

That's not quite it: An experimental investigation of (non-)exhaustivity in clefts*

Joseph P. De Vaugh-Geiss
Universität Potsdam

Swantje Tönnis
Universität Graz

Edgar Onea
Universität Graz

Malte Zimmermann
Universität Potsdam

Submitted 2017-06-02 / First decision 2017-07-17 / Revision received 2018-01-16 /
Accepted 2018-01-19 / Final version received 2018-02-02 / Published 2018-04-18 /
Final typesetting 2024-09-04

Abstract We present a novel empirical study on German directly comparing the exhaustivity inference in *es*-clefts to exhaustivity inferences in definite pseudoclefts, exclusives, and plain intonational focus constructions. We employ mouse-driven verification/falsification tasks in an incremental information-retrieval paradigm across two experiments in order to assess the strength of exhaustivity in the four sentence types. The results are compatible with a parallel analysis of clefts and definite pseudoclefts, in line with previous claims in the literature (Percus 1997, Büring & Križ 2013). In striking contrast with such proposals, in which the exhaustivity inference is conventionally coded in the cleft-structure in terms of maximality/homogeneity, our study found that the exhaustivity inference is not systematic or robust in *es*-clefts nor in definite pseudoclefts: Whereas some speakers treat both constructions as exhaustive, others treat both constructions as non-exhaustive. In order to account for this unexpected finding, we argue that the exhaus-

* We wish to thank Anna-Christina Boell and Julia Hensel for preparation of the German stimuli, and Friederike Buch, Ferdinand Kreuzkamp, Mareike Philipp, and Rico Winkel for their assistance in running the experiments at the Universität Göttingen and Universität Potsdam. This work has greatly benefited by the insightful comments and constructive criticisms of the editor and three anonymous reviewers at *Semantics & Pragmatics*, David Beaver, Judith Degen, Elsi Kaiser, and Maribel Romero as well as from the participants at *Experimental Pragmatics 2015* in Chicago, USA, the *XPrag.de* annual meeting in Göttingen, Germany, the *Experimental Semantics* workshop at ESSLI 2015 in Barcelona, Spain, the *Experimental Pragmatics* workshop at the Societas Linguistica Europaea 2015 in Leiden, Holland, as well as the PhD reading group at the Universität Konstanz, among many others too numerous to list here. This research is funded by the German Research Foundation (DFG) as part of the project “Exhaustivity in *it*-Clefts” in the priority program XPrag.de. All errors are our own.

tivity inference in both clefts and definite pseudoclefts — specifically those with the compound definite *derjenige* — is pragmatically derived from the anaphoric existence presupposition that is common to both constructions.

Keywords: experimental study, exhaustivity, *es*-clefts, definite pseudoclefts, anaphoric existence presupposition

1 Introduction

The sentence in (1) is the German counterpart of the English *it*-cleft provided in the translation. In this paper, we will simply refer to such German sentences as *clefts*, although they are only one of several possible cleft-structures in German and they are mainly known in the literature as *es*-Spaltsätze (*es*-Clefts) (Huber 2002, Altmann 2009). To define the terminology used here, we characterize clefts as constituted by the neuter pronoun *es*, which is arguably an expletive (Gazdar et al. 1985, Pollard & Sag 1994, É. Kiss 1999), a copula verb and a cleft pivot which agree in number and person, and a cleft relative clause with a relative pronoun which agrees in number and gender with the cleft pivot.¹

- (1) Es ist JOHN, der getanzt hat.
it COP.SG John REL.MASC.SG danced has
'It is John who danced.'

In addition to their so-called canonical inference, cleft sentences are frequently claimed to come with two inferences of particular interest for semantic theory: an existential inference and an exhaustivity inference. For example (1), we exemplify these inferences in (2).

- (2) a. Canonical inference: John danced.
b. Existential inference: Somebody danced.
c. Exhaustivity inference: Nobody other than John danced.

There is little controversy in the literature about the canonical and the existential inferences in clefts. While the existential inference is typically as-

¹ In this study, we only consider focus-background clefts, in which the cleft pivot carries the nuclear focal pitch accent and the cleft relative clause is de-accented. As opposed to this, so called topic-comment clefts have also been observed in the literature (for a recent discussion, see Hedberg 2013); however, we have nothing to say about them in this paper.

That's not quite it

signed the status of a presupposition (e.g., Horn 1981, Rooth 1996, Delin 1992, Hedberg 2000) and commonly considered obligatory (Dryer 1996, Rooth 1999; cf. Büring & Križ 2013, discussed further in Section 4.1), the canonical inference is generally taken to be part of the proffered content; that is, it is an at-issue semantic inference. By contrast, the interpretive status and obligatory presence of the exhaustivity inference is very much debated in the literature.²

The semantic literature offers two main sources for the origin of the exhaustivity inference. On the one hand, there is a pragmatic account of clefts, which was first proposed in Horn 1981 and more recently defended in Horn 2014. According to the pragmatic account, cleft exhaustivity is not conventionally coded in the cleft structure itself but is rather a generalized conversational implicature, derived from the fact that clefts also have an existential presupposition. There are variants of this idea which mainly build on the observation that the cleft pivot is focused. For instance, DeVeaugh-Geiss et al. 2015 argue that the source of pragmatic exhaustification in clefts lies in the non-canonical, unambiguous focus marking. A similar conclusion is put forward by the dynamic account of Pollard & Yasavul 2016, according to which *it*-clefts in English are anaphoric expressions that specify their anaphoric antecedent, whereby exhaustivity occurs as part of a question-answer paradigm.

At the same time, many scholars assume a semantic source of cleft exhaustivity, in which the exhaustive inference is conventionally coded in the cleft structure itself. A large part of such semantic accounts builds on the connection between definiteness and clefts and is thus dubbed by us the *semantic definite* account of clefts; see Akmajian 1970, Szabolcsi 1994, Percus 1997, Büring & Križ 2013. Such accounts hold that cleft sentences such as (1) contain a covert determiner element, or some more complex compositional derivation, that makes them semantically equivalent to definite descriptions such as (3), which in turn are assumed to be semantically exhaustive. In particular, the exhaustivity inference is typically modeled as a maximality presupposition (Percus 1997) or as a homogeneity presupposition (Büring & Križ 2013).

² The title of this article, “That’s not quite it: An experimental investigation of (non-) exhaustivity in clefts” is a reference to Büring & Križ’s (2013) *Semantics and Pragmatics* article “It’s that, and that’s it! Exhaustivity and homogeneity presuppositions in clefts (and definites),” which we take issue with here, in particular the claim that *it*-clefts are semantically exhaustive.

- (3) Derjenige, der getanzt hat, ist JOHN.
 DEF.MASC.SG REL.MASC.SG danced has COP.SG John
 ‘*The one that danced is John.*’

A further type of semantic account has recently emerged which suggests that exhaustivity in English clefts, and arguably in German clefts as well, is derived from a conventional interaction between clefts and the question under discussion (sensu Roberts 1996), hence being closer to the focus-based pragmatic accounts. Such accounts include Velleman et al. 2012, Destruel et al. 2015, Beaver & Onea 2015.

Even though this is common practice in the literature, we will not refer to examples like (3) simply as definite descriptions. Example (3) is a specificational construction involving a heavy definite description consisting of a complex definite determiner *der.MASC / die.FEM / das.NEUT -jenige* (involving the distal demonstrative stem *jen*) and a full relative clause. We will call such constructions *definite pseudoclefts*. We choose this theory-neutral label in order to signal two facts. First, the structure in (3) involves a clear definite description on the surface, which certainly plays a role in the semantic interpretation of such structures. But, second, the structure is related but not identical to German pseudoclefts, such as the one in (4), which are also known as *wh*-clefts. In German, such pseudoclefts are built around a *wh*-element in a free relative clause. Crosslinguistic observations, however, show that there are languages in which pseudoclefts obligatorily appear with a definite article, such as Romanian or Spanish (Romero 2005).

- (4) Wer am besten getanzt hat, ist JOHN.
 who the best danced has COP.SG John
 ‘*Who danced the best is John.*’

All three inferences shown in (2) for clefts are also typically attributed to definite descriptions in general, and to definite pseudoclefts like the one in (3) in particular. As with clefts, these inferences have been hotly debated for definite descriptions too. In particular, the presuppositional status of both the uniqueness inference (e.g., Szabo 2000, Ludlow & Segal 2004) and the existential inference (Coppock & Beaver 2015) has been challenged. Still, the mainstream view seems to be that definite descriptions presuppose both existence and uniqueness.

We do not have much to add in this paper to the debate about the existence and uniqueness presuppositions of definite descriptions. What we

That's not quite it

find is that the literature has never considered that definite pseudoclefts differ from more run-of-the-mill definite descriptions in any of these respects. Instead, scholars pointed out parallels between clefts and definite pseudoclefts, also as far as exhaustivity is concerned. They assume this to indicate that clefts are like definites in general. A particularly striking example for this is [Križ 2015](#). Anticipating the discussion to come, we will question this assumption in the discussion of our experimental findings and suggest that clefts may well be similar to definite pseudoclefts without apparently sharing properties of definiteness. In a nutshell, we claim that the exhaustivity inference in definite pseudoclefts occurs independently of the uniqueness presupposition of the definite article, namely as a by-product of resolving an anaphoric existence presupposition.

One general contention in the theoretical literature is whether cleft exhaustivity is conventionally coded and thus clefts should invariably give rise to exhaustivity inferences. This contention has recently been challenged by findings of several experiments on the interpretation of clefts. Studies such as [Onea & Beaver 2009](#), [Destruel et al. 2015](#), [DeVeugh-Geiss et al. 2015](#), [Destruel 2012](#), [Byram-Washburn, Kaiser & Zubizarreta 2013](#) and others report various ways in which cleft exhaustivity does not align with standard expectations raised by semantic inferences of any kind. In other words, the experimental findings point to a pragmatic analysis of the exhaustivity inference in which exhaustivity is not conventionally coded as part of the literal semantic meaning of clefts.

The main objective of this article is to bridge the empirical gap opening up between the vast majority of theoretical accounts, on the one hand, and the experimental findings, on the other, by presenting and discussing the results of two novel experimental studies on the nature of cleft exhaustivity. The experiments come in the form of verification and falsification tasks in an incremental information-retrieval paradigm, akin to the *incremental verification task* (IVT) used by [Conroy \(2008\)](#) to test for the available interpretations of scopally ambiguous strings and by [Franke, Schlotterbeck & Augurzky \(2016\)](#) for literal, local, and global readings of embedded scalars. Our studies improve upon existing experimental studies on *it*-clefts in systematically comparing the interpretation of clefts as in (1), definite pseudoclefts as in (3), plain intonational focus constructions as in (5), and exclusives as in (6).

- | | | | |
|-----|--|-----|--|
| (5) | JOHN hat getanzt.
John has danced.
<i>'John danced.'</i> | (6) | Nur JOHN hat getanzt.
Only John has danced.
<i>'Only John danced.'</i> |
|-----|--|-----|--|

This four-way comparison including uncontroversial instances of pragmatic exhaustivity (plain accent focus) and truth-conditional exhaustivity (exclusives) leads to a more complete view of the problem, and it allows to test for the predictions of a large array of different theories.

Our experimental results show, somewhat surprisingly, that the pragmatic implicature account as well as the semantic definite account are both right *and* wrong to a certain extent. As predicted by the definite account, clefts were interpreted exactly like definite pseudoclefts in the experiments, contrasting with plain foci and exclusives. Conversely, unlike what is predicted by the definite account and other semantic analyses of exhaustivity inferences in clefts, neither clefts nor definite pseudoclefts are obligatorily interpreted as exhaustive. This finding seems to call for a pragmatic account that treats the exhaustivity of clefts and of definite pseudoclefts in parallel ways.

In our analysis of the experimental findings, we suggest a version of Horn's (1981) original analysis, in which the exhaustivity of clefts rests essentially on their existential presupposition. While Horn (1981) had to stipulate a specific pragmatic rule in order to derive exhaustiveness from the existential presupposition, our proposal derives exhaustivity from the anaphoricity of clefts: Exhaustivity arises whenever the anaphoric antecedent of the existential presupposition is interpreted as maximal by the hearer, in a way similar to what Pollard & Yasavul (2016) propose. Crucially, our pragmatic account deviates from the earlier pragmatic implicature account in DeVeugh-Geiss et al. 2015 in that it does not derive an exhaustivity implicature from the explicit and unambiguous structural marking of focal alternatives (cf. Büring 2016). Moreover, we propose a similar analysis for definite pseudoclefts with *derjenige*, suggesting they differ from plain definite descriptions in that they must be interpreted as obligatorily anaphoric expressions and, moreover, these constructions pragmatically derive exhaustivity independently of the maximality semantics of the compound definite article.

The paper is structured as follows: In Section 2 we discuss previous theoretical accounts and their predictions for the behavior of clefts and definite pseudoclefts in semantic experiments. We also provide an overview of previous experimental studies on cleft exhaustivity, together with a brief dis-

That's not quite it

cussion of their shortcomings. We then introduce our own experimental approach and show why it has advantages over previous approaches. Section 3 and Section 4 form the heart of the paper. In Section 3 we describe the experimental set-up and results in detail. In Section 4 we put forward a pragmatic analysis in terms of the anaphoric existence presupposition of clefts and of definite pseudoclefts with *derjenige*, and we will demonstrate how this analysis can account for the experimental data. In Section 5 we conclude with a summary of the main findings and a brief discussion.

2 Theoretical and experimental approaches

In this section we will briefly introduce the predictions of the main theoretical approaches to cleft exhaustivity. We will also sum up the existing experimental findings and discuss some shortcomings of previous experimental studies. These shortcomings will motivate the experimental studies to be presented in the next section. In our studies, we focus on two important interpretive aspects, which constitute the central parameters against which we will evaluate experimental and theoretical approaches: (i) The *strength* of the exhaustivity inference across experimental conditions and speakers; and (ii) the *parallel* behavior of clefts and definite pseudoclefts regarding exhaustivity. The term *strength* is a cover term that refers to the overall robustness and systematicity of an exhaustivity inference. A robust inference is both obligatory and non-cancellable across all contexts for all speakers of a language group. The term systematicity is related to the notion of robustness but it refers specifically to the regularity of exhaustivity across experimental set-ups, experimental conditions, and also across speakers.³ Regarding the question whether exhaustivity in clefts is semantic or whether it is pragmatic, strength seems to be a key feature.

The second parameter, *parallel* behavior, is included because of the important research tradition that derives cleft exhaustivity from an underlying definite structure, which has been assumed to be intimately related if not identical to the structure of definite pseudoclefts. If such approaches, in-

³ Note that robustness and systematicity may occur independently of each other: A systematic inference can be context dependent but still uniform across speakers; however, in this case, it would not be a robust inference. In other words, a robust inference occurs in any context and it is non-cancellable, though possibly only for a sub-group of the population; a systematic inference can be context-dependent, detected with various experimental methods and, most importantly, it is not only attested with a (small) part of the speakers of the respective language.

cluding the recent proposal in [Büring & Križ 2013](#), are on the right track, clefts and definite pseudoclefts are expected to behave in fully parallel ways regarding exhaustivity. Importantly, the parallelism parameter first and foremost touches upon the question of how cleft (and pseudocleft) exhaustivity is structurally derived, and only indirectly upon the question of whether the exhaustivity inference is semantic or pragmatic in nature. Given the widespread contention in the literature that definites are semantically exhaustive, a parallel behavior of clefts and definite pseudoclefts could indeed be taken as evidence for semantic exhaustivity in clefts. However, our experimental findings will suggest that this hypothesis cannot be maintained, and, consequently, that the exhaustivity inference is not semantic in clefts, nor is it in definite pseudoclefts.

In the following, we will first consider the predictions of three types of theories regarding the two dimensions of possible variation. Building on these predictions, we then present some insights from existing experimental data.

2.1 Theoretical predictions

Theoretical approaches to cleft exhaustivity divide into the non-conventionally-coded pragmatic and the conventionally-coded semantic accounts. The most prominent pragmatic account is the implicature analysis in [Horn 1981, 2014](#). According to Horn, the implicature is triggered by the interaction of the obligatory existence presupposition of clefts and the additional use of a non-canonical and less economical cleft structure. In particular, [Horn 1981](#) proposes an idiosyncratic, structure-specific pragmatic principle of derivation according to which if a speaker uses an *it*-cleft of the form *it is α that P* which presupposes $\exists x.P(x)$ and asserts $P(\alpha)$, then she implicates $\forall x.x \neq \alpha \rightarrow \neg P(x)$ in the form of a generalized conversational implicature.

Regarding the two parameters $\pm strength$ and $\pm parallel$ discussed above, the pragmatic implicature account predicts the exhaustivity inference to be subject to cancellability or variability across contextual conditions and speakers. [Horn \(1981\)](#) provides naturally occurring examples such as (7) as points in case. However, [Horn \(1981: 133, ex. 18d\)](#) also notes that out-of-the-blue cancellation is not always possible, as shown in (8). The explanation given by [Horn \(1981\)](#) for cases in which cancellation of the implicature appears to be difficult or impossible is that in such cases the fact that the speaker uses

That's not quite it

a marked structure (the cleft as compared to the canonical sentence) would not be justified if exhaustivity did not hold.

- (7) It's the ideas that count, not just the way we write them.
(Richard Smaby, lecture, *via* Ellen Prince; Horn 1981: ex. 13d)
- (8) #It was a pizza that Mary ate; indeed it was a pizza and a calzone.

So, the main prediction of the pragmatic account is a lack of strength in the exhaustivity inference. There is no clear prediction regarding the parallel behavior of definite pseudoclefts, for the main reason that Horn's theory of cleft exhaustivity makes no claim about definite pseudoclefts at all. However, Horn & Abbott (2016) clearly argue that uniqueness is a conventional part of the meaning of definite descriptions; that is, in their terminology, it is a conventional implicature. This predicts a robust and systematic exhaustivity inference to obtain with definite descriptions, and hence no parallel behavior between clefts and definite pseudoclefts. While we consider this a prediction of Horn's theory, we note in passing that a possible parallel behavior between clefts and definite pseudoclefts in itself does not show that Horn's pragmatic theory of cleft exhaustivity is misguided. It could also be taken to show that his analysis is incomplete, and that Horn & Abbott's analysis of definite descriptions, is, independently, incorrect or at least it does not apply to definite pseudoclefts. More importantly, however, if clefts and definite pseudoclefts behave in a parallel fashion, Horn's theory must be extended to explain this fact irrespective of Horn & Abbott's views on definiteness in general.

Next to the pragmatic analysis, there are two prominent semantic approaches to cleft exhaustivity. The semantic definite accounts treat clefts and definite descriptions as sharing the logical form of identity statements in which a discourse referent is identified with the cleft pivot or restrictor predicate (e.g., Percus 1997, Büring & Križ 2013, Križ 2017). More specifically, in Percus 1997 clefts contain a covert definite operator and have the underlying syntax and semantics of a definite description, whereas in Büring & Križ 2013 clefts and definite descriptions can be treated in parallel in terms of their semantic contribution, although the analysis for clefts does not strictly depend on this. The exhaustivity inference in these approaches is either modeled in terms of a maximality presupposition (Percus 1997, Szabolcsi 1994)⁴ or a ho-

⁴ Although Percus (1997: p. 342) refers to a uniqueness presupposition, we think it is more appropriate to refer to maximality since Percus's analysis would apply to both singular and

mogeneity presupposition (Büring & Križ 2013). For an example like (1), the maximality account presupposes a maximal discourse referent that dances and asserts that this referent is identified with the pivot. By contrast, assuming a homogeneity presupposition, it is asserted that John dances and it is presupposed that John is not a proper mereological subpart of the maximal individual that danced; i.e., either nobody danced or John is the maximal individual that danced.⁵ Most clearly for Percus 1997, and potentially for Büring & Križ 2013 as well, definite pseudoclefts as in (3) are expected to share with their cleft counterparts the asserted and presupposed meaning. Regarding the two main parameters discussed above, the predictions of the definite semantic account for cleft exhaustivity are the clearest of all. If clefts may be considered definites in essence and if both are assumed to conventionally-code exhaustivity, clefts and definite pseudoclefts are expected to show parallel interpretive behavior.⁶ In particular, both sentence types are predicted to exhibit exhaustivity inferences in a robust and systematic manner.

Note, however, that definite descriptions are not a homogeneous class as far as exhaustivity is concerned. Some definites seem to be less exhaustive than others; e.g., weak definites (Schwarz 2009, Barker 2004, Carlson et al. 2006) or seemingly indefinite definites (Carlson & Sussman 2005) do not presuppose uniqueness. In any case, it is not obvious whether it is possible to treat all different kinds of definites alike. Other approaches such as Abbott 2014 distinguish semantic uniqueness and referential uniqueness, defining the latter as follows: “[T]he essence of definiteness in a definite description is that the speaker intends to use it to refer to some particular entity, and (crucially) expects the addressee to be able to identify that very intended referent.” This pragmatic notion of referential uniqueness incorporates the idea that uniqueness may refer to the discourse status of previously mentioned discourse referents or discourse referents entailed by the preceding

plural entities. In particular, Percus explicitly entertains the possibility that the covert NP-proform involved in clefts is unspecified for number, thereby allowing for both singular and plural reference.

⁵ For details on the theoretically-predicted presupposition failure when a plurality such as John and Mary danced, see Büring & Križ 2013 and Križ 2017.

⁶ However, since in Büring & Križ 2013 the analysis of cleft exhaustivity in terms of homogeneity does not depend on the same analysis being extended to definite descriptions, exhaustivity violations as in our experiments may not predict full parallelism between the two sentence types: the cleft will incur a homogeneity violation (assuming a homogeneity presupposition), the singular definite a uniqueness violation (assuming a uniqueness presupposition). We thank an anonymous reviewer for pointing this out.

That's not quite it

discourse. It allows for the use of definite descriptions with familiar (Heim 1982) rather than semantically unique referents, as long as they are identifiable in discourse. The semantic approach to clefts described above, however, analyzes definites, and in particular definite pseudoclefts, as presupposing uniqueness. In Section 4.2, by contrast, we will argue that definite pseudoclefts do not fall into the category of semantically unique definites and adopt a familiarity analysis instead.

The second prominent semantic account of cleft exhaustivity is the inquiry-terminating (IT) construction analysis of Velleman et al. (2012), in which clefts have a semantically predicative form just as their canonical counterparts, with an additional meaning component giving rise to exhaustivity. In this analysis, cleft structures are treated as as conventional devices to give a final and therefore complete answer to a question. In particular, they factor the meaning components of clefts into two components of different discourse-semantic status. At the at-issue level (e.g., Simons et al. 2010, Tonhauser et al. 2013), a cleft asserts the same as the respective canonical sentence would, namely that the predicate denoted by the cleft relative clause holds of the cleft pivot. At the same time, clefts express the not-at-issue inference that all stronger focus alternatives to the cleft prejacent are excluded. The at-issue truth of the prejacent and the exclusion of stronger alternatives are modeled by means of MIN- and MAX-operators, as shown in (9) for the cleft in (1). In this account, clefts have the same semantics as sentences with exclusives (*only*) except for the important difference that with exclusives the at-issue and not-at-issue status of the two components is reversed.

(9) It is John who danced.

At-Issue: $\text{MIN}(\llbracket \text{JOHN danced} \rrbracket) =$ There is a focus alternative that is at least as strong as the proposition *John danced* which is true.

Not-At-Issue: $\text{MAX}(\llbracket \text{JOHN danced} \rrbracket) =$ All stronger focus alternatives entailing *John danced* (e.g., *John and Bill danced*; *John, Bill, and Mary danced*; etc.) are false.

This account makes a clear prediction about the strength of exhaustivity, which is similar to the prediction of the definite account above: Exhaustivity in clefts is expected to be systematic and robust. We do not know of any case where the meaning of definites is treated by means of focus-sensitive MIN- and MAX-operators, and thus, the IT-construction account does not make any

predictions about the parallel behavior of clefts and definite pseudoclefts with regard to exhaustivity.

Summing up, the (A) pragmatic, (B) semantic definite, and (C) semantic IT-construction approaches differ in their predictions regarding the two parameters identified above, i.e., [\pm strength] (robustness and systematicity) and [\pm parallel] with respect to definite pseudoclefts. The predictions of each are schematically presented in Table 1. Strikingly, Table 1 contains only three of the four possible combinations of values for the two parameters. There is one logically possible combination that is not predicted by any formal account of cleft exhaustivity: [- strength] and [+ parallel]. On this setting, cleft sentences are expected to behave like definite pseudoclefts, but, crucially, the interpretive effect would not be robust nor systematic, but rather subject to contextual factors, experimental conditions, or inter-speaker variability. Eventually, we report in Section 3 that it is this hitherto unpredicted combination of parameter values which we find in our experiments. In other words, our experimental findings will show all existing formal theories of exhaustivity in clefts and definite pseudoclefts to be wrong, at least in part. We first turn to existing experimental research on the topic.

	\pm strength	\pm parallel def.pse.
(A) <i>pragmatic</i>	-	-
(B) <i>semantic definite</i>	+	+
(C) <i>semantic IT-construction</i>	+	+/-

Table 1 Predictions of three theoretical approaches to cleft exhaustivity.

2.2 Existing experimental approaches

Recent years have seen an increase in experimental approaches to the interpretation of cleft sentences in English, German, and French. For the most part, the experimental studies were motivated by the fact that the theoretical literature was incapable of settling the exact interpretive status of the exhaustivity inference on the basis of pure introspection and native speaker intuitions. One problem is that intuitions on cleft exhaustivity are often too shaky and variable, necessitating the need for controlled and quantifiable experimental methods; another problem is that different theories tend to focus on different subsets of the data and to disregard others, necessitating a more

That's not quite it

comprehensive study on the relevant aspects of exhaustivity. Notably, while the formal linguistic literature exhibits a preference for semantic analyses of cleft exhaustivity, all existing experimental studies point toward a pragmatic analysis, in line with the pragmatic implicature analysis of Horn 1981, 2014.

The study of Onea & Beaver 2009 (and replications thereof) used the *Yes, but...*-test comparing clefts, exclusives, and canonical sentences. They found that participants chose weaker continuations whenever exhaustivity was violated in a cleft, as compared to exclusives in which exhaustivity is at-issue. These findings indicate that the exhaustiveness of clefts is weaker than would be expected on a semantic account. However, Mayol & Castro-viejo (2013) and Xue & Onea (2011) claim that corrective *but*-responses are in fact contradictions of not-at-issue content in the sense of Simons et al. 2010 and Tonhauser et al. 2013. Hence, the results of Onea & Beaver 2009 just show that exhaustivity in clefts is not-at-issue, but would be in line with a pragmatic as well as a semantic account.

Byram-Washburn, Kaiser & Zubizarreta (2013) used written material and a dialogue-setting for testing the acceptability of clefts comparing exhaustivity violations and violations of contrastiveness inferences, which are also often attributed to clefts (Destruel & Velleman 2014, Destruel, Beaver & Coppock 2018). They found that a violation of contrastiveness leads to much lower acceptability ratings than the violation of exhaustiveness. Hence, they argue that cleft exhaustivity is not a semantically-coded presupposition, but rather a conversational implicature. They are, however, missing a direct comparison with maximality presuppositions, while the presumed interpretive status of the contrastiveness inference as a presupposition is not independently assessed.

Finally, DeVeugh-Geiss et al. (2015) report the results of an acceptability study. The aim of this study was to clarify whether or not the difference in at-issueness between the canonical inference and the exhaustivity inference of clefts is sufficient to explain the apparent weakness of the exhaustivity inference observed for clefts. The study showed that the exhaustivity inference in clefts is easier to cancel than, for instance, the prejacent of exclusive particles (*only*), even though both meaning components are commonly treated as not-at-issue (Horn 2014). Note that the acceptability ratings for exhaustivity cancellations in clefts were still quite low, though, with judgments in the mid-range of a 7-point scale. In a follow-up experiment with definite pseudoclefts in place of *it*-clefts, cancellations of uniqueness were by contrast treated in the same way as cancellations of the prejacent of exclusives, and

again, both inferences are not-at-issue. Hence, DeVeugh-Geiss et al. (2015) argued that at-issueness cannot be the sole factor responsible for the observed weakness. Rather, the experimental findings were taken in support of a pragmatic implicature account of cleft exhaustivity.

Summing up, the previous experimental studies have delivered ample evidence for the different status of the exhaustivity inference in clefts, on the one hand, and the at-issue exhaustivity expressed by exclusive particles, on the other. The experiments also provide some evidence in favor of a pragmatic nature of the exhaustivity inference, which mostly comes in the form of weakening effects (cancellability) and its sensitivity to contextual factors (non-robustness). At the same time, however, the experimental results do not provide conclusive evidence for the pragmatic implicature analysis of cleft exhaustivity. Either there are problems with linking the experimental findings to the nature of pragmatic or semantic inferences (Onea & Beaver 2009); or the exhaustivity effect in clefts is compared to inferences of equally unclear semantic or pragmatic status (Byram-Washburn, Kaiser & Zubizarreta 2013); or the graded values on the acceptability judgment scale are not as high as might be expected on a pragmatic account in which exhaustivity should be defeasible (DeVeugh-Geiss et al. 2015). Moreover, most experimental studies fail to make a direct comparison of the exhaustivity effect in clefts with the maximality presupposition of definites, even though the latter is considered the most likely semantic source of cleft exhaustivity, at least according to large parts of the theoretical literature. Given this state of affairs, we conclude that a more systematic experimental study directly aimed at examining the relevant parameters listed in Section 2.1 is required. In the next section, we describe the experimental setup of such a study.

3 The experiments: Verification and falsification in an incremental information retrieval setting

In light of the above discussion, we propose a novel experimental approach to the study of exhaustivity inferences in clefts. Our approach departs from its predecessors in two ways: First, our experiments involve an explicit comparison between clefts and definite pseudoclefts, as well as a comparison with two well-established control constructions, exclusives and plain intonational focus (i.e., focus marked *via* a pitch accent). The explicit comparison between clefts and definite pseudoclefts should provide useful evidence for establishing whether the source of the exhaustivity inference is the same in

That's not quite it

both structures or not.⁷ The explicit comparison with the control structures, and in particular with the plain focus condition, should provide evidence for establishing whether the exhaustivity inference is pragmatic in nature or not. Second, in order to overcome the observed difficulties in the interpretation of gradient acceptability ratings, we use an incremental information retrieval paradigm that involves decision-making and interpretation procedures, namely verification and falsification with the option of continuation. Given the different kinds of tasks involved in verification and falsification of inferences, we also directly test for the strength of the inference.

Recently, variations on verification and falsification tasks have been employed by [Abrusán & Szendrői \(2013\)](#) in a truth-value judgment task on reference failure for definite NPs, and by [Romoli & Schwarz \(2015\)](#) in a covered-box paradigm on local accommodation of presuppositions (or, conversely, contexts with global presupposition failure) for the trigger *stop* when embedded under negation. We extend such experimental methods here to two classes of (alleged) definite expressions, namely definite pseudoclefts and cleft sentences. Although experimental studies with presupposition contradiction are few, the above studies have in fact found that such sentences result in a majority of 'false' judgments (despite a 'can't say' option) ([Abrusán & Szendrői 2013](#)) and broad rejections (e.g., by selecting the covered box) ([Romoli & Schwarz 2015](#)). Moreover, verification and falsification experiments give rise to categorical judgments, which should in principle allow for an easier identification of non-gradient differences between the two target structures at hand. Finally, our experiments exhibit other important design features that allow for a controlled and systematic study of exhaustivity inferences in clefts and definite pseudoclefts.

- i. The experiments explicitly control for at-issue semantic exhaustivity triggered by exclusive particles and for *bona fide* pragmatic exhaustivity triggered by instances of *in situ* prosodic focus in auditory stimuli.
- ii. 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.

⁷ To avoid confusion, we should mention that when designing the experiment we shared the assumption made in the literature that definite pseudoclefts are plain definite descriptions and that they are, therefore, semantically exhaustive just like plain definites. Given the results of the experiments and further evidence we collected when analyzing the data, we consider this an inappropriate premise.

- iii. The experiments involve proper names referring to four individuals without additional specifications. Hence, there is no ordering of alternatives in terms of informational strength (for instance, scalar items such as *all* being logically stronger than *some*; see the extensive literature on scalar implicatures) in order to rule out attempts at explaining exhaustivity effects away by recourse to ordering on a logical or contextually-supplied scale.

3.1 Method

In this section we provide a general overview of the two mouse-guided sentence-picture verification/falsification experiments, which provide the empirical substance of this paper. Since the timeline and the stimuli in the two experiments were identical, we present the experiments together.

For Experiment I, we tested 32 native speakers of German, all students in Potsdam and Berlin, Germany (24 female, 8 male; average age: 25.6; age range: 20–48).⁸ For Experiment II we tested 32 native speakers of German—distinct from Experiment I—mostly students in Göttingen, Germany (20 female, 12 male; average age: 27.8; age range: 19–52). The experiments took part in a laboratory environment using self-programmed Python scripts (PSFL: GNU/Linux v.3.4.2; Windows v.3.3.5) with the PyGame module (LGPL: v.1.9.2ao, [Shinners 2011](#)). Participants were compensated for their time.

In the instructions to the experiments, participants were introduced to four roommates: Tom, Max, Jens, and Ben. Participants were told that these roommates undertake various activities together. At the start of each trial, participants were presented with a computer screen showing four covered boxes while an audio stimulus played in their headphones. The screen appeared as in picture (i) in Figure 1. After hearing the stimulus, participants were asked to uncover as many boxes as necessary to decide if the audio sentence they heard was true or false. Each box contained an illustration of one of the roommates and a written first person statement about which action this roommate carried out, as in picture (ii) in Figure 1, in which Max says *Ich habe einen Cocktail gemischt* “I mixed a cocktail” in the bottom left box.

⁸ There were 33 participants in Experiment I, but one participant was removed for erratic judgments on the exclusive control condition: 2 ‘continues,’ 2 ‘false,’ and 3 ‘true’ judgments as well as one missing data point at Box 2 due to already having made a judgment at Box 1; cf. the highly consistent decision to continue in the exclusive condition in Experiment I for the remaining 32 participants, seen in the left graph in Figure 2 on page 26.

That's not quite it

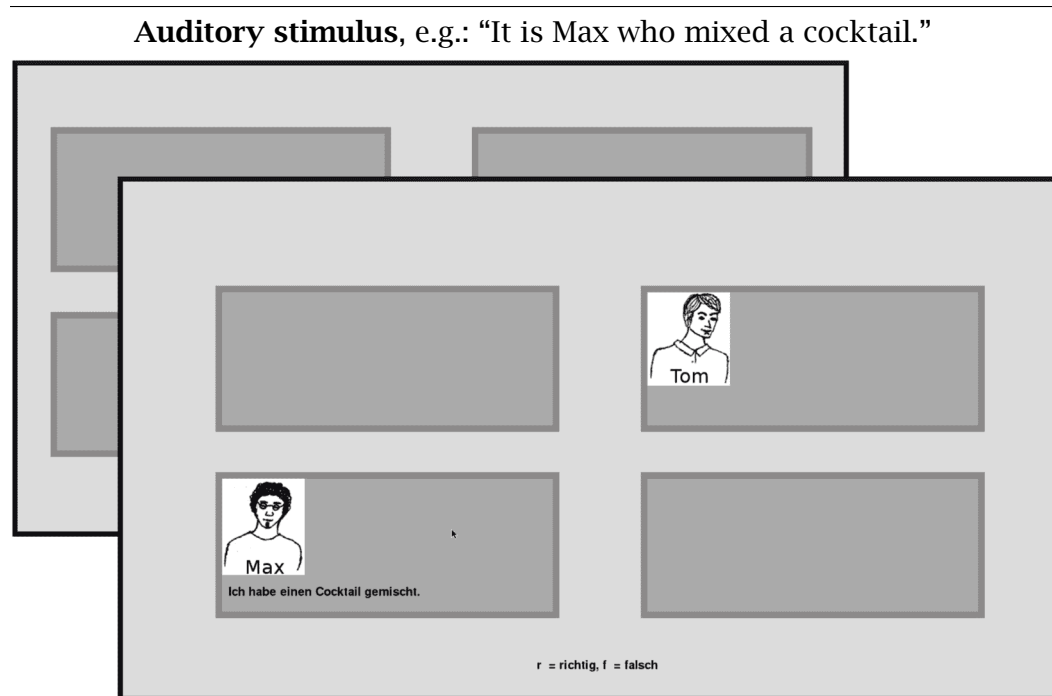


Figure 1 (behind) Start of each trial (front) Uncovering the 2nd box

At any time, participants could press *r* on the keyboard to signal that the sentence is *richtig* 'correct' or *f* to signal that the sentence is *falsch* 'false.' At Boxes 1-3 participants also had the choice of continuing by uncovering the next box.

Participants uncovered the boxes by moving the mouse over them. After entering the box, the cursor could not exit the box for at least 2000 ms. This procedure was intended to keep participants from the unnecessary uncovering of too many boxes, such as, e.g., by automatically mousing over all four boxes and then making a judgment. When the cursor eventually left a given box, the text disappeared while the picture remained visible, although it was possible to move the mouse back into an uncovered box at any point of the trial to see the text again.⁹ Hence, in picture (ii) in Figure 1, the participant

⁹ As pointed out by an anonymous reviewer, this makes the memory load heavier for participants, which may discourage continuing. Indeed, together with the 2-second delay, having the text disappear was intended to make unnecessary uncovering unfavorable. By having the text disappear, participants had to either commit this information to memory or re-uncover this information, thereby requiring an extra step backward; thus, a strategy of revealing un-

had already uncovered the top right box which presented information that is no longer visible, and is currently viewing the bottom left box.

Although participants were free to choose which box they uncovered, they did not know that their choice had no influence on what they saw: The order of uncovering in the experimental setup for each trial was pre-determined, and which location they uncovered did not matter. This was done to prevent any strategies when it came to revealing contextual information. After participants made a judgment, the boxes onscreen were re-covered and the next target or filler item played in their headphones.

Stimuli and presentation Both experiments began with three practice trials to make sure that the participants understood how to control the mouse with respect to the contextual information onscreen, and that their task was to uncover just as many boxes as necessary. If participants uncovered too many boxes in the practice trial, they were reminded not to do so.

For both experiments, the auditory test items consisted of 32 target stimuli and 32 filler items, all in German. The target sentence varied in four sentence-type levels involving (i) clefts, (ii) definite pseudoclefts, (iii) exclusives, and (iv) plain intonational focus constructions, as shown in (10)–(13).

- | | | |
|------|--|-------------|
| (10) | Es ist MAX, der einen Cocktail gemischt hat.
it is Max who a cocktail mixed has
<i>'It is Max who mixed a cocktail.'</i> | CLEFT |
| (11) | Derjenige, der einen Cocktail gemischt hat, ist MAX.
the.one who a cocktail mixed has is Max
<i>'The one who mixed a cocktail is Max.'</i> | DEF. PSE. |
| (12) | Nur MAX hat einen Cocktail gemischt.
only Max has a cocktail mixed
<i>'Only Max mixed a cocktail.'</i> | EXCLUSIVE |
| (13) | MAX hat einen Cocktail gemischt.
Max has a cocktail mixed
<i>'MAX mixed a cocktail.'</i> | PLAIN FOCUS |

In each of the four sentence types, the target sentence gives rise to an exhaustivity inference and a canonical inference, as discussed in the previous sections. These inferences are spelled out in (14).

needed information and only later making a judgment came at a higher cost. The control conditions ensure that, despite these costs, our measurements are nonetheless reliable.

That's not quite it

- (14) a. Exhaustivity: Nobody out of Tom, Jens, and Ben mixed a cocktail.
 b. Canonical inference: Max mixed a cocktail.

There were 32 lexicalizations, with 8 per sentence type distributed in a Latin square design across 4 lists and randomized during presentation. For the targets, grammatical subjects were proper names and grammatical objects were non-specific indefinite determiner phrases with an unspecific interpretation that either referred to an inanimate object or an animal. The reason for using non-specific indefinite object determiner phrases was to avoid any confounding uniqueness effects from additional definite articles in the clause. The non-specific construal was ensured by the absence of narrow pitch accent on the indefinite determiner in the auditory stimulus.

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).’ For all stimuli in the definite pseudocleft condition, the complex definite in subject position was singular and masculine, and it displayed singular nominative marking and gender agreement with the masculine proper name in predicative position.

Given one of the auditory stimuli in (10)–(13), Table 2 gives example stimuli in English for each possibility crossing all the factors for Experiment I and Experiment II. The different factors will be presented in the following.

	Exp. I (verifier)	Exp. II (falsifier)
Box 1	<i>(irrelevant information)</i> Jens: “I opened a bottle.”	
Box 2 (EARLY RESPONSE)	<i>(canonical verified)</i> Max: “I mixed a cocktail.”	<i>(exhaustivity falsified)</i> Ben: “I mixed a cocktail.”
Box 3 / Box 4 (LATE RESPONSE)	[+EXH] <i>(exh. verified)</i> Tom/Ben: “I fetched a straw.” or [-EXH] <i>(exh. falsified)</i> Tom/Ben: “I mixed a cocktail.”	[+CAN] <i>(can. verified)</i> Max: “I mixed a cocktail.” or [-CAN] <i>(can. falsified)</i> Max: “I fetched a straw.”

Table 2 Conditions of Experiment I (verifier) & Experiment II (falsifier).

Factorial design of Experiment I Experiment I involved a 4*2 factorial design, the two factors being SENTENCE TYPE and EXHAUSTIVITY. The EXHAUSTIVITY factor has two levels: [+EXH] and [-EXH]. In the [+EXH] condition no box provides information that would violate the exhaustivity inference triggered by the target sentence. Hence, for our example, Tom, Jens, and Ben report having performed other actions than having mixed a cocktail. By contrast, in the [-EXH] condition the third or fourth box uncovered by the participant provides a piece of information that contradicts the exhaustivity inference. In this case, for instance, the third box would contain the picture of Jens reporting that he (also) mixed a cocktail (see Table 2).

Dependent variables of Experiment I In Experiment I, the second box which was uncovered always entailed that the canonical inference triggered by the target sentence was true; that is why it is called “verifier” in Table 2. Hence, for Experiment I as shown in Table 2, the second box explicitly reveals that Max mixed a cocktail, which is identical to the canonical inference of (10)–(13). The first box never contained any information that would be relevant to the canonical or exhaustivity inference of the target sentence (see Table 2).

With this background, we measured two dependent variables. The first dependent variable was the response immediately following the uncovering of the second box, which had three possible values, i.e., whether the participant judged the sentence true or false immediately or opted to continue by uncovering one or more further boxes. We will call this variable the EARLY RESPONSE. The second dependent variable was the final evaluation of truth or falsity once all relevant information was available (i.e., at the third or fourth box). We will call this the LATE RESPONSE. Obviously, we only had data for the second dependent variable when the early response was to continue.

Factorial design of Experiment II Experiment II involved a 4*2 factorial design, just like Experiment I, the two factors being SENTENCE TYPE and CANONICAL. The four levels of the factor sentence type were identical to Experiment I. The CANONICAL factor has two levels: [+CAN] and [-CAN]. In the [+CAN] condition the third or fourth box reveals the information that the canonical inference triggered by the target sentence is true, e.g., the third box contained the information that Max mixed a cocktail. As opposed to this, the [-CAN] condition reveals the information either in the third or fourth box

That's not quite it

that the canonical inference is false. For our example, Max did something other than mixing a cocktail (see Table 2).

Dependent variables of Experiment II As opposed to Experiment I, in Experiment II the second box which was uncovered always entailed that the exhaustivity inference triggered by the target sentence was false (hence “falsifier”). Accordingly, for our example in Table 2 above, the second box explicitly reveals that one of Tom, Jens, or Ben (in our example it is Ben) mixed a cocktail. Again, the first box never contained any information that would be relevant to the canonical or exhaustivity inference. With this background, we measured exactly the same two dependent variables as in Experiment I, i.e., EARLY RESPONSE and LATE RESPONSE. Of course, the evaluation of these dependent variables is radically different from Experiment I given the different information in boxes 2-4.

Fillers As filler items, we had sentences with the universal quantifier *jeder* ‘everybody,’ as in (15); expletive expressions beginning with *es ist klar ...*, as in (16); subjects containing two conjoined proper nouns, as in (17); as well as the scalar expression *weniger als* ‘fewer than,’ as in (18). There were 8 lexicalizations per sentence type, randomized during presentation, and each participant heard the same 32 filler sentences. For the filler trials, the distribution of possible responses, i.e., verifiers of the canonical meaning and falsifiers of exhaustivity, was balanced across the four boxes with respect to the target stimuli. On top of deflecting participants’ attention from the target constructions at issue, the fillers served the overall purpose of quality control in measuring the reliability of the experimental method.

- (15) Jeder hat ein Buch ausgeliehen.
everyone has a book borrowed.
‘Everyone borrowed a book.’ UNIVERSAL
- (16) Es ist klar, dass Ben eine Geschichte erfunden hat.
it is clear that Ben a story invented has
‘It is clear that Ben invented a story.’ EXPLETIVE
- (17) Ben und Max haben einen Fehler korrigiert.
Ben and Max have a mistake corrected
‘Ben and Max corrected a mistake.’ CONJUNCTION

- (18) Weniger als drei Leute haben ein Bankkonto eröffnet.
fewer than three people have a bank.account opened
'Fewer than three people opened a bank account.' SCALAR

In all targets and filler items, the verb that described the activity was in the present perfect in German, which in English is often translated as simple past, as in the glosses here.

3.2 Summary of the logic of the experiments

In both experiments, we measure whether and at which point the participants decide that the target stimuli are true or false given incremental evidence. Specifically, in both experiments we are interested in how participants will respond at Box 2, which, crucially, is where the two experiments differ: Experiment I verifies the canonical inference at Box 2, whereas Experiment II falsifies the exhaustivity inference at Box 2. The specific questions associated with the Early Response variable for the two experiments are as follows:

Experiment I attempts to establish whether for a cleft or for any of the other analyzed sentence types the knowledge that the canonical inference is true suffices to decide that the cleft sentence is true simpliciter, or whether the exhaustivity inference is also considered by participants. Clearly, if a participant chooses to give a 'true' judgment at this early evaluation stage, this means that for this participant the exhaustivity inference does not matter (enough) to justify further investigation (i.e., *non-exhaustive* responses). As opposed to this, if a participant decides to continue after Box 2, it means that the exhaustivity inference is important enough to be checked against the upcoming incremental information (i.e., *exhaustive* responses), and hence, we expect that the participant will answer 'true' in the final evaluation in the [+EXH] condition and 'false' in the [-EXH] condition. Experiment II attempts to establish whether or not knowing that the exhaustivity inference is false at Box 2 will suffice for participants to judge the whole sentence as false (i.e., *exhaustive* responses), or whether the participants consider it possible for the sentence to still be judged true (i.e., *non-exhaustive* responses). Clearly, if a participant chooses to continue after Box 2, the only rational reason to do so is that for this participant the canonical inference is sufficient to assign the value true. Therefore, we expect that the late evaluation answer only depends on the canonical factor, and we expect that in the final evaluation

That's not quite it

the participant will answer 'true' in the [+CAN] condition and 'false' in the [-CAN] condition.

Theoretical predictions The theoretical predictions for plain intonational foci and exclusives are identical on any of the major theories discussed above and will be discussed together. Plain focus only gives rise to a weak pragmatic exhaustivity implicature, whereas exclusives give rise to a strong semantic and at-issue exhaustivity inference. Since we expect the exhaustivity inference to be frequently disregarded in the former and to be robustly present in the latter, the exhaustiveness patterns in the focus condition provide a baseline for *non-exhaustive* responses, and the exclusive condition provide a baseline for *exhaustive* responses. In terms of concrete experimental outcomes, this amounts to the following predictions for the early responses.

(19) Focus (implicatures calculated on the focus alternatives are not truth-functional)

Exp. I (verifier): Participants will give an early 'true' judgment at Box 2 since the exhaustivity implicature will not require further checking of the context to evaluate the asserted content.

Early Response \Rightarrow 'true'

Exp. II (falsifier): Participants will continue at Box 2 to evaluate the asserted content of the focus clause, since the exhaustivity implicature may not arise or may be cancelled if it is evidently false.

Early Response \Rightarrow 'continue'

(20) Exclusives (asserted exhaustivity is truth-functional)

Exp. I (verifier): Participants will continue until they have uncovered all the boxes to evaluate the asserted exhaustivity.

Early Response \Rightarrow 'continue'

Exp. II (falsifier): Participants will give an early false judgment at Box 2, since a sentence cannot be true if the semantic inferences it gives rise to are false.

Early Response \Rightarrow 'false'

For clefts and definite pseudoclefts, the predictions of the various theories are, naturally, different. Based on the discussion above we can reproduce the predictions from Table 1 (repeated below) in Table 3 including specific experimental outcomes associated with each type of theory.

<i>Theory</i>	\pm <i>strength</i>	\pm <i>parallel</i> <i>def. pse.</i>	<i>Exp. (Box 2)</i>	<i>Early Response at Box 2</i>	
				<i>Clefts</i>	<i>Def. Pse.</i>
(A) <i>pragmatic</i>	-	-	I (verifier) II (falsifier)	true continue	~ continue ~ false
(B) <i>sem. def.</i>	+	+	I (verifier) II (falsifier)	continue false	continue false
(C) <i>sem. IT</i>	+	+/-	I (verifier) II (falsifier)	continue or true false	~ continue ~ false

Table 3 Theoretical predictions for the early responses based on parameters of evaluation from Table 1. The symbol ~ indicates possible responses, since these approaches do not make any specific claims about definite pseudoclefts.

While the translation of Table 1 to the experimental predictions in Table 3 is for the most part straightforward — that is, [-strength] exhaustivity can generally be expected to show parallel response patterns to plain intonational focus, and [+strength] exhaustivity to show parallel response patterns to exclusives — not all cells are subject to clear predictions. For instance, for the (A) pragmatic and (C) semantic IT-construction approaches, a non-parallel behavior of clefts and definite pseudoclefts in these particular environments is not necessarily predicted. After all, these approaches do not make any specific claims about definite pseudoclefts (the ~ symbol merely indicates possible responses in Table 3), for which reason it is certainly compatible with theories (A) and (C) if clefts and definite pseudoclefts were to pattern alike in these particular circumstances. Moreover, theory (C) allows for the possibility that the [+strength] exhaustiveness inference of clefts, which they consider semantic but not-at-issue, will be disregarded in Experiment I because of its being not-at-issue; therefore, in Experiment I theory (C) is compatible with both ‘continue’ as well as ‘true’ responses at Box 2. In a sense then, except for the (B) semantic definite theory — which has a clear position on each of the slots — theories (A) and (C) are less directly tested by our design. Be that as it may, as will become obvious in the following sections, the experimental results we obtained go beyond the predictions of any of these three theories.

That’s not quite it

3.3 Results

Experiment I Data preparation: For Experiment I, there was 1/1024 potential judgments at Box 1 for the target items, which was treated as an error and removed from the statistical analysis, since there is no discernible reason at this point in the procedure to make a truth-value judgment.

We start by describing the results of the Early Response. Exclusives elicited a judgment at Box 2 only 1% of the time (2/256 responses): Almost all participants chose to continue uncovering Box 3 and Box 4. Plain focus, by contrast, elicited a high percentage of judgments at Box 2, namely 74% of the time (189/256 responses): In the majority of trials participants made a ‘true’ judgment without checking the contexts to see whether exhaustivity held. As compared to the two control conditions, clefts and definite pseudoclefts fell somewhere in the middle — at least in the overall numbers and proportions across all participants, but see the post hoc analysis discussed in Section 3.4 — with clefts eliciting a judgment 43% (110/255 responses), and definite pseudoclefts 41% of the time (105/256 responses). Since in all cases in which the Early Response was a judgment it was a ‘true’ judgment, we do not treat this as a three-valued parameter but as a two-valued parameter, i.e., whether a judgment happened. See Figure 2 for the observed proportions of Early Responses in Experiment I (left graph, triangles) made per sentence type.

We conducted a generalized linear mixed effects model for binomial data to compare statistically the likelihood of participants making a (‘true’) judgment.¹⁰ We used treatment contrasts encoded as numeric covariates for the SENTENCE TYPE condition, in which clefts were compared to each of the other levels. Crucially, no significant difference was found between clefts and definite pseudoclefts ($\hat{\beta} = -0.2831$, SE = 0.3076, $z = -0.920$, $p = 0.357$); by contrast, focus was significantly more likely to elicit ‘true’ judgments ($\hat{\beta} = 4.1125$,

¹⁰ We ran generalized linear mixed-effects models for binomial data in R (GPL-2|GPL-3: v.3.3.3, R Core Team 2017) with the *lme4* library (GPL-2|GPL-3: v.1.1-15, Bates, Mächler, et al. 2015). Following recommendations in Bates, Kliegl, et al. 2015, and utilizing the *rePCA* function in the *RePsychLing* library (MIT, v.o.o.4) (available at <https://github.com/dmbates/RePsychLing>), we included variance components/correlation parameters in the random-effects structure supported by the data. The resulting parsimonious models were as follows. Experiment I: *glmer*(TVJ.Box2 ~ DefPse + Foc + (1 + Foc || Participant) + (1 + Foc || Item), family = binomial). Experiment II: *glmer*(TVJ.Box2 ~ DefPse + Excl + Foc + (1 + DefPse + Excl + Foc || Participant) + (1 + Foc || Item), family = binomial). We ignored the late exhaustivity factor in our computation, because in this early stage of evaluation it plays no role.

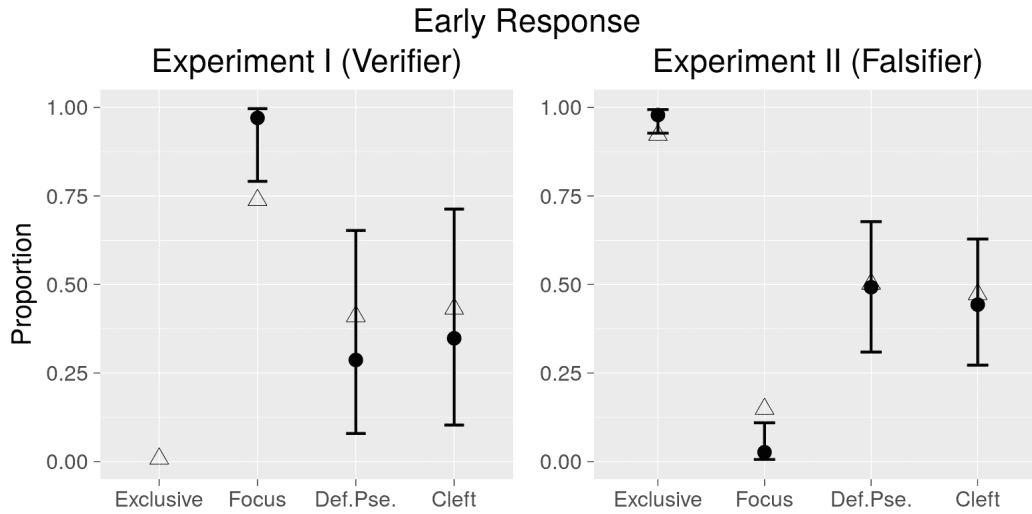


Figure 2 Observed (triangles) and back-transformed predicted proportions (dots, with 95% confidence intervals) for Early Responses for Experiment I (left) and Experiment II (right): judgment = 1, continue = 0. Given percentages close to zero for the exclusive condition in Experiment I, only the observed proportions are presented.

SE = 0.9120, $z = 4.510$, $p = 6.5e-06$). Note that given the difficulty of making meaningful comparisons with percentages close to zero, the exclusive condition was not included in the generalized linear mixed model for Experiment I. See Figure 2 for the back-transformed model-predicted proportions (left graph, dots with 95% confidence intervals) for Early Responses. As can be seen by the differences between the observed and model-predicted proportions — most notably in the plain focus conditions in both experiments, in which the observed proportions lie outside of the 95% confidence intervals — the model predictions very poorly match the observed data. This shows that the mixed-effects logistic regression is an inappropriate model for the data. However, once the participants are divided into responder groups based on the response patterns for clefts as in the post-hoc exploratory analysis presented below, the model predictions do match the data.

In the cases when participants chose to continue (Late Response): In the [+EXH] condition for Box 3 or Box 4, in which exhaustivity was not violated, the final judgment was consistently ‘true’; by contrast, in the [-EXH] condi-

That’s not quite it

			Exclusive	Focus	Def.Pse.	Cleft
Exp. I (verifier)	[+EXH]	‘true’	98% (123/126)	100% (32/32)	99% (76/77)	99% (71/72)
	[-EXH]	‘false’	99% (127/128)	91% (30/35)	91% (67/74)	93% (68/73)
Exp. II (falsifier)	[+CAN]	‘true’	14% (1/7)	96% (105/109)	85% (53/62)	87% (58/67)
	[-CAN]	‘false’	92% (12/13)	99% (108/109)	100% (65/65)	97% (65/67)

Table 4 Late responses as percentages (fractions in parentheses) in [+/-EXH] conditions in Experiment I and [+/-CAN] conditions in Experiment II.

tion, in which exhaustivity was violated, the final judgment was consistently ‘false.’ This is shown in Table 4.

Experiment II Data preparation: for Experiment II, there were 3/1024 ‘true’ judgments at Box 2 upon encountering a falsifier of exhaustivity, which were treated as errors and removed from the statistical analysis since there is no logical reason to make a ‘true’ judgment given the information revealed. Again, since in all cases in which the Early Response was a judgment it was a ‘false’ judgment, we treat this as a two-valued parameter, i.e., whether a judgment happened.

Exclusives elicited ‘false’ judgments 92% of the time at Box 2 (236/256 responses): Most participants chose not to continue uncovering further boxes. By contrast, plain focus elicited ‘false’ judgments only 15% of the time (38/256 responses), with most participants choosing to continue. Definite pseudoclefts elicited ‘false’ judgments 50% of the time (128/255 responses), and clefts were very similar in eliciting judgments 47% of the time (120/254 responses). See Figure 2 for the observed proportions of Early Responses in Experiment II (right graph, triangles) made per sentence type.

We conducted a generalized linear mixed-effects model for binomial data to compare the likelihood of participants making a (‘false’) judgment. Again, we used treatment contrasts encoded as numeric covariates: Clefts were the baseline comparison for all other sentence types. In both experiments, there was no significant difference found between clefts and definite pseudoclefts ($\hat{\beta} = 0.1978$, $SE = 0.2527$, $z = 0.782$, $p = 0.434$). By contrast, exclusives were significantly more likely to elicit ‘false’ judgments ($\hat{\beta} = 4.0413$, $SE = 0.5907$, $z = 6.842$, $p = 7.81e-12$), while focus was significantly more likely to elicit ‘continue’ ($\hat{\beta} = -3.3849$, $SE = 0.7151$, $z = -4.733$, $p = 2.21e-06$). See Figure 2 for

the back-transformed model-predicted proportions (right graph, dots with 95% confidence intervals) for Early Responses. Note again the differences between the observed and model-predicted proportions, namely in the focus condition, showing that the predictions from the model very poorly match the observed data; however, once participants are divided into responder groups as in the post-hoc analysis below, there is in fact a better match between the model predictions and the data.

In the cases when participants continued uncovering (Late Response), the final judgment in the [+CAN] condition was consistently 'true' (with the exception of exclusives in Exp. II; however, note the very low number of data points—i.e., most participants had made an early judgment), whereas the final judgment in the [-CAN] condition was consistently 'false.' This is shown in Table 4.

Post hoc analysis In both experiments the ratio of continue and true/false judgments for clefts and definite pseudoclefts as an early response were about 50-50, instead of the predicted 0-100 or 100-0 (modulo noise). A natural question is, then, whether the midway average arises due to differences in participant's behavior or whether the items created the variation. In a post hoc analysis, we found that when analyzing participant behavior individually two main groups emerged for clefts: In both experiments participants treated clefts either as exhaustively as they did exclusives (Experiment I: 19 participants; Experiment II: 14 participants) or as non-exhaustively as they did plain focus (Experiment I: 13 participants; Experiment II: 16 participants). Only two participants across both experiments responded at chance levels (Experiment II: 2 participants).

These categories were based on percentages for the response patterns in clefts, since after data preparation (in which erroneous judgments were removed; see above) not all participants had the same denominator for total possible judgments at Box 2. The two categories were calculated as follows. Participants who chose 'true' for clefts 60% or more of the time fell into the non-exhaustive interpretation group, generally treating clefts more like focus (i.e., they made a 'true' judgment upon verifying the canonical meaning of the sentence); and participants who in Experiment I chose 'true' for clefts 40% or less fell into the exhaustive interpretation group, treating clefts more like exclusives by continuing a majority of the time. Conversely, in Experiment II if participants made a 'false' judgment for clefts 60% or more of the time, they fell into the exhaustive interpretation group, treating the clefts as they did

That's not quite it

exclusives; and if they made a judgment 40% or less of the time, they were in the non-exhaustive interpretation group (i.e., they generally chose 'continue' upon falsifying exhaustivity, similar to focus). In both experiments, if participants made a judgment between 40–60% of the time, they fell into the chance group. Observed proportions (triangles) for each group are shown in Figure 3 for Experiment I and in Figure 4 for Experiment II.¹¹

The results presented in Figures 3 and 4 show that participants who treated clefts more like exclusives also treated definite pseudoclefts more like exclusives, and the same pattern was found for those who treated clefts like focus. Note again that the two experiments were run with different participants and these findings do not suggest that one and the same participant behaves in an erratic way across the experiments. On subsets of the data corresponding to the exhaustive and non-exhaustive groups for each experiment, we conducted a generalized linear mixed-effects model for binomial data to test the likelihood of making a judgment. We wanted to see in the non-exhaustive groups, whether clefts differed from focus, and in the exhaustive groups, whether clefts differed from exclusives.

For the non-exhaustive groups (left graphs in Figures 3–4) there was a significant difference found between clefts and focus in Experiment I ($\hat{\beta} = 1.6571$, $SE = 0.6992$, $z = 2.370$, $p = 0.0178$), with the focus condition more likely to elicit 'true' judgments in comparison to clefts; but by contrast there was no significant difference found between these two sentence types in Experiment II ($\hat{\beta} = -2.0136$, $SE = 1.1225$, $z = -1.794$, $p = 0.0728$). Inversely, for the exhaustive groups (right graphs in Figures 3–4), a significant difference was found between clefts and exclusives in Experiment II ($\hat{\beta} = 1.2883$, $SE = 0.4970$, $z = 2.592$, $p = 0.00953$), with exclusives more likely to elicit 'false' judgments than clefts.¹² Note again that given the difficulty of making mean-

¹¹ The late responses of those who continued, i.e., the exhaustive group in Experiment I and the non-exhaustive group in Experiment II, support our choice of labels for the two groups. The participants of the exhaustive group did actually judge false if exhaustivity was violated in Box 3 or 4 for Experiment I, and the participants of the non-exhaustive group in Experiment II judge true when the canonical was true in Box 3 or 4 even though exhaustivity was violated in Box 2 (see Table 4).

¹² An anonymous reviewer pointed out that such results could also point toward a three-way distinction between exhaustivity (with exclusives), partial exhaustivity (with clefts and pseudoclefts), and non-exhaustivity (with plain intonational focus), perhaps in parallel to the distinction between factive, semi-factive, and non-factive predicates (Karttunen 1971). However, an account along these lines would have nothing to say on the exact source of partial exhaustivity in clefts and pseudoclefts. In light of this, we favor an account in which the observed differences between clefts/pseudoclefts and focus (non-exhaustive group) or exclusives (ex-

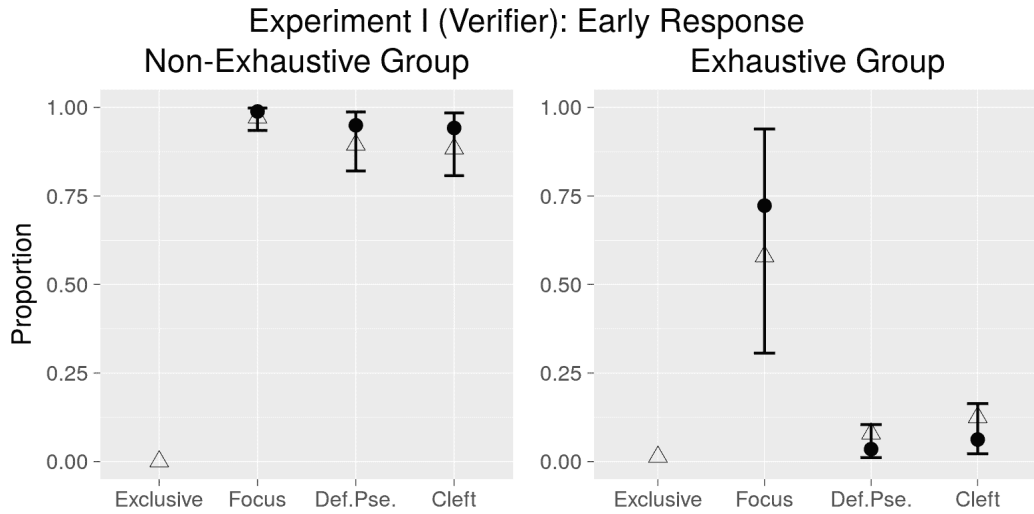


Figure 3 Observed (triangles) and back-transformed predicted proportions (dots, with 95% confidence intervals) for Early Responses for non-exhaustive (left) and exhaustive (right) groups: judgment = 1, continue = 0. Given percentages close to zero for the exclusive condition in Experiment I, only the observed proportions are presented.

ingful comparisons with percentages very close to zero, the exclusive condition was not included in the generalized linear mixed model for Experiment I. Crucially, the post hoc analysis found no statistical difference between clefts and definite pseudoclefts in both groups of participants in both experiments (Exhaustive Group, Exp. I: $\hat{\beta} = -0.5993$, $SE = 0.4206$, $z = -1.425$, $p = 0.154$, Exp. II: $\hat{\beta} = 0.1882$, $SE = 0.4156$, $z = 0.453$, $p = 0.65071$; Non-Exhaustive Group, Exp. I: $\hat{\beta} = 0.1479$, $SE = 0.4879$, $z = 0.303$, $p = 0.7618$, Exp. II: $\hat{\beta} = -0.1447$, $SE = 0.4965$, $z = -0.292$, $p = 0.7707$), albeit in two different ways.

haustive group) are accounted for on the basis of different interpretive processes underlying the exhaustivity inferences in exclusives (truth-functional entailments), clefts/pseudoclefts (accommodation of implicit discourse antecedent), and plain focus (scalar implicature), respectively. In section 4, following the approach in Pollard & Yasavul 2016, we will propose a pragmatic analysis of cleft/pseudocleft exhaustivity that sets it apart from the semantically entailed exhaustivity of exclusives, on the one hand, and the focus-driven scalar exhaustivity implicature of plain focus, on the other.

That's not quite it

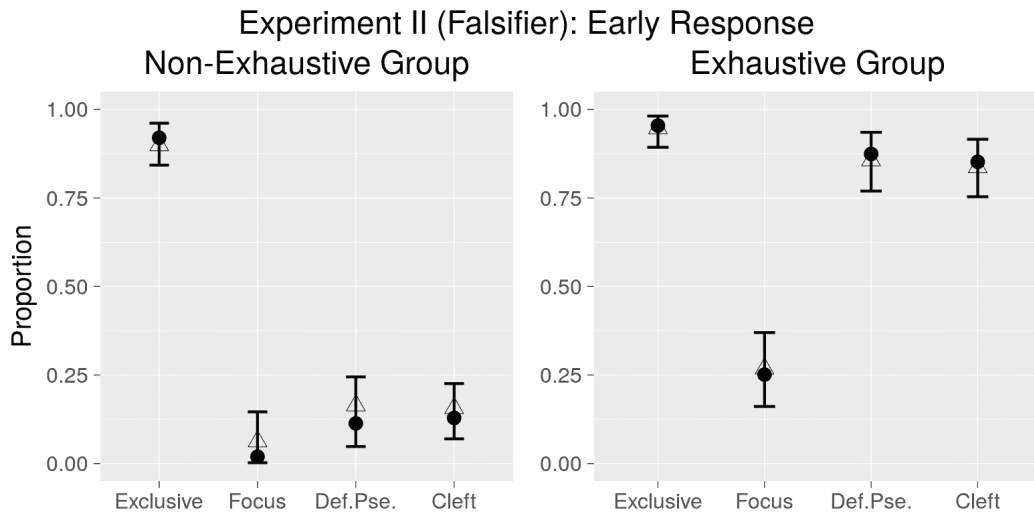


Figure 4 Observed (triangles) and back-transformed predicted proportions (dots, with 95% confidence intervals) for Early Responses for non-exhaustive (left) and exhaustive (right) groups: judgment = 1, continue = 0.

3.4 Discussion

In this section we revisit the logic of the experiments and theoretical predictions discussed in Section 3.2. First, we evaluate participant response patterns in terms of the logic of the experiments, discussing how the results show that the participants understood the task and acted accordingly. This will allow us to disregard from further discussion the late evaluation results, for given the logic of our experiments they are predictable from the early evaluation data. In the second step, we discuss how the results relate to the theoretical predictions. In doing the latter, we will also discuss the results of the post hoc analysis, which have somewhat surprising consequences that are unexpected in light of the existing theoretical literature.

Evaluation of the logic of the experiments In both experiments, we measured whether and at which point the participants decided that the target stimuli were true or false given the incremental evidence provided. Of particular interest was Box 2, which verified the canonical inference in Experiment I and falsified exhaustivity in Experiment II. The primary questions for the

early responses and expected response patterns for the late responses were as follows.

- Experiment I - Early Response (Box 2): Was it enough to verify the canonical inference to make a truth-value judgment (*non-exhaustive* response), or was exhaustivity also considered (*exhaustive* response)? Late Response (Box 3/4): In the latter case, i.e., for those participants for whom exhaustivity was important enough to continue uncovering, we expect ‘true’ responses in the [+EXH] condition and ‘false’ responses in the [-EXH] condition.
- Experiment II - Early Response (Box 2): Does falsifying exhaustivity suffice to judge the whole sentence as false (*exhaustive* response), or did participants consider it still possible to judge the sentence as true by continuing to uncover boxes (*non-exhaustive* response)? Late Response (Box 3/4): In the latter case, i.e., for those participants for whom violating exhaustivity was not sufficient to make a ‘false’ judgment, we expect ‘true’ responses in the [+CAN] condition and ‘false’ responses in the [-CAN] condition.

Indeed, when participants continued uncovering Box 3 and Box 4, the expected patterns for the late responses were precisely what we found (see Table 4 on page 27), and hence the late evaluation data substantiate that participants understood the logic of the experiment; beyond this, the late evaluation data are of no interest.

In a simple semantic model in which we have a sentence E licensing two inferences p and q , it should be obvious that experiments manipulating verification/falsification in the way reported above are expected to produce mirror image results. If a hypothetical speaker finds that p is true and decides that this information suffices to judge E as true, she considers that the inference q is, in some sense, irrelevant; accordingly, the speaker can be expected not to judge E as false if she was presented with the evidence that q is false. For such a speaker, sentence E simply means p , whereas q is not strictly entailed and therefore neglectable. Conversely, if a hypothetical speaker finds that p is true and yet decides to check the truth of q in order to evaluate whether E is true, this person is expected to judge E as false once she sees that q is false. For such a speaker E means at least the logical conjunction between p and q . (For the sake of clarity, recall that in our experiments the participants in Experiment I were distinct from participants in Experiment II.)

That's not quite it

Assuming that exhaustivity is an inference of some type acknowledged in the literature for all the sentence types tested in the experiment, it follows that both an early 'true' judgment in Experiment I, and an early continue decision corresponding to a late 'true' judgment in Experiment II, will indicate that the exhaustivity inference is not as strong as a semantic inference would be expected to be.

Even though the two experiments are mirror images of each other on a simple model, conducting both Experiment I and Experiment II in tandem is not superfluous. Consider the possibility that p is an at-issue inference of E whereas q is a not-at-issue inference. The literature, following [Tonhauser et al. 2013](#), seems by and large to converge in acknowledging that not-at-issue inferences do not form a homogeneous group. A conceivable class of not-at-issue inferences might be such that, on hearing E , the not-at-issue inference q simply does not come to mind as something that has been conveyed. Call this class *not-immediate* inferences and leave it open whether this is an empty class. Crucially, a hypothetical speaker could simply judge E as true without checking for a not-immediate inference q , precisely because q did not come to mind, but was potentially taken for granted or forgotten altogether. As opposed to this, when faced with the explicit falsity of q , a hypothetical speaker is no longer in a state of mind in which q can be disregarded. Hence, in this case, we would expect E to be judged largely true when verified and false when falsified. As the experimental results show, this was not the case, however: The exhaustivity inference of clefts did not behave like not-immediate inferences would be expected to behave.

Moreover, apart from eliminating the above-mentioned source of confound, including both verification and falsification in the experimental setup also serves the purpose to detect and overcome potential biases of participants toward judging sentences true rather than false. If such were the case, we would expect that participants judge the exhaustivity inference true more often in Experiment I than false in Experiment II. The results clearly show that this did not happen.

Evaluation of the theoretical predictions The results show that clefts and definite pseudoclefts behave in an unexpected way when compared to theories (A) to (C) from Table 3. In particular, in both experiments the ratio of continue and true/false judgments as an early response were about 50-50, instead of the predicted 0-100 or 100-0 (modulo noise). Moreover, clefts and

definite pseudoclefts show neither a similarity to exclusives nor a similarity to plain focus.

More interestingly, in the post hoc analysis it was found that participants fell into two groups, and about half of the participants acted as the semantic definite account would have it: These participants judged definite pseudoclefts and clefts almost as exhaustively as exclusives — that is, they cared about exhaustivity and behaved accordingly. By contrast, the other half of participants showed the exact opposite behavior — these participants were willing to identify the referent x in a way which was not exhaustive with respect to P . This constitutes a serious puzzle for the semantic definite account, as one would not expect a semantically hardwired inference to be available for only half of the population. At the same time, it is a serious problem for the pragmatic approach as well, since the exhaustive group did not interpret plain focus in a parallel way to clefts.

In light of this, it is implausible to assume that, for the exhaustive group, the exhaustivity inference in clefts is an implicature that happened to remain uncanceled, while it is subject to cancellation with the non-exhaustive group. More generally, the different behavior of plain focus vs. clefts suggests that the exclusion of salient focus alternatives, possibly per implicature, is not the driving force behind the exhaustivity inference in the latter. Finally, given that exhaustivity is a significant inference in communication, it will not do to assume that there are two dialects of German in order to explain the observable differences between participants. If there were such two dialects (that were not geographically separated by a natural border), their speakers would be expected to show a systematic failure of mutual understanding when a cleft or pseudocleft is used. Instead, a valid explanation for the observed pattern should rather involve some parameter of evaluation that can be reasonably taken to differ for the two participant groups, such that the exhaustivity inference is present only if that parameter is set to a certain value.¹³ Except for [Pollard & Yasavul 2016](#), none of the above mentioned accounts involve such a parameter.

¹³ Moreover, recall that the domain of quantification was explicitly fixed in the experiments, and there was also no scalar ordering of alternatives in terms of noteworthiness or unlikelihood. These features of the experimental set-up rule out typical attempts at explaining exhaustivity violations away as only apparent ([É. Kiss 2010](#), [Skopeteas & Fanselow 2011](#)). The explanation for the split behavior of participants in the cleft and definite pseudocleft conditions must lie elsewhere.

That's not quite it

In conclusion, no difference between clefts and definite pseudoclefts was found in the two experiments conducted. Critically, both sentence types lacked an exhaustive interpretation with about half of the participants. The exhaustivity inference in clefts and definite pseudoclefts was not found to be strong: It was neither robust nor systematic. This observation is incompatible with any of the three main theoretical approaches to cleft exhaustivity in the literature, and calls for an alternative account.

4 Analysis

In this section, we present our main analysis based on the experimental observations above. In particular, we claim that participants only got an exhaustive reading if they took clefts and definite pseudoclefts to anaphorically refer to an implicit question (following [Pollard & Yasavul 2016](#)). We present the core of the analysis in [Section 4.1](#). In [Section 4.2](#), we discuss how our proposal for clefts extends to the case of German definite pseudoclefts with *derjenige*.

4.1 Anaphoricity 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](#)); see in particular [Delin 1992](#) for ample empirical evidence for the anaphoricity of clefts. The anaphoric potential of clefts can be formally expressed in the form of an existence presupposition, following [van der Sandt 1989](#) and [Rooth 1996](#). We further assume that existential presuppositions must be licensed in discourse. Following [van der Sandt \(1992\)](#) and many others, we adopt a dynamic model of discourse and assume that there are two main options for licensing presuppositions, viz. accommodation and binding. Moreover, we follow [van der Sandt \(1992\)](#) and many others in assuming that, whenever possible, presupposition binding in context must be chosen over accommodation.

Empirically, this assumption amounts to the observation that whenever a cleft follows an explicit discourse referent with the relevant properties in the preceding discourse, the cleft must by necessity refer back to that discourse referent. This is shown in [\(21\)](#), which — on its only licit interpreta-

tion — suggests that the discourse referent introduced by the indefinite in the first sentence is anaphorically picked up and further specified in the cleft sentence. In other words, the existential presupposition of the cleft is necessarily dynamically bound by material in the first sentence.

- (21) A: Judy was looking for somebody all afternoon.
 B: It was her youngest daughter that Judy was looking for.

Turning to the experimental setting of our experiments, there is no linguistic context against which to evaluate the audio stimulus: As a result, we assume that the existential presupposition of the cleft condition must be accommodated.¹⁴ 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, say a maximality presupposition built into the structure of clefts. Our reasons for rejecting this common assumption for clefts are as follows: If paired with some sort of maximality effect, the existential presupposition will require the discourse to contain a bound or accommodated discourse referent

¹⁴ An anonymous reviewer asked how we can be sure that the existence presupposition was accommodated rather than simply ignored. Indeed, in the experimental literature there are examples of participants outright ignoring presuppositions in unembedded environments, such as, for instance, with the German iterative *wieder* ‘again’ (Tiemann 2014). In order to account for such cases, Tiemann proposed a maxim of interpretation called *Minimize Accommodation* (MA) — the only principled account for such data the authors are aware of — which dictates: “Do not accommodate a presupposition unless missing accommodation will lead to uninterpretability of the assertion!” (43). As Tiemann (2014: p. 44) writes: “this is a principle that *every interpreter* adheres to when faced with a situation in which s/he cannot ask for further information regarding the PSP” [emphasis added]. Assuming MA is even applicable here, we think such a maxim makes the wrong predictions in light of our data: Most or all participants would be expected to leave the existence presupposition unaccommodated in our experiment (cf. the ignored presupposition of *wieder* in Tiemann 2014). If that were the case, however, then clefts and focus would end up having the exact same semantic contribution and would be predicted to elicit identical response patterns, contrary to what we found; furthermore, we would have no satisfactory account for why half the population treated clefts as exhaustively as exclusives. Thus, we rather assume that participants accommodate the anaphoric existence presupposition, from which one can derive the exhaustive/non-exhaustive interpretation.

That's not quite it

x , such that x has the property P described in the cleft relative, and nobody other than x (in the relevant domain) has property P . Given this, there are two possibilities to consider.

The first possibility is to assume that the compositional semantics of clefts is built around an identity statement such that the cleft pivot x equals the discourse referent y described by the cleft relative ($x = y$). In this case, the presupposed maximal discourse referent with cleft relative property P , namely y , will be identical to the cleft pivot x . This in turn amounts to clefts being semantically exhaustive, in contradiction to our experimental findings. So, we must reject this possibility. The second possibility is to assume that the compositional semantics of clefts does not involve an identity relation, but a plain predication relation instead ($P(x)$). In this case, the cleft would presuppose there to be a maximal discourse referent x with cleft relative property P , and in addition it would assert that it is the cleft pivot x that has the property P . It is easy to see, as pointed out by [Büring & Križ \(2013\)](#), that maximality is vacuously satisfied in this scenario whenever the existence presupposition is satisfied. The reason is that mere existence already entails the existence of a maximal witness, such that the presence of an appropriate referent in the discourse will automatically satisfy maximality. So, postulating an additional maximality presupposition in clefts will either come out as empirically false or as semantically vacuous, depending on the compositional analysis of clefts chosen.

Observe that, up to this point, our analysis shares the fate of [Horn's \(1981\)](#). Horn also assumed that clefts come with an existential presupposition, but on top of this he was forced to invoke a general pragmatic principle in the form of a generalized conversational implicature in order to derive the exhaustivity inference; see [Section 2.1](#) for discussion. Again, the assumption of a general pragmatic principle cannot account for the experimental data at hand, as it would predict a uniform behavior of participants in the experiment, contrary to fact. Instead we propose that part of what the experiment participants did was to reason about the anaphoric antecedent of the cleft's existential presupposition. Building on an idea in [Pollard & Yasavul 2016](#), we will argue that there are two such reasoning procedures, resulting in an exhaustive or non-exhaustive interpretation, respectively. Importantly, both procedures are compatible with an underlying identificational semantics of clefts, in which the value of a variable x is equated with the denotation of the focused cleft pivot (see below). The relevant question is how the value for the variable is resolved to some salient discourse antecedent.

According to Pollard & Yasavul (2016), one way of constructing a suitable discourse referent x in the absence of explicit context consists in taking the cleft to answer an implicit *wh*-question. That is, participants may take a cleft of the form “It is α who P ” to address the question 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 $x = \alpha$, the result will be that the maximal individual x with property P equals the pivot α , which comes down to an exhaustivity claim. This account of cleft exhaustivity relies on the assumption first made by Hamblin 1957 that questions invariably denote sets of *complete* answers, the cleft serving to identify one of those complete answers. The second strategy, according to Pollard & Yasavul (2016), consists in accommodating a non-maximal discourse referent, as is the case, e.g., with indefinite antecedents; see our example (21). On this resolution of the discourse antecedent, the cleft simply expresses that there is some x with property P , and $x = \alpha$, which does not trigger an exhaustivity inference. However, given that indefinites have also been associated with (potential) questions in recent inquisitive semantic analyses (e.g., Onea 2016), it is not obvious to us whether the two resolution strategies should be tied to the presence or absence of a context question.

In view of this problem, we propose the following modified account of the behavior of the exhaustive and non-exhaustive groups in our two experiments, which retains the central insight of Pollard & Yasavul 2016. Members of the exhaustive group predominantly accommodate a discourse antecedent that is maximal with respect to the backgrounded property P , viz. (22a). When casting this in a question-based discourse analysis (Roberts 1996), the corresponding QUD could be either an exhaustively interpreted *wh*-question, or else an identification question (22b). The discourse referent x can be modeled with the iota-operator, and the meaning of the exhaustive-interpreted cleft is shown in (22c).

- (22) a. There is a maximal (sum) individual x that mixed a cocktail.
It's MAX that mixed a cocktail.
- b. Who_{COMPL} mixed a cocktail? / Who is the maximal x that mixed a cocktail?
- c. ASS: $x = \max$, PSP: $\exists x[x \text{ mixed a cocktail}]$
 $\Rightarrow \iota x[x \text{ mixed a cocktail}] = \max$

Members of the non-exhaustive group, by contrast, predominantly chose to accommodate an indefinite (non-maximal) discourse antecedent, viz. (23a),

That's not quite it

as in Pollard & Yasavul 2016. The indefinite gives rise to the *potential question* in (23b) (Onea 2016), an open complement question, resulting in a non-exhaustive interpretation. Technically, the non-maximal discourse referent x can be modeled by means of a choice function (Reinhart 1997, Winter 1997), which picks a random element from the backgrounded cleft property P , as in (23c).

- (23) a. Somebody mixed a cocktail.
It's MAX that mixed a cocktail.
b. Who is this somebody that mixed a cocktail? / Who was it?
c. ASS: $x = \max$, PSP: $\exists x[x \text{ mixed a cocktail}]$
 $\Rightarrow f(\llbracket \text{mixed a cocktail} \rrbracket) = \max$

The foregoing assumptions suffice to explain our experimental findings. On the proposed analysis, the exhaustivity inference is a pragmatic effect that can be reliably predicted in explicit contexts, but which is not mandatory in the absence of overt linguistic context.¹⁵ Depending on whether participants choose a maximal or an indefinite (non-maximal) discourse antecedent, the cleft triggers an exhaustive or a non-exhaustive interpretation, respectively—responses in the early and late measures will pattern accordingly (see Section 3.4 under *Evaluation of the logic of the experiments*). Importantly, the source of the exhaustive effect does not lie in the underlying identificational semantics of the cleft *per se*, but it lies in the different mechanisms for assigning a value to the variable x in the asserted identificational statement $x = \max$, i.e. iota-operator vs. choice-function. Following Reeve's (2012) analysis of the pronoun *it* in *it*-clefts as a referring expression, the underlying identificational semantics of clefts is derived by equating the meaning of the cleft pivot with the meaning assigned to this pronoun, i.e., a contextually salient discourse antecedent. As the literal meaning of *it*-clefts no longer makes reference to maximality/uniqueness, there is no longer a tension between the fact that such sentences express an identificational statement and the fact that they

¹⁵ One reviewer pointed out that given the contextual sensitivity of the exhaustivity interpretation one might instead model the exhaustivity inference as a particularized conversational implicature (PCI). We cannot exclude that possibility, but how to go about spelling out the analysis is not obvious to us. The main problems as we see it are determining which context would need to be assumed to derive the PCI, and moreover, what role the existence presupposition would play, since it would be necessary for such an analysis to predict that canonical sentences do not give rise to the same exhaustivity inference in these contexts.

can be non-exhaustive.¹⁶ Finally, while we take exhaustive inferences with clefts to be pragmatic in nature (Horn 1981, 2014), on our account, they have nothing to do with the exhaustification of focus alternatives, nor with scalar implicatures computed over focus alternatives, pace DeVeugh-Geiss et al. 2015.

4.2 The case of 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. As mentioned in Section 2.1 definites do not seem to constitute a homogeneous class. In the following we want to argue against definite pseudoclefts falling into the same category as semantically unique definites. In particular, we claim that for definite pseudoclefts in German, deriving exhaustivity with an anaphoric familiarity analysis à la Heim 1982 better captures the results reported here.

Following a long list of scholars ranging from Frege (1892) to Coppock & Beaver (2015), definite descriptions in general are commonly treated as triggering a uniqueness presupposition as in (24).

- (24) ‘The NP_{sg}’: Presupposes that the extension of NP has the cardinality smaller or equal to 1.

This predicts a strong exhaustivity effect with definite pseudoclefts, contrary to what we found and reported here. In order to account for the observed absence of exhaustivity effects with about half of the participants, we instead need to resort to a familiarity-based analysis of definiteness.

More precisely, we would like to propose that definite pseudoclefts do indeed express anaphoric reference as part of their conventional meaning, as evidenced by their discourse-semantic behavior and by their morpholexical make-up. These two aspects distinguish definite pseudoclefts from regular definite descriptions which we will discuss in the following. Observe that definite pseudoclefts are deviant as discourse openers, especially in comparison to their plain definite description counterparts, even if the two types

¹⁶ Alternatively, one could analyze *it*-clefts as structurally on a par with definite pseudoclefts (see Section 4.2), and assume, following Percus 1997, that both sentence types contain a (covert) strong, anaphoric definite determiner in the sense of Schwarz 2009. The individuals picked out by such determiner are unique in a weaker sense. They refer to the unique contextually salient discourse antecedent satisfying the backgrounded predicate *P*. As shown in the main text, such discourse antecedents can also be provided by indefinite NPs, resulting in non-exhaustive cleft interpretations.

That's not quite it

of definite expressions have the same descriptive content. The relevant contrast is illustrated in (25). Example (25b) 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 (25a), in contrast, resists such an interpretation. The most natural interpretation for (25a) 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 rather hard to accommodate, especially at the beginning of a story.¹⁷

(25) 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.'

Likewise, (26a) is only acceptable if it is already evident in the preceding discourse that there is a man standing behind the hearer, or at least that there is a salient group of men with the speaker intending to refer to one of them. (26b), in contrast, would be a perfectly natural statement in a general discussion about hats, signaling by way of random example that the man behind the hearer has a relevant kind of hat. In such contexts, (26a) is not licit.

- (26) a. #Derjenige, der hinter dir steht, hat einen tollen braunen
the.one who behind you stands has a fancy brown
Hut.
hat
'The one who's standing behind you is wearing a fancy brown hat.'
- b. Der Mann hinter dir hat einen tollen braunen Hut.
'The man behind you is wearing a fancy brown hat.'

¹⁷ Arguably, the meaning of *derjenige*-phrases may be more complex than described here. Possibly, they also require explicitly or implicitly mentioned alternatives to the DP in the preceding discourse. Since those alternatives were always given in our experiment (in the form of the four roommates) this issue does not influence our analysis. Furthermore, an anonymous reviewer pointed out to us that the observed difference in (25) does not seem to exist in English. The difference between German and English might arise from the fact that German *derjenige*-phrases contain a demonstrative element *jene*- as opposed to the pro-NP form *one* in English *the one who*-phrases.

These observations about regular definite descriptions and definite pseudoclefts again speak in favor of them falling into different categories.

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. To be concrete, we analyze the definite DP in definite pseudoclefts as a strong anaphoric determiner in the sense of Schwarz 2009, which evaluates uniqueness against some contextually salient discourse antecedent. This is shown in the following:

(27) ASS: $\iota y[y \text{ mixed a cocktail} \wedge y = \mathbf{x}] = \max$, PSP: $\exists x[x \text{ mixed a cocktail}]$

As the individual denoted by the definite pseudocleft DP is no longer just evaluated relative to the background predicate P , we do not expect strong uniqueness or exhaustivity effects for this construction. Same as for *it*-clefts, depending on whether the variable x in (27) is resolved to a maximal/unique or to some indefinite antecedent, the pseudocleft will end up with an exhaustive or non-exhaustive interpretation, respectively. This accounts for the parallel behavior of clefts and definite pseudoclefts in our experiments, irrespective of whether or not the two sentence types share the same underlying syntax; see footnote 16.

5 Conclusion

In this paper, we have 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 alike 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 strong 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 participants showed systematic differences and parallel behavior in the interpretation of clefts and definite pseudoclefts. About half of the participants treated both construction types consistently as exhaustive, while the other half treated both construction types as non-exhaustive. Again, this finding poses a challenge to semantic theories of cleft exhaustivity.

That's not quite it

In response to these data, we argue that there must be some pragmatic component in the derivation of cleft exhaustivity. Our approach provides a pragmatic analysis of exhaustivity in clefts and definite pseudoclefts which is based on the assumption that both sentence types are anaphoric and introduce an existence presupposition. The proposed analysis generates some interesting issues to be investigated in future research, such as the role of (implicit or explicit) questions in the (non-)exhaustivity of clefts, as well as possible cross-linguistic differences given the differing discourse-semantics for cleft constructions (e.g., German vs. French). The results reported here and the proposal to account for them provide a stepping stone for more theoretical, experimental, and cross-linguistic work in order to get a more fully detailed compositional analysis of the exhaustivity inference in clefts (and definite pseudoclefts).

References

- Abbott, Barbara. 2014. The indefiniteness of definiteness. In Thomas Gamer-schlag, Doris Gerland, Rainer Osswald & Wiebke Petersen (eds.), *Frames and concept types* (Studies in Linguistics and Philosophy 94), 323-341. Springer. https://doi.org/10.1007/978-3-319-01541-5_14.
- Abrusán, Márta & Kriszta Szendrői. 2013. Experimenting with the king of France: Topics, verifiability and definite descriptions. *Semantics and Pragmatics* 6(10). 1-43. <https://doi.org/10.3765/sp.6.10>.
- Akmajian, Adrian. 1970. On deriving cleft sentences from pseudo-cleft sentences. *Linguistic Inquiry* 1(2). 149-168. <http://www.jstor.org/stable/4177550>.
- Altmann, Hans. 2009. Gradpartikel. In Ludger Hoffmann (ed.), *Handbuch der deutschen Wortarten*, 357-386. Walter de Gruyter. <https://doi.org/10.1515/9783110217087.357>.
- Barker, Chris. 2004. Possessive weak definites. In Ji-yung Kim, Barbara Hall Partee & Yury A Lander (eds.), *Possessives and beyond: Semantics and syntax*, 89-113. MA: GLSA Publications.
- Bates, Douglas, Reinhold Kliegl, Shravan Vasishth & Harald Baayen. 2015. Parsimonious mixed models. arXiv Preprint. <http://arxiv.org/abs/1506.04967>.
- Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1-48. <https://doi.org/10.18637/jss.v067.i01>.

- Beaver, David & Edgar Onea. 2015. Cleftomania. Paper presented at the workshop *Pragmatics of Questions*, Amsterdam.
- Büring, Daniel. 2016. Unalternative semantics. *Semantics and Linguistic Theory (SALT)* 25. 550–575. <https://doi.org/10.3765/salt.v25i0.3634>.
- Büring, Daniel & Manuel Križ. 2013. It's that, and that's it! Exhaustivity and homogeneity presuppositions in clefts (and definites). *Semantics and Pragmatics* 6(6). 1–29. <https://doi.org/10.3765/sp.6.6>.
- Byram-Washburn, Mary, Elsie Kaiser & Maria Luisa Zubizarreta. 2013. The English *It*-cleft: No need to get exhausted. Poster at *Linguistic Society of America (LSA)*.
- Carlson, Greg & Rachel Sussman. 2005. Seemingly indefinite definites. In Stephan Kepser & Marga Reis (eds.), *Linguistic evidence: Empirical, theoretical and computational perspectives* (Studies in Generative Grammar (SGG) 85), 71–86. Berlin: De Gruyter Mouton. <https://doi.org/10.1515/9783110197549.71>.
- Carlson, Greg, Rachel Sussman, Natalie Klein & Michael Tanenhaus. 2006. Weak definite noun phrases. *North East Linguistic Society (NELS)* 36. 179–196.
- Conroy, Anastasia Marie. 2008. *The role of verification strategies in semantic ambiguity resolution in children and adults*. College Park, MD: University of Maryland dissertation.
- Coppock, Elizabeth & David Beaver. 2015. Definiteness and determinacy. *Linguistics and Philosophy* (38). 377–435. <https://doi.org/10.1007/s10988-015-9178-8>.
- Delin, Judy. 1992. Properties of *It*-cleft presupposition. *Journal of Semantics* 9(4). 289–306. <https://doi.org/10.1093/jos/9.4.289>.
- Destruel, Emilie. 2012. The French *c'est*-cleft: An empirical study on its meaning and use. In Christopher Piñón (ed.), *Empirical Issues in Syntax and Semantics: Selected papers from CSSP*, 95–112. http://www.cssp.cnrs.fr/eiss9/eiss9_destruel.pdf.
- Destruel, Emilie, David Beaver & Elizabeth Coppock. 2018. Clefts: Quite the contrary! *Sinn und Bedeutung (SuB)* 21. <https://ojs.ub.uni-konstanz.de/suub/index.php/sub/article/view/141>.
- Destruel, Emilie, Daniel Velleman, Edgar Onea, Dylan Bumford, Jingyang Xue & David Beaver. 2015. A cross-linguistic study of the non-at-issueness of exhaustive inferences. In Florian Schwarz (ed.), *Experimental perspectives on presuppositions*, 135–156. Springer. https://doi.org/10.1007/978-3-319-07980-6_6.

That's not quite it

- Destruel, Emilie & Leah Velleman. 2014. Refining contrast: Empirical evidence from the English *it*-cleft. In Christopher Piñón (ed.), *Empirical issues in syntax and semantics: Selected papers from CSSP*, 197–214. http://www.cssp.cnrs.fr/eiss10/eiss10_destruel-and-velleman.pdf.
- DeVeugh-Geiss, Joseph P., Malte Zimmermann, Edgar Onea & Anna-Christina Boell. 2015. Contradicting (not-)at-issueness in exclusives and clefts: An empirical study. *Semantics and Linguistic Theory (SALT)* 25. 373–393. <https://doi.org/10.3765/salt.v25i0.3054>.
- Dryer, Matthew. 1996. Focus, pragmatic presupposition and activated propositions. *Journal of Pragmatics* 26(4). 473–523. [https://doi.org/10.1016/0378-2166\(95\)00059-3](https://doi.org/10.1016/0378-2166(95)00059-3).
- É. Kiss, Katalin. 1999. The English cleft construction as a focus phrase. In Lunella Mereu (ed.), *Boundaries of morphology and syntax*, 217–229. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/cilt.180.14kis>.
- É. Kiss, Katalin. 2010. Structural focus and exhaustiveness. In Malte Zimmermann & Caroline Féry (eds.), *Information structure: Theoretical, typological and experimental perspectives*, 64–88. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199570959.003.0004>.
- Franke, Michael, Fabian Schlotterbeck & Petra Augurzky. 2016. Embedded scalars, preferred readings and prosody: An experimental revisit. *Journal of Semantics* 34(1). 153–199. <https://doi.org/10.1093/jos/ffw007>.
- Frege, Gottlob. 1892. Über Sinn und Bedeutung. *Zeitschrift für Philosophie und Philosophische Kritik*. 25–50. Translated as [Frege 1984](#).
- Frege, Gottlob. 1984. On sense and meaning. In Brian McGuinness (ed.), *Collected papers on mathematics, logic, and philosophy*, 157–177. Oxford: Basil Blackwell.
- Gazdar, Gerald, Ewan Klein, Geoffrey K. Pullum & Ivan A. Sag. 1985. *Generalized Phrase Structure Grammar*. Cambridge, MA & Oxford: Harvard University Press & Blackwell.
- Hamblin, Charles Leonard. 1957. *Language and the theory of information*. London School of Economics & Political Science (University of London) dissertation.
- Hedberg, Nancy. 2000. The referential status of clefts. *Language* 76(4). 891–920. <https://doi.org/10.2307/417203>.
- Hedberg, Nancy. 2013. Multiple focus and cleft sentences. In Katharina Hartmann & Tonjes Veenstra (eds.), *Cleft structures*, 227–250. Amsterdam: John Benjamins. <https://doi.org/10.1075/la.208>.

- Heim, Irene. 1982. *The semantics of definite and indefinite noun phrases*. Amherst, MA: University of Massachusetts dissertation. Republished as Heim 2011. <https://scholarworks.umass.edu/dissertations/AAI8229562>.
- Heim, Irene. 2011. *The semantics of definite and indefinite noun phrases*. <https://semanticsarchive.net/Archive/jA2YTJmN/Heim%20Dissertation%20with%20Hyperlinks.pdf>. Republished version of Heim 1982 typeset by Anders Shoubye and Ephraim Glick.
- Horn, Larry. 1981. Exhaustiveness and the semantics of clefts. *North Eastern Linguistic Society (NELS)* 11. 125-142. <https://scholarworks.umass.edu/nels/vol11/iss1/10>.
- Horn, Larry. 2014. Information structure and the landscape of (non-)at-issue meaning. In Caroline Féry & Shinichiro Ishihara (eds.), *The Oxford handbook of information structure*. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199642670.013.009>.
- Horn, Laurence R. & Barbara Abbott. 2016. <the, a>: (in)definiteness and implicature. In Joseph Keim Campbell, William Kabasenche & Michael O'Rourke (eds.), *Reference and referring* (Topics in Contemporary Philosophy 10), 325-356. Cambridge, MA: MIT Press. <https://doi.org/10.7551/mitpress/9581.003.0014>.
- Huber, Stefan. 2002. *Es-Clefts und det-Clefts: Zur Syntax, Semantik und Informationsstruktur von Spaltsätzen im Deutschen und Schwedischen*. Stockholm: Almqvist & Wiksell International.
- Karttunen, Lauri. 1971. Some observations on factivity. *Papers in Linguistics* 4. 55-69. <https://doi.org/10.1080/08351817109370248>.
- Križ, Manuel. 2015. *It-clefts, exhaustivity, and definite descriptions*. Institut Jean-Nicod, Paris. Unpublished manuscript.
- Križ, Manuel. 2017. Referentiality, exhaustivity, and trivalence in *it-clefts*. Institut Jean-Nicod, Paris. Unpublished manuscript. <https://semanticsarchive.net/Archive/TVjNDUoM/Kriz%202017%20Clefts.pdf>.
- Ludlow, Peter & Gabriel Segal. 2004. On a unitary semantical analysis for definite and indefinite descriptions. In A. Bezuidenhout & M. Reimer (eds.), *Descriptions and beyond: An interdisciplinary collection of essays on definite and indefinite descriptions*, 420-454. Oxford: Oxford University Press. <https://doi.org/10.1093/os0/9780199270514.003.0013>.
- Mayol, Laia & Elena Castroviejo. 2013. How to cancel an implicature. *Journal of Pragmatics* 50. 84-104. <https://doi.org/10.1016/j.pragma.2013.02.002>.
- Onea, Edgar. 2016. *Potential questions at the semantics-pragmatics interface*. Leiden: Brill. <https://doi.org/10.1163/9789004217935>.

That's not quite it

- Onea, Edgar & David Beaver. 2009. Hungarian focus is not exhausted. *Semantics and Linguistic Theory (SALT)* 19. 342–359. <https://doi.org/10.3765/salt.v19i0.2524>.
- Percus, Orin. 1997. Prying open the cleft. *North Eastern Linguistic Society (NELS)* 27. 337–351. <https://scholarworks.umass.edu/nels/vol27/iss1/25>.
- Pollard, Carl & Ivan Sag. 1994. *Head-driven Phrase Structure Grammar*. Stanford, CA: CSLI Publishing.
- Pollard, Carl & Murat Yasavul. 2016. Anaphoric *it*-clefts: The myth of exhaustivity. *Chicago Linguistic Society (CLS)* 50. <https://web.archive.org/web/20170514171536/http://www.ling.ohio-state.edu/~yasavul.1/cls50paperfinal.pdf>.
- Prince, Ellen. 1978. A comparison of *wh*-clefts and *it*-clefts in discourse. *Language* 54. 883–906. <https://doi.org/10.2307/413238>.
- R Core Team. 2017. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. <https://www.R-project.org/>.
- Reeve, Mathew. 2012. *Clefts and their Relatives*. Amsterdam: John Benjamins. <https://doi.org/10.1075/la.185>.
- Reinhart, Tanya. 1997. Quantifier-Scope: How labor is divided between QR and choice functions. *Linguistics and Philosophy* 20. 335–397. <https://doi.org/10.1023/A:1005349801431>.
- Roberts, Craige. 1996. Information structure in discourse: Towards an integrated formal theory of pragmatics, 1998 revision. In Jae Hak Yoon & Andreas Kathol (eds.), *Papers in semantics* (Working Papers in Linguistics 49). The Ohio State University, Department of Linguistics. A revised version was published in Roberts 2012.
- Roberts, Craige. 2012. Information structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics* 5(6). 1–69. <https://doi.org/10.3765/sp.5.6>.
- Romero, Maribel. 2005. Concealed questions and specificational subjects. *Linguistics and Philosophy* 28(6). 687–737. <https://doi.org/10.1007/s10988-005-2654-9>.
- Romoli, Jacopo & Florian Schwarz. 2015. An experimental comparison between presupposition and indirect scalar implicatures. In Florian Schwarz (ed.), *Experimental perspectives on presuppositions* (Studies in Theoretical Psycholinguistics (SITP) 45), 215–240. Cham: Springer. https://doi.org/10.1007/978-3-319-07980-6_10.

- Rooth, Mats. 1996. On the interface principles for intonational focus. *Semantics and Linguistic Theory (SALT)* 6. 202–226. <https://doi.org/10.3765/salt.v6io.2767>.
- Rooth, Mats. 1999. Association with focus or association with presupposition. In Peter Bosch & Rob van der Sandt (eds.), *Focus: Linguistic, cognitive, and computational perspectives*, 232–244. Cambridge: Cambridge University Press.
- van der Sandt, Rob. 1989. Presupposition and discourse structure. In Renate Bartsch, Johan van Benthem & Peter van Emde Boas (eds.), *Semantics and contextual expression* (Groningen-Amsterdam Studies in Semantics 11), 267–294. Dordrecht: Foris. <https://doi.org/10.1515/9783110877335-010>.
- van der Sandt, Rob. 1992. Presupposition projection as anaphora resolution. *Journal of Semantics* 9(4). 333–377. <https://doi.org/10.1093/jos/9.4.333>.
- Schwarz, Florian. 2009. *Two types of definites in natural language*. University of Massachusetts at Amherst dissertation. <https://doi.org/10.7275/1077454>.
- Shinners, Pete. 2011. *PyGame*. <http://pygame.org/>.
- Simons, Mandy, David Beaver, Judith Tonhauser & Craige Roberts. 2010. What projects and why. *Semantics and Linguistic Theory (SALT)* 20. 309–327. <https://doi.org/10.3765/salt.v20io.2584>.
- Skopeteas, Stavros & Gisbert Fanselow. 2011. Focus and the exclusion of alternatives: On the interaction of syntactic structure with pragmatic inference. *Lingua* 11. 1693–1706. <https://doi.org/10.1016/j.lingua.2011.05.005>.
- Soames, Scott. 1989. Presupposition. In Dov Gabbay & Franz Guenther (eds.), *Handbook of philosophical logic*, vol. IV, 553–616. Dordrecht: Reidel. https://doi.org/10.1007/978-94-009-1171-0_9.
- Szabo, Zoltan. 2000. Descriptions and uniqueness. *Philosophical Studies* 101. 29–57. <https://doi.org/10.1023/A:1026437211756>.
- Szabolcsi, Anna. 1994. All quantifiers are not equal: The case of focus. *Acta Linguistica Hungarica* 42(2/3): *Special issue: 5th symposium on logic and language*. 171–187. <http://www.jstor.org/stable/44306737>.
- Tiemann, Sonja. 2014. *The processing of Wieder ('again') and other presupposition triggers*. Eberhard Karls Universität Tübingen dissertation. <https://publikationen.uni-tuebingen.de/xmlui/handle/10900/57169>.
- Tonhauser, Judith, David Beaver, Craige Roberts & Mandy Simons. 2013. Towards a taxonomy of projective content. *Language* 89(1). 66–109. <https://doi.org/10.1353/lan.2013.0001>.

That's not quite it

- Velleman, Dan Bridges, David Beaver, Emilie Destruel, Dylan Bumford, Edgar Onea & Liz Coppock. 2012. *It*-clefts are IT (Inquiry Terminating) constructions. *Semantics and Linguistic Theory (SALT)* 22. 441–460. <https://doi.org/10.3765/salt.v22io.2640>.
- Winter, Yoad. 1997. Choice functions and the scopal semantics of indefinites. *Linguistics and Philosophy* 20. 399–467. <https://doi.org/10.1023/A:1005354323136>.
- Xue, Jingyang & Edgar Onea. 2011. Correlation between presupposition projection and at-issueness: An empirical study. In *Proceedings of ESSLLI 2011: Workshop on projective meaning*, 171–184. Ljubljana, Slovenia.

Joseph P. De Veugh-Geiss
Universität Potsdam
Department Linguistik
Karl-Liebknecht-Str. 24-25
D-14476 Potsdam
Deutschland
de.veugh-geiss@alumni.uni-potsdam.de

Swantje Tönnis
Universität Graz
510 Institut für Germanistik
8010 Graz, Mozartgasse 8/I
Österreich
swantje.toennis@uni-graz.at

Edgar Onea
Universität Graz
510 Institut für Germanistik
8010 Graz, Mozartgasse 8/I
Österreich
edgar.onea-gaspar@uni-graz.at

Malte Zimmermann
Universität Potsdam
Exzellenzbereich
Kognitionswissenschaften
Department Linguistik / SFB 632
Karl-Liebknecht-Str. 24-25
D-14476 Potsdam
Deutschland
mazimmer@uni-potsdam.de