<table>
<thead>
<tr>
<th>Poster session 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wednesday, September 4 - 15:00 – 16:00</td>
</tr>
</tbody>
</table>

| 1 | Elbert Jan Booij and Galit Weidman Sassoon  
Classification under 'big': The role of leaps in distributions and its delayed acquisition |
| 2 | Saskia van Putten  
The meaning of the Avatime additive particle: similarity, topic shift and set membership |
| 3 | Stephen Politzer-Ahles and Robert Fiorentino  
Sensitivity of online scalar inferencing to context and to processing load |
| 4 | Elena Castroviejo, Laia Mayol and E. Allyn Smith  
A cross-linguistic study of direct denial as a diagnostic for meaning types |
| 5 | Mirjana Mandic and Boban Arsenijevic  
The role of educational background in generating scalar implicatures |
| 6 | Judith Degen, Christine Gunlogson and Michael K. Tanenhaus  
Using corpora to study the distributional features of some (but not all) implicatures |
| 7 | Anna Pryslopska  
Implicatures in uncooperative contexts. Evidence from a visual world paradigm. |
| 8 | John Hoeks, Harm Brouwer, Petra Hendriks and Laurie A. Stowe  
How the Brain Responds to Partial Answers: An ERP Study |
| 9 | Rosa S. Gisladottir, Dorothee J. Chwilla and Stephen C. Levinson  
Early speech act comprehension in spoken dialogue: Evidence from ERPs |
| 10 | Kazuko Yatsushiro and Uli Sauerland  
When Children are more Pragmatic than Adults |
| 11 | Lyn Tieu and Jeffrey Lidz  
Testing for domain widening: Experimental evidence for obligatory exhaustification |
| 12 | Annika Deichsel and Klaus Von Heusinger  
Testing the scopal properties of three different indefinite articles in German |
| 13 | Emmanuel Chemla, Chris Cummins and Raj Singh  
Reading Hurford disjunctions: Processing evidence for embedded implicatures |
| 14 | Chia-Ying Chu and Utako Minai  
Comprehension of Demonstratives by English-speaking Preschool Children |
| 15 | Christina Kim, Ming Xiang and Chris Kennedy  
Context dependence and shiftability in two classes of gradable adjectives |
Big Differences: The standard for `big' as used by adults and children

Every time a relative adjective such as tall/short is used, it divides the domain of discourse – from candles to mountains – into those items that may be called, e.g. tall and the rest of them. To make communication possible, a hearer must know where the cutoff point between tall and non-tall entities lies. The relevant information must be extracted from the distribution of lengths among the entities in the comparison class. The question this study investigates is: What is this relevant information? One hypothesis (H1) is that a certain degree, dependent on range, but fixed across different distributions, is the standard. According to a second view (H2), speakers use the rank ordering of entities by their length, selecting as tall the n% tallest entities (Barner & Snedeker 2008). An even more widespread intuition, going back to Vennemann & Bartsch (1972), is that speakers make more complex calculations: the standard is based on e.g. the average in the domain, with a potential effect of the standard deviation as well (Solt 2011); e.g., those entities are tall which are taller than the average length plus a certain proportion of the standard deviation (H3).

We addressed these options, with special attention to a fourth one, whereby leaps in the distribution trigger cutoff points. The importance of leaps was already noted by Kennedy (2007) and Van Rooij (2011a), who discuss their role in the analysis of implicit comparison, as in This pencil is long, compared to that one, where the distinction between the two pencils must be easy to observe. Gaifman (2010) and Van Rooij (2011b) highlight the merits of leaps as means to avoid the Sorites paradox, by ‘breaking’ Sorites series. However, no previous studies have aimed to test how likely leap midpoints are to be chosen as standards, as opposed to, e.g. averages.

Following Booij (2012), we hypothesized that the use of gradable adjectives is subject to a principle of discriminatory economy. One consequence is that distinctions that are easy to make are privileged, for they minimize the risk of being wrong. A cutoff point for e.g. tall that coincides with a leap in the distribution, meets this condition. It reduces uncertainty, as most of the items will be conspicuously shorter or taller.

Methods: Two groups of Dutch-speaking subjects were tested, 28 adults and 26 primary school children. The adjective to be tested was the Dutch word groot (En.: big, large), as applied to pencils. Each subject was presented with two out of four randomly assigned sets of 12 pencils, one after another, and was asked to put the big pencils in a basket. The length distributions of the four sets, D1-D4, are displayed in Figure 1. The distributions were all symmetrical, and had the same average (10.5 cm), but differed in shape and standard deviation (SD_{D2} > SD_{D1} ≈ SD_{D3} > SD_{D4}). Pencil length in D1 was linearly increasing; D2 was clustered into two sets separated by a leap in the middle (bimodal); D3 was like D1, except for two leaps; D4 had one large cluster in the middle.

Predictions: If the criterion for big is to be bigger than a fixed size (H1), then the transition midpoint, i.e. the point in between the length of the first big pencil and that of the last non-big one (our estimate of the 'real' cutoff point), should be expected to be relatively invariant across distributions (P1). If the criterion for big is to be among a fixed percentage of the biggest pencils (H2), then the transition rank, i.e. the number of the first big pencil (counting from the smallest to the biggest) must be expected to be relatively invariant across distributions (P2). If the criterion for big is
to be of a size beyond the average plus a times the standard deviation \((m + a*sd)\) (H3), then, since the average is constant, the transition midpoint should be farther from the average when the standard deviation is bigger (P3). Finally, if the cutoff point for big should coincide with a leap (H4), then leaps in the distribution should enhance the likelihood for the transition midpoint (of the transition from non-big to big), to fall in that region (P4).

Results and discussion: Adults All respondents were consistent in their response, i.e. no big pencil was smaller than any non-big pencil in the same test. Mann-Whitney U tests yielded that, except for the combinations D1-D3 and D3-D4, the differences between the transition ranks for the four distributions were highly significant \((p < .01)\). With respect to the transition midpoint, the difference was only significant for the combination D2-D4. Thus, predictions P1-P2 were not borne out, contra the fixed size hypothesis (H1), and, in particular, the fixed-rank hypothesis (H2). P3 was not borne out, either: the transition midpoint was roughly closer to the average with bigger standard deviation, contra the average plus a times standard deviation hypothesis (H3).

The difference between D1 and D3 is of special interest because the mean and standard deviation of both are almost the same, but D3 has leaps, the rightmost of which is between 13 cm and 15.5 cm. As predicted, a peak of respondents had their transition from non-big to big in this interval. If we take the results for D1 (where there are no leaps) to be a fair estimate of the 'background' likelihood of the transition to be between 13 cm and 15.5 cm, then a binomial test gives a significance of \(p < .054\) for the peak associated with the leap in D3. Thus the leap indeed appears to trigger a cutoff point where it would otherwise have been less probable, in support of H4.

This result suggests that standard selection is affected by leaps, over and above any effect of average or deviation. This may be explained by a preference for distinctions that are easy to make, so as to minimize communication errors. Moreover, standard selection based on leaps may be less costly than one based on average, if, instead of using degrees, speakers encode the information by means of semi-orders (van Rooij 2011a,b). If so, the use of leaps is a natural consequence, appealing also in terms of processing costs.

Children In all four cases the results for the children differed significantly from those for the adults \((p < .05, \text{ except for D2 where } p < .085)\). The children, unlike the adults, in several cases showed a transition below the average length of the pencils. None of the comparisons of the transition midpoints/ranks in D1-D4 proved significant. Leaps did, however, affect the results for the transition rank, especially in D4 vs. D1, contra the fixed-rank hypothesis (H2). In contrast to what was seen with the adults, children tended to locate the transition at the scale midpoint, as a fixed-size hypothesis (H1) predicts. Leaps did not affect the transition midpoint, so the leap hypothesis (H4) found no support for children. Possibly, the tendency to minimize communication errors by using distinctions that are easy to make (leaps) develops later than age 8.

The meaning of the Avatime additive particle: similarity, topic shift and set membership

Avatime is a Kwa (Niger-Congo) language spoken in the South-East of Ghana. The language has an additive particle, *tsyɛ*, which associates with the immediately preceding constituent. When we look at examples such as (1), *tsyɛ* seems to be similar in meaning and use to the English additive particle *too*. However, it can also be used very differently. An example is (2), where using *too* in English would be infelicitous.

(1) ě́ɛ́-trɛ́ rrr lɛ́ ba-nùvɔ̀-wa *tsyɛ* be-sɛ́ bɛ́ɛ́-trɛ́
   C₁S.PROG-go ID then C₁P-child-DEF ADD C₁P-leave C₁P.PROG-go
   ‘He was going and the children, too, left and were going.’

(2) ka-drụ̀i-a kḕ-nēmi o-bídzɛ́ lɛ̀ àdzràmɔ-ɛ́ *tsyɛ* a-ɣɔ a-dɔ̀
   C₆S-dog-DEF C₆S.PROG-bite C₁S-girl and cat-DEF ADD C₁S-jump C₁S-land
   nǐ́ ọ̀-nyime su=i
   LOC C₁S-man side=CM
   ‘The dog is biting the girl and the cat (*tsyɛ*) jumped onto the man’s side.’

Examples such as (2) contradict two common assumptions about additive particles: firstly, that they associate with the focused part of the sentence (König, 1991) and secondly, that there has to be a “unique contrasting constituent in each conjunct” (Kaplan, 1984, 511), i.e. that everything else apart from the contrasted constituents has to be identical. In example (2), the additive particle associates with the topic of the sentence, not the focused part, and there is more than one contrasting constituent: the dog versus the cat, and biting the girl versus jumping onto the man’s side. It thus looks more like a contrastive topic or topic shift construction than like a typical example of additivity.

So how can we characterize the particle *tsyɛ* then? Is it a marker of topic shift rather than additivity, and does it impose constraints on the predicate? A possible analysis is that rather than requiring two predicates to be identical, *tsyɛ* requires them to be similar. This could work for example (2), as biting a girl and jumping on a man can be seen as similar. There might be other factors influencing the use of *tsyɛ*. One is to what extent the two events are perceived as belonging to a set. Another is whether the two events occur in sequence or simultaneously. Finally, if *tsyɛ* only marks topic shift, we do not expect it to associate with a referent which was also the topic of the previous sentence.

To look into these questions, we designed a production experiment. We developed video stimuli consisting of video clips in which two actors carry out parallel events. These two events vary on the dimension of similarity, from identical to completely unrelated or opposite. Each of these event pairs was recorded four times in different temporal presentations, once with the events happening sequentially (both actors are present all the time), once with the events happening simultaneously, once with the events happening in different video clips (with only one actor in each) and once with an intervening event between the two main events. Speakers of Avatime were asked
to describe what they saw in these videos. The descriptions were coded for whether or not an additive particle was used in the description of the second event.

Results show that additive particles can be used with all different event pairs. There are no contexts in which the particle is obligatory or in which the particle cannot be used. However, there are clear differences between the event pairs in how frequently the particle was used in their description. It was used most frequently to describe identical event pairs, such as a man and a woman who both write a letter or both sit down. Interestingly, identical events in which different objects were acted upon (e.g. cutting an onion versus cutting a tomato) were described using additive particles just as often as completely identical events, suggesting that differences between objects were less salient than differences between actions. Most non-identical event pairs had equally low numbers of additive particle use, except eating rice versus drinking water which was more frequently described using tsye, possibly because of higher similarity.

One thing that these results clearly indicate is that context is not deterministic in selecting the particle. Rather, whether or not the particle is used depends on how the speaker views the event. It also seems that the particle tsye does have an additive component to its semantics, as it occurs more frequently with identical predicates. As it also occurs with non-identical predicates, our hypothesis that the particle indicates similarity of two predicates, rather than identity, might work. However, we do need to account for the use of tsye in examples like (3) in which it is used even though the two events are completely different. Two analyses are possible to account for these examples: either the particle is polysemous and can be used to express both additivity and topic shift or the particle always indicates similarity and some of the participants simply viewed event pairs such as the one in (3) as similar.

(3) ó-nyime a-bite kɔl-yɛ ó-nyime tole tsye a-gba dɔmɛ
C₁S-man C₁S-make call-DEF C₁S-man one ADD C₁S-sweep thing
‘A man made a phone call, one man (tsye) swept (the floor).’

Another result is that the particle tsye is not necessarily used with a shifted topic: it can also be used if the two events are not adjacent and the actor of the second event was the topic of the previous sentence. However, the particle is used much less frequently in these cases. Additive particles were also used less frequently in the description of two events which took place in separate video clips. These results support the hypothesis that for the particles to be used, the two events have to be viewed as a set. Participants are less likely to see the two events as belonging to a salient set if they take place in different videos or if another event intervenes.

All in all, systematic elicitation using video stimuli has been shown to be helpful in uncovering the semantics and pragmatics of the additive particle tsye. Similarity of two predicates seems to play an important role in the use of the particle and even though it may look like a marker of topic shift, it does not necessarily associate with shifted topics. What seems to be more important is perception of the two described events as a set.

References


The effect of informational context and dual-task processing load on the realization of scalar inferences

**Scalar inference** refers to the interpretation of a less informative term (e.g., *SOME*) as meaning the negation of a more informative term (e.g., to mean *NOT ALL*). Various accounts exist to explain how comprehenders realize the interpretation of *SOME* as *NOT ALL* during online comprehension (Katsos & Cummins, 2010). Default accounts assume that the inference-based interpretation (*NOT ALL*) is realized automatically and without effort, whereas context-driven accounts assume that the process is effortful and only occurs in certain contexts. Many experiments have investigated the speed of scalar inferencing, but few have directly tested for processing costs incurred during the online realization of scalar inferences.

We report two experiments using a self-paced reading paradigm based on Breheny, Katsos, & Williams (2006). That study compared reading times to quantifiers in contexts where a scalar inference is realized to those in contexts where it is not. While the results suggested that the realization of scalar inferences involved a processing cost, the stimuli differed in more ways than just the bias towards or against scalar inference—e.g., one condition involved a repeated noun penalty in the critical region (see Huang & Snedeker, 2009). Thus, we use materials (see (1)) that are maximally similar across conditions, differing only in the presence of the quantifier *all* or *any* in the context sentence. *All* creates an upper-bounded context, which biases *some* to be interpreted as meaning *not all* (Katsos & Cummins, 2010), whereas *any* creates a lower-bounded context with no such bias. Control conditions involved the same manipulation but used *only some of* rather than *some of* as the quantifier, ensuring that the *not all* interpretation in these conditions is based on semantics rather than on pragmatic inference.

1) Mary was preparing to throw a party for John's relatives. She asked John whether *(all of them / any of them)* were staying in his apartment. John said that *(only some of them / some of them)* were. He added that the rest would be staying in a hotel.

Under context-driven accounts, slower reading times at or after "some of them" in the upper-bounded context *(all)* where the effortful inference is realized, compared to the lower-bounded context *(any)*. "the rest" should also be read faster in the upper-bounded context, because the interpretation of *some as not all* will make the reader aware of a remaining subset of referents, making "the rest" easier to integrate into the discourse model.

In the first experiment (N=28 participants; reading times for critical regions in *some of* sentences are shown in the figure below), linear mixed models showed that "the rest" was read faster after the quantifier *some of* in upper-bounded contexts compared to lower-bounded contexts—suggesting that the inference was realized in upper-bounded but not lower-bounded contexts—as predicted by context-driven models. This effect was specific to sentences with *some of* and was not observed in control conditions (sentences with *only some of*), which as expected were unaffected by context. Contrary to the predictions of context-driven accounts, however, context did not influence the reading times for "some of" or regions following it, providing no evidence that the realization of the inference was computationally costly.

The second experiment tested for processing costs in inference realization by having participants listen to irrelevant background speech (Martin et al., 1988) while reading passages such as those in (1). If the realization of scalar inferences requires processing resources, the additional processing load due to the background speech should inhibit scalar inferencing (De Neys & Schaeken, 2007) and thus attenuate the facilitation effect that was
observed on "the rest" in the first experiment. Preliminary results (N=26 participants who did not participate in the first experiment; reading times for critical regions in some of sentences are shown in the figure below) show that in this experiment no significant effect of context was obtained at "the rest" or following regions, suggesting that inferencing either did not occur in this experiment or that the inference was not utilized to help lexical access or integration for "the rest". This suggests that the presence of an additional processing load inhibited scalar inferencing. No difference emerged between sentences with relatively easy background speech (spoken lists of pseudowords) and relatively hard background speech (spoken lists of English words).

These results of these two studies suggest that scalar inferences are context-dependent but do not yield significant processing costs at the quantifier in this paradigm. Indirect evidence that that inferencing requires processing resources was found—insofar as inferences did not seem to be computed when processing resources were unavailable—but processing costs were not observed directly when inferences were actually computed. These results may be consistent with a constraint-based model (Degen & Tanenhaus, 2011) which assumes inferences are sensitive to context and processing costs, but when sufficiently facilitated can be realized cost-free. On the other hand, the results could also be accounted for under default models (if such a model were to postulate that the inference was automatically realized and then cancelled in the lower-bound context in the first experiment, but was not cancelled in the second experiment because processing resources were unavailable) or under context-driven models (if such models were to postulate that the kinds of processing costs invoked by inferencing are not detectable in self-paced reading times). The findings thus motivate further tests for possible processing costs associated with inferencing using alternate methods, and suggest potential refinements to the psycholinguistic models of scalar inferencing.

References
Breheny, Katsos, & Williams, 2006. *Journal of Memory and Language*, 100, 434-463.
A cross-linguistic study of direct denial as a diagnostic for meaning types

**Introduction:** One of the diagnostics that has been used to identify different types of meanings and, ultimately, to test the distinction between so-called at-issue and projective content (Simons et al. 2010) is direct denial, here defined as the ability to say No,... in response to a statement. The claim in the literature is that only at-issue content can be the target of direct denial (e.g. Bonami & Godard 2008, Jayez & Rossari 2004, Faller 2002, Papafragou 2006), while projective content (which includes presuppositions and conventional implicatures) can only be indirectly denied (e.g. von Fintel’s 2004 Hey, wait a minute test for presupposition, following Shanon 1976). However, the theoretical literature does not make a case for possible variations within the subtypes of projective content (with the exception of non-restrictive relative clauses, cf. Koev 2012), and different kinds of denials (i.e., the contrast between No and No, that's not true) is not taken into consideration. In this contribution, we report the results of an experiment testing the direct deniability of all of these meaning types as well as some subtypes thereof in English, Catalan and Spanish. We show that the type of direct denial interacts with meaning type and that the heterogeneity of categories such as conventional implicatures leads to heterogeneity in felicity ratings. We also discuss some unexpected cross-linguistic differences.

**Materials and methods:** Native-speakers of each of the 3 languages (88 for Catalan, 55 for English, and 32 for Spanish) participated in an on-line-based experiment where they were asked to listen to 88 two-turn dialogues (majority fillers) across 4 conditions depending on the type of direct denial, as shown in (1) (where x is the meaning of interest and y, an alternative to x). In the first turn, a statement was made that crucially contained one of 5 meaning types or subtypes in (2). These included three types of conventional implicature (CI); following Potts (2005), we included non-restrictive relative clauses and expressives, but we made an additional distinction within expressives, namely between emotive ones like fucking and descriptive ones like idiot. As for presuppositions, we included both lexical and structural ones. In the second turn, for the critical stimuli, participants heard one of the four direct denials. The experiment was a Likert judgment task, where participants needed to rate how strange would it be – if strange at all – to overhear someone utter that specific reply. Table 1 shows an example of each meaning type (MT) with denial [No], (1a).

(1) Kind of denial                          (2) Meaning Type (MT)
    a. No, not x. [No]                       a. Baseline (assertion)
    b. No, that’s not true. Not x. [NTNT]    b. CI-non-restrictive relative clause
    c. No, y. [No-alternative]               c. CI-emotive
    d. No, that’s not true. Y. [NTNT-alternative] d. CI-descriptive

(3) MT Example: Turn1 (statement)         Example: Turn2 (response)
    (2a) Daniel is at the library            No, he is not at the library
    (2b) Eric, who is from NYC, battled cancer No, he’s not from NYC
    (2c) Lisa found her fucking keys          No, you’re not irritated
    (2d) I told Bob he was cute and the idiot bought me dinner No, Bob is not an idiot
    (2e) Sylvia is late again                 No, she has never been late
    (2f) It wasn’t my son who broke the vase  No, nobody broke the vase

**English:** The distinction between assertions and projective meanings in terms of deniability is significant (p < .001). We show that assertions prefer denials without alternatives, while
all of the marginal meaning types are much better with No responses than NTNT responses. While lexical presuppositions were found acceptable, clefts were judged infelicitous. Non-restrictive relative clauses couldn’t be felicitously denied, especially for the NTNT replies. Finally, a striking distinction surfaces in the case of expressives. While emotives are the worst, descriptives are judged felicitous with No-answers.

Comparison with Spanish and Catalan: The distinction between assertions and projective meanings is also significant in Spanish and Catalan. There are, however, some interesting contrasts between Spanish and Catalan, on the one hand, and English, on the other hand. Both in Catalan and Spanish, denying a non-restrictive relative clause is judged much better than in English. For Spanish, the distinction between assertions and relative clauses is hardly significant (p = 0.04), and clefts are judged felicitous with No-answers. As in English and Catalan, NTNT replies are judged worse. Finally, even though Spanish and Catalan are very closely related, our results show some contrasts for the descriptive expressives: in Catalan they receive higher ratings (similarly to English), while the ratings are much lower for Spanish.

Conclusion: This set of novel data might prove to be relevant for any theories of dialogue which aim to propose a model that accounts for how information is stored, retrieved and structured, and how shared knowledge is updated in the course of conversation. A more specific outcome of this experimental work concerns the applicability of denial as a diagnostic. Especially, it may help us interpret direct denial properly in semantic fieldwork (Matthewson 2004). We have shown that, at least for English, NTNT-alternative fares better than the others as a test for at-issueness. Lastly, we also point out that the expression of direct denial and its effect on the targeted meaning is a locus of cross-linguistic variation.

The role of educational background in generating scalar implicatures

Theory of scalar implicatures (SI) has been subject to debate between pragmatic and semantic approaches. According to the former, SIs are considered to be a component of speaker meaning which is not linguistically encoded, but arises out of conversational norms (Horn 2006). Whether the scalar terms are interpreted with a SI or not depends on contextually induced expectations of informativeness, which suggests that SIs are not part of the semantic component itself (Katsos 2008). According to the semantic approach, SIs are considered incremental to the semantic component of grammar, and are generated automatically (Chierchia 2004).

Previous behavioral research in the domain of SIs has focused on the developmental perspective, indicating that children are not as capable of generating them as adults are, given that adult participants derive SIs at a rate of more than 80%, while children remain at chance level or lower (Noveck 2001, Papafragou & Musolino 2003, among others). However, several studies report on lower percentage of SIs in adult speakers of certain languages, Serbian being one of them: according to the preliminary results of the COST Action A33 project (Katsos, Andelković, Savić, Jošić 2009, Katsos 2011), only 54% of Serbian adult responses were implicature-based rejections in interpreting sentences with the quantifier neki ‘some’. Departing from here, we investigated different linguistic and extralinguistic factors affecting the availability of SIs. This study reports on the role of social factors, in particular educational background.

Previous studies show that children do derive SIs if properly trained (to detect the pragmatic anomaly, as a trigger to go beyond the explicit meaning of a sentence) (Guasti et al 2005). We hypothesized that a reverse process, imposed by certain educational contents, has an effect on adult speakers’ judgments (training in logic strengthens the logical entailment component of Serbian sentences involving neki ‘some’, and leads to its purely cardinal interpretation).

Twenty-two students (mean age: 20), native speakers of Serbian participated in the study. Based on the their field of studies, they were divided into Group 1 (Math, CompSci, Physics and Electric Engineering) and Group 2 ((Bio)chemistry, Pharmacy, Medicine and Biotechnological Engineering). Main criterion was the type of training they received: Group 1 attended an intensive course in set theory and first-order logic, while Group 2 had no comparable course. We took this formal training as an instantiation of the educational parameter in our hypothesis above, expected to yield an effect of strengthening of the cardinal interpretation of neki ‘some’.

In a modified version of the Truth-value judgment task, participants were shown a set of visual stimuli (e.g. five books in a shelf), preceded by an introductory sentence (e.g. *We have bought 5 books.*) and followed by an audio-recorded sentence containing the quantifier *some* (e.g. *Some books are in the shelf*). They were asked to evaluate whether the utterance was right or wrong in the visually presented situation. Participants received 8 critical trials, 8 control trials and 4 filler sentences administered in a pseudo-random order. In the critical trials, where the use of the quantifier *svi* ‘all’ was a more informative alternative to neki ‘some’ for the given situations, the participants that do generate the SI were expected to reject the sentence as pragmatically inappropriate, and those who do not to accept it as proper.

ANOVA analysis of the proportions of ‘No’ responses revealed a main effect of type of education (*F*=9.697, df=1, *p*=0.005). None of the Group 1 participants generated SI, whereas in Group 2 57% of participants did so. These results conform to the previous results for Serbian, regarding the small percentage of SIs in adult speakers in comparison to other languages. SIs arising from the interpretation of neki ‘some’ are not generated by default or automatically.
These results alone do not disseminate between the semantic and pragmatic approaches to SI, as the former can argue that the effect of inhibition is pragmatic in nature, imposed over the meaning that comes with a SI. They do, however, set ground for our next step: an investigation into the processing complexity of the two patterns, which should tell us whether the generation of SIs or its inhibition is the process that requires additional, pragmatic, measures to be taken.

The very fact that generating SIs is subject to variation sheds a new light on this phenomenon. It should be taken into account when discussing children’s low performance, not just in requiring a more careful approach to the adult control groups, but also in explaining the asymmetries.

References:
Scalar implicatures can arise from an utterance like (1), standardly taken to have a semantic lower-bound interpretation (at least one student failed) and a Gricean upper-bound interpretation (not all of the students failed).

(1) Some of the students failed. Not all occurrences of some, however, so easily lend themselves to the upper-bound interpretation. Some can appear as an indefinite determiner with a singular head noun, for instance, as in (2), where it does not give rise to a scalar implicature.

(2) Some (* but not all) guy walked into an exam.

Between the relatively clear-cut cases of (1) and (2) lie many uses of some that are harder to categorize. For instance, the standard lower- and upper-bound interpretations given above do not extend straightforwardly to the use of some with non-count nouns, as in (3). In (4), a not-all interpretation is plausible, but it does not seem to be the central point.

(3) The courts generally show some leniency to first-time offenders.

(4) Our program will discuss some causes of grief.

The literature on scalar implicatures – both theoretical and experimental – has concentrated on cases like (1), with the (often implicit) assumption that scalar inferences are nearly obligatory when some is encountered (e.g., Huang & Snedeker, 2009). However, this Frequency Assumption is untested to date. Its validity has far-reaching consequences. Theoretically, it is the foundation of accounts that consider scalar implicatures defaults (e.g., Levinson, 2000). In the experimental literature, delays in implicature processing have been attributed to a staged process of pragmatic inference: first the semantics of the utterance is computed, and pragmatic enrichment follows in a second costly step. However, these results might instead reflect an inverse frequency effect within a constraint-based system (Degen & Tanenhaus, 2011). Robustness and speed of computing a scalar implicature might be a function of degree of probabilistic support from contextually available cues. We present two studies that combine corpus analyses with web-based rating studies as a first step in the methodologically challenging task of identifying and quantifying the frequency of, and the factors that affect, upper-bound interpretations of some in naturally occurring utterances.

**Study 1.** We extracted all 1389 utterances containing some from the Switchboard corpus that were not part of a disfluency, did not have an empty NP head noun (some say flying is fun), and were not specific indefinites (head noun of some-NP was singular count noun). For each utterance we collected 10 implicature ratings on Mechanical Turk. Participants saw the printed target utterance preceded by 10 lines of dialog context. They were then presented with a minimally different utterance where some was replaced by some, but not all, thus making explicit the content of the candidate scalar inference (e.g., Some of the students failed was paired with Some but not all of the students failed). Participants rated the similarity of the two utterances on a 7-point scale: 1 = “very different meaning”; 7 = “same meaning”. Each participant first saw two practice items that were clear examples of both ends of the scale. If the implicature is part of the originally conveyed meaning, similarity ratings for that utterance should be high. The Frequency Assumption predicts that high ratings should be more frequent than low ratings.

**Results.** The mean similarity rating was 3.9 out of 7. Only 44% of ratings were higher than the midpoint 4; 47% were lower than 4. Moreover, of the 34% of utterances rated either 1 or 7, only 44% were rated 7. Fitting mixtures of Gaussians to the mean by-item similarity ratings revealed that the data are optimally generated by three underlying components (see Figure 1): two with low variance at either end of the scale (means: 2.3, 5.5.) and one in the middle of the scale with high variance (mean: 3.8). If the end-point components are interpreted as reflecting
lower-bound and upper-bound interpretations, and the intermediate component reflects uncertainty, then only 28% of the data are contributed by upper-bound interpretations. Contrary to the standard assumption, then, most uses of *some* do not strongly support a scalar implicature.

**Study 2.** As an initial exploration of some of the factors that are likely to mediate an implicature, we used the same dataset to extract information about: a) whether *some* occurred in the partitive; b) the NP head’s discourse accessibility (discourse-old or new); c) the NP’s topicality as measured by whether it was topicalized/in subject position vs. in object/other positions; and d) a gradient measure of quantifier strength (strong, presuppositional, quantificational *some* vs. weak, non-presuppositional, cardinal *some*).

*Results.* A mixed effects linear regression (by-speaker random intercepts) predicted each item's mean similarity rating from fixed effects of partitive, discourse accessibility, topicality, and quantifier strength. Partitive NPs, discourse-old heads, topical NPs, and stronger uses of *some* each significantly increased similarity ratings. Partitive results are shown in Figure 2.

These results suggest that contrary to what is standardly assumed, *some* often does not provide strong support for an implicature. More generally, the work reported here is a first step towards an understanding of how implicature rates are modulated by linguistic factors in natural utterances, which we argue is necessary for designing and interpreting experimental studies of when and how implicatures are computed in on-line processing.

---

**References**


Implicatures in uncooperative contexts. Evidence from a visual world paradigm.

Scalar implicatures are a hot topic among linguists and there is much controversy as to how they are computed. Neo-Griceans (Geurts, 2009) argue that semantic and pragmatic meanings are processed in two hierarchically ordered stages, causing implicatures to be delayed in comparison to purely semantic quantifiers like all. Defaultists (Storto & Tanenhaus, 2004) claim that listeners arrive at the pragmatic meaning automatically and not later than e.g. for all. Relevance theorists are divided in their assessment; some (Breheny et al., 2006) claim implicatures can be computed immediately and effortlessly, while others (Huang, Y.T & J. Snedeker, 2009) consider it a time and resource consuming process. The present study addresses this debate in a visual world experiment which is partially a replication of Grodner et al. (2010), but also investigates whether implicatures are computed when a prerequisite for communication in Gricean terms – cooperativity – is lacking. Will implicatures be computed immediately (defaultists), will they be delayed (relevance theorists) or even absent (Neo-Griceans)?

The pretest aimed at confirming that the German quantifier *einige* (some) carries an implicature, as it does not fully equal *some*. 64 fillers and 16 items in 4 conditions were distributed over 4 lists in an online WebEx questionnaire. The conditions were pairs of pictures (7x7 matrix with 4 colors) and quantified sentences (e.g. “In my bag are some red pebbles.”): quantifier *all* was paired with a picture where all pebbles were of the relevant color and *none* with a corresponding none-picture; *some* was paired with two pictures, some and all (see Fig. 1-3.) A condition of false pairs was added for control. 32 German native speakers judged the pairs’ acceptability on a 7-point scale with the option to refute the pair as unacceptable. *Some+some* conditions judged better that *some+all* ($F_{1,29}=132.4$). False controls were rejected 92.4%, items <11% of the time.

Materials in the eye-tracking experiment were almost identical to the pretest. The factors were quantifier type (*some, all, none*), picture (*some, all, none*), and speaker (cooperative, uncooperative). 2 randomized experimental lists were counterbalanced for speaker gender. 112 fillers and 24 items in 6 conditions: quantifier $\times$ speaker paired with the three pictures. The targets for *all* and *none* for both speakers were the corresponding pictures, whereas for *some* the target in for the cooperative speaker was *some* and *all* for the uncooperative one. 22 German native speakers participated in the study. The experiment consisted of 4 parts. First, the participants were tested for color blindness using 24 Ishihara plates. Secondly, they played a card game with a friendly and opposing confederate where they were forced to make and respond to utterances resembling the items. This phase ensured the plausibility of *all* and *none* and introduced speaker types. In the following eye-tracking part, they were first asked to view an image while listening to an utterance; the task was to decide whether the speaker was the cooperative or uncooperative player. Lastly, in the main experiment, preceded by two dummy trials, the participants were asked to view 3 randomly ordered pictures in a
visual world and listen to an utterance made by one of the speakers. They were then to choose the picture that best fits the utterance.

The all-picture for *some* quantifier was chosen significantly more often when the speaker was uncooperative (z=0.0003). A GLMM analysis revealed that for the cooperative speaker, the pragmatic target of *some* was fixated as early as the target of *all*. In uncooperative cases there was a substantial temporal delay between *some* and *all* quantifiers: picture (some, all) × quantifier (some, all) × bins (5-10, 11-20) × cooperative (yes, no), z=3.4.

To sum up, in standard, cooperative contexts implicature computation was fast and automatic. However, in uncooperative cases the implicature was computed later due to the cancellation and re-computation processes. The results replicate Grodner et al. (2010); they are consistent with both the defaultist and the relevance theoretic view Breheny et al. (2006), however go against Neo-Gricean predictions.

Figure 1. Matrix image for quantifier *all*.

Figure 2. Matrix image for quantifier *some*.

Figure 3. Matrix image for quantifier *none*.

**References**


How the Brain Responds to Partial Answers: An ERP Study

During conversation, speakers and listeners act on certain basic assumptions enabling them to communicate swiftly and seemingly effortlessly. When someone poses a question, for instance, the addressee is expected to give a full answer, and say no more but certainly no less than is asked for. However, people do not always behave as ideal conversational partners, and they often give answers that are not complete. In two experiments, we investigate how the brain responds to such incomplete answers.

Participants were presented with written mini-dialogues (word-by-word) while their ERPs (Event-Related brain Potentials) were being measured. The answers were always semantically and syntactically sound, but differed in pragmatic felicity with respect to the preceding question (critical word underlined). Participants were asked to read for comprehension, and answered a comprehension question about one in five sentences (always a filler). The experimental materials were in Dutch, which—unlike English—does not allow for a verb after the final adverb in Experiment 1.

<table>
<thead>
<tr>
<th>Experiment 1</th>
<th>Experiment 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Neutral condition</strong></td>
<td><strong>Neutral condition</strong></td>
</tr>
<tr>
<td>A: John kissed Mary and Jim heartily.</td>
<td>A: John kissed Mary.</td>
</tr>
<tr>
<td><strong>Violation condition</strong></td>
<td><strong>Violation condition</strong></td>
</tr>
<tr>
<td>A: John kissed Mary and Jim heartily.</td>
<td>A: John kissed Mary.</td>
</tr>
</tbody>
</table>

In both experiments, a violation of conversational expectations led to *positivities*. In experiment 2, the positivity started quite early (350 ms post-onset), had a broad distribution, and became more frontal as time elapsed (from 600 ms and further). In Experiment 1, the early part of the positivity was only present at occipital electrodes, and the later part of the effect was predominantly left-lateralized. Moreover, Experiment 1 also showed an early *negativity* (150-350 ms post-onset), which did not appear in Experiment 2.

Unlike in real life conversations, our readers knew that they would get only one sentence as an answer to the question. This means that at the end of those sentences, they knew the answer was not complete. How did participants deal with these incomplete answers? In Experiment 1, readers likely expected a verb to follow “Jim”, but instead got an adverb (“heartily”). This category violation may have caused the early negativity, which we interpret as an ELAN. In later time-windows, we found evidence for an early positivity (from 350 ms post-onset onwards) that evolved into a left-lateralized (Exp. 1) or a strongly frontal (Exp. 2) positivity. Despite their early onsets, and different scalp distributions, we would like to interpret these positivities as members of the P600-family, reflecting different subprocesses involved in the assembly of a discourse representation (e.g., Burkhardt, 2007). One of these processes may involve the computation of the pragmatic implication that Jim did not do anything, for instance because he was unfriendly or uninterested. We are working on a follow-up study with spoken dialogues.

Early speech act comprehension in spoken dialogue: Evidence from ERPs

Like other aspects of social behaviour, verbal communication involves actions. Everyday conversation contains a rich structure of speech acts such as offering, teasing and complimenting (Austin, 1976; Searle, 1969). These speech acts allow us to impact our world, using language. For communication to be successful, listeners have to recognize these speech acts – also called actions (Schegloff, 1996) – and plan an appropriate response. Bridging the literal meaning of utterances and the relevant speech act is far from straightforward, as the form and content of utterances is frequently underspecified for this level of meaning (Levinson, 2013). In a recent self-paced reading study it was demonstrated that sequential context is a powerful cue in speech act comprehension, as listeners can identify speech acts in action-underspecified utterances with very high accuracy based only on the prior turn (Gisladottir, Chwilla, Schriefers, and Levinson, 2012).

To date, very little is known about the time course of speech act processing in spoken conversation. To investigate speech act comprehension in real-time we performed an Event-Related Potential (ERP) study on spoken dialogues. Can listeners get the action early in the utterance, sidestepping the full propositional content, or is speech act comprehension a later, more controlled process?

The materials consisted of spoken dialogues in which critical utterances (e.g., „I have a credit card“) delivered the speech acts of Answer, Pre-offer and Declination depending on the prior turn. When responding to a question („How are you going to pay for the ticket?“), the assertion „I have a credit card“ functions as an Answer. When following an offer („I can lend you money for the ticket“) the critical utterance delivers a Declination. When responding to a complaint („I don’t have any money to pay for the ticket“), the utterance „I have a credit card“ functions as a prelude to an offer, called a Pre-Offer in Conversation Analysis (Schegloff, 1988; Schegloff, 2007). The Answer dialogues serve as the control condition since the conceptual distance between sentence meaning and the action intended is the smallest. In Declinations and Pre-offers an assertion is used as a vehicle for a more specific action (Schegloff, 2007). However, while the context heavily constrains the critical utterance in the Declinations, due to an adjacency pair structure (Schegloff, 2007), the Pre-offers project a continuation of the sequence (with a more direct offer), forming a pre-sequence (Schegloff, 1988; Schegloff, 2007). By comparing these three actions we can investigate a broader range of speech acts, while highlighting the role of sequential context.

Participants were asked to listen to each dialogue and categorize the critical utterance as doing answering, offering or declining. After the recording session subjects filled out the
Empathy Quotient (EQ) (Baron-Cohen and Wheelwright, 2004), allowing us to assess whether cognitive (theory of mind) and affective empathy influences speech act comprehension.

If speech act comprehension is an early, more automatic process there should be ERP differences between the conditions at the first word or the verb. On the other hand, if speech act comprehension involves later, more controlled processes the speech acts should differ at the final word. The main findings were as follows: Early ERP effects were observed to both Declinations and Pre-offers, which elicited a sustained positivity (relative to Answers) at right fronto-temporal sites starting between 400 to 600 ms after first word onset. Declinations and Pre-offers also elicited differential ERP effects. In Declinations a frontal positivity emerged at the midline and over the left hemisphere from 400 ms after the onset of the first word. This positivity was positively correlated with the EQ during an early time window, indicating that social-emotional processing is involved in speech act comprehension.

At the final word no ERP effects occurred in Declinations. We take this to show that in highly constraining contexts listeners can get the speech act before hearing the final word. In contrast, the Pre-offers elicited a late posterior negativity at the last word. This late effect may reflect the projective inference required in pre-sequences.

This is the first ERP study to directly explore speech act comprehension in spoken dialogues. The findings show that listeners tune in to the speech act meaning of utterances immediately, that is already at the first word. By going beyond the distinction between “direct” and “indirect” speech acts we have demonstrated that utterance interpretation is sensitive to specific action meanings. Speech act comprehension takes into account the nuances of the speech act exchange, including sequential context and social-emotional factors.

When Children are more Pragmatic than Adults

Introduction It is well established that children behave less pragmatic than adults with quantity implicatures in comprehension (Noveck 2001, Bott and Noveck 2004, Pouscoulos et al. 2007, Huang and Snedeker 2009, 2011, Katsos et al. 2011, among others). In this paper, we show that in a different domain, children are more pragmatic than adults. Specifically, we examine children’s production of passive sentences in two different discourse contexts: one where it is licensed to leave out the by-phrase, and another where it is not. Our experiment shows that five year old children are more sensitive to the pragmatic principle Be Brief than adults are and that children also show sensitivity to not be underinformative. Our results therefore suggest that children distinguish between pragmatic inference based on quantity and those based on speaker effort. Our result leads to several new directions to pursue to understand children’s pragmatic abilities.

This study We investigate children’s and adults’ production of full and short passives in German. Two prototypical examples are in (1). The German passive is constructed with the auxiliary verb werden and past participle. In the full passive (1a), the agent phrase appears as von-phrase (by-phrase), but this is optional and the short passive (1b) is fully grammatical.

(1) a. Das Mädchen wird von dem Papa geküsst.
   the.Neut girl become.3rd.sing by the.Masc.Dat Papa kiss.part
   ‘the girl is kissed by the father.’ Full Passive (FP)

   b. Das Mädchen wird geküsst.
      the.Neut girl become.3rd.sing kiss.part
      ‘the girl is kissed.’ Short Passive (SP)

We assume two pragmatic principles govern the choice between (1a) and (1b): One principle that prefers the more informative full passive (1a) and one that prefers the shorter, more economical (1b). Many different concrete theoretical proposals exist, e.g. Horn’s Q and R maxims or Relevance theory. For concreteness, we adopt principles of Grice (1989): Quantity I favors the more informative (2a) over (2b) in contexts where the agent is not already clear. And Be Brief! favors (2b) over (2a), since (2b) is shorter.

We consider two different discourse contexts. Both can be characterized by a question under discussion, namely the two questions in (2). In (2a), two potential agents are under discussion, while action kiss is given. In (2b), however, the agent is given, while two potential actions are under discussion.

(2) a. Did the father or the grandfather kiss the girl? Context 2 Agents
    b. Did the father kiss or hug the girl? Context 2 Verbs

Prediction We expect adults to produce the full passive in the 2 Agents context, but not in the 2 Verbs context where full and short passive don’t differ in their informativity. For children, our null hypothesis is that they should not differ from adults. In addition, we consider the hypotheses that children are not sensitive to Quantity at all. Then they should produce only short passives.

Method We conducted an elicitation experiment of relative clauses based on the method [novogrodsky06a]: The experimenter told stories about two girls (or boys for male subjects). At the end of each story, children had to tell the experimenter which girl/boy they’d rather be.
Two stories created either the 2 Agents or the 2 Verbs contexts. An example for the 2 Agents context would be: *The father kisses one girl, and the grandpa kisses one girl. Which girl would you rather be? Start with, ‘I’d rather be the girl . . . ’*. For the 2 Verbs context, the example would instead start with *The father hugs one girl and the father kisses one girl*. We used 5 contexts of each type. The three kinds of responses targeted by our design were the following:

(3) a. the girl that the mother kissed (object relative clause)
   b. the girl that was kissed by the mother (long passive)
   c. the girl that was kissed. (short passive)

Because object relative clauses are difficult for 5-year-old children, we expected that at least the children tested would produce passive sentence with high frequency (cf. Belletti 2011).

**Subjects** 20 monolingual German speaking children (5;1–5;11, mean = 5;6, 9 girls, 11 boys) living in Berlin participated this study and 20 adults. The children were tested at the day care center they attended.

**Results and Discussion:** Among the children’s 200 responses, 51 of the responses used a grammatical passive form, as did 121 of the adult’s 200 response. The graphs in (4) show for children (left) and adults (right) whether a FP (blue bars) or a an SP (pink bars) was used. As predicted, the adults produce SPs only in the 2 Verbs context (Fisher’s exact test (FET): $p < 10^{-7}$). The children also show sensitivity to informativity (FET: $p < 10^{-4}$), concurring with (Katsos et al. 2011). Thirdly, there is a significant difference between children and adults (chi-sq with pooling, $p = .021$), which refutes the null hypothesis.

(4)

The nature of the child-adult difference is, in our view, the most interesting finding: As columns 2 and 4 in the diagrams show, children show a significantly stronger preference for the SP in the 2 Verbs condition than adults do (FET: $p = .026$). Since in the 2 Verbs condition only the Maxim of Brevity is at issue, our results indicate that children show higher sensitivity to this maxim than adults. In sum, we see lesser sensitivity to quantity in children, but greater sensitivity to the brevity maxim. Further work is needed to establish whether the main difference between our results and those in the literature cited above is quantity vs. brevity or comprehension vs. production. Therefore, our results suggest two directions for further research: on the one hand, to consider the effects of brevity in comprehension, on the other hand, to test for the effects of quantity in production.
Testing for domain widening: Experimental evidence for obligatory exhaustification

**Introduction:** The domain-widening status of the indefinite *any* has been called into question in the theoretical literature, on the basis of debatable judgments. In this paper, we provide novel empirical evidence of *any*’s pragmatic contribution: given appropriate contexts, *any* is systematically interpreted as quantifying over wider domains than plain indefinites.

**Theoretical Background:** Polarity-sensitive *any* is licensed by downward-entailing (DE) operators such as negation (Ladusaw, 1979, a.o.) (1,2). One influential line of work explains the restricted distribution of *any* as follows: *any* must widen the domain of quantification and yield a stronger assertion; this succeeds only in DE environments (Kadmon & Landman, 1993; Chierchia, 2006). In (3), B’s final assertion widens the domain to include point-and-shoot cameras, resulting in a stronger assertion (as negatively quantifying over a larger domain yields a stronger assertion than the same quantification over a narrower domain).

(1) John doesn’t have any cameras. (2) *John has any cameras.

(3) A: Do you have a camera?
   B: No.
   A: Nothing too fancy, maybe a point-and-shoot one?
   B: I don’t have ANY cameras. [point-and-shoot or otherwise!]

Chierchia provides a compositional implementation of the domain widening hypothesis (DWH). *Any* is truth-conditionally equivalent to plain indefinites (like *some* or *a*) (4), but activates a set of alternatives corresponding to the indefinite associated with more restricted domains of quantification available in the context (5). Crucially, *any* triggers obligatory exhaustification of these alternatives via a pragmatic enrichment operator (6), such that the activated alternatives must be weaker than the one asserted.

(4) \[\|\text{any}\| = \|\text{some}\| = \lambda P . Q . \exists x \in D [P_w(x) & Q_w(x)]\]

(5) \[\|\text{any}\|^{D-\text{ALT}} = \|\text{some}\|^{D-\text{ALT}} = \{\text{some}_D : D' \subseteq D\}\]

(6) \[O_c(p) = p \& \forall q \in C [q \rightarrow p \subseteq C_q]\], where C=ALT (Chierchia, 2006; in press)

Chierchia shows that in non-DE environments (2), exhaustification yields a logical contradiction, since none of the alternatives are entailed by the assertion and must be eliminated; failure to exhaustify results in ungrammaticality. In a DE environment (1), the relevant alternatives are all entailed, so exhaustification does not yield a logical contradiction.

**The Problem:** Despite the theoretical appeal of the DWH, some have objected that when we go beyond the clean minimal pair examples in the literature (e.g., *I don’t have a camera* vs. *I don’t have any cameras*), judgments are not as clearcut as expected. Plain indefinites may not always restrict and *any* may not always widen as expected (Arregui, 2008). What is crucial under an account like Chierchia’s is that *any* must quantify over “larger” domains than plain indefinites. But we cannot begin to systematically compare domain sizes without controlling for what sets of individuals count as relevant. This is where careful semantic/pragmatic experimentation that can systematically control for possible subdomain alternatives is crucial.

**Experiment:** Using the truth value judgment task paradigm (Crain & Thornton, 2000), we had participants listen to a series of short stories told through animated pictures on a laptop computer. We set up contexts in which we could carefully define the domain of quantification, including its subdomain alternatives. For example, the domain might consist of a set of 9 diamonds which varied along the contextual dimension of colour (3 red diamonds, 3 green diamonds, 3 blue diamonds). The largest domain of quantification in this context is that containing all 9 diamonds (3 reds, 3 greens, 3 blues). Possible subdomain alternatives include: {reds}, {reds, greens}, {blues, reds}, {greens}, etc. In each such story, two characters have to
find the various shapes in order to finish a puzzle; at a critical juncture of the story, it is established that though they have been successful in finding the first two subsets, they cannot finish their puzzles because they cannot find the final subset. Thus what is at issue revolves around just one particular subdomain alternative (e.g., the blue diamonds); at this point, a puppet appears on the screen to answer a question about the story (Why can’t they finish their puzzles?) with a negatively quantified statement: “They can’t find any diamonds!”/“They can’t find a diamond!”/“They can’t find diamonds!” Participants were asked to decide if the puppet’s statement accurately described the sequence of events (Did he say the right thing?). If participants restricted the domain of the relevant indefinite to a smaller subdomain alternative (e.g., the blue diamonds), they were expected to accept the statement; if the indefinite had to quantify over a larger domain, they were expected to reject the statement. 145 adult native speakers of English were randomly assigned to one of 3 test conditions (any, a, bare plural) or 3 control conditions. Each participant received 2 training items followed by 4 test and 4 filler items, which were randomized and counterbalanced.

**RESULTS:** The results from the 72 adults tested on the critical test conditions are provided in Figure 1. A one-way anova revealed a significant main effect of indefinite type, $F(2,69)=54.8$, $p<.001$; adults were significantly more accepting of the a and bare plural statements than the any-statements (Tukey HSD; both $p<.001$). We also elicited follow-up justifications from the participants to ascertain their reasons for accepting/rejecting the statements. Justifications for accepting the plain indefinite statements made reference to the relevant restricted subdomain alternative (e.g., Yes, they couldn’t find the blue diamonds). In contrast, justifications for rejecting any-statements made reference to other subdomain alternatives (e.g., red or green diamonds), suggesting these could not count as exceptions to the domain of quantification.

**CONCLUSION:** We see that once we carefully control for the possible alternatives that any can trigger, any must quantify over the widest domain available in the context; once a possible subdomain alternative has been introduced, it cannot count as an exception to the domain - hence any’s observed reduced tolerance of exceptions (Kadmon & Landman, 1993). In contrast, there is no such restriction on the interpretation of plain indefinites, which are free to restrict to salient subdomains. Our study thus provides novel empirical evidence for the DWH; more importantly, it provides the first experimental corroboration of its kind for an explanation of why any (as a domain widener) ought to be licensed in DE environments.

**Fig. 1:** Acceptance of negative statements (n=72)

<table>
<thead>
<tr>
<th>Test condition</th>
<th>Acceptance (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>any</td>
<td>0</td>
</tr>
<tr>
<td>a</td>
<td>84</td>
</tr>
<tr>
<td>bp</td>
<td>72</td>
</tr>
</tbody>
</table>

Testing the scopal properties of three different indefinite articles in German

We aim to show that not only systems of indefinite pronouns may be quite complex, but that article systems can also express quite subtle differences with respect to their referential and scopal properties. One case in point is German which exhibits an interesting article system with three indefinite articles: (i) the unmarked indefinite article *ein* ‘a(n)’, (ii) the indefinite demonstrative *dieser* ‘this’ (similar in function to English indefinite *this* (Prince 1981, Ionin 2006) and (iii) and *son* (< so-*ein* ‘such-a’ (Hole & Klumpp 2000, Lenerz & Lohnstein 2005).

The German article system, roughly, exhibits a two-fold (in)definiteness contrast between the definite article *der/die/das* (with the respective plural forms) and the unmarked indefinite article *ein/eine/ein* (no form for plural). However, there is also a weak definite form (Schwarz 2008), and the two indefinite articles *dieser* and *son*, both unaccented and used primarily in informal registers. The three indefinite articles are licensed in existential contexts, one classical test (amongst others) for their indefiniteness.

(1)  Es gibt da einen / diesen / sonen / *den / *jeden Typen in meinem Englischkurs…
‘There is a / this / such-a / *the / *every guy in my English course…’

The three expressions can be assigned article status in German, since they are situated at the left edge of the nominal or determiner phrase [DP Art [(Adj) N (RC) ]]. Thus, other modifiers or determiners of the noun phrase cannot precede them (Hole & Klumpp 2000, Lenerz & Lohnstein 2005):

(2)  *ein / dieser / son schöne(r) Mann
‘a / this / such-a beautiful man’
(3)  *schön *ein / dieser / son Mann
‘beautiful a / this / such-a man’

According to this semantic and syntactic behavior, indefinite *dieser* and *son* can be categorized as indefinite articles (or determiners): they occupy the same position as the indefinite article *ein* and show the respective semantics, introducing new discourse referents.

Indefinite *dieser* and *son* appear to exhibit different referential properties compared to the unmarked indefinite article *ein*, with respect to specificity, scopal behavior, scope escaping properties and the possibility to allow for intermediate and functional readings. In the following we focus on the scopal behavior in intensional contexts only. We suppose that the unmarked indefinite article *ein* in (4) allows for a wide scope or narrow scope reading, while both indefinite *dieser* and *son* only allow for a wide-scope reading.

(4)  Eva will einen / diesen / son Film über Eliade sehen.
‘Eva wants to watch a/ this / such-a movie about Eliade.
 *ein:* specific, non-specific (wide scope, narrow scope)
 *dieser* wide scope
 *son* wide scope

In order to confirm these intuitions we conducted a pilot experiment in form of an online judgment task in which the acceptability (“sounds good” vs. “doesn’t sound good”) of sentences including *ein, dieser* and *son* on either forced wide-scope or forced narrow-scope readings was tested. 27 native speaker of German saw sentences of the following type (3 test items (lexicalizations) in total, varying with respect to the type of indefinite expression and +/- specificity):
Non-Specific (forced narrow-scope):
(5) Maria will einen / diesen / son Spanier heiraten. Sie hat aber noch keinen kennen gelernt.
'Maria wants to marry a / this / such-a Spaniard. She has not met one yet.'

Specific (forced wide-scope):
(6) Maria will einen / diesen / son Spanier heiraten. Er heisst Pedro und kommt aus Mallorca.
'Maria wants to marry a / this / such-a Spaniard. His name is Pedro and he is from Mallorca.'

The test sentences of the type “narrow scope” were used in order to check if the respective article allows for narrow-scope interpretations and the test sentences of type “wide scope” were used in order to check the general acceptability of wide-scope interpretations.

<table>
<thead>
<tr>
<th></th>
<th>narrow scope</th>
<th>wide scope</th>
</tr>
</thead>
<tbody>
<tr>
<td>ein</td>
<td>77,8%</td>
<td>81,5%</td>
</tr>
<tr>
<td>son</td>
<td>29,6%</td>
<td>51,9%</td>
</tr>
<tr>
<td>dieser</td>
<td>7,4 %</td>
<td>74,1%</td>
</tr>
</tbody>
</table>

Figure 1: Degree of acceptability of the sentences in wide scope / narrow scope contexts dependent on the respective indefinite determiner

The intuitions about the differences with respect to the referential behavior of the unmarked indefinite article *ein* and the indefinite *dieser* are confirmed: Sentences with *ein* were well-accepted in both contexts, confirming that *ein* is ambiguous between a wide scope and a narrow-scope reading. Indefinite *dieser* cannot be used in the narrow-scope contexts, which is mirrored in the low acceptability of the sentences in those contexts (7,4%). However, the results for *son* are not so clear: First, *son* is generally not well-accepted which may be due to the fact that it is assumed to be formal, colloquial and therefore “not-correct”. Furthermore, *son* in narrow-scope context was accepted by roughly a third of the participants which leads the conclusion that it allows for this interpretation, at least for those informants who find it acceptable in general.

This preliminary picture can be summarized as follows: The unmarked indefinite article *ein* and the indefinite *dieser* behaves like the “referential indefinites” of Fodor & Sag (1982), i.e. like directly referential expressions. The results for *son* ask for further investigation. In a follow-up study we tested the scopal behavior of the three articles under a universal quantifier. This environment allowed us to distinguish between a) wide scope, b) simple narrow scope readings and c) functional readings (type ‘his mother’), as we hypothesize that the functional reading might be the distinctive property of *son* indefinites. This design has 9 conditions, which were tested using 115 participants facing 18 lexicalizations, yielding 230 judgments per condition. The results of this study are in the process of being interpreted and we will present them in our paper.

References:
Reading Hurford disjunctions: Processing evidence for embedded implicatures

1. Disjunctions and Hurford’s constraint. Disjunctions are infelicitous if one disjunct entails the other, as in (1) (Hurford 1974, Gazdar 1979). It has been observed however that such disjunctions are saved in specific cases, such as (2), in which, at least at the descriptive level, the entailment relation can be broken by local pragmatic enrichments, as in (3).

   (1) *John lives in Paris or France.
   (2) John read some of the books or all of them.
   (3) John read [some but not all of the books] or all of them.

   These facts provide evidence that enrichments akin to scalar implicatures can occur at sub-sentential levels (Chierchia et al. 2013). We test a processing prediction of this approach.

2. Distinctive prediction of a local theory of exhaustification. In a theory à la Chierchia et al. (2013), scalar implicatures are generated by a covert operator, O, which has the semantics of only. In cases such as (2), insertion of O is required to avoid a violation of Hurford’s constraint by enabling the parse John read O(some of the books) or all of them. At the processing level, this may follow from two distinct Parsing Strategies (PS): (PS1) readers may interpret the first disjunct with O on the fly, in which case the second disjunct will not cause a violation; (PS2) readers may first interpret the first disjunct without an O-operator, in which case they will have to reanalyse and reparse the first disjunct with an O-operator upon reading the second disjunct (“all of them”) to avoid a Hurford’s constraint violation.

   Auxiliary hypotheses: Let us assume that there are processing costs associated with (a) insertion of an O-operator; (b) re-analysis of the first disjunct.

   Processing prediction: The O-approach then predicts a negative correlation between the reading times of the first disjunct and the reading time of the second disjunct, on the grounds that an extra processing cost will be incurred at the first disjunct or the second, but not both.

3. Note on alternative approaches. It is not clear whether other accounts can capture all the Hurford-related facts documented in Chierchia et al. (2013). However, our processing prediction is in any case not an automatic corollary of most feasible explanations. For instance, one plausible analysis, following Gazdar (1979) and Gajewski and Sharvit (2011), might predict a positive correlation: if the second disjunct “cancels” implicatures in the first, then a cost at the first (from computing an implicature by anticipation) would be met with a cost at the second (for cancelling it). By contrast, our negative correlation prediction crucially relies on the possibility of inserting the O-operator without further considerations (PS1).

4. Experiment. In the context of a story about aliens exploring a city, participants were asked to read sentences in a self-paced reading paradigm in which one word was presented at a time. The critical conditions required enrichment in order to avoid violating Hurford’s constraint, as in (4), or to avoid obtaining a tautologous reading, as in (5), which creates the same processing prediction in the O-approach.

   (4) The happy alien visited some of the bookshops or all of them.
   (5) The happy alien visited some of the bookshops or none of them.

   Non-critical conditions (6) and (7) were introduced; they do not create a problematic relation between the disjuncts and hence escape the negative correlation processing prediction.

   (6) The happy alien visited none of the bookshops or all of them.
   (7) The happy alien visited all of the bookshops or none of them.
Finally, to avoid predictability of the violation, we duplicated all of these conditions with a parallel condition involving the same structure (and most importantly the same quantifiers) up to the final word which, in place of “them,” introduces a new location, thus avoiding any potential Hurford’s constraint violation even without additional enrichment, as in (8).

(8) The happy alien visited some of the bookshops or all of the hospitals.

Different adjectives and locations were used to create a set of items, and these were rotated among the conditions to create eight different versions of the experiment. 20 participants were assigned to each of them at random via Mechanical Turk, and tested via Ibex.

5. Results. We analyzed reading times for “<quantifier> of the” in the first disjunct against “<quantifier> of the(m)” in the second disjunct (this was informed by results of a pilot study, but other regions of interest yield the same outcome). More precisely, we looked for correlations in the residuals of RTs after factoring out the general speed of reading of a given sentence by a given participant (by regressing the RTs for the critical regions on the RTs for the words that appeared in all sentences, “alien” and “visited”, see DeGutis et al. 2013). We excluded outliers (data points with RTs below 150ms or above 500ms for any of the critical words), and the responses of people who completed the experiment more than once or did not declare English as their native language.

Modelling the residuals in the second disjunct as a linear function of the residuals in the first disjunct and of whether the item was critical ((4) and (5) vs. (6) and (7)), we documented a significant main effect of the first disjunct residuals ($\beta = 0.602$, SE = 0.023, $t = 26.1$, $p < 0.001$), which we interpret as a general tendency for an item to be read at a consistent speed throughout (which persists above the first residual extraction analysis). Crucially, there was an interaction between the first disjunct residuals and whether the item was critical ($\beta = -0.085$, SE = 0.034, $t = -2.50$, $p < 0.05$), suggesting a negative correlation between the RTs of the two disjuncts, which – specifically in the critical cases – runs counter to the tendency towards consistent reading speed throughout the sentence.

6. Discussion. We interpret this result as suggestive that there is a possibility for readers to enrich the first disjunct anticipatorily, and that doing so pays off in terms of decreasing reading times in the second disjunct, where additional processing would otherwise be necessary in order to avoid a violation of Hurford’s constraint. This supports the O-operator theory. In the full paper we consider the implications for conceivable alternative approaches.
Comprehension of Demonstratives by English-speaking Preschool Children

The current study examines preschool children’s comprehension of demonstratives in light of their non-linguistic cognitive ability to incorporate contextual information that is required in demonstrative comprehension. The use of demonstratives, this and that, is highly context-dependent: this picks out an object near the speaker, while that picks out one that is far from the speaker; crucially, the speaker-object distance is determined based on the speaker’s perspective, which varies from context to context (e.g., Clark and Sengul, 1978; Murasugi, 1986). Thus, as learners of a language, children need to master both of these aspects of the use of demonstratives. Previous research has suggested that the mastery of demonstratives may be a challenge for children, reporting children’s non-adult-like use of demonstratives. The first non-adult-like response pattern observed with demonstratives is one in which children interpret demonstratives solely with respect to their own perspective, failing to comprehend demonstratives when the speaker’s perspective is not their own; this response pattern has been argued to reflect an egocentric bias in these children (Clark & Sengul, 1978; Webb & Abrahamson, 1976). This pattern was proposed to reflect children’s still-developing Theory of Mind (ToM) (de Villiers, 2007; Diesell, 2012), i.e., one’s ability to understand that others may have perspectives different from one’s own. The second non-adult-like pattern observed in the literature is one in which children associate the object nearest to the speaker as the referent of both this and that. Tanz (1980) analyzed this response pattern as salience-biased, claiming that children robustly selected the object closer to the speaker which is readily perceptible, hence, visually salient. Salience-based non-adult-like linguistic comprehension has also been reported in a recent study (Minai et al., 2012), attributing it to children’s still-developing Executive Function (EF) which hinders children’s disengaging themselves from salient information. Given this, children’s salience-biased demonstrative interpretation may also reflect their developing EF. Taken together, previous studies showed that children’s use of demonstratives is non-adult-like, with children’s response patterns strongly suggesting an association with their still-developing cognitive abilities, in particular, ToM and EF. Such a speculation, however, has not yet been directly tested.

The present study thus examines children’s interpretation of demonstratives directly with respect to their development of ToM and EF. To this end, four tasks were administered: two linguistic comprehension tasks (the Act-out Task and the Judgment Task) and two cognitive tasks (the Hiding Game and the Dimensional Change Card Sort). The Act-out Task measured children’s comprehension of distance contrast of this/that when the child shares the same perspective with the speaker. In the task, children were shown two boxes, one right in front of them and the other distant from them; the experimenter, serving as the speaker of the sentences with a demonstrative word, sat right next to the child so that the experimenter and the child shared the same perspective. The experimenter uttered the act-out instruction using demonstratives, e.g., “Put Pooh in this/that box”. The adult-like responses were (i) putting the character in the box right in front of them when children were told to put it in this box, and (ii) putting the character in the box apart from them when they were told to put it in that box. Children’s success in this task was taken as the confirmation that they are able to comprehend this and that contrastively based on their own perspective, serving as the foundation for us to investigate whether their success in the comprehension of this and that will extend to cases in which the speaker’s perspective is different from their own --- this was tested by the Judgment Task, in which children comprehended demonstratives uttered by the speaker whose perspective is not their own. The task was administered with visualized stories, with a King and a Servant as the main characters. In the story, there were two identical objects (e.g., two plates), one of which was located near the King while the other of which was located far from the King. The King demanded that the Servant paint one of the objects, by uttering
sentences with demonstratives, e.g., “Paint this/that plate blue”. The Servant tried to fulfill the King’s demand by painting one of the objects. The child was asked to judge whether the Servant painted the right object the King told him to paint: if the King told the Servant to paint this object, then the Servant was expected to paint the object near the King; if the King told the Servant to paint that object, then the Servant was expected to paint the object apart from the King. The ‘Hiding Game’ (Gale et al., 1996; Schick et al., 2007) was administered as a measurement of children’s ToM. In this game, three characters appeared: the Hider, the Knower and the Guesser. The Hider hid the object in one of the three boxes present in the experimental workspace; the Knower witnessed the entire hiding event (thus ‘knows’ where the object was hidden in the end); the Guesser was blindfolded while the hiding took place. Crucially, while the hiding was taking place, the screen blocked the scene and the child could not see the hiding event; after the hiding was completed, the screen was removed and the Guesser’s blindfold was also removed. The Knower and the Guesser selected the box each of them thought has the object inside; the Knower always selected the right box with the object (reflecting that he has the right ‘knowledge’ about the hiding outcome), while the Guesser selected a wrong box (not witnessing the hiding and thus not knowing the outcome). The child was asked to determine which of the box he/she thinks has the object inside. If the child correctly selected the Knower’s box, it was interpreted as an adult-like response, as the child could correctly infer that the Knower, not the Guesser, has the reliable knowledge about the position of the hidden object. A standard version of the Dimensional Change Card Sort (DCCS) (Zelazo, 2006) was also utilized to measure children’s EF; the child was asked to sort cards that can be classified based on the two dimensions, color and shape, first according to one dimension (Pre-switch Phase) and then according to the other (Post-switch Phase). Whether the child can successfully sort cards in the Post-switch Phase was measured.

Thirty-six English-speaking 3-, 4-, 5-, and 6-year-olds participated (3:0-6:2, mean=4.7). We first report the results of the Act-out Task. As success in this task indicates that children are able to comprehend this and that correctly based on their own perspective, we examined the individual children’s response patterns to see how many children indeed succeeded/passed; 23 out of 36 children were categorized as successful ‘passers’, with correct responses in 4 out of 6 trials. Crucially, we also examined whether there was an across-task correlation among the 23 Act-out passers’ responses in the Judgment Task, the Hiding Game and the DCCS as follows. We first calculated the mean percentage of correct comprehension in the Judgment Task (57.61% for this; 61.96% for that). In the cognitive tasks, we clustered children into ‘passers’ and ‘failers’; in the Hiding Game, children were categorized as passers if they gave correct responses in 4 out of 5 trials (18 passers); in the DCCS, children were categorized as passers if their responses were correct in 5 out of 6 trials in the Post-switch Phase (12 passers). In order to examine whether there is a correlation among the results of the three tasks, we calculated Spearman’s Rho among the following variables: (1) the mean percentage of correct responses for the demonstrative this in the Judgment Task, (2) the mean percentage of correct responses for the demonstrative that in the Judgment Task, (3) the pass/fail category in the Hiding Game, and (4) the pass/fail category in the DCCS. The results revealed that there was a significant positive correlation between the mean percentage of the correct comprehensions of this and the pass/fail category in the DCCS ($r_s=.381, p < .05$); there was also a marginal positive correlation between the mean percentage of the correct comprehensions of this and the pass/fail category of the Hiding Game ($r_s=.329, p=.062$). These results are consistent with the hypothesis that children’s success/failure in comprehending demonstratives is indeed related to the development of non-linguistic cognitive abilities including Executive Function and Theory of Mind, which are required to appropriately incorporate the necessary contextual information about the speaker’s perspective into demonstrative interpretation.
Context dependence and shiftability in two classes of gradable adjectives

The current study investigates two subclasses of gradable adjectives: Relative gradable adjectives (RAs) like tall, which use open scales, and Absolute gradable adjectives (AAs), which have their standards fixed at the maximum (full) or minimum (bent) point on the relevant scales. Both classes exhibit context sensitivity; in sentences like (1)-(2), what counts as tall or full varies with aspects of the sentences are embedded in. However, it has been noted that AAs lack the characteristic behaviors of vague predicates—susceptibility to the Sorites paradox, and borderline cases that resist classification (e.g. as either tall or not tall)—which RAs systematically exhibit. Prior research has proposed that this points to a real difference between RAs and AAs associated with the types of scales used by each class [1-2]. According to such an account, the context dependence that arises for each adjective type has a distinct source. RAs have lexical representations which require a contextually-determined comparison class for semantic computation; salience in the local discourse representation determines the comparison class, which sets a standard of comparison (semantic context dependence). By contrast, AAs do not have semantically indeterminate lexical representations. Instead, their variability arises because they are sensitive to features of the global communicative context (although AAs, to a certain extent, might still be interpreted with respect to comparison classes, as [3-4] suggest): what counts as e.g. a full glass varies by how high a standard of precision is set in a particular situation, which itself depends on the goals and intentions associated with interlocutors (pragmatic context dependence). For instance, (2) may describe a glass that is only 95% full in circumstances where the standard of precision is set relatively low (e.g. ordering drinks at a bar), whereas a high precision context (e.g. a laboratory experiment) may require that the liquid reach the brim for the glass to be considered full. Regardless of how this idea is formally implemented ([1-2] analyze it as slack regulation; [3] treats it in terms of granularity), the dynamic profiles of RAs and AAs provide another way to tease the two classes apart. In particular, shifting a standard of precision or granularity at the global context level may differ from shifting standards of comparison, as suggested by [6] (for supporting experimental evidence, see [3] and [7]). We present four experiments designed around the differences between RAs and AAs described above. First, Experiments 1-2 show that RAs are sensitive to salience in the local context in a way that AAs are not. Experiments 3-4 show that RAs and AAs exhibit different dynamic patterns in a manner consistent with e.g. [6]: specifically, AAs show an asymmetry in the direction of shiftability, while RAs show general facilitative effect of repeated exposures regardless of shift direction.

Experiments 1-2 (n1=28,n2=20) showed that RAs but not AAs depend on the local discourse context. We created sets of images representing six points on a continuum characterized by an adjective pair (tall-short ladder). Participants indicated for each image and adjective pair whether they considered the pictured object to be e.g. tall, short, or neither. In addition, they either saw items in isolation, interspersed with trials with other adjective-noun pairings so that each trial represented its own local discourse context (Experiment 1), or with the images associated with the same adjective-noun continuum presented as a group (Experiment 2), thereby providing a comparison class in context. Responses were fitted with mixed-effects regression models with Subject, Adjective, and Noun as random effects, and Scale position (position on the continuum) as fixed effect [8]. Only for RAs, we found: (i) isolated presentation increased dependence on item-specific prototypes: the identity of the head noun accounted for 75.4% of the random effects variance in a mixed effects regression model, as compared to 67.4% when items were presented in groups; and (ii) the position of an item on an adjective-noun continuum was a reliable predictor of response for grouped (β=.40, p<.0001), but not for individual presentation (β =-.05, p>.1). By contrast, AAs showed no differences in interpretation as a function of presentation type. Together, Experiments 1-2 demonstrate that RAs heavily rely on the local discourse context, while AAs exhibit relative stability despite changes to the local context.
Experiments 3-4 (n_A=36, n_G=36) examined effects of presentation order, in order to establish how RAs (Experiment 3) and AAs (Experiment 4) pattern over repeated uses, and in particular whether they differ in terms of the shiftability of their standards. Using the procedure described for Experiment 1, participants judged images in trials like (3) (RAs) or (4) (AAs). As in Experiment 1, the images were presented in isolation, but taken from sets of matching images that varied along the relevant dimension (e.g. tallness, fullness). Responses were fit to mixed-effects regression models including Subject, Adjective, and Noun as random effects, fixed effect of Scale position, and the following additional fixed effects reflecting aspects of the preceding trials in the experimental session: (i) number of prior instances of the same adjective, and (ii) whether preceding trials included an instance of an extreme exemplar for RAs (the tallest candle in the experimental items) or a perfect exemplar for AAs (perfectly full cup). (Predictors that did not significantly improve model fit were excluded using model comparison.) Of interest was whether setting a maximally precise standard of precision for an AA would decrease the likelihood of accepting the same adjective on a subsequent trial, which would require relaxing the previously set standard, and whether prior high standard trials for RAs would have a similar effect on subsequent interpretation.

The AA model showed that prior exposure to a maximal precision exemplar decreased the likelihood of accepting a subsequent use of the same adjective (β=-2.6, p<.0001); further, this effect was conditioned on the current object's position on the scale (e.g. how close to perfectly full a cup was): objects distant from maximally precise were more strongly affected by a prior maximally-precise exemplar (β=-0.59, p<.0001). In addition, the effect of having seen a maximal precision exemplar did not decrease as the number of intervening uses of the same adjective increased (β =.0038, p>.1). RAs displayed a different pattern of effects over multiple occurrences: prior exposure to an extreme exemplar increased the likelihood of accepting a subsequent use of the same adjective (β =.60, p<.05), irrespective of the current object's scale position. This facilitative effect decreased as the number of intervening uses of that adjective increased. The pattern of effects observed for RAs suggests something akin to similarity-based facilitation (i.e. priming) which decays in strength over time (or intervening exposures). These diverge from the AA data, which appear more categorical in nature, at least with respect to the possibility of standard shifts.

Conclusions: These results provide additional experimental evidence for a distinction between Relative and Absolute GAs. We show that these classes of expressions can be differentiated with respect to (i) their sensitivity to different aspects of the context in a way that suggests that standards of comparison for RAs are closely tied to the local discourse representation, while standards of precision for AAs are tightly linked to broader features of the context at large, and (ii) their dynamic profiles, including the kinds of standard shifts that are possible for each class from one usage of an adjective to the next. Crucially, this does not mean that comparison classes for AA do not exist at all (in fact, the issue is still debated, see e.g. [2-3]), but rather that AAs and RAs differ in their sensitivity to different aspects of the context.

(1) Alex is tall. (2) The glass is full. (3) image: ladder; This item is {tall, short, neither} (4) image: glass; This item is {full, empty, neither}