts. On this analysis, the exhaustivity inference is a pragmatic effect that can be reliably predicted in a number of explicit contexts, but which leads to ambiguity in the absence of overt linguistic context. Still, an analysis along these lines is not without problems. A potential issue is that the proposed analysis relies explicitly on an identificational, as opposed to, e.g., a predicational, semantic analysis of clefts, working in tandem with the presupposed anaphoric content.

Nevertheless, we do not want to dismiss the above sketch of an analysis altogether. At present, though, our experimental data do not provide sufficient and conclusive evidence in favor of this particular spell-out. Fortunately, there is no need for basing the pragmatic analysis of cleft exhaustivity on the semantic assumption that clefts are identificational. In particular, there is an additional implicature giving rise to exhaustivity even on a predicational semantics for clefts. This second implicature interacts with the existence presupposition and is triggered by the choice between singular and plural clefts. The fact that German clefts show consistent and transparent semantic marking of number is assumed to make the effect even stronger.

Number implicature of German clefts As discussed in Büring and Križ 2013, there are plural and singular clefts, as shown in (10).

(10) a. Es waren Georg und Friedrich, die Wilhelm verprügelt haben.
it COP.**PL** Georg and Friedrich REL.MASC.**PL** Wilhelm beaten have
'*It was George and Frederick that beat William.*'

b. Es war Georg, der Wilhelm und Friedrich verprügelt hat.
it COP.**SG** Georg REL.MASC.**SG** Wilhelm and Friedrich beaten has
'*It was George that beat William and Frederick.*'

It is reasonable to assume that the presupposed discourse referent inherits the singular or plural feature from the cleft sentence. Put differently, plural clefts presuppose the existence of a non-atomic sum-individual, while singular clefts at least give rise to the implicature that the anaphoric antecedent is not a sum individual, but rather an atomic individual.

Especially in situations with zero context, as found in our experiments, we expect this implicature to give rise to an exhaustivity inference along the following lines of reasoning: By using the singular cleft, the experimental speaker implicates that there is no sum-individual antecedent with the cleft relative property P. If there actually were sum-individuals with the cleft

relative property P, and the speaker were aware of that fact, the listener would be in no position to identify the singular discourse referent the speaker meant to refer to, given the absence of further linguistic context. This in turn would make it impossible to assign the cleft sentence a proper interpretation, as it remains unclear which individual the cleft sentence is about. Hence, being cooperative, the speaker must have meant to convey that there are no potential plural antecedents at all. Hence, we are safe to assume that there is exactly one individual with property P in the context, which is then identified by the pivot. The core ingredients of the number-based analysis are summed up in (11):

(11) Components of number-based pragmatic analysis of cleft exhaustivity:
   a. Asserted content: P(focus) or focus = x
   b. Existence presupposition: $\exists x[P(x)]$, atomic(x)
   c. Singular-based implicature: $\neg\exists Y[\text{sum}(Y) \wedge P(Y)]$

Note that participants have the option to reason that the speaker is not committed to exhaustivity to a degree that is sufficient to judge the cleft sentence as true or false based on the truth or falsity of this implicature. This can happen, for instance, if a speaker has no relevant knowledge concerning the other individuals under consideration, thereby restricting the scope of the existence presupposition to the singular individual she is certain about. This explains the behavior of the non-exhaustive group.

We would like to stress that the analysis sketched in the previous subsection and the number-based analysis presented here are by no means mutually exclusive. It may very well be possible that both interpretive processes are simultaneously active: The number-contrast would create an exhaustivity implicature based on the anaphoric existential presupposition plus some reasoning over the speaker's intention in choosing a singular over a plural cleft. And, additionally, it would also be possible to maximize the anaphoric antecedent.

Definite pseudoclefts What remains to be done is to show how the pragmatic analysis developed for clefts can be extended in order to capture the parallel interpretive properties of definite pseudoclefts in our experiments. Following a long list of scholars ranging from Frege (1892) to Coppock and Beaver (2015), definite descriptions in general are commonly treated as triggering a uniqueness presupposition. However, at least for the particular definite expressions found in definite pseudoclefts in German, we argue that deriving exhaustivity in an anaphoric familiarity-based analysis à la Heim 1982 better captures the results reported here.

We argue that definite pseudoclefts in German cannot be analyzed as run of the mill definite descriptions, given our experimental findings. Instead, we would like to propose that definite pseudoclefts express anaphoric reference as part of their conventional meaning, as evidenced by their discourse-semantic behavior and by their morpholexical make-up. First, note that the form *-jenige* shows the weak inflectional properties of prenominal modifying elements in definite contexts, but, more importantly, as a deictic expression it must anaphorically relate to a salient discourse referent in the preceding context (or at least to a perceptually salient individual in the utterance situation). We therefore propose that the strong bias for interpreting definite

pseudoclefts as anaphoric expressions, rather than as referentially unique expressions, follows from the explicit presence of the demonstrative (anaphoric) element as part of the complex definite determiner.

Second, observe that definite peudoclefts are deviant as discourse openers, especially in comparison to their plain definite description counterparts, even if the two types of definite expressions have the same descriptive content. The relevant contrast is illustrated in (12). Example (12b) allows for easy accommodation of the fact that the lord, whoever that may be, has been murdered by someone, thereby triggering the interpretation that the gardener was the murderer. Example (12a), in contrast, resists such an interpretation. The most natural interpretation for (12a) is that it presupposes that the murder of the lord has already been the topic of discussion in the preceding discourse, either explicitly or implicitly. This being a condition on discourse structure, and not on the external world as such, it is very hard to accommodate, especially at the beginning of a story.

(12) Out Of The Blue
   a. #Derjenige, der den Lord umgebracht hat, war der Gärtner.
      the.one who the lord murdered has was the gardener
      '*The one who murdered the lord was the gardener.*'
   b. Der Mörder des Lords war der Gärtner.
      '*The murderer of the lord was the gardener.*'

Having established that definite pseudoclefts express an anaphoric relationship in the form of an existence presupposition rather than uniqueness in the utterance situation, we can apply the same reasoning as for the cleft-case, which gives us precisely the same predictions: With uniqueness no longer part of the semantic meaning of the definite pseudocleft expressions, we do not expect systematic or robust uniqueness or exhaustivity effects to show up with this construction. Furthermore, the singular-plural contrast is also observed with definite pseudoclefts, although oddly only in the masculine and feminine paradigm. The masculine and feminine forms *derjenige* and *diejenige* are morphosyntactically and semantically marked for singular, as opposed to their plural counterpart *die-jenigen* 'the.PL-ones.' This accounts for the parallel behavior of clefts and definite pseudoclefts in our experiments. Since we only referred to masculine referents in the experiments, the use of the singular form of the definite pseudocleft should give rise to the same number-based pragmatic reasoning procedure as laid out for the case of clefts above.

By contrast, the neuter form *dasjenige* resembles the neuter determiner/relative form *das* 'the/which' and the neuter *wh*-expression *was* 'what' in being semantically number-neutral (Zimmermann, 2011; Bayer, 2002). As a result, the neuter forms can refer to singular and plural forms alike, as illustrated in (13a), unlike their masculine and feminine counterparts, as in (13b):

(13) a. Dasjenige, was Peter gekauft hat, ist Brot und Rosen.
      DEF.**NEUT** WH.**NEUT** Peter bought has COP.**SG** [bread and roses]$_{\textbf{PL}}$
      '*What Peter bought is bread and roses.*'

b. *Derjenige, den Peter eingeladen hat, ist Klaus und Peter.
DEF.MASC.**SG** REL.MASC.**SG** Peter invited has COP.**SG** [Klaus and Peter]$_{\mathbf{PL}}$
(lit.) '*The one that Peter invited is Klaus and Peter.*'

The difference between masculine/feminine and neuter forms in terms of semantic number marking gives rise to another interesting prediction to be checked in future research: If the number-based pragmatic analysis of pseudoclefts is on the right track, we predict neuter definite pseudoclefts in German to be less exhaustive than their masculine or feminine counterparts. Given that the neuter singular form can refer to both singularities and sum individuals, there is no comparable contrast in the semantic number paradigm, and singular-based implicatures should be much weaker, or altogether absent.

5. Conclusion

We reported the results of two offline experiments on cleft exhaustivity in the incremental information-retrieval paradigm. It was shown that clefts and definite pseudoclefts are treated on a par by the participants of a verification and a falsification experiment, in contrast to sentences with plain intonation foci and to sentences with exclusive particles. In particular, the exhaustivity inference in clefts and definite pseudoclefts is more pronounced than with plain focus, while being less systematic and less robust than with exclusive particles. We have argued that the non-systematic and non-robust nature of the exhaustivity effect is not accounted for by existing theoretical accounts, be they semantic or pragmatic. Moreover, a post hoc analysis further unveiled that about half of the participants treated both clefts and pseudoclefts systematically as exhaustive, while the other half treated both as non-exhaustive. Again, this finding poses a challenge to semantic theories of cleft exhaustivity.

In response to the novel data, we argue that there must be some pragmatic component in the derivation of cleft exhaustivity. We then sketched a pragmatic analysis of cleft exhaustivity in clefts and definite pseudoclefts, which is based on two central assumptions: Both sentence types are anaphoric and introduce an existence presupposition. The choice of singular or plural in clefts and definite pseudoclefts triggers a pragmatic implicature, which is further supported by a systematic contrast in the marking of semantic singular and plural in German. The proposed analysis makes a number of interesting predictions to be investigated in future research, such as crosslinguistic differences or the special case of neuter definite pseudoclefts in German. At the same time, in its present form the proposed analysis is only a sketch, and more theoretical, experimental, and crosslinguistic work needs to be done before a fully detailed compositional analysis of exhaustivity in clefts and definite pseudoclefts is in reach.

References

Abbott, B. (2006). Where have some of the presuppositions gone? In B. Birner and G. Ward (Eds.), *Drawing the Boundaries of Meaning: Neo-Gricean Studies in Pragmatics and Semantics in Honor of Laurence R. Horn*, pp. 1–20. John Benjamins.

Abrusán, M. (2016). Presupposition cancellation: Explaining the 'soft–hard' trigger distinction. *Natural Language Semantics 24*, 165–202.

Atlas, J. and S. Levinson (1981). It-clefts, informativeness and logical form: Radical pragmat-

ics. In P. Cole (Ed.), *Radical Pragmatics*, pp. 1–61. Academic Press.
Bayer, J. (2002). Minimale Annahmen in Syntax und Morphologie. *Linguistiche Arbeitsberichte 79*, 277–297.
Büring, D. and M. Križ (2013, August). It's that, and that's it! Exhaustivity and homogeneity presuppositions in clefts (and definites). *Semantics and Pragmatics 6*(6), 1–29.
Byram-Washburn, M., E. Kaiser, and M. L. Zubizarreta (2013). The English it-cleft: No need to get exhausted. Poster: Linguistic Society of America (LSA).
Chierchia, G. and S. McConnell-Ginet (1996). *Meaning And Grammar: An Introduction To Semantics*. MIT Press.
Coppock, E. and D. Beaver (2015). Definiteness and determinacy. *Linguistics and Philosophy 38*, 377–435.
Delin, J. (1992). Properties of *it*-cleft presupposition. *Journal of Semantics 9*(4), 289–306.
Destruel, E., D. Velleman, E. Onea, D. Bumford, J. Xue, and D. Beaver (2015). A cross-linguistic study of the non-at-issueness of exhaustive inferences. In F. Schwarz (Ed.), *Experimental Perspectives on Presuppositions*, pp. 135–156. Springer.
Destruel, E. and L. Velleman (2014). Refining contrast: Empirical evidence from the English *it*-cleft. In C. Piñón (Ed.), *Empirical Issues in Syntax and Semantics 10*, pp. 197–214.
DeVeaugh-Geiss, J. P., M. Zimmermann, E. Onea, and A.-C. Boell (2015). Contradicting (not-)at-issueness in exclusives and clefts: An empirical study. In *Semantics and Linguistic Theory 25*, pp. 373–393.
Dryer, M. (1996). Focus, pragmatic presupposition and activated propositions. *Journal of Pragmatics 26*(4), 473–523.
É. Kiss, K. (1998). Identificational focus versus information focus. *Language 74*(2), 245–273.
Frege, G. (1892). Über Sinn und Bedeutung. *Zeitschrift für Philosophie und Philosophische Kritik 100*, 25–50. English Translation: *On Sense and Meaning*, in B McGuinness (Ed.), *Frege: Collected Works*, pp. 157–177. Blackwell.
Halvorsen, P. K. (1976). Syntax and semantics of cleft sentences. In *Proceedings from the 12th Chicago Linguistic Society*.
Halvorsen, P. K. (1978). *The Syntax and Semantics of Cleft Constructions*. Ph. D. thesis, University of Texas at Austin.
Hedberg, N. (2000). The referential status of clefts. *Language 76*(4), 891–920.
Heim, I. (1982). *On the Semantics of Definite and Indefinite Noun Phrases*. Ph. D. thesis, University of Massachusetts at Amherst.
Horn, L. (1981). Exhaustiveness and the semantics of clefts. In V. E. Burke and J. Pustejovsky (Eds.), *North Eastern Linguistic Society (NELS) 11*, pp. 125–142.
Horn, L. (2014). Information structure and the landscape of (non-)at-issue meaning. In C. Féry and S. Ishihara (Eds.), *The Oxford Handbook of Information Structure*, pp. 108–127. Oxford University Press.
Karttunen, L. (1971). Some observations on factivity. *Papers in Linguistics 5*, 55–69.
Link, G. (1983). The logical analysis of plurals and mass terms: A lattice-theoretic approach. In R. Bäuerle, C. Schwarze, and A. von Stechow (Eds.), *Meaning, Use, and the Interpretation of Language*, pp. 303–323. de Gruyter.
Onea, E. and D. Beaver (2009). Hungarian focus is not exhausted. In E. Cormany, S. Ito, and D. Lutz (Eds.), *Semantics And Linguistic Theory (SALT) 19*, pp. 342–359.
Percus, O. (1997). Prying open the cleft. In K. Kusumoto (Ed.), *North Eastern Linguistic*

Society (NELS) 27, pp. 337–351.

Pollard, C. and M. Yasavul (2016). Anaphoric it-clefts: The myth of exhaustivity. In *Proceedings of Chicago Linguistic Society (CLS) 50*.

Prince, E. (1978). A comparison of *wh*-clefts and *it*-clefts in discourse. *Language 54*, 883–906.

Rooth, M. (1996). On the interface principles for intonational focus. In T. Galloway and J. Spence (Eds.), *SALT VI*, pp. 202–226.

Rooth, M. (1999). Association with focus or association with presupposition. In P. Bosch and R. van der Sandt (Eds.), *Focus: Linguistic, Cognitive, and Computational Perspectives*, pp. 232–244. Cambridge University Press.

Saur, E.-M. (2013). Clefts: Discourse function and the nature of exhaustivity violation effects. Master's thesis, Universität Potsdam.

Shinners, P. (2011). Pygame. `http://pygame.org/`.

Soames, S. (1989). Presupposition. In D. Gabbay and F. Guenther (Eds.), *Handbook of Philosophical Logic*, Volume IV, pp. 553–616. Reidel.

Stalnaker, R. (1974). Pragmatic presuppositions. In M. Munitz and P. Unger (Eds.), *Semantics and Philosophy*, pp. 197–214. New York University Press.

Szabolcsi, A. (1994). All quantifiers are not equal: The case of focus. *Acta Linguistica Hungarica 42*, 171–187.

van der Sandt, R. (1989). Anaphora and accommodation. In R. Bartsch, J. van Benthem, and P. van Emde Boas (Eds.), *Semantics and Contextual Expression*. Foris.

Velleman, D. B., D. Beaver, E. Destruel, D. Bumford, E. Onea, and L. Coppock (2012). *It*-clefts are IT (Inquiry Terminating) constructions. In A. Chereches (Ed.), *Semantics And Linguistic Theory (SALT) 22*, pp. 441–460.

Zimmermann, M. (2011). Discourse particles. In P. Portner, C. Maienborn, and K. von Heusinger (Eds.), *Semantics: Handbücher zur Sprach- und Kommunikationswissenschaft HSK 33.2*, pp. 2011–2038. Mouton de Gruyter.

A. Appendix: Target auditory stimuli

Target auditory stimuli in German for narrow focus condition. In order to recreate the cleft, definite pseudocleft, and exclusive conditions, follow the examples in (6)–(9) on pages 352–3.

1. Tom hat einen Pullover angezogen.
2. Max hat einen Cocktail gemischt.
3. Jens hat einen Reifen gewechselt.
4. Ben hat einen Koffer gepackt.
5. Jens hat einen Flyer gedruckt.
6. Ben hat eine Katze gestreichelt.
7. Tom hat ein Hemd gebügelt.
8. Max hat ein Gedicht aufgesagt.
9. Jens hat einen Teppich gekauft.
10. Ben hat einen Kuchen gebacken.
11. Tom hat einen Kinderwagen geschoben.
12. Max hat einen Kaktus gepflanzt.
13. Tom hat eine Ziege gefüttert.
14. Max hat eine Schürze genäht.
15. Jens hat ein Regal getragen.
16. Ben hat ein Märchen erzählt.
17. Tom hat einen Weihnachtsbaum geschmückt.
18. Max hat einen Brief geschrieben.
19. Jens hat einen Ball geworfen.
20. Ben hat einen Berg bestiegen.
21. Jens hat eine DVD eingelegt.
22. Ben hat eine Orange ausgepresst.
23. Tom hat ein Steak gebraten.
24. Max hat ein Zimmer aufgeräumt.
25. Jens hat eine Karte gebastelt.
26. Ben hat einen Ofen befeuert.
27. Tom hat einen Tumor entfernt.
28. Max hat einen Fleischspieß gegrillt.
29. Tom hat eine Rechnung bezahlt.
30. Max hat eine Salbe aufgetragen.
31. Jens hat ein Loch gebohrt.
32. Ben hat ein Schwein beobachtet.