Scalar Items in Embedded Position: An Experimental Revisit

Fabian Schlotterbeck, Michael Franke and Petra Augurzky

Abstract

. . . make concrete . . .

Contents

1 Introduction

The existential quantifier *some* is usually assumed to receive a semantic interpretation similar to logical ∃, so that the sentence *Some boys cried* is literally true in a situation where all boys cried. But it is also usually considered to be a scalar item in that its use invites comparison with (at least) the semantically stronger universal quantifier *all* (c.f. Horn, 1972; Gazdar, 1979; Atlas and Levinson, 1981). This comparison can lead to an upper-bounding meaning enrichment, e.g., when an utterance of (1a) is taken to invite the inference in (1b).

(1) a. Hans solved some of the problems.

b. ⇝ Hans solved some but not all of the problems.

c. Hans solved all of the problems.

The classical explanation of this inference, following the pioneering work of Grice (1975) (see Geurts, 2010, for recent overview), is that (1b) is a pragmatic inference, a so-called *quantity implicature*, derived by an abductive inference as the best explanation of why an informed, knowledgable and cooperative speakers have uttered (1a) when they could also have uttered the semantically stronger and relevant (1c).

This paper deals with the interpretation of two types of sentences, where the scalar item *some* occurs in the scope of other logical operators. In AS-sentences (short for ALL-SOME) as in (2) the scalar item *some* is embedded under universal quantifier *all*. In ES-sentences (short for EXACTLYONE-SOME) like (3) *some* takes scope under the non-monotonic quantifier *exactly one*.[a] According to current pragmatic theory, there are at least three relevant candidate readings for AS- and ES-sentences: (i) a *literal reading* like in (2a) and (3a) where *some* has only its literal meaning; (ii) a *global reading* like in (2b) and (3b) where, according to Gricean intuition, we enrich utterances of (2) and (3) with the negation of alternative utterances of the corresponding sentences (4) and (5) where *some* is replaced by *all*; and also (iii) a *local reading* like in (2c) and (3c) where *some* is interpreted as *some but not all* in the scope of the embedding quantifier.

a. explain what non-monotonic means? here or later?

(2) All of the students read some of the papers. (AS)

a. All of the students read some and maybe all of the papers. (AS-LIT)

b. All of the students read some and maybe all and it's not the case that all of the students read all of the papers. (AS-GLB)

c. All of the students read some but not all of the papers. (AS-LOC)

(3) Exactly one of the students read some of the papers. (ES)

a. Exactly one of the students read some and maybe all of the papers. (ES-LIT)

b. Exactly one of the students read some and maybe all and it's not the case that exactly one of the students read all of the papers. (ES-GLB)

c. Exactly one of the students read some but not all of the papers. (ES-LOC)

(4) All of the students read all of the papers.

(5) Exactly one of the students read all of the papers.

Given this relative abundance of theoretically conceivable readings, two interlocked empirical questions arise:

Q1: which of these three conceivable readings are available to naïve subjects; and

Q2: which of the available readings are preferred.[1]

Addressing these questions empirically is relevant because they lie at the heart of the current debate about the exact location and nature of the interface between semantics and pragmatics. The controversy concerns the question to what extent the upper-bounding inference from *some* to *some but not all* is conventionalized within the compositional computation of semantic values. For that matter, it is particularly relevant to assess the availability and relative preference of local readings. (Section 3 will discuss the relevant theoretical positions and the significance of alleged local readings in more detail.)

A number of empirical studies have already examined the availability of readings for AS- and ES-sentences (e.g. Geurts and Pouscoulous, 2009; Clifton and Dube, 2010; Chemla and Spector, 2011). However, results have not been as clear-cut as one might have hoped for: for instance, the available empirical evidence is inconclusive as to whether local readings exist (see also van Tiel, 2012, for related discussion) and do not clearly address all preference relations between readings relevant to distinguish between theoretical positions. In the following we will argue that these heterogeneous results are partly due to the specific tasks used in previous studies, i.e., complexity differences in variants of the picture-verification paradigm. Moreover, previous studies have only accumulated limited evidence pertaining to the second question that may help decide between theoretical positions, namely which of the attested readings subjects prefer. Finally, previous studies presented target sentences visually. But as it is often argued that intonational stress on an embedded scalar item can favor a local reading (e.g. Horn, 2006; Geurts, 2009; Chemla and Spector, 2011; Geurts, 2010; van Tiel, 2012), it seems important to present target sentences auditorily, so as to be able to control for effects of "silent intonation" that subjects may apply when reading a sentence (see Bader 1998, Fodor 2000)[b].

b. provide references

In reaction to this situation, we therefore presented sentence material auditorily and also employed a different kind of visual presentation of the pictorial material: subjects were presented with pictures that were initially covered and could incrementally be uncovered at the subjects' request; at each step of uncovering, subjects had to decide whether they could already give a truth-value judgement or needed more information (Comry ????)[c]. This way we obtained behavioral data that is both indicative of the reading subjects assumed and independent of the complexity of the visual stimulus. At the same time, we hypothesized that the incremental nature of this task would shed light on the preferences over readings, because the temporal distribution of truth-value judgments would give away which reading subjects were waiting to evaluate, so to

c. insert references

[1] It may ultimately be impossible to keep unavailability and strong dispreference strictly apart. For our purposes this is not important. Every time we speak of availability, the so-inclined reader may think of non-negligible preference.

speak. Adding to previous studies, we therefore made sure that our method is able to reveal information about preferential readings by including ambiguous test items like (6) which are known to preferentially receive the late-closure reading in (6a) and not the dispreferred, but attested early-closure reading (6b) (Frazier 1987 XYZ).[d]

d. insert proper references

(6) The letter is connected with circles and squares with suns.

a. The letter is connected with squares with suns and circles. (LC)

b. The letter is connected with circles with suns and squares with suns. (EC)

- summarize results

The paper is structured as follows. Section 2 elaborates on the three kinds of relevant readings for our target sentences. Section 3 works out the different theoretical positions and their predictions about availability and preference. Section 4 recaps the results of previous studies on this subject, arguing for the need of a more refined methodology. Section 5.1 describes our experimental design. Section 5.5 states our results, which we discuss in Section 5.6.[e]

e. rephrase eventually

2 Get to know your readings

Three readings are *prima facie* conceivable for the AS- and ES-sentences in (2) and (3). These are logically dependent in intricate ways.

AS-sentences. An AS-sentence like (2), repeated below, has a literal reading (LIT) as in (2a), a global reading (GLB) as in (2b) and a local reading (LOC) as in (2c).

(2) All of the students read some of the papers.

a. All of the students read some and maybe all of the papers. (AS-LIT)

b. All of the students read some and maybe all and it's not the case that all of the students read all of the papers. (AS-GLB)

c. All of the students read some but not all of the papers. (AS-LOC)

These readings stand in a strict entailment relation: the local reading asymmetrically entails the global reading, which asymmetrically entails the literal reading:

(7) LOC $\subset$ GLB $\subset$ LIT

GLB entails LIT because, in general, global readings are defined as the conjunction of the literal reading and the negated (relevant/feasible) alternative(s) of the to-be-interpreted utterance. This entailment is asymmetric, because the information that not all of the students read all of the papers is not entailed by the literal reading. To see that LOC $\subset$ GLB, notice that the case where all of the students read some but not all of the papers is a special case of the case where all of the students read some (and maybe all), while not all of the students read all of the papers.

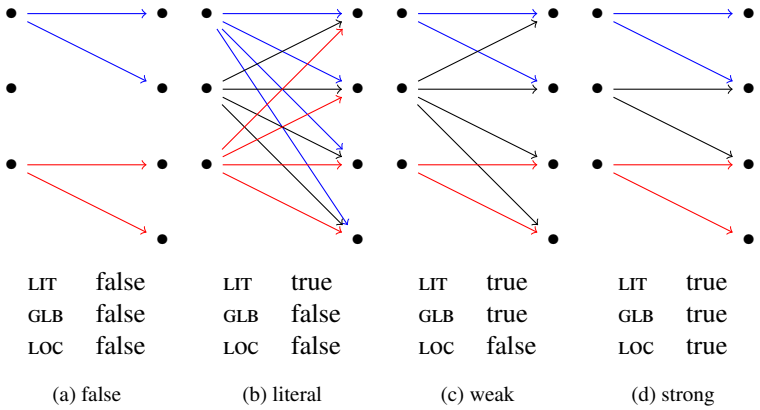
situation	truth value		
	LIT	GLB	LOC
false	0	0	0
literal	1	0	0
weak	1	1	0
strong	1	1	1

(a) AS-sentences

situation	truth value		
	LIT	GLB	LOC
false	0	0	0
literal	1	0	0
local	0	0	1
all	1	1	1

(b) ES-sentences

Table 1: Possible truth-value distributions for readings of AS- and ES-sentences



(a) false (b) literal (c) weak (d) strong

Figure 1: Distinguishing scenarious for AS-sentences

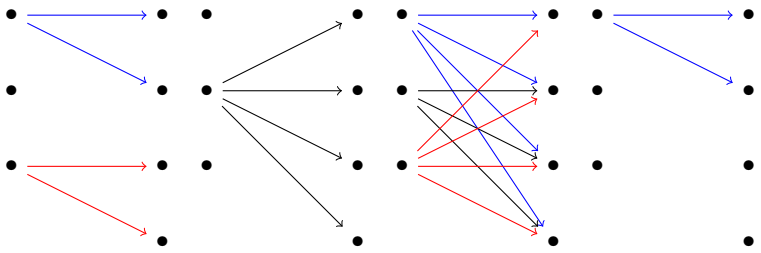
Given these entailment relations, there are four kinds of situations, the names for which we borrow from Chemla and Spector (2011), that we can distinguish based on different truth values for our candidate readings. These are given in Table 1a. Examples of these kinds of situations are given in Figure 1, where the dots on the right of each diagram represent students, the dots on the left represent papers and an arrow from a student to a paper indicates that the student read the paper.[f] Notice that other arrangements of arrows might equally well serve as examples for the various situations.

f. improve graphics

ES-sentences. The situation for ES-sentences like (3) is similar but a little more complicated because the embedding quantifier is non-monotonic. Again, we consider a literal reading as in (3a), repeated below, a global reading as in (3b) and a local reading as in (3c).

(3) Exactly one of the students read some of the papers.

a. Exactly one of the students read some and maybe all of the papers. (ES-LIT)



(a) false (b) literal (c) local (d) all

Figure 2: Distinguishing scenarious for ES-sentences

b. Exactly one of the students read some and maybe all and it's not the case that exactly one of the students read all of the papers. (ES-GLB)

c. Exactly one of the students read some but not all of the papers. (ES-LOC)

Entailment relations in this case are non-linear:

(8) LIT ⊃ GLB ⊂ LOC

By definition of global readings GLB entails LIT. This entailment is asymmetric because the extra information that it is not the case that exactly one student read all of the papers is not entailed by the literal reading (3a). However, unlike for AS-sentences, LOC is not stronger than GLB, but asymmetrically entailed by the latter. To see this, notice that the global reading is equivalent to:

(3b′) Exactly one of the students read some but not all and everybody else read none of the papers.

Finally, LOC and LIT are logically independent: all possible combinations of truth-values for LOC and LIT are possible.

Given these entailment relations, there are again four different situations corresponding to the four possible distributions of truth values for candidate readings. These are given in Table 1b and named following Chemla and Spector (2011). Examples for each situation are given in Figure 2.

3 Theories and predictions

We consider three main theoretical positions which make different predictions about the readings of AS- and ES-sentences (c.f. Horn, 2006; Geurts, 2010; Sauerland, 2012, for overview). We will refer to these here as traditionalism, conventionalism and grammaticalism and treat each one in turn. Since there is some leeway in assessing the

predictions for some of these positions (depending on which of several reasonable additional assumptions we should adopt), we will distinguish different varieties of each position. For convenience, the predictions of each (variety of each) position are also summarized in Table 2 at the end of this section.[g]

g. micha says: I'll leave all the varieties in here until we know which ones we need and which ones we don't need

3.1 Traditionalism

We refer to traditionalism as traditionalism because of its conservative stance towards Grice's original theory of conversational implicatures (Grice, 1975). Many author's have defended traditionalist positions in this sense. Of the more recent literature, we would consider as traditionalist, among others, contributions by Spector (2006), Sauerland (2004), Russell (2006), Schulz and van Rooij (2006), Geurts (2010) or Franke (2011).

According to Grice, conversational implicatures, of which quantity implicatures are a special case, are to be thought of as rationalizations of speaker behavior. Central in this reasoning is the assumption that the speaker's behavior is efficient (if not optimal) and goal-oriented. Usually, the assumed goal of conversation is the cooperative exchange of helpful information from the speaker to the hearer.

Consequently, the Gricean recipe (c.f. Geurts, 2010) for deriving a simple scalar inference like that in (1b) from an utterance of (1a) is as follows: if the issue whether Hans solved only some or all of the problems is relevant, then a cooperative and knowledgable speaker would utter (1c) if in a position to do so; hence, one of the most natural explanations of why such a speaker has not uttered (1c), but only (1a) is that she is uncertain of whether (1c) is true; but on the assumption that she is knowledgeable (competent, opinionated, informed . . .) it follows that (1b) should in fact be true.[2,h]

h. do we need to enlarge on epistemic implicatures?

(1) a. Hans solved some of the problems.

b. $\rightsquigarrow$ Hans solved some but not all of the problems.

c. Hans solved all of the problems.

The Gricean recipe applies also to AS- and ES-sentences and derives the global reading in a straightforward way. Consequently, traditionalism predicts that both literal and global readings are available: literal readings, because these form the starting point of pragmatic reasoning; global readings because these may be arrived at by the Gricean recipe.

Which of these readings, if any, does traditionalism predict to be preferred? This depends on whether the auxiliary assumptions necessary to derive global readings by the Gricean recipe are plausibly met in the particular case of utterance of AS- and ES-sentences. These auxiliary assumptions include relevance of the extra information provided in the global reading, mutual awareness that the stronger alternative has been a speaker option, the speaker's competence about the issue, etc. Normally, traditionalist

[2] We are glossing here somewhat swiftly over the more nuanced details of the derivation of implicatures targeting the speaker's epistemic state (e.g. Gazdar, 1979; Soames, 1982), as this is not crucially relevant for the issues we are interested in here.

accounts would assume that these extra assumptions are met. In that case, traditionalism would predict that the global readings should be preferred over the literal readings. On the other hand, it might also be hypothesized that, for example, a competence assumption is harder to justify for AS- and ES-sentences in general than for simpler sentences such as (1), because of the additional quantificational element: it might be less clear that the speaker knows exactly how many students solved how many problems, than that the speaker knows exactly how many problems, e.g., Hans solved. In that case, or if any other assumption of the Gricean recipe cannot be maintained, traditionalism would predict that the literal reading would be preferred over the global one. But that means that there are at least two varieties of traditionalism that, depending on which additional assumptions we would make, yield slightly different predictions: the strong variety of traditionalism maintains that the auxiliary assumptions of the Gricean recipe hold usually/strongly/unless-completely-intenable and therefore predicts that global readings are preferred over the literal ones; the weak variety of traditionalism holds that the auxiliary assumptions are more fragile and predicts that literal readings are preferred over the global ones.[3] **i**

i. acknowledge comment by Philippe Schlenker here;possibly expand

Does traditionalism also predict local readings to be available? The answer is slightly different for AS- and the ES-sentences. For ES-sentences, traditionalism does not predict that a local reading is available, at least not as a quantity implicature (c.f. Geurts and Pouscoulous, 2009; Chemla and Spector, 2011). This is because traditionalism assumes that quantity implicatures are pragmatic enrichments of the literal meaning of an utterance, obtained by conjoining the literal meaning with a suitable set of negated alternatives. But since the literal and the local reading of ES-sentences are logically independent, there is no way that local readings can be derived in a traditionalist manner as a quantity implicature: if X and Y are logically independent propositions, then there is no proposition Z such that X would be equivalent to $Y \wedge \neg Z$. Traditionalism often does concede that (something like) local readings can occur if scalar items are marked with *special intonation*, albeit then as a signal of a different pragmatic process (e.g. Horn, 2006; Geurts, 2009, 2010). We will come back to this issue in Section **??**.

On the other hand, as for AS-sentences, there is a traditionalist way of deriving the local reading, namely by assuming that not only (4) is an alternative to (2), but also the sentence in (9).

(2) All of the students solved some of the problems.

(4) All of the students solved all of the problems.

(9) Some of the students solved all of the problems.

Clearly, if we conjoin a literal reading of (2) with the negation of (9) we obtain exactly the local reading. (Notice that the negation of (9) entails the negation of (4), so it makes no difference (not) to add it in this case.) Consequently, traditionalism does predict that the local reading of AS-sentences is available if (9) is an available alternative. We should

[3] We mention for clarity that although weak traditionalism does not predict the global reading to be preferred, it might still predict an *epistemically weak implicature*, similar to the global reading, that the speaker is uncertain whether the stronger alternative is true. Whether it does predict that depends on which auxiliary assumption are assumed to be met and which are not.

therefore introduce another distinction: restricted traditionalism assumes that (9) is not available and therefore predicts that the local reading for AS-sentences is not available; in contrast, unrestricted traditionalism assumes that (9) is available and so is the local reading.

Still, even unrestricted traditionalism would not predict that the local reading is preferred over the global reading. This is because the derivation of the local reading via (9) hinges on yet another auxiliary assumption, namely the availability of the alternative in (9), which seems much less obvious and therefore presumably is less readily available than the alternative in (4). Consequently, we derive the following predictions about preferences for unrestricted traditionalism: weak unrestricted traditionalism prefers the literal over the global over the local reading, while strong unrestricted traditionalism prefers the global over the local over the literal reading (see also Table 2).

3.2 Conventionalism

Lexical conventionalism is a position that has been defended for instance by Levinson (2000) and Chierchia (2004).[4] The general idea is that if the upper-bounding inference from *some* to *some but not all* occurs frequently, then it would be inefficient to execute the Gricean reasoning that traditionalists propose over and over again. Conventionalism therefore assumes that scalar inferences have become routinized, hard-wired, so much so that there are two lexical entries for a scalar item like *some*, one with meaning *some and maybe all* and one with the meaning *some but not all*. Levinson (2000) proposed that the latter is the preferred default meaning, a position which we will refer to as defaultism.[5] **j** For completeness sake, we will call the opposite view that the literal meanings are *ceteris paribus* preferred literalism **k**. We are not aware that anyone has defended this option, but we will come back to this later on. **l**

Both varieties of conventionalism make interesting predictions about available readings for AS- and ES-sentences. Unaided, conventionalism does not predict that global readings are available. The lexically stored upper-bound for scalar *some* only allows to derive literal and local readings. Moreover, defaultism predicts that local readings are preferred over literal ones. Contrary to that, our dummy position of literalism predicts that literal readings are preferred over local ones. **m**

j. Petra says: empirical stuff in footnote irrelevant; I agree, let's move it to discussion section

k. better name?

l. where exactly

m. Petra says that his is too short; suggests skipping the whole section; I'm not sure; could do that too; anyway, it must be mentioned because conventionalism turns out to be one of the best predictors of our data;

3.3 Grammaticalism

The third and final kind of theoretical position we consider is grammaticalism. Recently, a lot of evidence based on intuitive judgements and more theoretical considerations has been discussed in favor of this position (c.f. Chierchia, 2006; Fox, 2007; Magri, 2011; Sauerland, 2012; Chierchia et al., forthcoming; Chierchia, forthcoming).

[4] See also Sauerland (2012) who argues quite favorably for a pragmatically informed conventionalism, although eventually dismissing it in favor of grammaticalism.

[5] Levinson's defaultism has strong empirical evidence against it, because several processing studies show that the computation of scalar inferences appears to be time- and effort-consuming, when they *do* arise, not when they *don't*, as defaultism would have it (c.f. Breheny et al., 2006; Breheny and Katsos, 2008; De Neys and Schaeken, 2007). But there is also competing evidence that scalar inferences are immediate, low-level processes (Grodner et al., 2010).

Grammaticalism approaches quantity implicatures by postulating a silent operator that can be variably applied during compositional computation of a sentence's truth-value (Chierchia, 2006), if necessary multiple times (Fox, 2007). This silent operator is variably referred to as O(·) or Exh(·), because it is assumed to be similar in effect to the meaning of particle *only* or of the mechanism of *exhaustive interpretation* (Groenendijk and Stokhof, 1984; von Stechow and Zimmermann, 1984; van Rooij and Schulz, 2004; Schulz and van Rooij, 2006; Fox, 2007). For our purposes, it is enough to note that Exh(·) is a poly-typed function that enriches an expression X, which crucially need not be a full proposition, based on a set, Alt(X), of (suitable, relevant) alternatives to X that yields an enriched meaning of the form:[6]

(10) $\text{Exh}(X, \text{Alt}(X)) = X \bigwedge_{A \in \text{Alt}(X)} \neg A$.

As this operator can apply at various scope sites, the grammatical approach predicts that all three readings of AS- and ES-sentences are available: literal readings arise if no exhaustification operator is applied; global readings arise if the exhaustification operator takes sentence-wide scope as in (11); and local readings arise if the exhaustification operator takes scope under the respective quantifiers as in (12).

(11) a. Exh(All of the students solved some of the problems).
b. Exh(Exactly one of the students solved some of the problems).

(12) a. All of the students Exh(solved some of the problems).
b. Exactly one of the students Exh(solved some of the problems).

The grammatical approach, as described so far, is rather flexible in that it makes many readings available. It is so flexible, in fact, that it must be constrained in some way or other to shield the approach against overgeneration. The easiest examples where this is crucial are occurrences of scalar items in downward-entailing environments, such as in:

(13) It's not the case that Hans solve some of the problems.

An assertion of (16) normally would be taken to convey (14a), not the disjunctive meaning in (14b).

(14) a. Hans solves none of the problems.
b. Hans solved none or all of the problems

The latter, however, is a reading that grammaticalism predicts to be available by letting the Exh(·)-operator take scope under negation, as in:

(15) It's not the case that Hans solved Exh(some of the problems)

Indeed, this reading is not unattested, for witness a continuation of (13) as in:

[6] More sophisticated formulations of exhaustive interpretation have been proposed (e.g. Schulz, 2005; Schulz and van Rooij, 2006; Spector, 2006; Fox, 2007) but this simple formulation is sufficient for the purposes of this paper. Also, we gloss here over non-trivial details in the way this operation is to be specified exactly within a compositional semantics (c.f. Chierchia, 2006).

(16) It's not the case that Hans solve *some* of the problems. He solved *all* of them.

But this reading seems to be marked, in that it requires exceptional contextual circumstances and a non-standard intonational stress.

To account for the markedness of certain examples, grammaticalism is therefore aided by a principle that specifies which readings are to be preferred if several readings are generated by optional applications of Exh(·) in the compositional computation of semantic values. The principle that grammaticalists adhere to (c.f. Fox and Spector, 2009; Chierchia et al., forthcoming; Chierchia, forthcoming) is the strongest meaning hypothesis of Dalrymple et al. (1998). Roughly speaking, if a given sentence has several conceivable readings, the strongest meaning hypothesis selects for the strongest of these. Chierchia et al. (forthcoming) discuss two concrete variants of the strongest meaning hypothesis. Towards a definition that is just precise enough for our current purposes, fix a sentence with propositional content S with candidate readings $C(S)$, where $C(S)$ contains S and all readings derivable from inserting Exh(·) at suitable scope sites.[7] Then define *ceteris paribus* preferences among members of $X, Y \in C(S)$ as either:

(17) $X >_w Y$ iff $X \subset Y$ (Chierchia et al., forthcoming, (104))

or:

(18) $X >_s Y$ iff $X \subset Y$ and X and Y are readings obtained from applications of Exh(·) at exactly the same scope sites, except for one.
(Chierchia et al., forthcoming, (105))

Since $X >_s Y$ entails $X >_w Y$, we will speak of weak grammaticalism and strong grammaticalism, depending on whether (17) or (18) is used.

These two varieties of grammaticalism give rise to different predictions about preferred readings. The preference for readings that weak grammaticalism predicts mirrors the entailment relations in (7) and (8): for AS-sentences weak grammaticalism predicts that local readings are preferred over global readings which in turn are preferred over literal readings; for ES-sentences strong grammaticalism predicts that global readings are most preferred, while literal and local readings are not ranked with respect to preference. Strong grammaticalism, on the other hand, predicts that both AS- and ES-sentences preferably get either a local or a global reading rather than a literal reading. But strong grammaticalism does not rank these former two with respect to each other because they differ in more than one application of the Exh(·) operator.

Taking stock, traditionalism, conventionalism and grammaticalism predict different patterns of availability and preference for AS- and ES-sentences. Most notably, traditionalism does not predict local readings for ES-sentences to be available, while conventionalism does not predict any global readings to be available. Grammaticalism predicts that all readings are available, but, given the strongest meaning hypothesis, also comes with a clear indication of preference: literal meanings are dispreferred. All predictions are summarized in Table 2.

[7] Strictly speaking, we should compare *parses* of sentences with Exh(·) at various scope sites with respect to the readings that these parses give rise to (c.f. Chierchia et al., forthcoming), but we may ignore this detail here for ease of exposition.

	preference & availability	
theoretical position	AS	ES
traditionalism		
weak restricted	LIT > GLB	LIT > GLB
weak unrestricted	LIT > GLB > LOC	LIT > GLB
strong restricted	GLB > LIT	GLB > LIT
strong unrestricted	GLB > LOC > LIT	GLB > LIT
conventionalism		
defaultism	LOC > LIT	LOC > LIT
literalism	LIT > LOC	LIT > LOC
grammaticalism		
weak	LOC > GLB > LIT	GLB > LIT, LOC
strong	GLB, LOC > LIT	GLB, LOC > LIT

Table 2: Predictions of the relevant theoretical positions. Listed are the predicted preference relations. If a reading is not listed, it is predicted to be unavailable.

4 Previous studies

In order to disentangle the divergent predictions of these opposing theoretical positions, a number of empirical studies have been carried out. In this section, we will focus on the influential studies by Geurts and Pouscoulous (2009), Clifton and Dube (2010) and Chemla and Spector (2011). Unfortunately, as we argue here, the conjoined evidence from all of these studies is inconclusive as to the availability and preference of relevant readings.

4.1 Geurts and Pouscoulous (2009)

Geurts and Pouscoulous (2009) conducted a picture-verification task to find out whether local readings of AS- and ES-sentences are available.[8] The critical conditions of their study presented subjects with pictures like those in Figure 1c and 2c where the local reading gets a different truth-value from the literal and the global reading. In particular, for AS-sentences the local reading is false for the critical picture in Figure 1c whereas the literal and global readings are true; for ES-sentences the local reading is true for the critical picture in Figure 2c, whereas the literal and global readings are false.[n]

n. Petra wants more detail in the description of this study; I'm not sure that this is needed...?

The results of Geurts and Pouscoulous were strikingly unambiguous: there were *no* responses indicative of a local reading; *all* of the subjects judged AS-sentences true in a situation like in Figure 1c and *all* of the subjects judged ES-sentences false in situations like 2c.

These results were criticized on theoretical grounds (e.g., Sauerland (2010), but see Ippolito (2010) for supporting evidence), as well as based on empirical observations

[8] Actually, Geurts and Pouscoulous (2009) did not use ES-sentences, but sentences where scalar *some* was embedded under non-monotonic quantifier *exactly two* (with appropriate pictures, of course). This case is a little more complex, but we will gloss over this here, treating their data, as if it was obtained for ES-sentences.

(Clifton and Dube, 2010; Chemla and Spector, 2011). In the following, we will review these empirical studies in some detail, as they constitute an important background for our own experiment. We will first consider a comment by Clifton and Dube (2010), which explicitly distinguishes between the availability of a reading and its preference.

4.2 Clifton and Dube (2010)

In a reply to Geurts and Pouscoulous's study Clifton and Dube raised the question whether the use of a picture-verification paradigm might have been infelicitous for testing the availability of strong readings, at least in the case of AS-sentences. Asking whether a sentence fits a picture might have created a bias for accepting sentences also on a weaker and probably dispreferred reading. Clifton and Dube's study was therefore aimed at finding out about a potential preference relation between local and literal readings. To this end, Clifton and Dube developed a picture-choice task where subjects were presented with an AS-sentence and a pair of pictures, hence introducing the option to choose between different alternatives. Subjects were asked to "indicate which shape is best described by the sentence" and could choose either picture, or options 'both' and 'neither.' There were two versions of this experiment, differing in which kind of picture pairs were presented on critical trials. In version 1, the picture pair consisted of the weak and strong situations in Figures 1c and 1d. The response percentages observed by Clifton and Dube were:

weak	strong	both	neither
3	39	57	1

That the majority answer is "both" could be taken as evidence that the literal reading is the preferred one. But the almost 40% of choices for the strong situation, so Clifton and Dube argue, might be indicative of the availability of the local reading. In version 2 of the experiment, the picture pair consisted of the literal and weak situations in Figures 1b and 1c. In this case, response percentages were:

weak	literal	both	neither
28	6	50	17

Again the majority response "both" might speak for a preference for the literal reading, but, as Clifton and Dube propose, the 17% of "neither" answers in this case again suggest that the local reading is available. Taken together, Clifton and Dube take these results to contradict Geurts and Pouscoulous's findings. Local readings are, after all, attested if subjects are given a choice as to which situation they consider most fitting for an AS-sentence.

Taken together, the results by Clifton and Dube demonstrate that participant' choices for AS-readings might be strongly affected by the specific experimental paradigm used. If task demands are only felicitous for a literal reading, the *absence* of choices of local readings cannot be accounted for by assuming an inavailability of these readings *per se*.[o] However, at present it is less clear whether these results can also be generalized to further situations. For instance, subjects rejected the ES-sentences in Geurts and

[o] micha does not understand this sentence

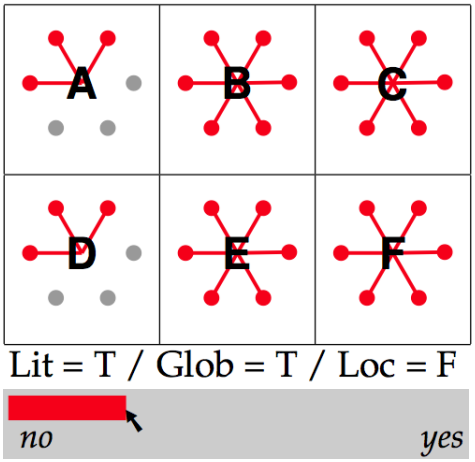


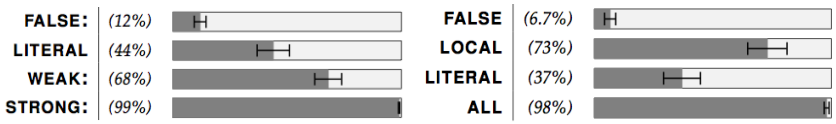
Figure 3: Example of critical trial for AS-condition by Chemla and Spector's (2011)

Poscoulous' original study, despite the fact that they could have accepted it based on the local reading. It is thus necessary to test both universal and non-monotonic quantifiers within the same experiment using a task that offers alternatives. Additionally, it would be informative to also probe into the availability and the relative preference of global readings. The study of Chemla and Spector (2011) did both of that.

4.3 Chemla and Spector (2011)

Chemla and Spector (2011) also took issue with Geurts and Pouscoulous's design, arguing that, firstly, the pictorial material used in Geurts and Pouscoulous's study was unduly difficult; that, secondly, these pictures also may have failed to make the local reading sufficiently relevant; and that, thirdly, the restriction to a categorial choice (whether the sentence fits the picture of not) may induce a bias against non-preferred readings in cases where candidate readings stand in entailment relations (c.f. Sauerland, 2010, for this latter criticism). To meet these potential problems, Chemla and Spector (2011) presented subjects with pictorial material like that in Figure 3, which was assumed to be easier to assess and better at highlighting the relevance of the local readings. Additionally, subjects were asked, not for categorial judgements, but for graded judgements: subjects could freely click on a scale, as shown in Figure 3, to indicate how much they considered a picture fitting for a given sentence (c.f. Chemla, 2009, for more on this method).

Albeit in a different format, the pictures used by Chemla and Spector were all different instantiation of the situation types in Figures 1 and 2. The results reported in this study are averaged clicking positions, as shown in Figure 4. According to Chemla and Spector, the crucial piece of evidence for the availability of local readings for AS-sentences was that these sentences yielded higher graded acceptability scores for the strong situation than for the weak situation (although these differ only with respect to the truth value of the local reading). In a similar way, evidence for the availability of



(a) AS-condition (b) ES-condition

Figure 4: Results from Chemla and Spector's (2011) study

the local reading for ES-sentences comes from the difference between the local and the literal situation. Strikingly, in the latter case, the ES-sentences received an average of 73% acceptability in the local situation although the literal and global readings are false in this case.

4.4 Reflection

Summing up, the mentioned studies have presented diverging evidence concerning the availability of local readings. Various problematic aspects of these studies have already been discussed elsewhere (e.g. Clifton and Dube, 2010; Chemla and Spector, 2011; van Tiel, 2012). Here, we want to point out two particular methodological problems that do impede their interpretation. However, we also want to make clear that these problems follow almost directly from the nature of the phenomenon under investigation and therefore seem nearly impossible to avoid.

Firstly, a mapping between responses in a picture-verification task and readings of the relevant sentences is problematic due to the logical dependencies between the relevant readings discussed in Section 2. Let us consider AS-sentences in the following because these have been addressed by all three aforementioned studies. (Similar arguments apply to ES-sentences.) It is impossible to construct situations that clearly correspond to local readings of AS-sentences, since the local reading of such sentences implies their global and local readings. Every situation which is consistent with such a sentence under its local reading is also consistent with the other two readings. Therefore, we cannot test directly and independently all the cases we might wish to using a simple picture-verification paradigm. It would be especially desirable to test situations that are consistent with the local reading, but inconsistent with the global and literal reading. In that case *yes*-responses would unambiguously indicate availability of local readings. However, since this is impossible we are left with suboptimal choices. In the experiment of Geurts and Pouscoulous (2009) situations that are inconsistent with local readings but consistent with the other two readings were tested. In that case *no*-responses would be decisive – when compared to a suitable baseline. However, *yes*-responses were observed exclusively. These can be attributed to one of the other readings. Therefore, assuming a principle of charity (Davidson, 1973) to be at play in picture verification tasks, the data of Geurts do not provide evidence against the existence of local readings. In the experiments of C&S and C&D on the other hand situations that are consistent with all three readings were tested. Here again 'yes'-judgments cannot be interpreted confidently because they can always be attributed to

global or literal readings. Both C&S and C&D attempt to surpass this problem by interpreting relative preferences. This does, however, bring us to the second problem.

Preferences may be affectd by graphical properties of the picture materials. For example, verification complexity or typicality of the pictures with respect to the test sentences may differ between conditions and thereby induce certain preferences. In his careful critique van Tiel has eleborated on this point. Van Tiel points out that the relative preferences C&S and C&D interpret can be explained solely in terms of typicality, because the 'strong'-situation (see e.g. Figure 1d) are more typical instances of the sentence meaning than the competing pictures, irrespective of local implicatures.

To conclude, in order to decide about the availablity of local implicatures we would ideally like to test situations that can unambiguously be mapped to the different readings and do not differ in typcality or other crucial properties. Given the logical dependencies inherent to the relevant readings and the considerations concerning typicality presented by van Tiel this seems impossible to achieve. As we willl argue below we do, however, believe it is.

4.5 To-Be-Included Here

- establish independence (as much as possible) of pictorial complexity/stereotypicality and assessment of reading
  - not desirable to raise the level of relevance of local readings, because we want to test the base-line / default (cite Geurts / van Tiel)
  - unclear what graded truth-value judgements measure exactly (re-read Chemla (2009) if needed)
  - cannot separate judgements about sentences from judgements about pictures (van Tiel)
  - unclear whether design assesses preferences of readings or targets pictorial complexity / typicality (van Tiel)
- provide a benchmark from which to deduce preference relations
- control for intonational pattern (present auditorily)
  - intonational effects discussed by Horn (2006); Geurts (2009, 2010), (c.f. Chemla and Spector, 2011)
  - mention Schwarz et al. (2008) (re-read, check for newer version)
- different languages

5 An Incremental Verification Task

5.1 Design

General Idea. In order to test the availability and preferences of different readings of AS- and ES-sentences, we were looking for a way to unambiguously map responses

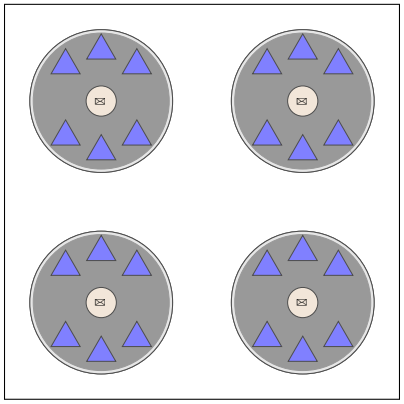


Figure 5: Initial picture presented to participants where all possible connections from each letter to its surrounding triangles are covered.

from a picture verification task to specific readings. We were particularly interested in the question whether local readings would be observed at all, but we were also interested in obtaining information about the relative preferences of the relevant readings. In addition, we wanted to minimize effects of any graphical differences between the picture materials. To achieve this we designed an experiment employing a modified version of picture verification, namely the incremental verification task (IVT, see Conroy ???)[p]. The general idea is that subjects are to judge sentence material based on pictures that do not necessarily contain all the information necessary to judge a certain reading true or false. In that case, participants can demand that more information is revealed. Participants are instructed to make a truth-value judgement as soon as they are able to.[q]

p. reference

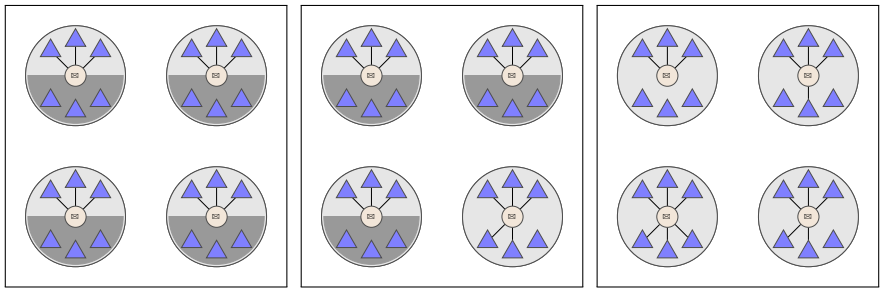
q. need to mention somewhere that trial ends with each TV judgement

The Target Conditions. More concretely, our study presented participants with pictures, like those in Figures 5, 6 or 7, depicting a set of four identical central elements (e.g., letters), which could be connected to surrounding elements (e.g., triangles). Initially, any potential connections between central and surrounding elements were covered by dark gray color (see Figure 5). Sentences to be judged were in German (see (19) and (20) below) and presented auditorily, and participants were asked to listen to a given sentence and then uncover the potential links step by step until they felt able to give a truth value judgment. Three options were available for participants at each step: (i), demand more information, (ii), judge the sentence as true, and (iii), judge the sentence as false. Importantly, each of the three potentially available readings corresponded to a specific step in the uncovering process where the truth value of that reading (and only of that reading) could be assessed for the first time. The location in the sequence and the corresponding truth-value judgement differed between AS- and ES-sentences, as described presently. (This is, again, because of the logical dependencies between readings.)

Consider the AS-sentence in (19).

(19) Alle diese Briefe sind mit einigen ihrer Dreiecke verbunden.
All these letters are with some their triangles connected.
All of these letters are connected to some of their triangles.

Figure 6 shows the three corresponding critical positions for this sentence in the order of the uncovering process. First, the situation in Figure 6a becomes available, which is true under a literal reading, while the local and global readings cannot be evaluated yet. Second, in Figure 6b, the literal reading is still true, but now the global reading can be additionally confirmed by a *true* response, as all the connections of one of the letters are now uncovered, and it is now visible that one of the letters is not connected to all of its triangles. Finally, decisions concerning the local reading are possible as soon as all connections have been uncovered. In this case, the critical position shown in Figure 6c is incompatible with a local reading: One of the central elements is linked to all of its surrounding elements, thus yielding to *false* answers if participants indeed adopted the local reading. Note that associating the final position with *false* answers allows us to exclude both literal and global readings, which correspond to a *true* answer at this position. In sum, *true* or *false* answers on particular positions in the incrementally revealed picture sequence can be mapped uniquely to candidate reading. All other *true* or *false* answer were counted as errors. The mapping between responses and readings is summarized in Table 3a.



(a) Step 1 (b) Step 2 (c) Step 3

Figure 6: Example sequence for AS-sentences

	more info	true	false
Step 1	error	LIT	error
Step 2	error	GLB	error
Step 3	-	error	LOC

(a) AS-Sentences

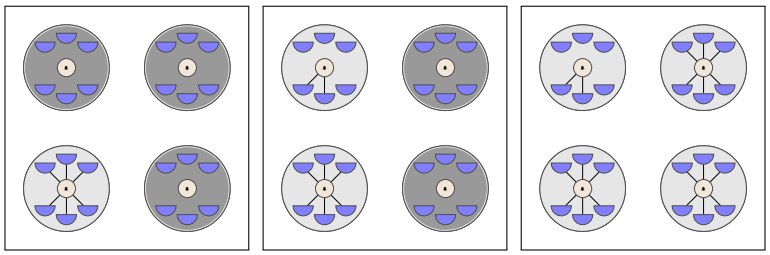
	more info	true	false
Step 1	error	error	GLB
Step 2	error	error	LIT
Step 3	-	LOC	error

(b) ES-Sentences

Table 3: Mapping from response types to readings. Notice that on steps 3 the option "more info" is not available because the whole picture has been revealed.

ES-sentences, due to the non-linear relationship of candidate readings described in

Section 2, require a slightly different sequence of unfolding which yields a different order in which truth-value judgements can be made. For a sentence like (20) we ob-



(a) Step 1 (b) Step 2 (c) Step 3

Figure 7: Example sequence for ES-sentences

tain an unambiguous mapping between truth values and readings with respect to the sequence in 7.

(20) Genau eine der Glocken ist mit einigen ihrer Halbkreise verbunden.
Exactly one of-the bells is with some its semicircles connected.
Exactly one letter is connected to some of its semicircless.

At the first critical position shown in Figure 7c, the global reading corresponds to a *false* response, as one of the bells is already linked to all of its semicircles. The literal reading becomes available in 7b, corresponding to a *false* response, as there are two bells that are connected to at least some of their surrounding elements. Finally, confirming 7c by a *true* response corresponds to the local reading, as one of the bells is linked to some but not all of its semicircles, while all of the other bells are linked to all of them. Note that again, global and literal readings differ from local readings with respect to their truth values. Mappings between responses and readings are summarized in Table 3b.

In addition to the unambiguous mapping from readings to truth values, the present design minimizes graphical effects because there is no comparison between different pictures (c.f. van Tiel, 2012). Moreover, we controlled for pictorial effects by introducing control conditions, which are described in the next section.[r]

r. Are we sure that this paragraph needs to go here? this should be mentioned in the motivation preceding this section and in the discussion at the end, no?

Controls. In order to make sure that subjects understood the task, i.e., gave truth-value judgements neither sooner not later than at the first possible position in a sequence, we included a set of control conditions. These also controlled for response biases that are independent of the sentence meaning, but only depend on the experimental procedure or on the picture materials. For each of the three readings in the AS- and ES-conditions, we constructed an unambiguous control sentence as in (21) and (22), requiring the same judgement at the identical position in the same sequence used for the respective targets. For instance, example (21a) requires a *true*-response analogous to the AS-sentence in (19) under its literal reading, (21b) corresponds to the AS-sentence under its global reading and (21c) corresponds to its local reading. With regard to the ES-sentences, controls like (22a) correspond to the global reading, (22b) to the literal

and (22c) to the local reading. If, independently of sentence meaning, there was any bias to respond in a certain way at any point in the sequences, this should affect control sentences to the same degree as it affects target conditions.

(21) a. Alle Briefe sind mit mindestens drei ihrer Dreiecke verbunden.
All letters are with at-least three their triangles connected.
All letters are connected to at least three of their triangles.

b. Mindestens ein Brief ist mit genau fünf seiner Dreiecke verbunden.
At-least one letter is with exactly five his triangles connected.
At least one letter is connected with exactly five of its triangles.

c. Jeder Brief ist mit mindestens vier seiner Dreiecke verbunden.
Every letter is with at-least four his triangles connected.
Every letter is connected to at least four of its triangles.

(22) a. Alle Glocken sind mit weniger als vier ihrer Halbkreise verbunden.
All bells are with fewer than four their semicircles connected.
All bells are connected with fewer than four of their semicircles.

b. Alle Glocken sind mit allen ihren Halbkreisen verbunden.
All bells are with all their semicircles connected.
All bells are connected to all of their semicircles.

c. Mindestens drei Glocken sind mit allen ihren Halbkreisen verbunden.
At-least three bells are with all their semicircles connected.
At least three bells are connectd with all of their semicircles.

Controlling ambiguity resolution. In order to control for artificial effects of presentation order on the availability of readings, we included a second type of control conditions. In particular, we tested (i), whether the order of given readings affects preferences for ambiguous sentences and (ii), whether prosodic information can, in principle, shift reading preferences in the present task. Note that for our target sentences, the logical entailment relations between readings always require local readings to be evaluated at the end of a sequence. It could thus in principle be possible that no subject ever reaches this point, because they gave truth-value judgements earlier, thereby ending the trial. In that case it would be unclear if these decisions have been affected by a general inavailability of local readings or by the fact that one of the earlier presented readings is the preferred one. By including globally ambiguous structures with a known preference over available readings, we thus intended to test whether participants in our task occasionally choose dispreferred readings, even if they were presented following preferred readings.

We therefore included globally ambiguous *late-closure structures* as in (23), which have been repeatedly shown to exhibit interpretive preferences.

(23) Der Brief ist mit Kreisen und Vierecken mit Sonnen verbunden.
The letter is with circles and squares with suns connected.
The letter is connected with circles and squares with suns.

a. The letter is connected with squares containing suns, and it is also connected with circles. (LC)

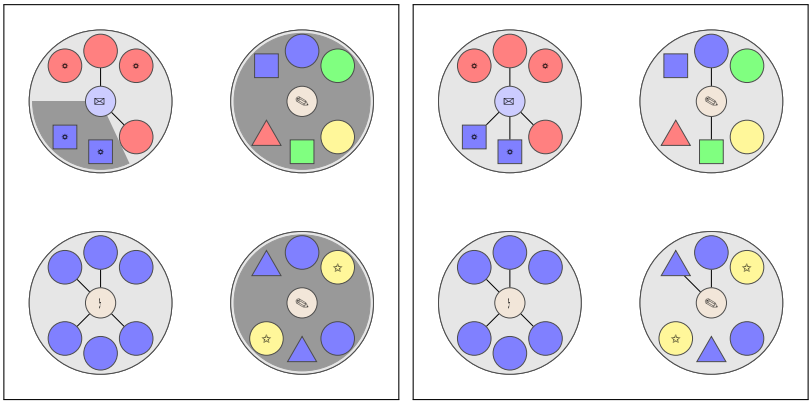
b. The letter is connected with circles and squares, both of which are containing suns. (EC)

In late-closure sentences, a modifier can be attached to one of two preceding hosts. More specifically, in (23), the prepositional phrase *with suns* can be attached to the preceding noun *squares*, resulting in the so-called *late-closure* or *low-attachment* reading ((23a), see Frazier, 1987, for details)[s]. Alternatively, the PP can be attached to the whole conjoined NP *circles and squares*, corresponding to the *early-closure* or *high-attachment* reading (23b). Generally, the late-closure reading is preferred over the early closure reading for sentences comparable to those in (23) (REFS)[t].

s. references

t. references

For each sentence, two sequences were designed. In the first, the critical position associated with the dispreferred high-attachment reading preceded the critical position for the low-attachment reading. In this case, a partly covered graph as in Figure 8a preceded a graph as in 8b. If participants adopted an early-closure reading as in (23b),



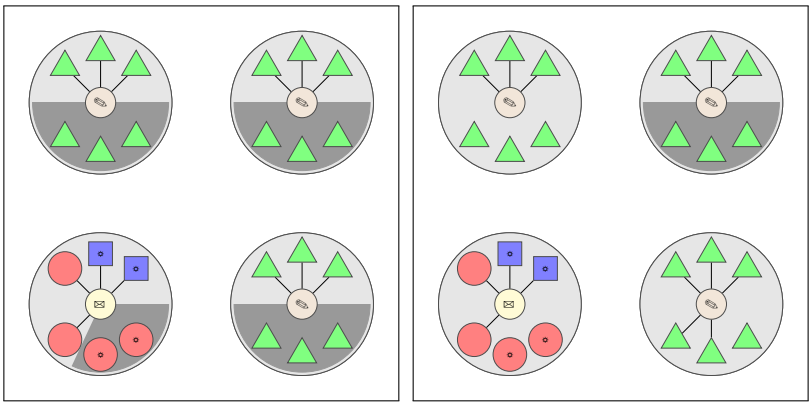
(a) Step 1 (b) Step 2

Figure 8: Critical steps of a sequence for sentence (23) where early-closure/high-attachment reading (23b) can be judged first.

a *false*-judgment would be expected at the position illustrated in 8a, as none of the circles connected to the letter contain any suns. Under the late-closure reading in (23b), in contrast, a *true*-response would be expected no sooner than in the step shown in Figure 8b. The order of possible judgements is reversed in the second kind of sequence including pictures like 9a followed by pictures like 9b. Here, the late-closure reading in (23a) can be judged first requiring a *true*-response on 9a. The early-closure reading in (23b) can only be judged later in the sequence. When we reach 9b a *false*-response is expected under an early-closure/high-attachment reading.

Testing Effects of accentuation. [u] It has been claimed that local readings only emerge

u. These two paragraphs still need a proper home; Petra opts for "materials" section.



(a) Step 1 (b) Step 2

Figure 9: Critical steps of a sequence for sentence (23) where late-closure/low-attachment reading (23a) can be judged first.

in the case of special accentuation of the the scalar item under consideration (REFERENCES). Further, it is generally assumed that the required type of accentuation is a pitch accent (REFS). To test for this possibility we decided to present the sentences (eg., (19) and (20)) auditorily in two versions each. In the accented version a pitch accent was placed on *einige*. The second version had neutral prosody. If accentuation is the driving force for local readings, we would expect a higher proportion of local readings in the former than in the latter version of the sentences.

We also wished to make sure that differences in prosody may generally lead to a shift in interpretation preferences in the ITV task. Sentences like (23) are well suited to test this assumption since it is well-known that their interpretational preferences are sensitive to prosody. If an utterance of (23) includes a prosodic phrase boundary between *squares* and *with*, this intonational pattern strongly corresponds to the high-attachment reading. If, instead, a prosodic phrase boundary after *circles* is present, the acceptance of the low-attachment reading is boosted. To test whether prosodic information is taken into account in the ITV we presented both types of prosodic structures expecting the described boost in acceptance for the low-attachment reading. In addition, we also included a neutral version containing no phrase marking whatsoever. This should receive intermediate acceptance rates. Further, we also presented unambiguous sentences corresponding to the early-closure reading. As in the case of the AS- and ES-sentences the unambiguous counterparts served as a baseline to control for response biases that are independent of sentence meaning.

5.2 Procedure

5.3 Materials

A detailed description of our auditory material is found in Appendix A.

5.4 Participants

41 native speakers of German took part in our study, none of which had any prior exposure to logic or formal semantics. We excluded 3 subjects due to insufficient performance on controls ($\leq$ 50% correct answers).

- possibly mention ages, backgrounds and sexes?

5.5 Results

Performance on basic control conditions was high (92% correct responses on average), indicating that subjects understood the task, giving truth-value judgements at the "right" moment in a sequence without being influenced in general by inessential features of the picture material or showing general response biases.

The judgments obtained for the four target conditions are depicted in Figure 10.[v] We coded the judgments as *literal*, *global* or *local* if they were as expected under one of these readings and as *error* if not. The distribution of readings as coded by us is presented in Figure 11. Participants mostly gave judgments indicating literal or local readings. Judgments compatible with a global reading were seldom obtained. In the AS-conditions participants made hardly any errors, whereas the amount of errors was slightly greater in the ES-conditions.

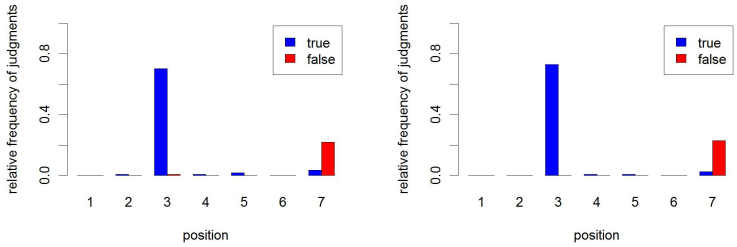
v. we should introduce the actual sequences that we presented so that it is clear what it means to say *true* or *false* on some position or other

In the AS-conditions (see Figure 11a) judgments were consistent with literal readings in 65.0% of the trials with neutral prosody and in 67.5% of the trials with accented prosody.[w] The number of global readings was very low, with 1.6% of the neutral and 0.8% of the accented trials (corresponding to a total of 3 answers across all trials). Judgments indicating local readings were given in 22.0% of both the neutral and accented trials. In the ES-conditions (see Figure 11b) judgments consistent with literal readings were given in 47.2% and 44.7% of the trials with neutral and accented prosody, respectively. Global readings were observed in 0.8% of the trials in both the neutral and the accented condition. Judgments indicating local readings were given in 22.8% of the neutral and 27.6% of the accented trials.

w. IMPORTANT: Figure contains different numbers (even if rounded)! Fabian, did you exclude the 3 bad performers when you obtained this data?

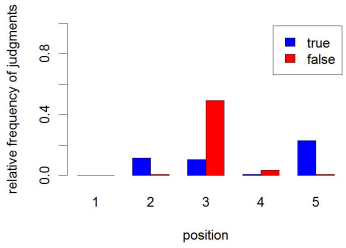
In order to test wether local readings exist we tested whether the number of local responses in each condition was higher than expected by chance. Local responses had to be given on the last position in each trial. If local readings didn't exist we would expect participants, who reached the last position (ie. who did not abort the trial prior to the last position) to give local judgments not more often than expected by chance. On the last position two responses were possible. One was compatible with a local reading the other was not. Therefore, we would expect 50% local responses by chance. As it turns out local responses were given significantly more often than 50% in all target conditions (all Bonferoni corrected $p<.05$). Note that this finding cannot be explained by some kind of general response bias on the last position since local readings required a yes-judgment in the ES-conditions but a no-judgment in the AS-conditions.

In order to test whether accentuation or the quantifier has an influence on the distribution of readings, log-liner models were computed (see Schepers 2003). The factors *Reading*, *Accentuation* and *Quantifier* were included in these models. Two variants of these models were computed. In the first, the factor *Item* was added to the above mentioned. In the latter *Participants* was included as a factor. Inclusion of these two factors

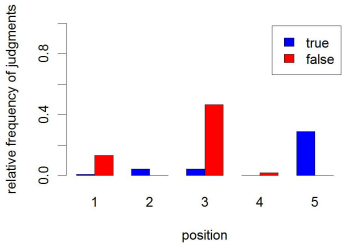


(a) AS-accented

(b) AS-neutral

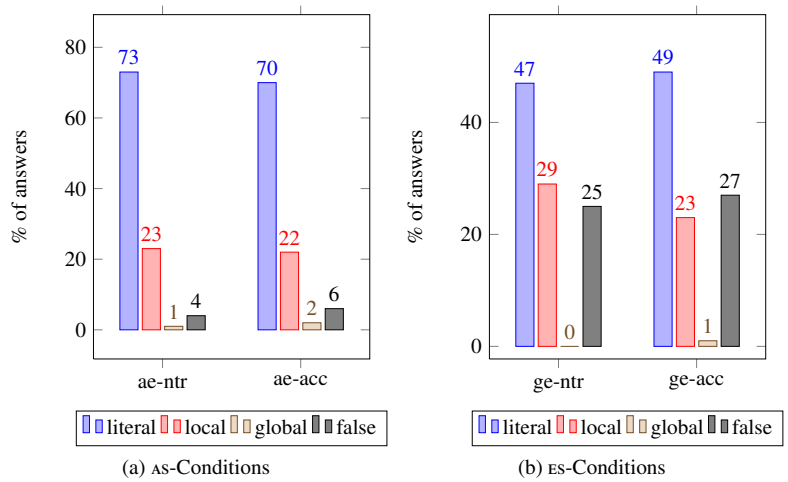


(c) ES-accented



(d) ES-neutral

Figure 10: Judgments for the four target conditions in Experiment 1. Relative frequency of yes/no-judgments is plotted against the position within the trial. In the AS-conditions there were seven positions. On the third position a yes-judgment corresponded to a literal reading, while on the fith position a yes-judgment corresponded to a global reading. A no-judgment on the seventh position was only consistent with a local reading. The ES-conditions had five positions. No- judgments on positions two and three corresponded to global and literal readings, respectively. A yes-judgment on the last segment corresponded to a local reading.



(a) AS-Conditions (b) ES-Conditions

Figure 11: Judgments

allowed us to test whether the distribution of readings was identical accros items and participants. We report log-likelihood ratio Chi-squares ($LRCS_1$ and $LRCS_2$), degrees of freedom (df_1, df_2) and significance levels (p_1 and p_2).

Unsurprisingly, there was an effect of reading ($LRCS_1 = 364.77, df_1 = 3, p_1 < .001, LRCS_2 = 364.77, df_2 = 3, p_1 < .001,$) because overall the readings were distributed inhomogeneously (sse Figure 12d). The quantifier had an influence on the distribution of readings as revealed by a reliable interaction of *Reading* and *Quantifier* ($LRCS_1 = 49.32, p_1 < .01, LRCS_2 = 32.16, p_1 < .01,$). Looking at the data it seemed unlikely that the quantifier affected the amount of local readings. We suspected that this interaction was due to the higher number of errors and lower number of literal readings in the ES-conditions as compared with the AS-conditions. A difference in the distribution of judgment types between participants was revealed by a reliable interaction of *Reading* and *Participant* ($LRCS_1 = 487.70, p_1 < .01$). We were interested in whether this effect was due to some of the participants being more likely to choose literal readings than others. Finally, there was a three-way interaction of *Participants*, *Construction* and *Reading* ($LRCS_1, = 143.92, df_1 = 117, p_1 < .05$) which could stem from to the fact that some participants were more prone than others to make errors in the ES-conditions.

In order to find out whether the interactions just reported affected the amount of local readings, the log-linear models were computed again with a different coding of the readings. Here, we only considered the amount of local readings versus all other kinds of judgments. Judgments were coded as *local* and *other*. If the reported interactions affect the amount of local readings they should show up again. The expected but irrelevant effect of *Reading* was again significant ($LRCS_1 = 134.55, df_1 = 1, p_1 < .001, LRCS_2 = 144.64, df_2 = 1, p_1 < .001,$). In addition, the interaction of *Participant* and *Reading* ($LRCS_1 = 487.70, p_1 < .01$) as well as the three-way interaction of

Construction, *Participant* and *Reading* ($LRCS_1 = 487.70, p_1 < .01$) were significant. No other effects were significant. In particular, the interaction of *Construction* and *Reading* was not significant ($LRCS_1 = 2.10, p_1 = .15, LRCS_2 = .74, p_2 = .39$).

The interaction of *Participant* and *Reading* shows that there are indeed varying preferences for local readings among German speakers. Further examination of the distribution of local readings revealed a clear pattern (see Figure **??**). In all four conditions, about seven out of the 40 participants (16%) were consistently giving judgments that indicate local readings. Further, the relative frequency of local judgments per participant strongly correlated between all four conditions (all $r > .6, p < .001$). Appart from the localists about half of the participants (and 41.5% of all participants) were consistently giving judgments that indicate literal readings. Only a minority of participants exhibited inconsistency. Reading preferences were more clear-cut in the AS-conditions than in the ES-conditions. The number of consistent litaralists was lower in the ES-conditions (25.6%) as compared to the AS-conditions (57.3%). Also, the number of participants that showed inconsitency was higher in these conditions. However, the number of consistent localists did hardly differ between constructions.

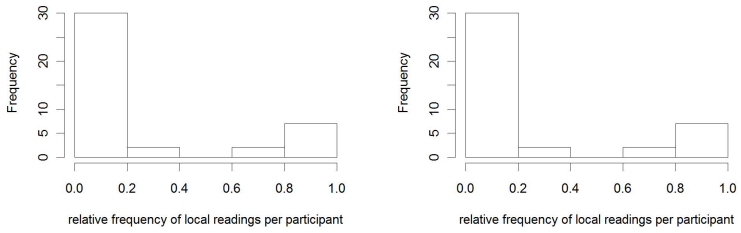
The three way-interaction of *Participant*, *Construction* and *Reading* is due to the fact that for some participants preferences for local over other readings deviated between the two construction types. These deviations were, however, not systematic to any degree. How much local preferences deviated between the two construction types per participant is depicted in Figure **??**. Most participants had exactly the same preferences in the two constructions. However, a few had a stronger and a few others a weaker preference for local readings in the ES- than in the AS-conditions. Since we did not find any systematic patterns here, we speculate that the three-way interaction is due to the ES-construction beeing understood non-standardly by a few participants. This speculation is plausible given the higher number of errors in the ES- as compared to the AS-conditions.

The absence of the interaction between *Construction* and *Participant* indicates that the type of construction does not affect the amount of local readings. We assumed based on the observed distribution of judgments that the type of contsruction affected the amount of literal but not local and global readings. To further test this assumption, we also considered literal and global readings versus other judgments in separate analyses. Comparing global readings versus other judgments *Construction* and *Reading* were found to interact ($LRCS_1 = 42.80, df = 1, p < .001; LRCS_2 = 22.14, df = 1, p < .001$). Comparing global readings to other judgments no such interaction was obeserved ($LRCS_1 = 1.24, df = 1, p = .29, LRCS_1 = .21, df = 1, p = .65$).

5.6 Discussion

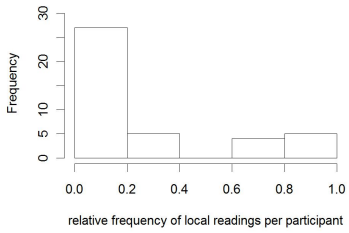
5.6.1 Bits and Pieces

- The distribution of responses found in previous studies could in part be explained by differences in typicality of the pictures with respect to the possible readings (see van Tiel). In the present case this kind of reasoning would have to be extended to the typicalilty of parts of pictures. On intuitive grounds, we take this kind of reasoning to be implausible. We will now shortly discuss how graphical

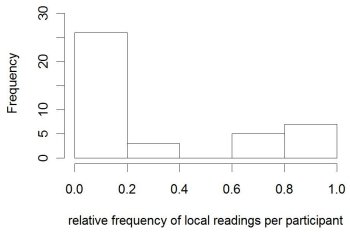


(a) AS-accented

(b) AS-neutral



(c) ES-accented



(d) ES-neutral

Figure 12: Judgments for the four target conditions

properties might still affect judgments in the worst case. In the first step of the AS-condition local and global readings cannot be judged, yet. Therefore, graphical properties should not affect judgments if participants have one of these readings in mind. If participants have a literal reading in mind, the worst case scenario is that they refrain from judging the sentence as true at this point due to some, eg. atypical, graphical property of the picture. The situation is similar at the second step. The local reading is still not decidable. Therefore, graphical properties should not play a role if participants have a local reading in mind. A positive judgment might, however, be a delayed judgment of a literal reading. Again, the worst case scenario for our purpose is that participants refrain from answering because of graphical properties. At the last step the local reading requires a negative judgment. Now, we could obtain delayed negative responses from literal or global readings because the graph as a whole is, e.g., logically true but atypical with respect to these readings. These kinds of judgments would be indistinguishable from local responses. Note, however, that we have no reason to believe that typicality has large effects, because, given the results from van Tiel, the picture as a whole is expected to be one of the more typical situations for AS-sentences. Still, we cannot exclude the possibility that graphical properties of the pictures might have the described effect. In the ES-conditions this type of effect is, how-

ever, entirely impossible since the literal and global reading is plainly false given the initial parts as well as the whole graph. Summing up, in the worst case, which we consider implausible, graphical effects might lead to an enhanced number of local responses in the AS- as compared to the ES-condition. We have to keep this possibility in mind. We want to stress, however, that the design presented here improves on previous studies with respect to possible graphical effects.

6 Conclusions

A The auditory sentence material

A phonetically trained female native speaker of German was instructed to realize two prosodic versions of each target sentence, and four versions of the target-related fillers. In addition to these sentences, a set of unrelated fillers and neutral control structures was also recorded. For these constructions, the speaker was instructed to realize prosodic contours as neutral and natural as possible.

For all target sentences (AS and ES), the determiner *some* was produced with a contrastive pitch accent as well as with a neutral accent, respectively. Contrastive accents in German are realized by an H*L contour (XYZ)[x], thus employing a generally falling pattern. A set of 15 experimental items was recorded for each condition (*accented* vs. *unaccented*), resulting in a total number of 60 target sentences.

x. insert references

In contrast to the accent manipulation, target-related fillers differed with respect to prosodic phrasing. Prosodic phrase boundaries in German are realized by a rise in F0 as well as by a durational increase on the final part of the constituent preceding the boundary (prefinal lengthening) plus an optional pause (XYZ)[y]. Boundaries for these filler sentences were either realized at the position separating the second PP from the preceding material (*late boundary*, corresponding to a high-attachment reading) or directly preceding the second conjunct in the first PP (*early boundary*, corresponding to a low-attachment reading). As the prosodic realization of the targets involved the comparison between an accented and a neutral variant, we also included a third version of target-related fillers, in which our speaker produced the sentences without any pronounced boundaries. Finally, an additional condition was recorded, using disambiguated sentences that were prosodically neutral (XYZ)[z]. For each prosodic variant, a set of 30 items was read, yielding 150 target-related fillers. Together with 30 unrelated fillers and 90 control sentences, a total number of 300 sentences was recorded. The session was recorded in an acoustically shielded booth (44.1 kHz sampling rate, 16 bit amplitude resolution).

y. references: Vaissiere, 1983; Fery, 1993

z. proper reference to explanation above

Before entering the judgment task, experimental items and target-related fillers were analyzed with respect to their acoustic properties. As both accented elements as well as prosodic boundaries were expected to differ with respect to their F0 and/or durational properties, we calculated durational values as well as difference values between minimal and maximal F0 for each word. As targets slightly differed with respect to the total number of words as well as with respect to certain lexical properties (i.e., *seinen* vs. *ihren*, see above (XYZ)[a]), we considered the following analysis regions:

a. proper reference

(24) a. $|_{R1}$ Alle $|_{R2}$ dieser $|_{R3}$ NP$_1$ $|_{R4}$ sind $|_{R5}$ mit $|_{R6}$ einigen $|_{R7}$ ihrer $|_{R8}$ NP$_2$ $|_{R9}$ verbunden.

b. $|_{R1}$ Genau einer $|_{R2}$ der $|_{R3}$ NP$_1$ $|_{R4}$ ist $|_{R5}$ mit $|_{R6}$ einigen $|_{R7}$ seiner $|_{R8}$ NP$_2$ $|_{R9}$ verbunden.

Note that differences between Regions 1, 2, 4 and 7 can be expected due to lexical differences between the AS and ES-conditions. As target-related fillers did not differ with respect to their lexical properties, we carried out word-by-word analyses for these conditions. Durational values included the respective word plus any following silent interval. Note that we did not include the disambiguated fillers in these analyses as they involved very different sentence types (i.e., constructions involving prepositional phrases vs. relative clauses).

(25) $|_{R1}$ D $|_{R2}$ NP$_1$ $|_{R3}$ ist $|_{R4}$ mit $|_{R5}$ NP$_2$ $|_{R6}$ und $|_{R7}$ NP$_3$ $|_{R8}$ mit $|_{R9}$ NP$_4$ $|_{R10}$ verbunden.

For the durational analyses, constituents were automatically labeled by the *Aligner* tool (XYZ),[b] and the obtained values were manually corrected afterwards. For the targets, two-factorial ANOVAs with the factors QUANTIFIER (all vs. exactly one) and PROSODY (accented vs. unaccented) were carried out. For the target-related fillers, we carried out one-factorial ANOVAs with the factor PROSODY (early boundary, late boundary, neutral prosody). F0 values were extracted by means of special Praat scripts (http://www.fon.hum.uva.nl/praat/). For the present analyses, differences between minimal and maximal F0 values for each region or word were calculated. Again, two-factorial ANOVAs were carried out for statistical comparison of the fillers, and one-factorial ANOVAs were carried out for target-related structures.

b. reference! Rapp, 1998

Targets. Table 4 shows the durational values for each of the single regions in the sentence. Crucially, durational values of accented determiners were significantly increased as opposed to their non-accented counterparts. Overall, QUANTIFIER effects were observed, which might be attributed to the above-mentioned lexical differences, as well as to the fact that ES-structures generally contained fewer words, thus leading to a tendency of a durational decrease for each of the single words.

Differences between maximal and minimal F0 values for each of the single words in the sentence are depicted in Table 5. As is descriptively evident, accented determiners showed a larger F0 range compared to unaccented versions, an effect which is confirmed by statistical analyses. An additional QUANTIFIER *PROSODY interaction indicates that these differences are even more pronounced in the AS-condition. Finally, comparable to the durational analyses, QUANTIFIER effects were observed when conditions exhibited lexical differences.

Finally, differences between maximal and minimal F0 values for each of the single words in the sentence are given in Table 7. Again, the most reliable differences occur at the boundary regions.[c]

c. provide caption for table D

Target-related fillers. Table 6 shows the durational values of each of the single words in the sentence. As is descriptively evident, the largest durational differences

	AS		ES			STATISTICS
	acc	**ntr**	**acc**	**ntr**		
R1	225.8	216.9	563.0	530.0		**QUANTIFIER: $F = 2251.3$; $p < .001$** **PROSODY: $F = 15.2$; $p < .01$** QUANTIFIER*PROSODY: $F = 3.6$; $p = .09$
R2	245.3	234.7	128.7	128.0		**QUANTIFIER: $F = 525.5$; $p < .001$** PROSODY: $F = 2.5$; $p = .13$ QUANTIFIER*PROSODY: $F = 3.2$; $p = .10$
R3	397.0	414.6	398.6	400.0		QUANTIFIER: $F = .3$; $p = .57$ PROSODY: $F = 1.2$; $p = .29$ QUANTIFIER*PROSODY: $F = 1.5$; $p = .25$
R4	172.3	175.6	164.7	152.7		**QUANTIFIER: $F = 16.1$; $p = .001$** PROSODY: $F = 2.1$; $p = .17$ **QUANTIFIER*PROSODY: $F = 8.7$; $p < .05$**
R5	230.2	170.5	226.7	170.0		QUANTIFIER: $F = .2$; $p = .7$ **PROSODY: $F = 79.0$; $p < .001$** QUANTIFIER*PROSODY: $F = .0$; $p = .90$
R6	445.8	422.0	420.6	383.0		**QUANTIFIER: $F = 31.4$; $p < .001$** **PROSODY: $F = 23.8$; $p < .001$** QUANTIFIER*PROSODY: $F = 1.0$; $p = .34$
R7	199.7	194.7	235.3	249.3		**QUANTIFIER: $F = 19.9$; $p = .001$** PROSODY: $F = .9$; $p = .36$ **QUANTIFIER*PROSODY: $F = 9.6$; $p < .01$**
R8	503.0	494.7	494.6	494.7		QUANTIFIER: $F = .04$; $p = .84$ PROSODY: $F = 2.2$; $p = .17$ QUANTIFIER*PROSODY: $F = 1.7$; $p = .2$
R9	722.0	748.0	704.0	764.7		QUANTIFIER: F = .0; p = .98 **PROSODY: F = 10.9; p < .01** QUANTIFIER*PROSODY: F = 2.9; p = .11

Table 4: Durational values in ms for each of the single regions in the target sentences. Region 6 corresponds to *einigen*.

	AS		ES			STATISTICS
	acc	**ntr**	**acc**	**ntr**		
R1	17.6	19.6	52.7	31.6		**QUANTIFIER: $F = 14.6$; $p < .01$** PROSODY: $F = 2.5$; p = .14 QUANTIFIER*PROSODY: $F = 3.8$; $p = .07$
R2	27.4	36.4	19.0	17.1		**QUANTIFIER: $F = 17.1$; $p = .001$** PROSODY: $F = 1.9$; $p = .19$ QUANTIFIER*PROSODY: $F = 3.4$; $p = .09$
R3	82.6	115.1	67.6	91.7		**QUANTIFIER: $F = 8.7$; $p < .05$** **PROSODY: $F = 20.5$; $p < .001$** QUANTIFIER*PROSODY: $F = .5$; $p = .50$
R4	32.6	60.6	17.6	28.4		**QUANTIFIER: $F = 16.4$; $p = .001$** **PROSODY: $F = 8.6$; $p < .05$** QUANTIFIER*PROSODY: $F = 1.9$; $p < .19$
R5	62.4	46.0	39.8	65.1		QUANTIFIER: $F = .1$; $p = .80$ PROSODY: $F = .3$; $p < .59$ **QUANTIFIER*PROSODY: $F = 5.2$; $p < .05$**
R6	211.3	69.9	174.3	75.1		QUANTIFIER: $F = 2.9$; $p = .11$ **PROSODY: $F = 94.8$; $p < .001$** **QUANTIFIER*PROSODY: $F = 13.7$; $p < .01$**
R7	57.6	51.7	30.4	24.6		**QUANTIFIER: $F = 9.4$; $p < .01$** PROSODY: $F = .6$; $p = .45$ QUANTIFIER*PROSODY: $F = .1$; $p = .80$
R8	36.7	71.4	32.5	53.5		QUANTIFIER: $F = 3.1$; $p = .1$ **PROSODY: $F = 45.5$; $p < .001$** QUANTIFIER*PROSODY: $F = .4$; $p = .51$
R9	47.1	61.5	42.3	60.7		QUANTIFIER: F = .2; p = .69 PROSODY: F = 3.3; p = .09 QUANTIFIER*PROSODY: F = .1; p = .7

Table 5: Difference between minimal and maximal F0 values in Hz for each of the single words in the target sentences. Region 6 corresponds to *einigen*.

were realized at the boundary regions (i.e. Region 5 and Region 7). Interestingly, small differences between conditions also yielded significance at other positions, suggesting that our speaker produced the different conditions very consistently (i.e., with little variance).

References

Atlas, Jay David and Stephen Levinson (1981). "It-clefts, Informativeness, and Logical Form". In: *Radical Pragmatics*. Ed. by Peter Cole. Academic Press, pp. 1–61.

Breheny, Richard and Napoleon Katsos (2008). *Experimental Investigations of the Semantics/Pragmatics Interface*. Course Material for ESSLLI.

Breheny, Richard et al. (2006). "Are Generalised Scalar Implicatures Generated by Default? An On-Line Investigation into the Role of Context in Generating Pragmatic Inferences". In: *Cognition* 100.3, pp. 434–463.

Chemla, Emmanuel (2009). "Presuppositions of Quantified Sentences: Experimental Data". In: *Natural Language Semantics* 17.4, pp. 299–340.

Chemla, Emmanuel and Benjamin Spector (2011). "Experimental Evidence for Embedded Scalar Implicatures". In: *Journal of Semantics* 28, pp. 359–400.

Chierchia, Gennaro (2004). "Scalar Implicatures, Polarity Phenomena and the Syntax/Pragmatics Interface". In: *Structures and Beyond*. Ed. by Adriana Belletti. Oxford University Press, pp. 39–103.

— (2006). "Broaden Your Views: Implicatures of Domain Widening and the Syntax/Pragmatics Interface". In: *Linguistic Inquiry* 37.4, pp. 535–590.

— (forthcoming). *FC Nominals and FC Disjunction: The Identity Thesis*. Ed. by Anamaria Fălăuş. Palgrave Macmillan.

Chierchia, Gennaro et al. (forthcoming). "The Grammatical View of Scalar Implicatures and the Relationship between Semantics and Pragmatics". In: *Semantics. An International Handbook of Natural Language Meaning*. Ed. by Claudia Maienborn et al. de Gruyter.

Clifton, Charles and Chad Dube (2010). "Embedded Implicatures Observed: A Comment on Geurts and Pouscoulous (2009)". In: *Semantics & Pragmatics* 3.7, pp. 1–13.

Dalrymple, Mary et al. (1998). "Reciprocal Expressions and the Concept of Reciprocity". In: *Linguistics and Philosophy* 21.2, pp. 159–210.

Davidson, Donald (1973). "Radical Interpretation". In: *Dialectica* 27, pp. 313–328.

Fox, Danny (2007). "Free Choice and the Theory of Scalar Implicatures". In: *Presupposition and Implicature in Compositional Semantics*. Ed. by Uli Sauerland and Penka Stateva. Hampshire: Palgrave MacMillan, pp. 71–120.

Fox, Danny and Benjamin Spector (2009). "Economy and Embedded Exhaustification". Handout for a talk given at Cornell University.

Franke, Michael (2011). "Quantity Implicatures, Exhaustive Interpretation, and Rational Conversation". In: *Semantics & Pragmatics* 4.1, pp. 1–82.

Gazdar, Gerald (1979). *Pragmatics: Implicature, Presupposition, and Logical Form*. New York: Academic Press.

	early	late	ntr		STATISTICS: EFFECT OF PROSODY
R1	105.7	100.7	98.1		**$F = 5.0$; $p < .05$** Early vs. late: $F = 3.4$; $p = .07$ **Early vs. ntr: $F = 9.1$; $p < .001$** Late vs. ntr: $F = 1.7$; $p = .21$
R2	332.3	306.6	313.6		**$F = 28.1$; $p < .001$** **Early vs. late: $F = 70.4$; $p < .001$** **Early vs. ntr: $F = 28.7$; $p < .001$** Late vs. ntr: $F = 3.0$; $p = .09$
R3	168.0	167.0	169.3		$F = .2$; $p = .77$
R4	135.3	139.7	141.3		$F = .7$; $p = .52$
R5	561.3	1268.3	642.6		**$F = 819.5$; $p < .001$** **Early vs. late: $F = 1110.4$; $p < .001$** **Early vs. ntr: $F = 94.4$; $p < .001$** **Late vs. ntr: $F = 680.1$; $p < .001$**
R6	143.3	176.3	150.0		**$F = 16.1$; $p < .001$** **Early vs. late: $F = 25.3$; $p < .001$** Early vs. ntr: $F = 1.1$; $p = .30$ **Late vs. ntr: $F = 22.4$; $p < .001$**
R7	1183.7	535.3	557.3		**$F = 1920.3$; $p < .001$** **Early vs. late: $F = 2251.4$; $p < .001$** **Early vs. ntr: $F = 2108.1$; $p < .001$** **Late vs. ntr: $F = 9.5$; $p < .01$**
R8	167.0	150.3	151.0		**$F = 12.8$; $p < .001$** **Early vs. late: $F = 17.7$; $p < .001$** **Early vs. ntr: $F = 15.8$; $p < .001$** Late vs. ntr: $F = .0$; $p = .85$
R9	384.7	381.0	402.7		**$F = 11.7$; $p < .001$** Early vs. late: $F = .7$; $p = .48$ **Early vs. ntr: $F = 11.2$; $p < .01$** **Late vs. ntr: $F = 22.7$; $p < .001$**
R10	691.3	681.3	730.0		**$F = 10.0$; $p < .001$** Early vs. late: $F = .8$; $p = .39$ **Early vs. ntr: $F = 11.8$; $p < .01$** **Late vs. ntr: $F = 16.8$; $p < .001$**

Table 6: Durational values in ms for each of the single regions in the target-related fillers. Regions 5 and 7 correspond to the nouns preceding the boundaries.

	early	late	ntr		STATISTICS: EFFECT OF PROSODY
R1	31.2	23.1	31.7		$F = .9$; $p = .43$
R2	332.3	306.6	313.6		**$F = 7.3$; $p < .01$** **Early vs. late $F = 14.1$; $p = 001$** Early vs. ntr. $F = 1.1$; $p = .30$ **Early vs. late $F = 11.1$; $p < .01$**
R3	168.0	167.0	169.3		**$F = 4.2$; $p < .05$** **Early vs. late $F = 14.9$; $p = .001$** Early vs. ntr. $F = 1.7$; $p = .21$ Early vs. late F 3.7 = ; $p = .07$
R4	135.3	139.7	141.3		$F = 1.3$; $p = .39$
R5	561.3	1268.3	642.6		**$F = 129.3$; $p < .001$** **Early vs. late $F = 314.1$; $p < .001$** **Early vs. ntr. $F = 84.4$; $p < .001$** **Early vs. late $F = 39.2$; $p < .001$**
R6	143.3	176.3	150.0		**$F = 6.7$; $p < .01$** Early vs. late $F = 1.4$ $p = .24$ **Early vs. ntr. $F = 14.8$; $p = .001$** **Early vs. late $F = 5.1$; $p < .05$**
R7	1183.7	535.3	557.3		**$F = 132.5$; $p < .001$** **Early vs. late $F = 145.1$; $p < 001.$** **Early vs. ntr. $F = 303.9$; $p < .001$** Early vs. late $F = 1.5$; $p = .23$
R8	167.0	150.3	151.0		$F = .8$; $p = .47$
R9	384.7	381.0	402.7		$F = 3.1$; $p = .06$
R10	691.3	681.3	730.0		**$F = 25.8$; $p < .001$** **Early vs. late $F = 39.7$; $p < .001$** Early vs. ntr. $F = .2$; $p = .7$ **Early vs. late $F = 60.4$; $p < .001$**

Table 7: ???

Geurts, Bart (2009). “Scalar Implicature and Local Pragmatics”. In: *Mind and Language* 24.1, pp. 51–79.
— (2010). *Quantity Implicatures*. Cambridge University Press.
Geurts, Bart and Nausicaa Pouscoulous (2009). “Embedded Implicatures?!?” In: *Semantics & Pragmatics* 2.4, pp. 1–34.
Grice, Paul Herbert (1975). “Logic and Conversation”. In: *Syntax and Semantics, Vol. 3, Speech Acts*. Ed. by Peter Cole and Jerry L. Morgan. Academic Press, pp. 41–58.
Grodner, Daniel J. et al. (2010). ““Some,” and Possibly All, Scalar Inferences are not Delayed: Evidence for Immediate Pragmatic Enrichment”. In: *Cognition* 166, pp. 42–55.
Groenendijk, Jeroen and Martin Stokhof (1984). “Studies in the Semantics of Questions and the Pragmatics of Answers”. PhD thesis. Universiteit van Amsterdam.
Horn, Laurence R. (1972). “On the Semantic Properties of Logical Operators in English”. PhD thesis. UCLA.
— (2006). “The Border Wars: A Neo-Gricean Perspective”. In: *Where Semantics Meets Pragmatics*. Ed. by Klaus von Heusinger. Vol. 16. Current Research in the Semantics/Pragmatics Interface, pp. 21–48.
Ippolito, Michela (2010). “Embedded Implicatures? Remarks on the debate between globalist and localist theories”. In: *Semantics & Pragmatics* 3.5, pp. 1–15.
Levinson, Stephen C. (2000). *Presumptive Meanings. The Theory of Generalized Conversational Implicature*. Cambridge, Massachusetts: MIT Press.
Magri, Giorgio (2011). “Another Argument for Embedded Scalar Implicatures based on Oddness in Downward Entailing Environments”. In: *Semantics & Pragmatics* 4.6, pp. 1–31.
De Neys, Wim and Walter Schaeken (2007). “When People Are More Logical Under Cognitive Load: Dual Task Impact on Scalar Implicature”. In: *Experimental Psychology* 54.2, pp. 128–133.
van Rooij, Robert and Katrin Schulz (2004). “Exhaustive Interpretation of Complex Sentences”. In: *Journal of Logic, Language and Information* 13, pp. 491–519.
Russell, Benjamin (2006). “Against Grammatical Computation of Scalar Implicatures”. In: *Journal of Semantics* 23.361–382.
Sauerland, Uli (2004). “Scalar Implicatures in Complex Sentences”. In: *Linguistics and Philosophy* 27, pp. 367–391.
— (2010). “Embedded Implicatures and Experimental Constraints: A Reply to Geurts & Pouscoulous and Chemla”. In: *Semantics & Pragmatics* 3.2, pp. 1–13.
— (2012). “The Computation of Scalar Implicatures: Pragmatic, Lexical or Grammatical”. In: *Language and Linguistics Compass* 6.1, pp. 36–49.
Schulz, Katrin (2005). “A Pragmatic Solution for the Paradox of Free Choice Permission”. In: *Synthese* 147, pp. 343–377.
Schulz, Katrin and Robert van Rooij (2006). “Pragmatic Meaning and Non-monotonic Reasoning: The Case of Exhaustive Interpretation”. In: *Linguistics and Philosophy* 29, pp. 205–250.
Schwarz, Florian et al. (2008). “Strengthening ‘or’: Effects of Focus and Downward Entailing Contexts on Scalar Implicatures”. Unpublished manuscript; retrived from semanticsarchive March 11 2011.

Soames, Scott (1982). "How Presuppositions are Inherited: A Solution to the Projection Problem". In: *Linguistic Inquiry* 13.3, pp. 483–545.

Spector, Benjamin (2006). "Scalar Implicatures: Exhaustivity and Gricean Reasoning". In: *Questions in Dynamic Semantics*. Ed. by Maria Aloni et al. Amsterdam, Singapore: Elsevier, pp. 229–254.

von Stechow, Arnim and Thomas Ede Zimmermann (1984). "Term Answers and Contextual Change". In: *Linguistics* 22.1, pp. 3–40.

van Tiel, Bob (2012). *Embedded Scalars and Typicality*. Talk held at XPrag 2011, Barcelona.