Despite 40 years of research, quantification is currently again a central topic of research at the interface between logic and language. Generalized Quantifier Theory (GQT) and Discourse Representation Theory (DRT) still are standard bearers when it comes to the representation of quantifier meaning. But a new frontier in quantification research is to account for the use of quantifiers, in addition to the meaning they contribute to sentence and discourse.

In GQT and DRT, many quantifiers are semantically identical to one another: most and more than half; at least three, more than two and three; and about 100 and some 100. Even when there is no identity, the meanings of quantifiers overlap, and many quantifiers could be used truthfully to describe the same situation. Nevertheless, the choice of a specific quantifier by a speaker has semantic and pragmatic consequences, and listeners are sensitive to the subtleties of a speaker’s choice. The goal of a theory of quantifier use is an account of how speakers choose one of the many true quantifiers and how listeners interpret a speaker’s choice.

The PUQOL workshop aims to bring together current research within semantics, logic, pragmatics, and cognitive science that addresses quantifier use within the perspective of model theoretic semantics and pragmatics. Abstracts were invited for 30-minute talks (including discussion) on any topic relating to this theme, including:

- the cognitive processes underlying quantifier production and understanding
- models of vagueness and approximation in quantifier meaning
- the link between information structure and quantification
- the interaction of quantification with pragmatic processes
- repercussions of work on quantifier use for semantic theory
- first language acquisition of quantifier meaning and use
- formal experimental results of particular use to the investigation of quantification
# PUQOL PROGRAM

*The Proper Use of Quantification in Ordinary Language* workshop at ESSLLI 2011
financed by the LogiCCC Eurocores of the European Science Foundation

<table>
<thead>
<tr>
<th>DAY 1: Monday, August 8th, 2011</th>
<th>chair: D. Westerstahl</th>
</tr>
</thead>
<tbody>
<tr>
<td>14:00 – 14:30 Uli Sauerland and Stephanie Solt, ZAS Berlin</td>
<td>The Proper Use of Quantifiers in Ordinary Language: Introduction</td>
</tr>
<tr>
<td>14:30 – 15:00 Natalia Zevakhina, Radboud University Nijmegen</td>
<td>Availability of quantifier scalars</td>
</tr>
<tr>
<td>15:00 – 15:30 Bob van Tiel, Radboud University Nijmegen</td>
<td>The typicality structure of quantifiers</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DAY 2: Tuesday, August 9th, 2011</th>
<th>chair: tba</th>
</tr>
</thead>
<tbody>
<tr>
<td>14:00 – 14:30 Hadas Kotek, Yasutada Sudo, Edwin Howard and Martin Hackl, MIT</td>
<td>A superlative reading for most</td>
</tr>
<tr>
<td>14:30 – 15:00 Aviya Hacohen, Dana Kozlowski and Ariel Cohen, Ben-Gurion University of the Negev</td>
<td>Superlative quantifiers as speech act modifiers: Experimental evidence</td>
</tr>
<tr>
<td>15:00 – 15:30 Marcin Zajenkowski, Rafa Styla and Jakub Szymanik, University of Warsaw and University of Groningen</td>
<td>Schizophrenic patients find proportional quantification difficult due to working memory deficits</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DAY 3: Wednesday, August 10th, 2011</th>
<th>chair: S. Solt</th>
</tr>
</thead>
<tbody>
<tr>
<td>14:00 – 14:30 Dany Jaspers, Hogeschool University Brussel</td>
<td>Different degrees of naturalness of quantifiers and colour terms</td>
</tr>
<tr>
<td>14:30 – 15:30 Bart Geurts, Radboud University Nijmegen</td>
<td>Quantifiers and pronouns: an inferential approach</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DAY 4: Thursday, August 11th, 2011</th>
<th>chair: U. Sauerland</th>
</tr>
</thead>
<tbody>
<tr>
<td>14:00 – 14:30 Robert Henderson, Scott Anderbois and Adrian Brasoveanu, University of California, Santa Cruz</td>
<td>The pragmatics of quantifier scope: A corpus study</td>
</tr>
<tr>
<td>14:30 – 15:30 Cornelia Ebert, University of Stuttgart</td>
<td>Variations among long-distance readings in English and German – an experimental study</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DAY 5: Friday, August 12th, 2011</th>
<th>chair: C. Roschger</th>
</tr>
</thead>
<tbody>
<tr>
<td>14:00 – 14:30 Chris Cummins, University of Cambridge</td>
<td>Constraint-based modelling of numerical quantifier usage and interpretation.</td>
</tr>
<tr>
<td>14:30 – 15:30 Laurence Horn, Yale University</td>
<td>Contrariwise: Disjunctive syllogism and pragmatic strengthening</td>
</tr>
</tbody>
</table>
Abstracts

(in alphabetical order)
Constraint-based modelling of numerical quantifier usage and interpretation

Chris Cummins

RCEAL, University of Cambridge

When selecting a numerically quantified expression, speakers choose between arbitrarily many truth-conditionally correct alternatives. In this presentation, I discuss a proposal for characterising this decision process as one of multiple constraint satisfaction. Using heuristics applied in Optimality Theory, I propose a set of constraints, and motivate these individually by appeal to the psycholinguistic literature. I then consider how the interaction of these constraints predicts that speakers will behave in particular quantificational environments, and present evidence that these predictions are borne out experimentally. Then I examine the system from the hearer’s perspective, and articulate the resulting predictions concerning the interpretation of given quantified expressions. With particular reference to the case of scalar implicatures from expressions such as ‘more than n’, I show that this model yields novel predictions that contradict the assumptions of the previous literature, and show new experimental evidence that supports these predictions. I conclude by considering the potential role of this model as a source of hypotheses about quantifier usage and interpretation, and its relation to other current pragmatic approaches.

My work is motivated by the observation that the speaker’s choice of quantified expression is not uniquely determined by semantic considerations. Due to the structure of the number system, quantified expressions participate in extensive entailment relations, with the consequence that (in the following examples) if (a) can be used truthfully to describe a situation then so can any of the corresponding expressions in (b).

1. (a) more than 20
   (b) more than 19/18/17…
2. (a) fewer than 20
   (b) fewer than 21/22/23…
3. (a) between 20 and 25
   (b) between 20 and 26/19 and 25/19 and 26…

Intuitively it is clear that the speaker’s preference depends upon various other considerations. In this presentation I formalise this intuition by arguing for the relevance of individual constraints governing this choice. Specifically, these constraints concern informativeness, granularity, quantifier simplicity, numeral salience, and numeral and quantifier activation. I present experimental evidence supporting the relevance of each one of these constraints individually. Then I consider the interaction of these constraints. Crucially, it is typically not possible for an utterance to satisfy all these constraints at once. For instance, if 4(a) is maximally informative, it violates numeral salience, while 4(b) uses a more salient numeral but violates informativeness.

4. (a) more than 101
   (b) more than 100
I therefore propose a treatment of numerical quantification in terms of a (unidirectional, speaker-referring) Optimality Theory analysis. This model predicts the speaker’s choice of expression (output) given the relevant contextual factors (input) and the speaker’s own constraint ranking. In particular, I provide experimental data to show that this model correctly predicts the way in which contextual manipulations influence the speaker’s choice of utterance in paradigms involving the correction of infelicitous statements, such as those used by Katsos et al. (2010).

Having motivated the model in this way, I consider its effects on interpretation. From the hearer’s perspective, utterances selected under this model are predicted to convey additional pragmatic effects. Alongside the semantic meaning, each expression admits probabilistic pragmatic enrichments of meaning, based on the observation that the expression can only surface in certain contexts. In particular, I explore the nature of the enrichments that are predicted to arise from expressions ‘more/fewer than \( n \)’, where the numeral \( n \) exhibits different levels of roundness. Contrary to the assumptions of Fox and Hackl (2006), following Krifka (1999), the model predicts that implicatures to the effect ‘not more/fewer than \( m \)’, for some \( m \) dependent on \( n \), are conveyed. I demonstrate empirically that these implicatures are calculated by hearers, and moreover – also in accordance with the predictions of this model – are attenuated by prior mention of the numeral.

I discuss the potential of this model in accounting for outstanding questions of quantifier distribution. Developing the line suggested by Krifka (2009), I argue that this model can account for the use of round numerals as approximations. Following Solt (2010), I consider the treatment of ‘most’ versus ‘more than half’, arguing that the distributional differences she observes might be accounted for as pragmatic enrichments, as a consequence of the salient nature of the level ‘half’.

Finally, I consider the relation of this account to other current theories of quantifier choice. From a relevance theory perspective, it could be argued that this account represents a way of calculating and evaluating the optimally relevant utterance for a given situation, and thus augments relevance theory with quantitative predictive power. This model also shares with game theoretic accounts its way of relating usage and interpretation, and could be seen as a principled relaxation of game theoretic assumptions as applied to numerical quantification.

References
The focus of my talk is on the relationship between the scope-taking possibilities of indefinites, and the form of the indefinite determiner in English and German. It has been known since at least Fodor & Sag (1982) that unlike other quantifiers, indefinites can scope out of an island such as a relative clause, as in (1a), taking long-distance (LD) scope (1b) in addition to local scope (1c).

(1) a. The teacher put away every toy that a/one/ONE child played with.
   b. LD scope (∃ ≥ ∀): There exists a specific child, such that the teacher put away every toy that this child played with.
   c. local scope (∀ ≤ ∃): The teacher put away every toy that any child played with.

I focus on weak indefinites with a, unstressed one and stressed ONE in English and unstressed ein and stressed EIN in German. Unlike indefinites with overt specificity marking, such as a certain (cf. Hintikka 1986, Kratzer 1998, Ionin 2010), indefinites with a/one and ein exhibit a range of scope interpretations. I will show that nevertheless, there are clear differences in the preferred interpretation for a vs. one/ONE and ein vs. EIN indefinites.

Most accounts of indefinite scope assume LD-readings to be freely available to different types of indefinites, regardless of determiner form (e.g., Reinhart 1997, Kratzer 1998, Schwarzschild 2002). In contrast, Endriss (2009) argues that LD-scope is linked to topicality: while topical indefinites take LD-scope, non-topical ones do not. Tying topicality to a certain stress pattern, Endriss shows that in German, indefinites with stressed EIN occur in Topic position more readily than those with unstressed ein; consequently, EIN indefinites also take LD-scope more readily than ein indefinites. English, unlike German, has two separate lexical items, a vs. one, and the question is open as to how determiner form and stress pattern relate to LD-scope.

An experimental study examines the link between LD-scope, determiner form, and stress pattern in both English and German (as part of an experiment series by Ebert, Ionin & Stolterfoht 2011). A web-based Truth-Value Judgment Task (TVJT) was used, with picture/sentence pairs; English and German versions of the TVJT were created, and administered to 44 native English speakers and 30 native German speakers, respectively. Sentences like (1a) and its German equivalent in (2) were presented auditorily, with determiner form varied between one, ONE and a for English and ein, EIN, and ’n (the reduced, cliticized version of ein) for German.

(2) Der Lehrer hat jedes Spielzeug weggeräumt, mit dem ’n/ein/EIN Kind gespielt hat.

The sentences were presented in the context of two-panel pictures, which came in two types: those matching the LD reading of the indefinite (Figure 1) and those matching the local reading (Figure 2), as explained in (3), with target truth-values spelled out in (4)-(5). The TVJT thus had a 2X3 design (2 picture types by 3 determiners), with 6 tokens per condition, for a total of 36 target items (plus 36 controls/fillers; at-ceiling performance on control items indicates that participants were paying attention). Three test scripts, each with 2 test orders, were used to avoid repetition of contexts and to control for ordering effects.
Figure 1: LD-scope scenario for (1a/2):         Figure 2: local-scope scenario for (1a/2):

(3) Explanation of figures:
Panel 1 (both figures): girl plays with duck, boy plays with blocks and train, both children play with ball, neither child plays with doll or star. [relationship between children and toys is expressed using Venn diagrams; participants are trained on this format]
Panel 2 (Figure 1): the teacher has put away the blocks, train, and ball
Panel 2 (Figure 2): the teacher has put away the duck, blocks, train, and ball
The location of the duck above is indicated by a circle, to highlight the difference between the two figures; there were no such circles in the actual test items.

(4) Truth-values for sentence (1a/2) in the context of Figure 1:
   a. LD scope (∃ > ∀)
      TRUE: there’s 1 child (the boy) such that the teacher put away every toy he played with
   b. local scope (∀ > ∃)
      FALSE: the teacher did not put away every toy played with by any child – the teacher failed to put away the duck, which the girl played with

(5) Truth-values for sentence (1a/2) in the context of Figure 2:
   a. LD scope (∃ > ∀)
      if the indefinite is interpreted as at least one: TRUE (there is at least one child – in fact, there are two – such that the teacher put away every toy they played with)
      if the indefinite is interpreted as exactly one: FALSE (there are two children – not exactly one – such that the teacher put away every toy they played with)
   b. local scope (∀ > ∃)
      TRUE: the teacher put away all four toys that any child played with

The results were the following: LD-scope was more readily available for one/ONE than for a, and for EIN than for ein and ‘n. Local scope was more readily available for a than for one/ONE, and for ‘n/ein than for EIN. Thus, the relevant factor facilitating LD-scope is determiner form in English, but stress on the determiner in German. We plan a follow-up study of the behavior of stressed and unstressed some in comparison to a.
We also found that for the tested indefinites, scope readings are a matter of preference rather than absolute judgments. This is unexpected on theories that freely allow all indefinite types to have both local and LD-readings and especially on those that assume a lexical ambiguity (e.g., the approach by Fodor & Sag 1982 or choice function approaches like Kratzer 1998 or Matthewson 1999); the findings are more compatible with Endriss (2009), which ties LD-scope to information-structure.
We propose the following account of the findings. The default determiner for singular indefinites is *a* in English and *ein* in German; the preferred interpretation for these default indefinites is non-topical, hence local, narrow-scope (cf. similar findings in Ionin 2010). In order to take LD-scope, the indefinite needs to be topical (per Endriss 2009); topicality is marked by stress in German, and by the use of *one* in English.

It should be further noted that *one/ONE* and *EIN*, in addition to taking LD-scope, were often interpreted as *exactly one* (evidenced by lowered rates of ‘true’ responses in the context of Figure 2, see (5)). We will discuss a planned follow-up study, designed to tease apart scope preferences and *at least* vs. *exactly* readings. Another interesting finding was that LD-scope was, overall, more accepted in English than in German. Possible reasons for this will be discussed.

**References:**

Quantifiers and pronouns: an inferential approach
Bart Geurts

By adopting a dynamic construal of quantifying expressions, we can account for various kinds of anaphora in a quite natural way, but this solution comes at a cost: we have to relinquish the intuitively compelling idea that, e.g., "All A's are B's" just means that the set of all A's is a subset of the set of all B's. If we want to stick to this idea, we could adopt a descriptivist analysis of anaphoric pronouns, but apart from the fact that such analyses quickly tend to become enmeshed in rather daunting systems of epicycles, they would force us to give up on the intuitively appealing idea that pronouns are semantically attenuate expressions that merely serve to pick up discourse referents made salient by the context in which they occur.

There is a third alternative. We can eat our cake and have it: if we allow pronouns to pick up discourse referents that can be inferred from the context, we can do justice to our intuitions about quantifiers and pronouns alike. The inferential approach allows us to stick to the classical treatment of quantifiers, forego descriptivist analyses of pronouns, and it will even solve a number of puzzles that thus far resisted a straightforward solution, like modal subordination and Partee's bathroom sentences.
Introduction: We provide experimental support for Cohen & Krifka’s (to appear) account of superlative quantifiers (henceforth SQs), according to which they are speech act modifiers. Traditionally (Keenan & Stavi 1986), superlative quantifiers are treated in the same way as comparative quantifiers (more than and fewer than), so that (1) is assumed to be equivalent to (2).

(1) Mary petted at least three rabbits
(2) Mary petted more than two rabbits.

More recently, however, Geurts & Nouwen (2007) have demonstrated that this account is inadequate, and proposed an alternative, according to which SQs express epistemic modal statements; specifically, (1) is claimed to mean (3).

(3) It is epistemically necessary that Mary petted three rabbits, and it is epistemically possible that she petted more.

Alternative theories are proposed by Büring (2007) and Cummins & Katsos (2010) who argue that (1) means the disjunction in (4).

(4) Mary petted exactly three rabbits or Mary petted more than three rabbits.

Cohen & Krifka (to appear) propose a different theory, according to which SQs are illocutionary operators. Specifically, (1) is interpreted as (5).

(5) For all n<3, the speaker denies that Mary petted exactly n rabbits.

Note that a crucial element of this theory is that speech acts can be under the scope of logical operators and contribute to the truth conditions of the sentence. For example, suppose Mary petted exactly two rabbits, and John utters (1). Now, according to (5), John is denying three statements (for n=0,1,2), one of which is true (for n=2), hence asserting a falsehood, and (1) is false, as desired.

Turning from falsity to truth, suppose Marypetted exactly four rabbits. It follows from (5), by way of conversational implicature, that the speaker refrains from denying that Mary petted exactly n rabbits for n≥3. Since one of the options entertained by the speaker is, in fact, true (for n=4), (1) is true, as desired.

Importantly, then, the falsity of (1) follows semantically, whereas its truth follows pragmatically, through conversational implicature. Cohen & Krifka argue for this claim on linguistic grounds; in this paper we provide experimental support for Cohen and Krifka’s proposal.

Processing: All three theories (epistemic, disjunction, and illocutionary) predict that the processing of SQs will take longer than that of comparative quantifiers, and indeed this prediction has been confirmed experimentally (Geurts et al 2010; Cummins and Katsos 2010). However, Cohen & Krifka’s illocutionary theory makes a further prediction: since only judgments of truth, but not judgments of falsity, require an implicature; and since implicatures require additional time to process (e.g. Bott & Noveck 2004), it follows that judgments of true SQ sentences will take longer than judgments of false ones. Importantly, this prediction does not follow from any of the competing theories, which assume that both truth and falsity are evaluated...
semantically, with no need for implicature. We conducted two online experiments, using the sentence verification task to test this prediction.

**Experiment A - methods:** Reaction times were recorded as participants judged sentences of the form *I see Q N Xs*, where Q is a quantifier (superlative or comparative); N is a number (3, 4 or 5); and X is one of two everyday objects, e.g., “I see at least 4 glasses”. The stimuli sentences were in Hebrew, which allowed us to control for frequency effects: in English, *at least* is much more frequent than *at most*, a fact that could conceivably affect behavior, resulting in a frequency confound. In contrast, Hebrew has two forms (*lexol hapaxot* ‘at least’ and *lexol hayoter* ‘at most’) with roughly the same (low) frequency. For completeness, we also added the much more frequent form *lefaxot* ‘at least’.

There were five experimental conditions, corresponding to three superlative quantifiers and two comparative ones, each consisting of 12 trials. Each trial included a written sentence presentation on a computer screen, which was simultaneously accompanied by a picture. All stimuli were counterbalanced and randomly distributed. We tested 28 Hebrew speaking adults.

**Experiment A - results and discussion:** Findings from previous studies demonstrating that SQs require substantially longer reaction times were replicated (Geurts et al. 2010, Cummins and Katsos 2010). More importantly, within superlative quantifiers, we observed a significant difference (*P*<0.02) between mean reaction time for judgments of truth (2613ms) and falsity (2360ms), as predicted by the illocutionary theory. The interaction between quantifier (*lexol hapaxot, lexol hayoter, or lefaxot*) and truth judgment (true or false) was not significant, indicating that all three SQs demonstrate a similar effect. No effect was observed for comparative quantifiers. These findings support the illocutionary theory.

One might conceivably claim, however, that our results could also be made compatible with the epistemic or disjunction theory, if it could somehow be demonstrated that the judgment of the logical form assumed by the theory requires more time for truth than for falsity. In order to rule out this possibility, we carried out a second experiment, in which we gave subjects sufficient time to process the sentence, including its implicature, before presenting them with the picture. The illocutionary theory now predicts that the effect will disappear, since the computation of the implicature can be carried out prior to the (truth) judgment itself. In contrast, the competing theories, even with the added assumption made above, would predict the effect to remain essentially unchanged.

**Experiment B - methods:** The same materials and procedure were used, with the exception that the sentence preceded the corresponding picture by 2 seconds. 27 Hebrew speaking adults were tested in this experiment.

**Experiment B – results and discussion:** A small difference between mean reaction times for truth (1750ms) vs. falsity (1883ms) was recorded. This difference is not statistically significant (*P*<0.8), which provides further support for the illocutionary theory.

**Conclusion:** Cohen & Krifka’s illocutionary theory predicts that reaction time to judgments of true SQ sentences will be longer than that of false SQ sentences, and that this difference will disappear if subjects are given sufficient time to compute the implicature prior to the judgment. These predictions are borne out by our experiments. Since Cohen & Krifka’s proposal crucially involves speech act operators in the scope of quantifiers, our findings can be seen as further evidence for the thesis that speech acts are full-fledged participants in the semantic game.
References
Bott, L., and Noveck, I. A. 2004. `Some utterances are underinformative:The onset and time course of scalar inferences. Journal of Memory and Language 51.437-457
Language83:533-559.
I. Introduction Most investigations of quantifier scope are concerned with the range of possible scopes for sentences with multiple quantifiers like (1)-(2) below. Instead, this study examines the actual scopes, i.e., the pragmatics of quantifier scope disambiguation. Although actual usage facts fall outside of semantics proper, we are interested in them because they too are facts about natural language interpretation. Moreover, they are facts which provide indirect evidence about the semantics of quantifiers, e.g. if some has a stronger preference for narrow scope than a, then their semantic representations should arguably be different.

1. Each &1_S_each# tape is to be assigned to a different &2_to_a_differen# time slot.
2. Each &1_S_each# professor has one or more &2_O_one_or_more# specialities.

Based on introspective judgments and previous work (e.g. [1], [5], [4]), we expect structural factors such as linear order (LIN.ORD) and grammatical function (GRAM.FUN) to influence scope. The lexical realization of quantifiers (LEX.REAL) has been overlooked in computational work (e.g. [2]), but we also expect it to contribute to scope disambiguation (e.g. [3]).

While introspection can provide evidence for these 3 individual factors, it cannot readily address the interactions between them. For example, (i) does GRAM.FUN affect preference for wide scope independently of, i.e., while controlling for, LIN.ORD – and vice-versa? This is particularly hard to answer in English since LIN.ORD and GRAM.FUN are highly correlated. In addition, (ii) do the two factors LIN.ORD and GRAM.FUN interact when they affect wide-scope preferences? Finally, (iii) does the lexical realization (LEX.REAL) of a quantifier, e.g., each vs. every vs. all have a significant contribution to preference for wide scope independently of LIN.ORD and GRAM.FUN? These questions are impossible to answer by introspection and impractical to answer via psycholinguistic methods unless the investigation is restricted to particular contrasts, e.g. focusing solely on every vs. each or on a vs. some.

II. The Corpus To address these questions, we assembled a corpus of sentences from LSAT logic puzzles and tagged it for quantifier scope. Logic puzzles are ideal for such an investigation because sentences with two or more quantifiers are frequent in this register. In addition, ambiguity must be minimal because test takers are expected to select a single correct answer. Finally, the LSAT explicitly states assumptions that might be left to world knowledge or shared discourse context in ordinary conversation. This aspect of the corpus allows us to control for the role of world knowledge to a great extent.

As shown in (1)-(2) above, each quantifier in a multiple-quantifier sentence was tagged for: (i) SCOPE – 1, 2, . . . ; in cases where no truth-conditional difference was clear, the felicity of “such that” paraphrases provided the ultimate criterion; (ii) GRAM.FUN – we distinguished Subject, Object, Pivot and Adjunct, as well as individual prepositions; (iii) LEX.REAL – we tagged the entire complex determiner in cases like more than two or a different. Though not explicitly tagged, LIN.ORD was recoverable from relative tag order.

III. Modeling and Results We focus on sentences where: there are exactly two quantifiers, the quantifiers interact scopally (i.e., no cumulative readings), and at least one is S or O. Given that some sentences have both an S and an O quantifier and the scope of one completely determines the scope of the other, we randomly sample one quantifier from each of these sentences to avoid double counting. Final dataset: 348 quantifiers / observations.

The paper compares several models, but we focus here on one (fitted with the lme4 package in R and in WinBUGS with vague priors): a mixed-effects logistic regression with (i) SCOPE (2 levels: narrow, wide; “success” level: wide) as the response variable, (ii) two fixed effects LIN.ORD (2 levels: first, last; reference level: first) and GRAM.FUN (2 levels: S,
O; reference level: S) and (iii) one random effect LEX-REAL (17 levels: a, a.different, …). Lexical realizations are modeled as random effects because our sample does not exhaust the population of English quantifiers and, in addition, various lexicalizations are closely related to one another, e.g., each, every and all or modified numerals like exactly 2 and at most 2.

The frequentist estimates for the fixed effects (formula for the glmer() function: SCOPE ~ LIN.ORD + GRAM.FUN + 1|LEX-REAL) are: INTERCEPT (i.e., first and S) 1.45 (p = 0.0108), LIN.ORDLAST −3.75 (p = 2.23e−08) and GRAM.FUNO −1.04 (p = 0.0502). As expected, being first and a Subject increases preference for wide scope, while being last or an Object decreases it. There was no significant interaction between LIN.ORD and GRAM.FUN.

We can more easily understand the estimates and their associated uncertainty in terms of wide-scope probabilities. The 4 figures below display the posterior distributions (obtained in WinBUGS) of the preference for, i.e., probability of, wide scope for the two grammatical functions S and O and the two linear-order positions first and last—together with the median probability (red) and the 0.025 and 0.975 quantiles (dark red); for presentational convenience, the figures display wide-scope preferences for the lexical item both.

The median preference for, i.e., probability of, wide scope for all 17 quantifiers instantiated in the corpus and the 0.025 and 0.975 quantiles are provided in the leftmost figure below (based on the posterior samples obtained in WinBUGS). We also provide the full probability distributions for the two quantifiers at the extremes of the spectrum, a.different and each (for presentational convenience, the figures below display wide-scope preferences when the lexical items occur first and as Objects): a.different (on its sentence-internal reading) has a strong preference for narrow scope, while each has a strong preference for wide scope.

The model with LEX-REAL random effects in addition to the LIN.ORD and GRAM.FUN fixed effects has a higher predictive adequacy than the model with fixed effects only (Somers’ Dxy is 0.9 with random effects and only 0.68 without), indicating a much more important role for lexical semantics in scope disambiguation than previously assumed (e.g. by [6]).

The symmetry of the standard post-Aristotelian square of opposition conceals two significant asymmetries along both its horizontal and vertical axes. First, as Aristotle recognized, the negative values (none, not all) are less basic or informative than their affirmative counterparts (all, some), an asymmetry directly reflected by the systematic imbalance in the lexicon of quantificational and modal operators in natural languages. More significantly, while the CONTRARIETY relation between all and none is well established in the canonical texts, the corresponding “subcontrary” relation between some and some not/not all was not originally recognized as an opposition, much less assigned a label. In fact, natural languages tend to maximize contrariety, the relation between two propositions that can be simultaneously false but not simultaneously true. On the other hand, SUBCONTRARIETY, the relation between two propositions that can be simultaneously true but not simultaneously false, tends not to be directly represented, in part due to the tendency for the corresponding subcontraries (Some F are G, Some F are not G) to “collapse together” pragmatically in terms of information conveyed (Fogelin 1967, Horn 1972).

This presentation focuses on the tendency for contradictory (apparent wide-scope) negation to be semantically or pragmatically strengthened to contrary readings whenever possible, instantiating the dictum that “The essence of formal negation is to invest the contrary with the character of the contradictory” (Bosanquet 1911). Contrariety in this sense naturally extends from the “universal” values corresponding to the A and E vertices of the traditional square to other values that are INTOLERANT in the sense of Löbner (1987), while subcontrariety extends to TOLERANT operators: An operator is (in)compatible with its inner negation. Thus, the operators in (1a) are tolerant, while those in (1b) are intolerant.

(1) a. **Some** of my friends are linguists and **some** of them aren’t.
   Many of my friends are linguists and **many** of them aren’t.
   He **often** goes to church on Sunday and he **often** doesn’t.
   It’s **possible** that she’ll win, and **possible** that she won’t.
   It’s **50-50** that it’ll land heads, and **50-50** that it won’t.

   b. **#All** of my friends are linguists and **all** of them aren’t.
   **#Most** of my friends are linguists and **most** of them aren’t.
   #He **usually** goes to church on Sunday and he **usually** doesn’t.
   #It's **likely** that it’ll land heads, and **likely** that it won’t.
   #It's **certain** that she’ll win, and **certain** that she won’t.

I first review some effects of this tendency touched on in earlier work (Horn 1972, 1989):
(i) The failure of outer negation of an intolerant operator to incorporate lexically. Thus, while we can have Eng. nobody = ‘everybody ¬’ (~somebody), nor = ‘and not’ (~not or)
seldom or rarely = 'usually ¬' (¬ often'), unlikely = 'likely ¬', there are no lexicalizations of ¬ all (= nall), ¬ usually, ¬ and, ¬ likely.

(ii) The tendency for negative subcontraries (O values and corresponding tolerant negatives) to strengthen to contrary values, as with il ne faut pas partir—literally = 'It is not necessary to leave', but which can now mean only that one must not leave.

I then move on to consider the role of the disjunctive syllogism—

\[
\begin{align*}
\phi \lor \psi \\
\neg \phi \\
\therefore \psi
\end{align*}
\]

—in garbing contrary negation in contradictory clothing, as observed by Bosanquet. While the key disjunctive premise is typically suppressed, the role of the disjunctive syllogism can be detected in a variety of pragmatic (or semantic) strengthening effects in natural language, where the disjunction in question is pragmatically presupposed in relevant contexts. The phenomena in question include:

(iii) The tendency for negation outside the scope of (some) negated propositional attitude predicates (e.g. a believes that \(\phi\)) to be interpreted as associated with the embedded clause, i.e. NEG-RAISING, given the assumed disjunction that the subject will have made up her mind as to whether \(\phi\) or \(\neg \phi\) is the case (Bartsch 1973):

\[
\begin{align*}
F(a, \phi) \lor F(a, \neg \phi) & \quad \text{the assumed or covert disjunction} \\
\neg F(a, \phi) & \quad \text{the proposition asserted} \\
F(a, \neg \phi) & \quad \text{the proposition conveyed, via disjunctive syllogism}
\end{align*}
\]

(But see Horn 1978, 1989; Klooster 2003; Gajewski 2007 for complications).

(iv) The tendency for a semantically contradictory negation of an unmarked positive value, whether affixal (x is unfair/unhappy) or clausal (x is not fair/happy), to be strengthened to a contrary, whether via pragmatics (LITOTES) or via lexical semantics.

(v) The tendency for a negated plural definite (The papers aren't in Italian) to be understood as a contrary rather than a contradictory of the corresponding affirmative (The papers are in Italian), reflecting the HOMOGENEITY PRESUMPTION (Löbner 2000), another form of the covert disjunction.

(vi) The tendency for a weak (primary) implicature—

Given a scale \(<S, W>\), a speaker a uttering “...W...” is taken to Q-implicate \(\neg B_a(\ldots S\ldots)\) to be strengthened to the corresponding strong (secondary) implicature—

Given a scale \(<S, W>\), a speaker a uttering “...W...” is taken to Q-implicate \(B_a(\ldots S\ldots)\).

(See Mill 1867, Horn 2006, and Geurts 2009 on the COMPETENCE ASSUMPTION.)

(vii) The extension of this argument to fend off “localist” attacks (e.g. Chierchia 2004) against a standard Gricean account of so-called embedded implicature (Russell 2006; Geurts 2009, 2010), given that “we may assume that the competence assumption holds not only for the speaker but for the subject of the belief report” (Geurts).

(viii) Children’s adoption in word learning tasks of MUTUAL EXCLUSIVITY, i.e. the preference for mapping a novel lexical item to an unfamiliar object, one lacking a familiar label (Clark 1987, Markman & Wachtel 1988, Bloom 2000), a preference that can be seen as an instance of the disjunctive syllogism (Grassmann to appear).
It will be shown that these apparently quite diverse phenomena—constraints on lexical incorporation, semantic change, usage shifts, pragmatic inference, rhetorical force, on-line scope adjustments, word learning strategies—can be reconciled as instances of the general preference for contrariety, characteristically involving a deployment of the disjunctive syllogism. Nor should this be surprising in the light of the celebrated argument of Chrysippus (3d c. BCE) that a dog chasing a rabbit, after realizing that the rabbit didn’t go down Paths A or B, will unhesitatingly pursue it down the remaining Path C, thus employing the disjunctive syllogism (cf. e.g. Aberdein 2008).

References
Horn, Laurence R.


In his “Remarks on Colour” (1977: III-46), Ludwig Wittgenstein writes: “Among the colours: Kinship and Contrast. (And that is logic).” In this talk on quantifiers, colour percepts and colour concepts, proof will be provided that Wittgenstein’s intuition about the logical nature of colour relations has to be taken literally. Relations of opposition between primary and secondary chromatic percepts will be shown algebraically to be the same as those between the three primary items (all, some, none) in the lexical class of quantifiers and their contradictories. An important area where evidence can be found for this intriguing case of homology is that of lexicalisation asymmetries between different kinds of colour terms and different types of quantifiers. On the one extreme, certain lexicalisations are nonexistent and arguably cognitively nonnatural, such as Horn’s (1989) O-corner quantifier *nall and chromatic nonce-formations such as *reen (= “reddish green”), while at the other end of the lexicalisation spectrum there are items which are fully natural ordinary language terms (all, white, etc.), universally lexicalised by what will be identified as innate conceptual pressure. In between those two extremes, so it will be argued, there is a middle category with differences in lexicalisability. Items in this category, including such specialised terms as nand in logic and such nonbasic colour terms as cyan or magenta, reveal that differences in utility determine whether such items are lexicalised or not. Take the example of nand, which is an established term in the context of electronic circuits in engineering but unattested in everyday natural language. In principle, all types of logic gates (OR, AND, etc.) could be created equally as well from a network of NOR gates as from a NAND gate network. Yet the latter carried the day because the available technology made it easier to implement than a NOR gate network. It is for this extra-linguistic reason that the language of science successfully adopted the useful electronic engineering term “nand”, which has nonetheless remained equally as nonnatural in the natural logic of ordinary language as before. Thus, no natural language has structures such as (the equivalent of) *John nand Bill ate a piece of cake, meaning “either John, or Bill, or neither of them ate a piece of cake”.

Let us first describe the nature of the isomorphism postulated between quantifier and colour oppositions in more detail and then distil the different degrees of naturalness of lexicalisations from the two isomorphic models, with a special focus on the in-between category where utility helps overcome cognitive nonnaturalness for usage within a more restricted domain of application. This is a distinction reminiscent of the technical vs. natural-approximative opposition between more than half and (one of the meanings of) most.

The most straightforward historical starting point for a comparison of chromatic and predicate logical oppositions is the colour wheel on the one hand and the Aristotelian-Boethian logical square of oppositions on the other. Since the colour wheel and the square have different geometries, however, more converging models have to be called upon. On the logical side, the model in question is a logical hexagon based on work by Blanché (1953, 1969). This hexagon is in fact the bitriangle (1a), which is derived by adding to an AYE-triangle of contraries (A: all; Y: some, but not all; E: no(ne)) its dual IOU (I: some or all; O: some or none; U: all or no), i.e. a triangle of subcontrary relations (Moretti 2009, 142).
Algebraic definitions for the logical oppositions in this Blanché-star – i.e. for entailment (arrows), contradiction (red lines), contrariety (blue lines) and subcontrariety (green lines) – carry over to the realm of colour percepts, which can be mapped onto the same hexagon (1b). Let us see how.

In view of Thomas Young’s trichromatic theory of colour vision and its later refinements, the triangle of contraries will be shown to involve the primary colours red-green-blue. Experiential associations of blue with dark and cold (negative) versus yellow with light and warm (positive) and red with warmest, but also subject-independent asymmetries such as wavelength properties will provide arguments to link LW red with the A-corner, MW green with the Y-corner and SW blue with the negative E-corner. The other three colours in the star are secondary and composite: at the physical and perceptual level they represent the combined wavelengths – in set-theoretical terms: the union of the wavelengths – of green and red (= yellow), green and blue (= cyan) and red and blue (= magenta) respectively. As is known from the phenomenon of afterimages and from insights in Hering’s opponent process theory, the secondary triad yellow-cyan-magenta consists of complementaries of the primary colours, a type of opposition that is the colour equivalent of contradictoriness in the logical square (and hexagon) of oppositions. Thus, two quantifiers (e.g. no and some, maybe all) are contradictories if the intersection of their denotations is the null set and the union the universe; correspondingly, two colour percepts (e.g. blue and yellow) are complementaries if the intersection of the wavelengths (or activated cone types) involved is the null set – yielding a black percept – and the union is the universe of wavelengths (or cone activation), i.e. white.

Returning to the issue of the different degrees of naturalness of lexicalisations, two facts stand out in the stars. First of all, the operators that resist natural single-item lexicalization in logic (O-corner not all, for which there is no natural lexicalization *nall and U-corner all or no, for which there is no *allno) have counterparts in the realm of colour terms which are not naturally lexicalized either, in casu O-corner cyan and U-corner magenta, cognitively dyadic and less natural than red, green, blue, yellow. Yet, although in both stars the occupants of the relevant corners are nonnatural, there is a marked difference. The occupants do exist as specialised terms in the colour star, whereas they are completely absent in the case of the quantifiers. This is attributable to the difference in nature between colour percepts and logical concepts. While the former are cone activation patterns reflexively generated by the vision system, the latter involve more abstract entities. That the lexicalisation difference is due to this difference is reinforced by the fact that the colour star has to be expanded to accommodate the two achromatic colours black and white. These two correspond to neither all, nor some, nor no (whose denotation is the null set) and some or all or no (the universe) respectively in the predicate logical system. While the two logical operators involved are trivial and hence useless for any contingency, the null set of cone activation in the colour system does generate a
functional colour percept (black), which is not a useless or noninformative representation at all. In sum, the natural language lexicon contains at least a quantificational but probably a wider range of closed lexical fields (kinship terms, person, numerals) susceptible to a decompositional analysis in terms of the system of oppositions that characterizes colour percept discrimination. The nature of the opposition system results in a distinction between cognitively natural and nonnatural lexical items, the latter of which can only make it into specialised language, and then only in those cases where their adoption is facilitated by extra-linguistic utility.

References

Moretti, A. (2009), The Geometry of Logical Opposition, PhD dissertation, Université de Neuchatel, Switzerland.
A Superlative Reading of Most
Hadas Kotek, Yasutada Sudo, Edwin Howard and Martin Hackl

Introduction. Recent studies ([1], [2]) have proposed competing analyses of the determiner most, using results of verification studies as key evidence (based on the ITT, 1). [1] argues that most is a superlative construction and that verifying most statements effectively involves the comparison of |A ∩ B| and |A − B|. For expository purposes, we take (2) to be a sufficient approximation of this view. [2] proposes an alternative, (3), based on the observation that under extremely short presentations of a visual scene the complexity of A−B does not affect the accuracy of most statements. (3) accounts for this insensitivity because A−B is not directly used in the verification of most(A)(B); rather, its cardinality is computed from |A| and |A ∩ B|.

(1) The Interface Transparency Thesis (ITT)
The verification procedures employed in understanding a declarative sentence are biased towards algorithms that directly compute the relations and operations expressed by the semantic representation of that sentence.

(2) \[ \text{[most]}(A)(B) = 1 \text{ iff } |A \cap B| > |A − B| \]

(3) \[ \text{[most]}(A)(B) = 1 \text{ iff } |A \cap B| > |A| − |A \cap B| \]

This study provides new experimental evidence in favor of an analysis of most as a superlative. The evidence comes in the form of a hitherto unnoticed latent superlative reading of sentences with most in subject position, like most of the dots are blue.

Our study. We conducted an experiment that compared the verification of quantified statements containing most and more than half. We take more than half as a baseline against which the two competing representations for most can be compared. Specifically, we assume that the semantics of more than half explicitly references A and A ∩ B, but not A−B, (4). If the correct representation of most is as in (3) then the same entities are involved in the verification of most and more than half, and we would expect them to be similarly affected by a manipulation of the complexity of A−B. If the correct representation of most is as in (2), we expect the manipulation to affect most differently than more than half.

(4) \[ \text{[more than half]}(A)(B) = 1 \text{ iff } |A \cap B| > \frac{1}{2} |A| \]

Methods and Design. We used the Self-Paced Counting method as in [1]: 51 participants verified statements such as (5)-(6) relative to arrays containing 13-18 dots displayed on a computer screen. The dots were uncovered in groups of 2 or 3 as participants pressed the spacebar. The dependent measure is the accuracy rates of the answers. To test whether the complexity of A−B affects verification we followed the approach in [2] and varied the number of colors that dots in A−B may have. Arrays thus contained either two colors (2-Color condition) or three colors (3-Color condition).

(5) Are most of the dots blue?
(6) Are more than half of the dots blue?

Results. We observe two surprising results: 1. The accuracy rates for the true items of both of the proportional determiners are surprisingly low. 2. Participants answered True more often in the 3-Color condition than in the 2-Color condition for most, but the accuracy rates for more than half remain the same across both Color conditions. That is, we observe a Truth×#Colors interaction for most (p<0.05, contrast-coded Mixed Logit Model with random subject and item effects) but not for more than half. Figure 1 shows the accuracy rates in the experiment for most, more than half and more than n broken down by the Color condition and by True vs. False.
Discussion. We suggest that the increase in the number of True answers to most statements in 3-Color arrays can be explained by a superlative analysis of most as in [1] and is unexpected under [2]. When more than two distinct subsets of dots are salient in the context, most is ambiguous between a dominant proportional reading, which is equivalent to more than half, and a latent superlative reading, according to which a sentence like (5) is true iff $|A \cap B|$ is greater than the cardinality of all contextually salient subsets of $A\!\!B$. As a result, in a 3-color array with the ratio 7:4:4, (5) is considered false under a proportional reading but true under a superlative reading. A corresponding 7:8 array in a 2-color condition is always judged false because the sentence is false under both readings.

Our experiment indicates that there is tension between the bias to choose the interpretation that makes the sentence true (the Principle of Charity, which states that when faced with an ambiguous sentence that is true on one reading and false on the other, language users prefer the reading of the sentence that makes it true) and the difficulty in revising the initial parse of a sentence (Dominance), [3]. It appears that in the competition between Dominance and Charity, one principle wins on some occasions and the other wins on other occasions, explaining the increase in true answers in the 3-Color condition.

We suggest (7), an extension of the analysis in [1], as the meaning of most of the As are Bs. It will return true iff there is a plurality X that satisfies both A and B and that has more atomic parts than any non-overlapping alternative plurality Y in a comparison class C. Our findings can be explained under the assumption that the setting for C is by default as in (7a) and that a setting as in (7b) can be achieved in rich enough contexts that suggest a natural partition of the domain of quantification. Under extremely short presentations of a scene the partitioning of the domain is not supported and hence [2] do not observe sensitivity to the number of colors in their arrays.

(7) $\exists X \left[ A(X) \land B(X) \land \forall Y \in C \left[ Y \neq X \rightarrow |X| > |Y| \right] \right]$

Presupposition on C: C contains at least two non-overlapping elements.

a. $C = A$ (A closed under i_sum-formation) proportional reading

b. $C = \{X: X \in A \land (X \in D_1 \lor X \in D_2 \lor \ldots)\}$ superlative reading

(where $D_1, D_2, \ldots$ = subsets in a salient partitioning of $A$, C is a cover of A)


The typicality structure of quantifiers

Bob van Tiel

A speaker who says that he ate some of the cookies usually conveys that he didn’t eat all of them. How does this scalar inference come about? The classic, Gricean answer is that scalar inferences involve reasoning about the speaker’s beliefs and alternative utterances. In the example at hand, if the speaker ate all of the cookies, he should have said so. Since he didn’t, we may infer that he didn’t eat all of the cookies. Recently, some authors have challenged the Gricean account, proposing that scalar implicatures are caused by a syntactic operation instead (e.g. [2]). Two pieces of evidence have been launched against the Gricean account of scalar inferences.

First, scalar inferences seem to occur in embedded positions, contrary to the predictions of the Gricean account.

(1) Every square is connected to some of the circles.

?~> Every square is connected to some but not all of the circles.

Conventionalist experiments have confirmed that participants prefer the above sentence to describe a situation that verifies the embedded scalar inference that no square is connected to all of the circles [1, 3]. According to the conventionalists, this is ample evidence that embedded scalar inferences exist, thus necessitating a departure from the Gricean account.

Second, scalar inferences seem to be derived during sentence processing. On a Gricean account, the input for pragmatic reasoning is the proposition expressed by the whole sentence. Hence, scalar inferences should not occur before the whole sentence has been processed. Nonetheless, when participants hear the word some, they focus on the item that verifies the scalar inference even before the sentence has ended [5, 6].

Both arguments tacitly assume that a preference for a representation and visual focusing are reliable indicators of the derivation of an inference. I contest this assumption and argue that the results of the preference and focusing tasks are due to a typicality effect, rather than an inference. Though a situation in which no square is connected to all of the circles is most typical of (1), people don’t infer that this situation holds upon hearing the sentence. Three arguments support my interpretation.

First, participants don’t consider a sentence false in case the embedded scalar inference is falsified [4]. In this, it behaves different from normal inferences. Second, some significantly different preferences are manifestly not due to a scalar inference. For example, in the case of (1), a situation in which two squares are connected to all of the circles and four to only some of them is preferred over a situation in which the distribution is reversed. Similarly, participants focus on the bigger item when they hear the word big, though this is clearly not a scalar inference [7]. Third, I replicated Clifton & Dube’s [3] preference task, using category nouns (bird, vehicle, . . . ) and continuous adjectival scales (big, rich, . . . ). Such words give rise to typicality effects but not to scalar inferences. That is, a robin is more representative (i.e. more typical) of the category associated with bird than an ostrich, but nobody would infer upon hearing the word bird that the speaker is talking about a robin. Nonetheless, these words yield robust preference orderings exactly like those found for (1).

So a typicality explanation has at least some evidential plausibility. It remains to be shown that a situation in which no square is connected to all of the circles is indeed the most typical representation of (1). The first step towards this conclusion involves the
I argue that a set containing less than half of the circles is most representative of this category. Since representativeness is defined operationally by means of subjects’ ratings of how good a representation is of its category [8], I conducted a rating task in which participants were asked to indicate how well the following situations represent the sentence *Some of the circles are dark*.

As expected, situation A was judged the most representative instance, followed by B, C, and ultimately D. This finding, which can be backed up with more theoretical arguments, makes sense of the results of the focusing tasks. Since participants tend to focus on more representative instances, they pay attention to sets that don’t contain all elements upon hearing some [7]. This, however, is not a genuine inference.

Participants were subsequently asked to rate the following four situations for the sentence *Every circle is black*. For this sentence, it was found that D is the most representative situation, followed by C, B, and ultimately A. This suggests that the typicality structure of a universally quantified sentence is approximated by the mean typicality of the elements with respect to the predicate.

First off, these findings show that even the categories denoted by quantifiers have a typicality structure. In this, it bridges the gap between the typicality literature, which has focused almost exclusively on categories denoted by nouns, and the linguistic literature, which has largely ignored the issue of typicality effects. It is widely known that typicality effects have an impact on practically every dependent variable used in psychological research: reaction times, ease of learning, eye-tracking, production probability, et cetera [8]. Hence, many psychological experiments using some as a target have attributed their results too eagerly to scalar inferences. To begin with, it can be shown to follow from the typicality structure of some and every that a situation in which no square is connected to all of the circles is most representative of (1). Hence, the preference for this situation comes as no surprise.

Importantly, these differences in representativeness aren’t inferences. After all, nobody would conclude from the sentence *This is a bird* that the bird in question is a robin. The preference and focusing effects, taken to disprove the Gricean approach to scalar inferences, can be fully explained as typicality effects. This explanation, which is moreover able to deal with the three problems identified earlier, salvages the Gricean account from its conventionalist competitor.

**References**

van Tiel, Bob

Schizophrenic patients find proportional quantification difficult due to working memory deficits

Marcin Zajenkowski¹, Rafał Styla¹, and Jakub Szymanik²

¹ Department of Psychology, University of Warsaw
² Institute of Artificial Intelligence, University of Groningen

Research on normal language comprehension suggests that working memory and comprehension are strongly linked (Just and Carpenter, 1992). A specific relationship between working memory and language comprehension has also been found in studies devoted to quantifier processing. For example, McMillan et al. (2005) examined the pattern of neuroanatomical recruitment while subjects were judging the truth-value of statements containing natural language quantifiers. The authors considered two standard types of quantifiers: first-order (e.g., ‘all’, ‘some’, ‘at least 3’), and higher-order (e.g., ‘an even number of’, ‘more than half’). They concluded that all quantifiers recruit the right inferior parietal cortex, which is associated with numerosity, but only higher-order quantifiers recruit the prefrontal cortex, which is associated with executive resources, like working memory.

Recently, the processes underlying quantifier verification have received increased attention (see e.g. Geurts et al., 2009). Szymańek (2007, 2009) proposed that the cognitive difficulty of quantifier processing could be assessed on the basis of the complexity of the corresponding minimal automata that could handle the computational task. Szymańek and Zajenkowski (2010ab) investigated this hypothesis in empirical studies by comparing the processing of various classes of quantifiers with respect to their computational complexity. The authors concluded that proportional quantifiers are the hardest to verify and engage working memory to the highest degree. From a theoretical perspective, they require a recognition mechanism with unbounded internal memory (Van Benthem, 1986). During computation, the sizes of two sets need to be compared that may be simulated by a push-down automata. For instance, in order to verify the sentence “More than half of the cars are red”, one has to count and hold in the short-term memory the number of red cars and then compare it with the total number of cars. No such memorization/comparison is necessary when processing other quantifiers (cf. Hackl 2009, Pietroski et al. 2010). In that paper we provide more evidence on the relationship between cognitive difficulty and computational complexity in quantifier verification.

In parallel, research on cognitive impairments indicates the existence of working memory deficits among patients with schizophrenia (Lee and Park, 2005). Condrey et al. (1996) found that patients with schizophrenia did not perform as well as a control group in a reading span test (which tests working memory) and a language comprehension task. Moreover, they observed that the two tasks were strongly correlated.

Taking working memory deficits in schizophrenic for a starting point, we predicted that when asked to verify quantifier sentences with Aristotelian, numerical, parity, and proportional quantifiers a group of patients with schizophrenia and a healthy control group would diverge to the greatest extent when dealing with proportional sentences, because they engage working memory to the highest degree. Moreover, McMillan et al. (2005) found that higher order quantifiers (including proportional quantifiers) recruited the right dorsolateral prefrontal cortex, the same area that is associated with memory deficits in schizophrenia (Seidman et al., 1994).

We observed that the greatest difference between the controls and the patients appeared when they were faced with proportional quantifiers. The patients generally took more time to verify...
the sentences. However, they were less accurate only when dealing with proportional sentences. Presumably, the longer processing allowed the patients to verify Aristotelian, numerical and parity quantifiers at the same level as the controls. This is consistent with the theory that memory deficits in schizophrenia may be partly accounted for by the slowing of processing speed (Brebin et al., 1998). However, in terms of proportional quantifiers, slower processing did not enable the patients to match the controls’ scores, as the verification of the statements required a different cognitive mechanism. According to the computational theory, the high engagement of working memory, which is necessary for comparing sizes of two sets, hinders the verification of proportional sentences. Switching between processing and sustaining stored information may be too distracting for individuals with schizophrenia, especially given that they show deficits in such executive functions as control or the supervision of cognitive processes (Hutton et al., 1998). In other words, it seems that patients are unable to use more complex cognitive strategies associated with high demands on working memory to an adequate extent.

References
Availability of Quantifier Scalars

Natalia Zevakhina

Over the last 40 years, quantifier use has become an essential topic in formal and experimental pragmatics. It is widely assumed that quantifiers some, all, like many other linguistic expressions, form a scale based on informativeness (< some, all >). Use of a less informative (weaker) scale-mate implies negation of a more informative (stronger) one. Such an inference is called scalar implicature (SI).

There has been a tacit consensus in the theoretical literature that all scalar terms behave similarly with respect to the derivation of SIs, i.e., they constitute a unified class. This is in line with the defaultist assumption that SIs are associated with certain lexical expressions and computed by default in “normal” contexts. However, these implicit assumptions were only based on the theoretical observations of the two scales (quantifiers and connectives).

Starting with Noveck (2001), quantifiers have been highly frequently tested scalars. Since no significant difference was found between quantifiers and neither epistemic modals <might, have to> (see, e.g., Noveck 2001), nor aspectuals <start, finish> (see, e.g., Papafragou and Musolino 2003), the intuition that scalars are alike was maintained. However, the rates of recovered SIs (from 25% to 65%) in different verification tasks overviewed in Geurts (2010) already suggest potential diversity between the scalars. It is very likely that testing various kinds of scalars, both already experimentally examined (e.g., quantifiers, modals) and still being discarded (e.g., adjectives), will reveal significant differences between them.

Larson et al. (2009, in press) showed the mean for gradable adjectives lower than 25% whereas the mean for both quantifiers and modals was within the indicated bounds (from 25% to 65%). These results suggested diversity between the scalars though the aim the studies pursued was far from discovering that. They investigated the contribution of different kinds of inferences to the truth-conditional meanings of sentences. The methodology adapted to the aim of the experiments was not neutral for the issue in question.

Exploiting an inference task, I undertook an experiment with 50 English native speakers. Critical items were 2 pairs of quantifiers (<some, all>, <sometimes, always>), 2 pairs of modals (epistemic <may, will> and deontic <may, have to>), and 10 pairs of gradable adjectives (e.g., <warm, hot>, <intelligent, brilliant>). Participants had to answer “Yes” or “No” to questions about particular sentences that belonged to a fictitious person. Sentences contained weaker scale-mates (John says: “The water in the lake was warm.”), whereas questions had stronger scale-mates under negation (Would you infer from this that, according to John, the water in the lake was not hot?). I constructed 5 questionnaires, with 10 subjects assigned to each. An item was given a different instance in each questionnaire. 32 fillers were entailments and contradictions.

An ANOVA showed that the mean for gradable adjectives (M = .16, SD = .37) was significantly lower than the mean for quantifiers and modals (M = .87, SD = .33), F(1,654) = 527.76, p < .001. These two general patterns revealed diversity between the scalars and, in this sense, conformed with Larson et al.’s data. Furthermore, a post-hoc Tukey’s HSD test showed a significant difference between the adjective warm and each of the rest of the adjectives (all p’s < .001). Half of the rest of the adjectives received rates even close to 0% (e.g., <intelligent, brilliant>), while quantifiers and modals were evaluated at rates higher than 80%. These results supported the hypothesis that scalars exhibit a considerable diversity.

What is the reason for that? A likely explanation would be that some scale pairs are more available than the others. If a scale-mate of an available scale emerges, its stronger mate becomes relevant for a topic of conversation, as well, and, therefore, it is likely to show up,
too. Significant results of the data tested in minimal contexts make it even strong and plausible.

This stimulated me to undertake another experiment where I collected spontaneously generated alternatives to a targeted word. If there is a significant correlation between the two experiments, then it would suggest that some scalars are associated with each other and, therefore, are available. To verify this hypothesis, 60 English native speakers (12 subjects assigned per questionnaire) participated in the experiment. They had to give three alternative words to an underlined critical item (e.g., *The water in the lake was warm*). Critical items and fillers were the same as in the first experiment.

I did two correlation tests between Experiments 1 and 2. The first correlation test showed that there was a positive significant relationship between the numbers of SIs in Experiment 1 and the numbers of those alternatives in Experiment 2 that were identical to the stronger scale-mates in Experiment 1, $r = .673$, $p < .01$. The second correlation test revealed that there was a positive significant relationship between the numbers of SIs in Experiment 1 and the numbers of those alternatives in Experiment 2 that were stronger scale-mates but not necessarily identical to stronger scale-mates in Experiment 1 (e.g., *<many/much, most, all>* for *some*, *<hot, scalding/sweltering/steaming>* for *warm*), $r = .941$, $p < .01$. Remarkably, half of the participants responded with *hot* to *warm*, and this scale was again evaluated at the highest rate among the adjectives, as in Experiment 1. Therefore, Experiment 2 confirmed the results of Experiment 1.

Both experiments supported the hypothesis that SIs contingent on the availability of scale-mates are more determinate than the others. Such a distinction resembles the dichotomy of Qc- and Qo-implicatures proposed in (Geurts 2010). Qc-implicatures are triggered on the base of a closed set of relevant alternatives, whilst the derivation of Qo-implicatures might not be dependent on it. The distinction between the two categories of implicatures seems to be a plausible explanation of the SIs heterogeneity. The data suggest that quantifiers, modals (and aspectuals) yield Qc-implicatures, whereas adjectives tend to trigger Qo-implicatures. Speakers have various kinds of clues to establish a set of relevant alternatives. Quantifiers, modals (and aspectuals) exhibit a closed set of expressions and belong to special semantic fields not easily extended. Besides, they occur in special syntactic positions.

References