

What a Rational Interpreter Would Do: Building, Ranking, and Updating Quantifier Scope Representations in Discourse

Adrian Brasoveanu and Jakub Dotlačil*

¹ UC Santa Cruz, abrsvn@gmail.com

² Utrecht University/University of Groningen, j.dotlacil@gmail.com

Abstract

We frame the general problem of ‘rationally’ (in the sense of Anderson et al’s ACT-R framework) integrating semantic theories and processing, and indicate how this integrated theory could be explicitly formalized; an explicit formalization enables us to empirically evaluate semantic and processing theories both qualitatively and quantitatively. We then introduce the problem of quantifier scope, the processing difficulty of inverse scope, and two types of theories of scope, and discuss the results of a self-paced reading experiment and its consequences for these two types of theories. Finally, we outline how probabilities for LF construction rules could be computed based on the experimental results.

1 Introduction: ‘Rational’ theories of cognition

Anderson (1990) and much subsequent work argues for the following ‘rational cognition’ hypothesis (a.k.a. general principle of rationality): the cognitive system operates at all times to optimize the adaptation of the behavior of the organism. ‘Rationality’ is not used here in the sense of engaging in logically correct reasoning when deciding what to do. It is used in the sense of ‘*adaptation*’: human behavior is optimal in terms of achieving human goals. A ‘rational’, as opposed to ‘mechanistic’, approach to cognition is closely related to aiming for explanatory adequacy in addition to descriptive adequacy.

Developing a theory along the lines of the rational cognition hypothesis requires one to follow the six steps discussed in Anderson (1990: 29-30): **(1)** begin by precisely specify the goals of the cognitive system; **(2)** develop a formal model of the environment to which the system is adapted; **(3)** make minimal assumptions about computational limitations; **(4)** derive the optimal behavioral function given steps **1-3**; **(5)** examine the empirical literature to see if the predictions of the behavioral function are confirmed (if available; else do the empirical investigation); **(6)** finally, if the predictions are off, iterate. The theoretical commitments are made in steps **1-3**. They provide the “framing of the information-processing problem”. Steps **4-5** are about deriving and dis/confirming predictions. Finally, theory building is iterative: if one framing does not work, we try another.

Our goal in this paper is to get started with the first iteration of our rational analysis for a classical problem in formal semantics: *quantifier scope ambiguities*. In particular, we will study how interpreters deal with scope ambiguities during actual comprehension. The specific questions we are interested in are as follows. **(Q1)** How are quantifier scope ambiguities *represented*

*We want to thank Pranav Anand, Nate Arnett, Amy Rose Deal, Donka Farkas, John Hale, Roger Levy, Anna Szabolcsi, Matt Wagers and the UCSC S-Circle audience (Nov. 15, 2013). Adrian Brasoveanu was supported by a UCSC CoR SRG grant for part of this research. Jakub Dotlačil was supported by a Rubicon grant from the Netherlands Organization for Scientific Research for part of this research. The usual disclaimers apply.

by the interpreter? **(Q2)** How are these representations *built and maintained/updated* as the discourse is incrementally processed/interpreted? **(Q3)** Finally, how are these representations *ranked* so that the ambiguities are *resolved*? But what would it mean to provide a rational analysis for the problem of processing quantifier scope ambiguities? Paraphrasing the title of Hale (2011): what would a rational interpreter do?

In §2, we introduce the problem of quantifier scope and the difficulty of inverse scope, and we describe the results of a self-paced reading experiment targeting questions **Q1-Q3** above. In §3, we pick up the ‘rational’ analysis thread again and frame our information-processing problem, i.e., the parsing/interpretation problem, in detail. The main payoff of the detailed ‘framing’ is a much clearer understanding of the relation between semantic theories and the processor, so clear that explicit formalization of the connection between semantic theory and processing, as well as ways to do quantitative empirical evaluation, will be within reach. Finally, we will briefly outline how probabilities for LF construction rules could be computed.

2 Experimental investigations of quantifier scope

Consider the sentence in (1) below. The surface scope interpretation of this sentence is that there is a boy that lifted every box (the same boy lifted all of them); the inverse scope interpretation is that every box is such that a boy lifted it (a possibly different boy for each box).

- (1) A boy lifted every box.

A working definition of inverse scope that will suffice for this paper is that the interpretation of a quantifier is dependent on another quantifier that was introduced later (see Szabolcsi 1997 a. o. for a more precise definition). Importantly for us, inverse scope is costly relative to surface scope: it is harder to process (Pylkkänen and McElree 2006 and references therein). This is shown, for example, by the fact that a plural follow-up to (1) above, e.g., *The boys were looking for a marble* – which forces the inverse-scope reading – leads to increased reading times (RTs; Tunstall 1998 a.o.). The inverse scope interpretation is costly, hence the increase in RTs.

The previous literature leaves several issues open. Crucially, it focuses on sentences with only 2 quantifiers, as in (1) above. This might suffice to establish the cost of inverse scope readings but it doesn’t substantially help us understand how quantifier scope ambiguities are represented and maintained/updated by the interpreter. One could imagine at least two possibilities, which are often assumed in the literature: (a) the interpreter builds an LF representation that disambiguates scope readings; if the continuation is incompatible with it, the LF representation is revised accordingly (Pylkkänen and McElree 2006 and references therein); or (b) the interpreter builds a (mental/discourse) model structure, which is revised if the continuation is incompatible with it (Fodor 1982). One way to specify the model-based approach is to take indefinites to denote Skolem functions (or Skolemized choice functions) of variable arity (Steedman 2012): what gets revised then is the arity – and consequently the function.¹

We conducted two new experimental studies (eye-tracking and self-paced reading) to decide between these two possibilities. Here, we report only the self-paced reading experiment (see Dotlačil and Brasoveanu 2013 for the other experiment and details about the experimental designs of both experiments). The main novelty of the tasks: we examined the interaction of 3 quantifiers, 2 singular indefinites and 1 universal, in two-sentence discourses like (2) below:

¹The interpreter could also operate with underspecified structures/models (Ebert 2005 and references therein), but these theories have no clear way to explain inverse scope difficulty unless something else is added, e.g., that specifying scope relations is sometimes forced (mid-sentence) and is at least sometimes costly, so we’ll set them aside. See Radó and Bott (2012) for an experimental investigation of underspecification theories.

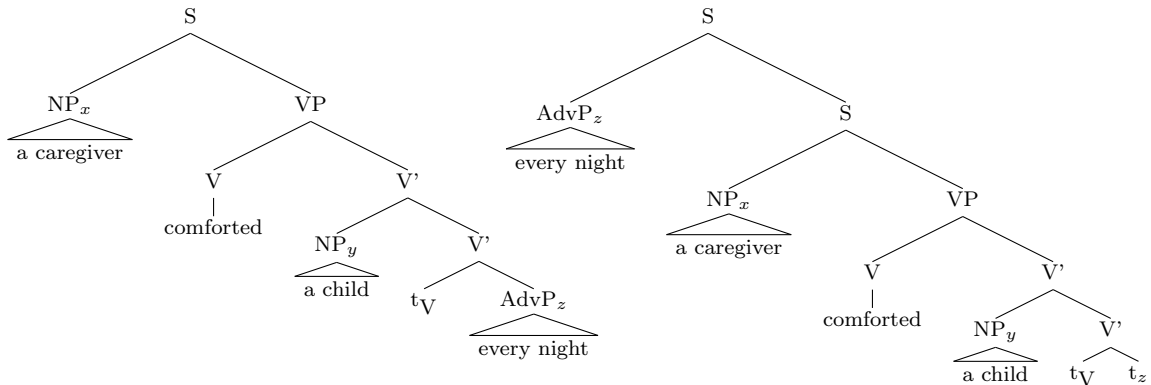
- (2) a. A caregiver comforted a child every night.
 b. The $\left\{ \begin{array}{l} \text{caregiver} \\ \text{caregivers} \end{array} \right\}$ wanted the $\left\{ \begin{array}{l} \text{child} \\ \text{children} \end{array} \right\}$ to get some rest.

The first sentence has 2 indefinites in SU and DO position and a universal quantifier as a sentence-final adverb. The second sentence elaborates on the entities brought to salience by the 2 indefinites. The only manipulation (that is relevant for our purposes; see Dotlačil and Brasoveanu 2013 for a much more detailed discussion) is morphological number on the SU and DO definites in the second sentence (2×2 design): the idea is that *singular* definite \Rightarrow *wide-scope* indefinite,² while *plural* definite \Rightarrow *narrow-scope* indefinite. The two theories of (inverse) scope we outlined above make the following predictions for this type of examples.

- (3) Predictions of the covert LF operations theory: (a) assume a base-generated structure with the universal adverb in the lowest position (Frazier and Fodor 1978, Larson 1988); (b) assume that the more operations we need to apply to obtain an LF, the less plausible/salient it is (Frazier 1978); (c) then: narrow scope SU \Rightarrow narrow scope DO.

Wide scope SU, wide scope DO:

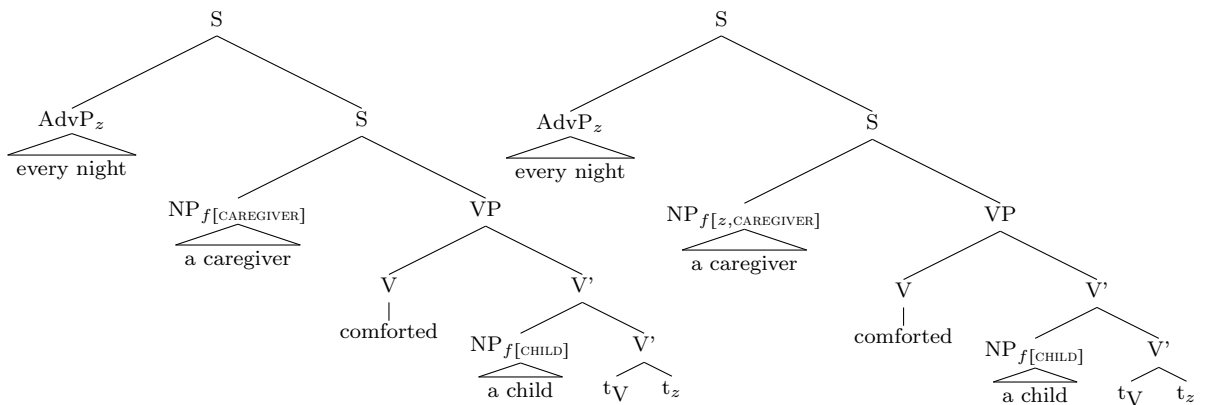
Narrow scope SU \Rightarrow narrow scope DO:



- (4) Predictions of the model revision theory: (a) assume that giving widest scope to the universal is costless, but setting the arities of the two Skolem functions is costly; (b) assume that the arities of the two Skolem functions are independently specified; (c) then: narrow scope SU \Rightarrow narrow scope DO.

Wide scope SU, wide scope DO:

Narrow scope SU \Rightarrow wide scope DO:



²Not necessarily wide scope: maybe narrow with 'accidental' coreference; we ignore this complication here.

We note that for presentational clarity, we postulated a very specific, LF-based theory but any theory that assumes a *scope hierarchy* (a strict total order: asymmetric, total and transitive) that has the DO by default in the scope of the SU will predict that narrow scope SU \Rightarrow narrow scope DO. This prediction is not made by theories that directly operate on models since DO scope is then independent of SU scope (as exemplified above using Skolem functions).

The experiment examined two-sentence discourses like (2a-2b) (henceforth CONTEXT:YES), but also their one-sentence counterparts consisting only of the second sentence (2b) (henceforth CONTEXT:NO). The main finding relevant for us is that in the CONTEXT:YES condition, the narrow-scope reading of the SU or the DO led to increased RTs, but there was a clear facilitation (observable in decreased RTs) when a narrow-scope DO followed a narrow-scope SU. That is, the inverse scope of the universal over the SU makes it easier to also interpret the DO as taking narrow scope. This facilitation cannot be due to the repetition of two plural forms because there is no facilitation in the CONTEXT:NO condition (in fact, this condition showed a borderline-significant *slowdown* in the region following the object for SU:PL & DO:PL condition). Thus, PL on the SU facilitates PL on the DO *but only* when the PL disambiguates scope. So the facilitation is (likely) due to the disambiguation role played by PL morphology.

These results are *incompatible*: (a) with the assumption that readers do not use disambiguating information quickly to reanalyze scope, (b) with (discourse/mental) model based theories of inverse scope – to the extent these theories do not keep track of some basic remnant of a grammatical/thematic scope hierarchy, (c) with related theories of scope, e.g., theories that take indefinites to denote Skolem functions/Skolemized choice functions of variable arity, or underspecification theories of scope – again, to the extent that specifying the scope of the DO is independent of specifying the scope of the SU in these theories.

The results are *compatible*: (a) with the assumption that the reanalysis is done on scope representations that can be specified in terms of LF/grammatical/thematic/linear order hierarchies, and (b) more generally, with the assumption that the processor builds hierarchical scope representations and updates/maintains them across sentential boundaries. Because of this, the results favor dynamic systems that have rich interpretation contexts like DRT (Kamp and Reyle 1993) rather than ‘less representational’ systems like DPL (Groenendijk and Stokhof 1991).

3 Framing the parsing/interpretation problem

These experimental results and their consequences help us understand how the interpreter builds and maintains scope representations, but we might want to do better. Theoretically, we left the connection between semantic theories and processing implicit but our conclusions/generalizations relied on a fairly tight connection between semantic theory and processing – how else could we link *behavioral measurements* in the experimental task and the *mental representations* postulated by our semantic theories? We don’t need to make this connection formally explicit for the conclusions to be acceptable, but it would be good to do it for all the usual reasons. Empirically, we only focused on whether the RTs for the different conditions are different or not (while taking into account sampling error etc.), but *the relative magnitudes of the RTs contain additional information that we largely ignored*. They might tell us something about the relative likelihood of the different scope representations investigated in the experiment.

So let’s ‘frame’ our information-processing problem, i.e., the parsing/interpretation problem, in more detail. A rational analysis of this problem is a *minimal* formally explicit theory of parsing/interpretation: it explicitly tries to make minimal assumptions about processing mechanisms and syntactic/semantic theories. We start with some basic, and largely uncontroversial, assumptions about the human processor (Marslen-Wilson 1973, Frazier and Fodor

1978, Tanenhaus et al. 1995, Steedman 2001, Hale 2011 a.o.): the human processor (*a*) is incremental – syntactic parsing and semantic interpretation do not lag significantly behind the perception of individual words, (*b*) is predictive – the processor forms explicit representations of words and phrases that have not yet been heard, and (*c*) satisfies the competence hypothesis – understanding a sentence/discourse involves the recovery of the structural description of that sentence/discourse on the syntax side, and of the meaning representation on the semantic side.

We will now go through the first 3 steps of rational theory construction for parsing/interpretation (see Hale 2011: 403 et seqq). First, we take the goal of the parser/interpreter to be that it should rapidly arrive at the syntactic and meaning representation intended by the speaker. This goal weaves together two competing demands: *be quick* and *be accurate*. Given the competence hypothesis, we can formulate this as follows:

- (5) The goal of the the parser/interpreter (step **1**): search through the space of syntactic structures and meaning representations quickly (the end state is reached fast) and accurately (the end state is the interpretation intended by the speaker).

This formulation is an instance of a general approach (Newell and Simon 1972): cognition as problem solving, and problem solving as search through a state space.

We turn now to step **2**: identifying a formal model of the environment to which the parser/interpreter is adapted. Since (*i*) sentence/discourse comprehension occurs in a speech community, and (*ii*) grammars describe the knowledge shared by the community, we take grammars to be models of the environment to which comprehenders are adapted (Hale 2011):

- (6) The formal model of the environment (step **2**): the parser/interpreter is adapted to categorical and gradient information specified in *the grammars (syntax and semantics)* of particular languages.

This step says nothing about what counts as a grammar (a syntactic or a semantic theory), which theory is best, etc. But it provides a clear link between processing and grammar: this step and the competence hypothesis provide the two central assumptions we relied on when interpreting our experimental results.

We finally turn to step **3**: computational limitations/specifications. Given a grammar (let's focus on syntax and semantics only), the parser/interpreter has to:

- (7) Computational limitations/specifications (step **3**): the parser/interpreter has to
 - a. define a way of *applying* the syntax and semantics *rules*;
 - b. define a way of *resolving conflict* when more than one rule is applicable.

Conflicts should be resolved in such a way that the estimated distance to completion is minimized (be quick) and the estimated correctness of the analysis is maximized (be accurate).

What does it mean to apply a rule? As already indicated, we take parsing/interpretation to be search through a state space. For syntax, a state is a partially completed syntactic structure (Hale 2011 and references therein). For semantics, a state is a partially constructed DRS (more broadly, LF) and/or a partially evaluated DRS/LF. Applying a grammar rule takes us from one state to another (strong competence hypothesis): *rule applications are transition/accessibility relations between states*. For syntax, we apply phrase structure rules. For semantics, we can take transitions to consist of (*i*) applying a DRS/LF construction rule and/or (*ii*) evaluating a sub-DRS/sub-LF and updating the current interpretation context in the process.

How do we resolve conflict to minimize distance to completion and maximize accuracy? Trying to maximize accuracy is very hard because we can't guess what the speaker intends to

say in the future. That is, it is hard to define heuristic values to maximize accuracy (Hale 2011): an analysis for the first few words may be very good if they’re followed by one continuation, very bad if followed by another. So let’s focus conflict resolution on minimizing distance to completion: assume that the current partial analysis is right; now let’s choose between two paths of analysis. We can estimate how far we are from completion based on previous experience, i.e., based on analyses that we completed before and that have the same initial subpart.

But how do we estimate distance to completion? For syntax, we can do it empirically: we can use a treebank, simulate the actions of a given parser (e.g., left-corner) for the sentences in the treebank and record how far particular intermediate states are from the correct end state. We can use those average distances to resolve conflict: select the analysis path with the smallest expected distance to completion. Hale (2011) uses *A* search*, which is *best-first* – try the best path first, keep a priority queue of alternates, and *informed* – it uses problem-specific knowledge (heuristic values) rather than a fixed policy (e.g., breadth first, depth first). The heuristic value at a state has 2 components: how far we traveled from the initial state + estimated distance to the goal; using both components minimizes *overall* path length.

But the empirical way is not really possible for semantics (yet). So let’s look at alternatives. An alternative way to compute heuristic values for the syntactic processor is to assume that our phrase structure rules are weighted (probabilistic grammars) and to derive expected distances to the end state based on those weights. The basic idea: the more uncertain an analysis path is, the more likely that path is to be far from the end state. Uncertainty is based on the weights themselves, but also on how many choices we have at a particular point. For example, big/complex phrases are more ‘uncertain’, hence avoided, because they can be expanded in many ways – and the more alternatives, the longer it takes to disconfirm the incorrect ones.

The exact procedure is less important, see Hale (2011: 430-432) for more details and a syntactic application. But the main moral for semantics is that estimating probabilities for DRS/LF construction and/or evaluation rules enables our semantic theories to make (more) precise predictions about processing. Thus, our proposal is as follows: (i) we can estimate probabilities experimentally based on RTs; (ii) once we estimate probabilities from one experiment, we can derive predictions for another;³ (iii) we can then evaluate these predictions with respect to their overall qualitative pattern, but we can also quantitatively evaluate them; (iv) things will probably not work out the first time around, so we go to step **6**: iterate.

Here’s how estimating probabilities could go. Take a simple two-sentence discourse with 2 quantifiers in the first sentence: *A boy climbed every tree. The boy/boys wanted to catch blue jays.* And suppose we measure the RTs on the word *wanted*. Assume (following Hale 2001 and Levy 2008) that the RTs vary according to how unexpected/surprising the SG *boy* is relative to the PL *boys*. In particular, assume that the difference in difficulty between SG, i.e., *SS* (surface scope), and PL, i.e., *IS* (inverse scope) is proportional to the difference between the surprisal of *SS*, i.e., $-\log(\mathbf{Pr}(SS))$, and the surprisal of *IS*, i.e., $-\log(\mathbf{Pr}(IS))$. Let’s make this precise by taking the difficulty of SG/PL to be measured in $\log(\text{RTs})$, with RTs measured in ms.:

$$(8) \quad \log(\mathbf{RT}(SS)) - \log(\mathbf{RT}(IS)) \propto (-\log(\mathbf{Pr}(SS))) - (-\log(\mathbf{Pr}(IS))), \text{ i.e.,} \\ \log(\mathbf{RT}(SS)) - \log(\mathbf{RT}(IS)) = c \cdot [\log(\mathbf{Pr}(IS)) - \log(\mathbf{Pr}(SS))], \text{ hence:} \\ \log\left(\frac{\mathbf{RT}(SS)}{\mathbf{RT}(IS)}\right) = \log\left(\left(\frac{\mathbf{Pr}(IS)}{\mathbf{Pr}(SS)}\right)^c\right), \text{ i.e., } \frac{\mathbf{RT}(SS)}{\mathbf{RT}(IS)} = \left(\frac{\mathbf{Pr}(SS)}{\mathbf{Pr}(IS)}\right)^{-c} \text{ (where } c > 0)$$

That is, RTs and probabilities are inversely related: the higher the probability of *SS* is relative to *IS*, the shorter the RTs for *SS* relative to *IS* because *SS* is less surprising/more

³Using probabilities does not necessarily mean that we commit to the fact that they are part of mental representations. They are useful theoretical constructs, just like possible worlds are in formal semantics.

predictable. The free parameter c allows for a flexible relation between ratios of RTs and odds (ratios of probabilities) and should be estimated from the data. Now let's take the RTs from the CONTEXT:YES condition of the self-paced reading experiment and estimate probabilities. We estimate 6 probabilities, 2 for the SU and 4 for the DO: $\Pr(SU = SS)$ (*caregiver*) – the probability that the SU takes wide scope (call it *SS* for uniformity) relative to the universal; $\Pr(SU = IS)$ (*caregivers*) – the probability that the SU takes narrow scope (call it *IS* for uniformity) relative to the universal; $\Pr(DO = SS|SU = SS)$ (*child|caregiver*) – the probability that the DO takes wide scope given that the SU takes wide scope; $\Pr(DO = IS|SU = SS)$ (*children|caregiver*) – the probability that the DO takes narrow scope given that the SU takes wide scope; $\Pr(DO = SS|SU = IS)$ (*child|caregivers*) – the probability that the DO takes wide scope given that the SU takes narrow scope; $\Pr(DO = IS|SU = IS)$ (*children|caregivers*) – the probability that the DO takes narrow scope given that the SU takes narrow scope.

To keep things simple, we will sum the RTs for the relevant regions of interest and obtain one measurement for each of the 42 participants by averaging over items. A serviceable basic Bayesian model with low information priors to estimate these probabilities can be constructed as follows. The data y consists of 42 $\frac{\mathbf{RT}(SS)}{\mathbf{RT}(IS)}$ ratios (one per participant), and $y_i \sim \text{Gamma}(\alpha, \beta)$. *Gamma* is a convenient distribution to use because the RT ratios are always positive. We reparametrize it in terms of its mean μ and standard deviation σ so that we can link it to probability ratios: α (shape) = $\frac{\mu^2}{\sigma^2}$ and β (rate) = $\frac{\mu}{\sigma^2}$. The mean of the *Gamma* distribution is then specified as $\mu = \left(\frac{\Pr(SS)}{\Pr(IS)}\right)^{-c}$, and we assume a *Unif*(0.01, 10) prior for c . Furthermore, we assume a uniform *Beta*(1, 1) prior for $\Pr(SS)$ and take $\Pr(IS) = 1 - \Pr(SS)$. Finally, we assume an *IGamma*($10^{-3}, 10^{-3}$) prior for the variance σ^2 . We also add random effects for participants, not listed in the model above for simplicity.

These are the means of the posterior distributions estimated using this model:⁴

$$\begin{array}{ll}
 (9) \quad \Pr(SU = SS) = 0.59 & \Pr(SU = IS) = 0.41 \\
 \Pr(DO = SS|SU = SS) = 0.55 & \Pr(DO = SS|SU = IS) = 0.51 \\
 \Pr(DO = IS|SU = SS) = 0.45 & \Pr(DO = IS|SU = IS) = 0.49
 \end{array}$$

We can now calculate joint probabilities, i.e., the probabilities of the 4 scope configurations for the initial sentence. In general, $\Pr(X, Y) = \Pr(X|Y) \cdot \Pr(Y)$, hence:

$$\begin{array}{ll}
 (10) \quad \Pr(SU = SS, DO = SS) = 0.33 & \Pr(SU = IS, DO = SS) = 0.21 \\
 \Pr(SU = SS, DO = IS) = 0.26 & \Pr(SU = IS, DO = IS) = 0.20
 \end{array}$$

We see that $SU = SS, DO = SS$ is about 6% more likely than $SU = SS, DO = IS$, which in turn is about 6% more likely than the two configurations in which $SU = IS$. It seems that every quantifier movement up the tree makes the resulting configuration 6% less likely.

There is basically no difference between the last two configurations $SU = IS, DO = SS$ and $SU = IS, DO = IS$. This unexpected result is due to the fact that we did not take into account the ‘baseline’ RTs provided by the CONTEXT:NO condition. But note that taking CONTEXT:NO into account would only make the probability of $SU = IS, DO = SS$ lower, *definitely not null*. In fact, our model assumed that $SU = IS, DO = SS$ is *a priori* possible: we did not build a probability of 0 for this configuration into the prior. This is right for the LF theory since we can imagine $SU = IS, DO = SS$ being derived from $SU = IS, DO = IS$ via an additional movement of the DO indefinite.

However, once we assume we have weights for LF rules that are reflected in RTs (because the heuristic values for the processor are derived from those weights), Skolem-function approaches

⁴We used R (R Core Team 2013) and JAGS (Plummer 2013) to estimate the posterior distributions.

and related approaches, e.g., Dependence Logic, become a viable option again. If covert LF operations are weighted, why not add weights to the arity specification procedure? We can specify the weights so that if a Skolem function is relativized to a variable x , Skolem functions lower in tree are by default also relativized to x . But note that the Skolem approach really needs the processor to enforce an ordering over scope configurations. In contrast, the LF approach provides the ordering on its own, and the processor only specifies particular weights.

The last observation shows that the theoretical relevance of experimental data is hard to assess without being even minimally explicit about processing, i.e., the structure of the parser/interpreter. Our minimal rational analysis indicated that we need some heuristic values/weights for the processor. But once we have those, we do not need semantic theories to induce orderings over scope representations, since the weights themselves can induce such orderings.

4 Conclusion

In sum, we outlined a rational (in the sense of ACT-R) analysis of the interpretation problem: we indicated how the relation between semantic and processing theories could be explicitly formalized. We introduced the specific problem of quantifier scope and the processing difficulty of inverse scope, and discussed two types of theories of scope. We presented the results of one experiment and its consequences for these two types of theories. We outlined how probabilities for scope representations, and for the LF rules used to build them, could be computed based on the experimental results. Associating weights/probabilities with our semantic representations enables our theories to make quantitative, not only qualitative, predictions. In addition, being formally explicit about processing can have a substantial impact on the interpretation of experimental results, and their (presumed) consequences for semantic theories.

References

- Anderson, John R.: 1990, *The adaptive character of thought*. Lawrence Erlbaum Associates, Hillsdale, NJ.
- Dotlačil, Jakub and Adrian Brasoveanu: 2013, ‘The manner and time course of updating quantifier scope representations in discourse’, *Language and Cognitive Processes*. accepted with revisions.
- Ebert, Christian: 2005, *Formal investigations of underspecified representations*, Doctoral Dissertation, King’s College, London.
- Fodor, Janet Dean: 1982, ‘The mental representation of quantifiers’, in S. Peters and E. Saarinen (eds.), *Processes, Beliefs and Questions*, 129–164. Reidel, Dordrecht.
- Frazier, Lyn: 1978, *On comprehending sentences: Syntactic parsing strategies*, Doctoral Dissertation, University of Connecticut.
- Frazier, Lyn and Janet Dean Fodor: 1978, ‘The sausage machine: A new two-stage parsing model’, *Cognition* **6**, 291–325.
- Groenendijk, Jeroen and Martin Stokhof: 1991, ‘Dynamic Predicate Logic’, *Linguistics and Philosophy* **14**, 39–100.
- Hale, John: 2001, ‘A Probabilistic Earley Parser as a Psycholinguistic Model’, in *Proceedings of the 2nd Meeting of the North American Association for Computational Linguistics*, 159–166.
- Hale, John: 2011, ‘What a rational parser would do’, *Cognitive Science* **35**, 399–443.
- Kamp, Hans and Uwe Reyle: 1993, *From Discourse to Logic. Introduction to Model theoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*. Kluwer, Dordrecht.
- Larson, Richard K.: 1988, ‘On the Double Object Construction’, *Linguistic Inquiry* **19**, 335–391.
- Levy, Roger: 2008, ‘Expectation-based syntactic comprehension’, *Cognition* **106**, 1126–1177.
- Marslen-Wilson, William: 1973, ‘Linguistic Structure and Speech Shadowing at Very Short Latencies’, *Nature* **244**, 522–523.
- Newell, Alan and Herbert A. Simon: 1972, *Human Problem Solving*. Prentice-Hall, Englewood Cliffs, NJ.
- Plummer, Martyn: 2013, ‘rjags: Bayesian graphical models using MCMC’. R package version 3-10.
- Pyllkkänen, Liina and Brian McElree: 2006, ‘The syntax-semantic interface: On-line composition of sentence meaning’, in *Handbook of Psycholinguistics*, 537–577. Elsevier, New York.

- R Core Team: 2013, 'R: A Language and Environment for Statistical Computing'. R Foundation for Statistical Computing, Vienna, Austria.
- Radó, Janina and Oliver Bott: 2012, 'Underspecified representations of quantifier scope?', in M. Aloni, V. Kimmelman, F. Roelofsen, G. W. Sassoon, K. Schulz, and M. Westera (eds.), *Logic, Language and Meaning: 18th Amsterdam Colloquium*. Springer, The Netherlands.
- Steedman, Mark: 2001, *The Syntactic Process*. MIT Press, Cambridge, MA.
- Steedman, Mark: 2012, *Taking Scope*. MIT Press, Cambridge, MA.
- Szabolcsi, Anna: 1997, *Ways of scope taking*. Kluwer, Dordrecht.
- Tanenhaus, M. K., M. J. Spivey-Knowlton, K. M. Eberhard, and J. C. Sedivy: 1995, 'Integration of visual and linguistic information in spoken language comprehension', *Science* **268**, 1632–1634.
- Tunstall, Susanne: 1998, *The interpretation of quantifiers: Semantics and processing*, Doctoral Dissertation, University of Massachusetts, Amherst.