What a Rational Interpreter Would Do:

Building, Ranking, and Updating Quantifier Scope Representations in Discourse

Adrian Brasoveanu – joint work with Jakub Dotlačil

Amsterdam Colloquium, ILLC · December 19, 2013

Introduction: ‘Rational’ theories of cognition

Anderson (1990) and much subsequent work argues for the following ‘rational cognition’ hypothesis:

General principle of rationality

The cognitive system operates at all times to optimize the adaptation of the behavior of the organism.

‘Rationality’ in what sense?

- not in the sense of engaging in logically correct reasoning when deciding what to do
- but 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.

Introduction: ‘Rational’ theories of cognition

How to use the principle of rationality to develop a theory of cognition (Anderson 1990, p. 29):

I. Precisely specify the goals of the cognitive system.

II. Develop a formal model of the environment to which the system is adapted.

III. Make minimal assumptions about computational limitations.

IV. Derive the optimal behavioral function given steps I.-III.

V. Examine empirical literature to see if the predictions of the behavioral function are confirmed (if literature available; else do the empirical investigation).

VI. If the predictions are off, iterate.

The goal of the talk today

Summary of rational theory construction:

- The theoretical commitments are made in steps I.-III.
- They provide the "framing of the information-processing problem" (Anderson 1990, p. 30).
- Steps IV.-V. are about deriving and dis/confirming predictions.
- The process of theory building is iterative: if one framing does not work, we try another.

Our goal today:

- Get started with the first iteration of our rational analysis. But for what problem?
- A classical problem in formal semantics: quantifier scope ambiguities.

The goal of the talk today

The specific questions we are interested in:

1. How are quantifier scope ambiguities represented by the interpreter?
2. How are these representations built and maintained / updated as the discourse is incrementally processed / interpreted?
3. How are these representations ranked so that the ambiguities are resolved?

Our particular strategy: a 'rational' analysis.

- 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?

Road map for the talk

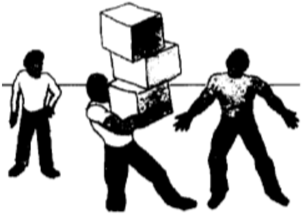
- introduce the problem of quantifier scope and the difficulty of inverse scope
- introduce two types of theories of scope and their predictions
- describe the results of an eye-tracking and a self-paced reading experiment and discuss their consequences for the two types of theories of scope
- pick up the ‘rational’ analysis thread again and ‘frame the information-processing problem’ (parsing/interpretation) in detail
- the main payoff of the detailed ‘framing’: 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

- briefly outline how probabilities for LF construction rules could be computed

Surface/inverse scope

(1) A boy lifted every box.



Surface scope

Inverse scope

Inverse scope

(2) A policeman stood on every corner.

(3) A tablecloth covers twenty tables.

(4) An American flag was hanging in front of every building.

Basic definition of inverse scope

The interpretation of a quantifier is dependent on another quantifier that was introduced “later”. (Szabolcsi 1997, 2011 a.o.)

The cost of inverse scope

- inverse scope is harder to process (Tunstall 1998, Anderson 2004, Filik et al. 2004, Reinhart 2006, Radó and Bott 2012 a.o.)
- it is the less likely interpretation (Ioup 1975, AnderBois et al. 2012 a.o.)

The cost of inverse scope

Establishing processing cost

(5) Kelly showed a photo to every critic last month.
The photo(s) was/were of a run-down building.

(6) Kelly showed every photo to a critic last month.
The critic(s) was/were from an art gallery.

(Tunstall, 1998)

The processing cost:

- signaled by increased reading times (RTs) associated with the plural continuation – but only in (5)
- taken as evidence that people posit a surface-scope interpretation and have to reanalyze
- taken as evidence that reanalysis is costly

Two explanations for the cost of inverse scope

a. Explanation in terms of covert logical form (LF) operations. (Tunstall 1998, Anderson 2004, Reinhart 2006)

b. Inverse scope requires revising (mental / discourse) model structure. (Fodor 1982, Crain and Steedman 1985, Johnson-Laird et al. 1989)

   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 is the arity (and consequently the function).

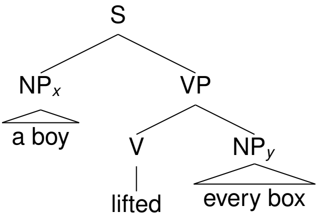
[c.] How about underspecification theories of scope? (Reyle 1993, Bos 1995, Muskens 1999, Muskens 2001, Ebert 2005)

- 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

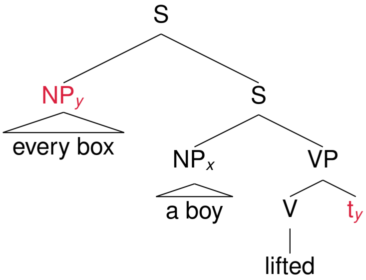
a. Inverse scope via covert operations

(7) A boy lifted every box.

Surface scope:

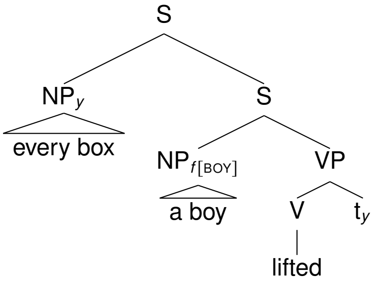


Inverse scope:

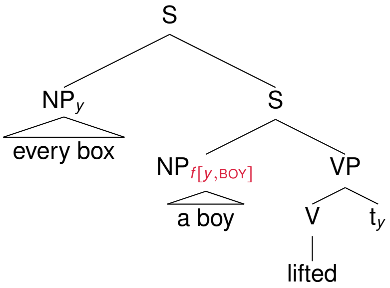


b. Inverse scope via model revision

Surface scope:



Inverse scope:



Open issues and two new experiments

- Very hard to distinguish between these accounts when we look at sentences with only 2 quantifiers.
- Also, we do not know what happens beyond the point of disambiguation:
  - do people really reanalyze their interpretation?
  - if so, how do they reanalyze towards inverse scope?

[‘Reanalysis’ is just a suggestive metaphor. We don’t use it to implicitly favor serial over ranked parallel parser models.]

So: two new experiments (eye-tracking, self-paced reading) that study the reanalysis of quantifier scope.

They provide evidence:

- against a model-based approach, and also against a Skolem function approach to the semantics of indefinites (also against underspecification theories)
- for particular surface/syntax-oriented approaches to scope

Main novelty of the experimental task

Examine the interaction of 3 quantifiers, 2 singular indefinites + 1 universal. Two-sentence discourses:

(8) A caregiver comforted a child every night.

(9) The $\left\{\begin{array}{c}\text{caregiver}\\ \text{caregivers}\end{array}\right\}$ wanted the $\left\{\begin{array}{c}\text{child}\\ \text{children}\end{array}\right\}$ to get some rest.

- first sentence: 2 indefinites in SU and DO position and a universal quantifier as a sentence-final adverb
- second sentence: elaborates on the entities brought to salience by the 2 indefinites
- the only manipulation is morphological number on the SU and DO definites in the second sentence (2 $\times$ 2 design)
- singular definite $\Rightarrow$ wide-scope indefinite

  not necessarily wide-scope: it might be narrow scope with ‘accidental’ coreference; we ignore this (w.l.o.g.).
- plural definite $\Rightarrow$ narrow-scope indefinite

Predictions of the two theories of (inverse) scope

a. Predictions of the covert LF operations theory:

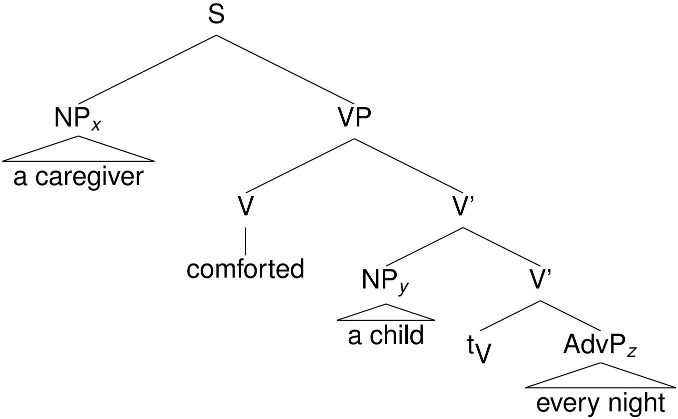
- Assume a base-generated structure with the universal adverb in the lowest position (Larson 1988 style; see also Kimball 1973 and Frazier and Fodor 1978).
- Assume that the more complex an LF is – i.e., the more operations we need to apply to obtain it – the less plausible/salient it is for interpreters.
- Then: if SU indefinite takes narrow scope $\Rightarrow$ the DO indefinite also takes narrow scope.

b. Predictions of the model revision theory:

- Assume that giving widest scope to the universal is costless, but setting the arities of the two Skolem functions is costly.
- Assume that the arities of the two Skolem functions are independently specified.
- Then: revising the model so that the SU indefinite takes narrow scope does not affect the scope of the DO indefinite.

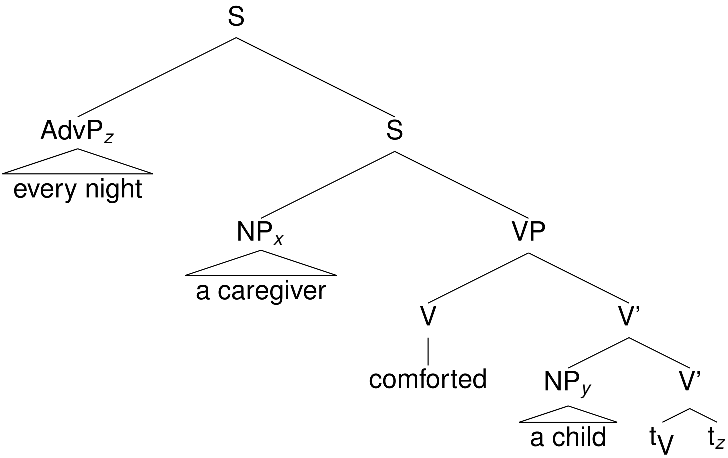
a. Predictions of the covert LF operations theory

Wide scope SU, wide scope DO:



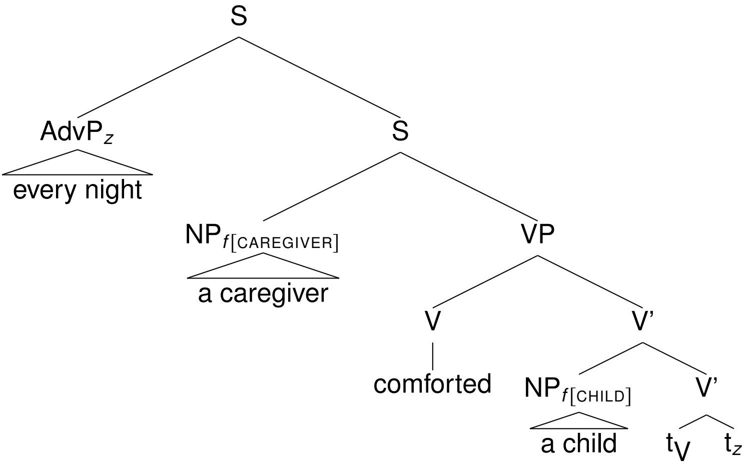
a. Predictions of the covert LF operations theory (ctd.)

Narrow scope SU ⇒ narrow scope DO:



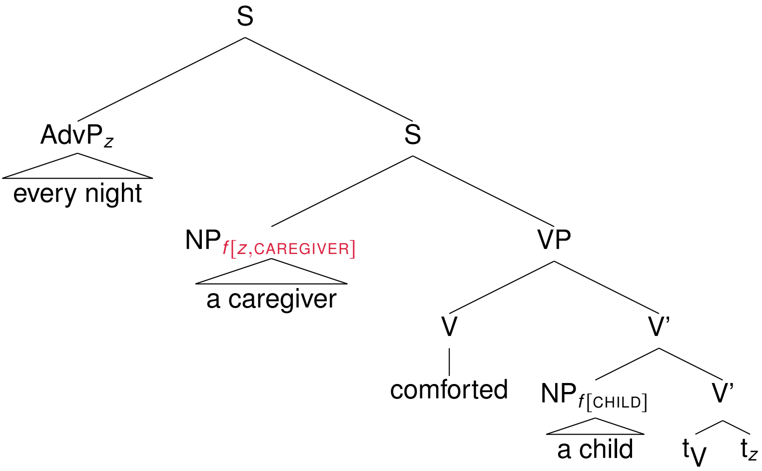
b. Predictions of the model revision theory

Wide scope SU, wide scope DO:



b. Predictions of the model revision theory (ctd.)

Narrow scope SU $\not\Rightarrow$ wide scope DO:



Hierarchical scope representations

The covert LF theory is not the only theory that predicts that narrow scope SU $\Rightarrow$ narrow scope DO.

Any theory that assumes a scope hierarchy (strict total order: asymmetric, total and transitive) will do:

- Thematic hierarchy (Jackendoff, 1972; Kurtzman and MacDonald, 1993):
  Agent $\gg$ Experiencer $\gg$ Theme $\gg$ Goal $\gg \ldots$
- Grammatical hierarchy (Ioup 1975; Reinhart 1983; the LF theory is an instantiation of this):
  Subject $\gg$ Indirect object $\gg$ Direct object $\gg$ Adjunct $\gg \ldots$
- Linear order (Fodor, 1982)

Inverse scope is obtained by promoting a quantifier in the hierarchy – e.g., using the grammatical hierarchy:

SS: Subject $\gg$ Indirect object $\gg$ Direct object $\gg$ Adjunct $\gg \ldots$

IS: Adjunct $\gg$ Subject $\gg$ Indirect object $\gg$ Direct object $\gg \ldots$

Experiment 1

An eye-tracking experiment testing the reanalysis of quantifier scope:

(10) A caregiver comforted a child every night. The caregivers wanted the children (NARROW, NARROW) ...

(11) A caregiver comforted a child every night. The caregiver wanted the children (WIDE, NARROW)...

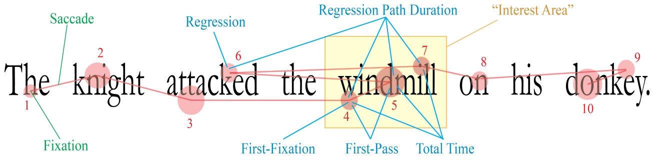
(12) A caregiver comforted a child every night. The caregivers wanted the child (NARROW, WIDE)...

(13) A caregiver comforted a child every night. The caregiver wanted the child (WIDE, WIDE)...

- 7 practice items, 39 test items, 67 fillers
- 33 comprehension questions; 27 participants; on average, 88% questions answered correctly (no participant less than 74%)

Measures

- First pass: time spent in the region for the first time
- Second pass: time spent re-reading the region
- Prob. of regression: how often do people regress back from the region to a previous part?
- Total time: total time spent in the region



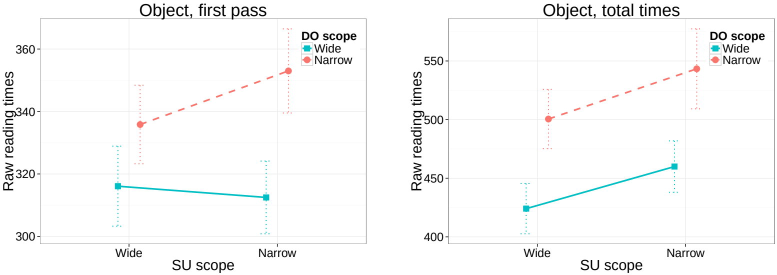
Assumptions:

- Higher reading time ← greater processing difficulties
- More regressions ← greater processing difficulties

Experiment 1: Results

In the DO region, effects of NARROW on reading times are additive:

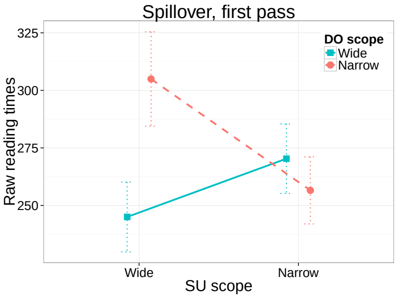
> The caregiver(s) wanted the child(ren) to get. . .



Experiment 1: Results (ctd.)

But the effects of NARROW are not additive in the spillover:

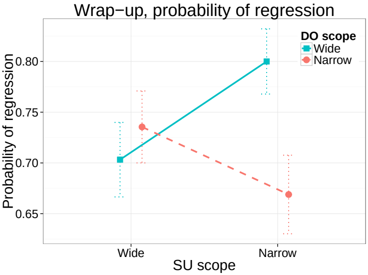
The caregiver(s) wanted the child(ren) to get...



Experiment 1: Results (ctd.)

The effects of NARROW are not additive wrt regression probability in both the DO and the wrap-up regions:

The caregiver(s) wanted the child(ren) to get . . . (wrap-up)



Experiment 1: Summary

- The inverse scope of the universal over the SU makes it easier to interpret the DO as taking narrow scope
- This follows if:
  a. readers quickly reanalyze their scope interpretation
  b. readers reanalyze their interpretation by updating a scope hierarchy (which would entail the narrow scope of the DO)

Could this be a lower level (morphological) issue?

- Maybe the plural on the SU primes the plural on the DO.

Follow-up: a self-paced reading task adding a +/- Context manipulation.

Experiment 2

CONTEXT

(14) A caregiver comforted a child every night.
The caregivers wanted the children (PL, PL) ...

(15) A caregiver comforted a child every night.
The caregiver wanted the children (SG, PL)...

(16) A caregiver comforted a child every night.
The caregivers wanted the child (PL, SG)...

(17) A caregiver comforted a child every night.
The caregiver wanted the child (SG, SG)...

NO CONTEXT

(18) The caregivers wanted the children (PL, PL) ...

(19) The caregiver wanted the children (SG, PL)...

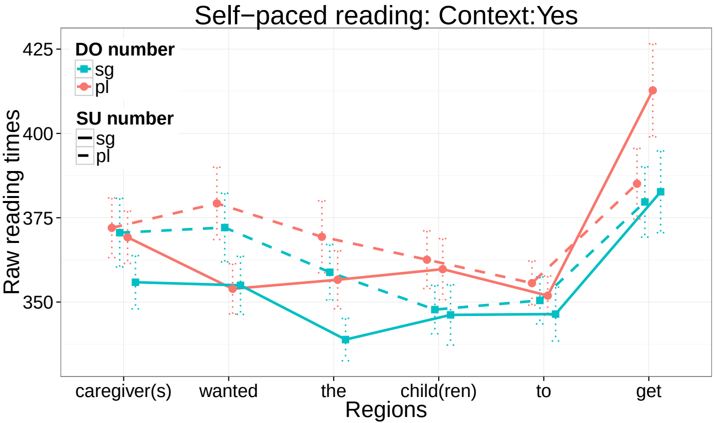
(20) The caregivers wanted the child (PL, SG)...

(21) The caregiver wanted the child (SG, SG)...

Experiment 2: Method

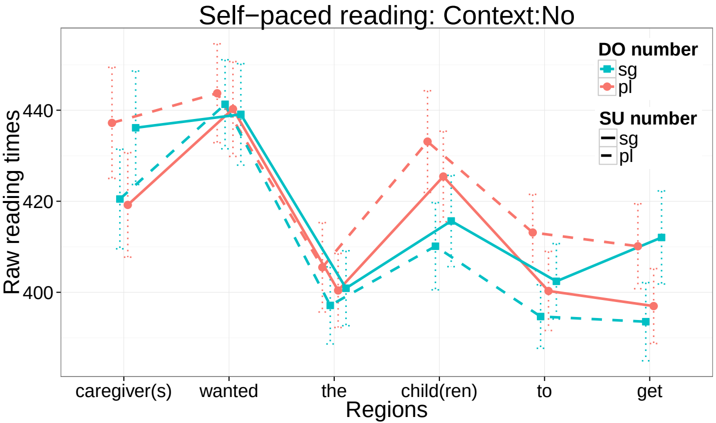
- 4 practice items, 39 test items, 65 fillers, 32 comprehension questions
- 88 participants (44 in CONTEXT, 44 in NO CONTEXT)
- 3 participants excluded because they answered less than 75% questions correctly

Experiment 2: Results for CONTEXT:YES



- A borderline-significant slowdown on the for SUBJECT:PL
- A significant slowdown on get for SUBJECT:SG & OBJECT:PL

Experiment 2: Results for CONTEXT:NO



- A borderline-significant *slowdown* on get for SUBJECT:PL & OBJECT:PL
- A significant three-way interaction: SUBJECT:PL × OBJECT:PL × CONTEXT:YES leads to speed up

Experiment 2: Summary

- PL on the subject facilitates PL on the object but only when the PL disambiguates scope (we reproduce the main result from Experiment 1)
- So the facilitation cannot be attributed to morphological priming, but is (likely) due to the disambiguation role played by PL morphology

Consequences of the experimental results

The results are incompatible:

- with the assumption that readers do not use disambiguating information quickly to reanalyze scope (Filik et al. 2004 a.o.)
- 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 etc. scope hierarchy

  e.g., theories that take indefinites to denote Skolem functions / Skolemized choice functions of variable arity
- with underspecification theories of scope

  to the extent that specifying the scope of the DO is independent of specifying the scope of the SU in these theories

Consequences of the experimental results (ctd.)

The results are compatible:

- with the assumption that the reanalysis is done on syntactic structures

  whether through the promotion of a structure in a parallel processing model (Gibson 1991, Jurafsky 1996) or as a repair strategy in a serial model (Frazier and Rayner, 1982)
- more generally, with the assumption that the processor builds hierarchical scope representations and updates / maintains them across sentential boundaries
- with dynamic systems that have rich interpretation contexts like DRT (Kamp 1981; Kamp and Reyle 1993), but not with systems like DPL (Groenendijk and Stokhof 1991) that are ‘less representational’

Integrating semantics and processing

The experimental results and the consequences we just summarized are substantial, 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 implicit connection formally explicit for the conclusions to be acceptable, but it would be good to do it for all the usual reasons

Integrating semantics and processing (ctd.)

Empirically:

- we only focused on whether the reading times for the different conditions are different or not (while taking into account sampling error etc.)
- but the relative magnitudes of the reading times contain additional information that we largely ignored
- they might tell us something about the relative likelihood of the different quantifier scope representations investigated in the experiment

So let's 'frame our information-processing problem' – 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

Basic assumptions about the human processor

Properties of the human processor:

1. incremental
   syntactic parsing and semantic interpretation do not lag significantly behind the perception of individual words
2. predictive
   the processor forms explicit representations of words and phrases that have not yet been heard
3. 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

(Marslen-Wilson 1973, Frazier and Fodor 1978, Tanenhaus et al. 1995, Steedman 2001, Hale 2011 a.o.)

Framing the parsing/interpretation problem: step I

We now go through the first 3 steps of rational theory construction for parsing/interpretation (see Hale 2011, 403 et seqq).

I. Goal of the parser/interpreter: rapidly arrive at the syntactic and meaning representation intended by the speaker.

→ two competing demands: be quick and be accurate

But given the competence hypothesis, we can reformulate this:

I. The goal of the the parser/interpreter:

> 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).

An instance of a general approach (Newell and Simon 1972): cognition as problem solving, and problem solving as search through a state space (for ‘well-defined’ problems).

Communicative uncertainty

Achieving this goal is difficult because there are many sources of communicative uncertainty (Sag 1992, pp. 7-10):

- Ambiguity: structural (*I forgot how good beer tastes*), lexical (*pen*), scopal, ‘of ellipsis’ (*Jones likes Smith more than Parker*)
- Uncertainty of reference: *She ran home afterwards* (who is she? whose home? after what?)
- Uncertainty of relation: *The nail is in the bowl* (nailed to the bowl or resting inside of it), *The Amsterdam book* (about Amst.? in Amst.? first discovered / read in Amst.?)
- Vivification (general meanings narrowed in context): *Craig cut the lawn/hair/cocaine/record/rookie*, *Coffee?* (one of Columbia’s most valuable cash crops, or ‘do you want coffee?’, or ‘is this coffee?’)
- Coercion: *The Boston Office called*
- Uncertainty of Import: *I thought Jones was a spy* (‘I was right all along’, or ‘I was mistaken’)

Framing the parsing/interpretation problem: step II

II. A formal model of the environment to which the parser/interpreter is adapted:

> the parser/interpreter is adapted to categorical and gradient information specified in the grammars (syntax and semantics) of particular languages.

Sentence/discourse comprehension occurs in a speech community. Grammars describe the knowledge shared by the community, i.e., the environment to which comprehenders are adapted. (Hale 2011)

- This step says nothing about what counts as a grammar (a syntactic or a semantic theory), which theory is best etc.
- But provides a clear link between processing and grammar.

→ this step and the competence hypothesis: two assumptions we relied on when interpreting our experimental results.

Framing the parsing/interpretation problem: step III

III. Computational limitations / specifications.

Given a grammar (let's focus on syntax and semantics only), 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);
- the estimated correctness of the analysis is maximized (be accurate).

III.a. What does it mean to apply a rule?

- parsing/interpretation is search through a state space to solve a problem (Newell and Simon 1972)
- for syntax: a state is a partially completed syntactic structure
- 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:
  - applying a DRS (more broadly, LF) construction rule and/or
  - evaluating a subexpression/sub-DRS and updating the current interpretation context as part of that evaluation

III.b. How do we resolve conflict?

How do we resolve conflict to minimize distance to completion and maximize accuracy?

Issue with maximizing accuracy (Hale 2011): hard to guess what the speaker intends to say in the future.

→ hard to define heuristic values to maximize accuracy: 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 focus conflict resolution on minimizing distance to completion:

- assume that the current partial analysis is right; now choose between two paths of analysis
- we can estimate how far we are from completion based on previous experience,

  i.e., based on analyses we completed before that have the same initial subpart

III.b. 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: best-first – try the best path first, keep a priority queue of alternates; informed – 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 minimizes overall path length
- the empirical way: not really possible for semantics

III.b. How do we estimate distance to completion?

Alternative:

- assume our phrase structure rules are weighted (probabilistic grammars) and derive expected distances to the end state based on those weights
- 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 – big / complex phrases are more 'uncertain'
- big / complex phrases are avoided because they can be expanded in many ways – and the more alternatives there are, the longer it takes to disconfirm the incorrect ones
- the exact procedure is less important, let's look at an example instead: Hale (2011, pp. 430-432) uses it to capture the following phenomenon . . .

Note that we are now moving from steps I.-III. (theory construction) to steps IV.-V.: computing predictions and dis/confirming them.

A syntactic example: NP/S vs. NP/Z ambiguities

Mild local syntactic ambiguity (easy garden path), known as **NP**/**S** ambiguity (example from Sturt et al. 1999):

(22)	the Australian woman saw	the famous doctor had been drinking quite a lot

initially, saw + **NP** more plausible; by the end, only saw + **S** possible; easy to recover (slightly higher RTs than controls)

Severe local syntactic ambiguity (hard garden path), known as **NP**/**Z** ambiguity (again, example from Sturt et al. 1999):

(23)	before the woman visited	the famous doctor had been drinking quite a lot

initially, visited + **NP** more plausible; by the end, only visited + **Z(ero)** possible; hard to recover (much higher RTs than controls)

A syntactic example: NP/S vs. NP/Z ambiguities

- use a weighted grammar where fronted PPs are unlikely
- the grammar formalizes a speech community that rarely fronts their PPs (about 25% of the time)
-

weight	rule
75	$S \rightarrow NP\ VP$
13	$S \rightarrow PP\ ,\ S$
12	$S \rightarrow PP\ S$
1	$SBAR \rightarrow S$
1	$SBAR \rightarrow that\ S$
...	...

- the resulting model correctly derives the greater severity of NP/Z relative to NP/S
- the search is lead down the garden path (as needed) in both cases, but it requires more search steps to recover in the NP/Z case than in the NP/S case

Main moral for semantics

Estimating weights/probabilities for DRS construction and/or DRS evaluation rules enables our semantic theories to make (more) precise predictions about processing.

Proposal:

- we can estimate probabilities experimentally based on reading times
- once we estimate probabilities from one experiment, we could derive predictions for a different experimental task
- we can check the predictions: the overall qualitative pattern; but we can also quantitatively evaluate them
- things will probably not work out the first time around; so we go to step VI.: iterate
- using probabilities does not mean that we commit to the fact that they are part of mental representations; they are useful theoretical constructs – just like possible worlds.

Here's how estimating probabilities could go . . .

From RTs to probabilities

Take a simple two-sentence discourse with 2 quantifiers in the first sentence:

(24) A boy climbed every tree.

(25) The $\left\{\begin{array}{c}\text{boy}\\ \text{boys}\end{array}\right\}$ wanted to catch blue jays.

- 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:

  difference in difficulty between SG, i.e., *SS* (surface scope), and PL, i.e., *IS* (inverse scope)

  $$\propto$$

  difference between the surprisal of *SS*, i.e., $-\log(\mathbf{Pr}(SS))$, and the surprisal of *IS*, i.e., $-\log(\mathbf{Pr}(IS))$.

From RTs to probabilities (ctd.)

Make this precise by taking the difficulty of SG/PL to be measured in log(RTs) (nat. log. of RTs measured in ms.):

- $\log(\mathbf{RT}(SS)) - \log(\mathbf{RT}(IS))$
  $\propto (-\log(\mathbf{Pr}(SS))) - (-\log(\mathbf{Pr}(IS)))$
  $= c \cdot [\log(\mathbf{Pr}(IS)) - \log(\mathbf{Pr}(SS))]$
- hence: $\log\left(\frac{\mathbf{RT}(SS)}{\mathbf{RT}(IS)}\right) = \log\left(\left(\frac{\mathbf{Pr}(IS)}{\mathbf{Pr}(SS)}\right)^{c}\right)$
- finally: $\frac{\mathbf{RT}(SS)}{\mathbf{RT}(IS)} = \left(\frac{\mathbf{Pr}(SS)}{\mathbf{Pr}(IS)}\right)^{-c}$ (where $c > 0$)
- 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 predictable.
- c is a free parameter that allows for a flexible relation between ratios of RTs and odds (ratios of probabilities)
- c should be estimated from the data

From RTs to probabilities: A simple example

Now take RTs from the CONTEXT:YES condition of the self-paced reading exp. and estimate probabilities.

(26) A caregiver comforted a child every night.

(27) The $\left\{\begin{array}{c}\text{caregiver}\\\text{caregivers}\end{array}\right\}$ wanted the $\left\{\begin{array}{c}\text{child}\\\text{children}\end{array}\right\}$ to get some rest.

We estimate 6 probabilities, 2 for the SU:

- **Pr**$(SU = SS)$ (caregiver): the prob. that the SU takes wide scope (we call it *SS* for uniformity) relative to the universal
- **Pr**$(SU = IS)$ (caregivers): the prob. that the SU takes narrow scope (we call it *IS* for uniformity) relative to the universal

From RTs to probabilities: A simple example (ctd.)

And 4 for the DO:

- $\mathbf{Pr}(DO = SS|SU = SS)$ (child|caregiver): the prob. that the DO takes wide scope given that the SU takes wide scope
- $\mathbf{Pr}(DO = IS|SU = SS)$ (children|caregiver): the prob. that the DO takes narrow scope given that the SU takes wide scope
- $\mathbf{Pr}(DO = SS|SU = IS)$ (child|caregivers): the prob. that the DO takes wide scope given that the SU takes narrow scope
- $\mathbf{Pr}(DO = IS|SU = IS)$ (children|caregivers): the prob. 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
- obtain one measurement for each of the 42 participants by averaging over items

From RTs to probabilities: A simple example (ctd.)

A serviceable basic Bayesian model with low information priors to estimate the probabilities:

- y (data): 42 $\frac{\mathbf{RT}(SS)}{\mathbf{RT}(IS)}$ ratios (one per participant)
- $y_i \sim Gamma(\alpha, \beta)$
  - *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) $= \frac{\mu}{\sigma^2}$
- $\mu = \left(\frac{\mathbf{Pr}(SS)}{\mathbf{Pr}(IS)}\right)^{-c}$ and assume a $Unif(0.01, 10)$ prior for c
- assume a uniform $Beta(1, 1)$ prior for $\mathbf{Pr}(SS)$ and take $\mathbf{Pr}(IS) = 1 - \mathbf{Pr}(SS)$
- finally, 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.

From RTs to probabilities: A simple example (ctd.)

And these are the means of the posterior distributions we can estimate using this type of model:

- $\mathbf{Pr}(SU = SS) = 0.59$
- $\mathbf{Pr}(SU = IS) = 0.41$
- $\mathbf{Pr}(DO = SS|SU = SS) = 0.55$
- $\mathbf{Pr}(DO = IS|SU = SS) = 0.45$
- $\mathbf{Pr}(DO = SS|SU = IS) = 0.51$
- $\mathbf{Pr}(DO = IS|SU = IS) = 0.49$

[we used R (R Core Team 2013) and JAGS (Plummer 2013) to estimate the posterior distributions]

From RTs to probabilities: A simple example (ctd.)

We can now calculate joint probabilities, i.e., the probabilities of the 4 scope configurations for the initial sentence.

- In general: $\mathbf{Pr}(X, Y) = \mathbf{Pr}(X|Y) \cdot \mathbf{Pr}(Y)$

- $\mathbf{Pr}(SU = SS, DO = SS) = 0.33$
- $\mathbf{Pr}(SU = SS, DO = IS) = 0.26$
- $\mathbf{Pr}(SU = IS, DO = SS) = 0.21$
- $\mathbf{Pr}(SU = IS, DO = IS) = 0.20$

- $SU = SS, DO = SS$ is about 6% more likely than $SU = SS, DO = IS$.
- And $SU = SS, DO = IS$ is about 6% more likely than the two configurations in which the SU takes narrow scope.
- It looks like every quantifier movement up the tree makes the resulting configuration 6% less likely.

From RTs to probabilities: A simple example (ctd.)

$\mathbf{Pr}(SU = IS, DO = SS) = 0.21$ $\mathbf{Pr}(SU = IS, DO = IS) = 0.20$

- But: there is basically no difference between the last two configurations $SU = IS, DO = SS$ and $SU = IS, DO = IS$.
- Unexpected; due to the fact that we did not take into account the 'baseline' RTs provided by the CONTEXT:NO condition.
- But this would only make the probability of $SU = IS, DO = SS$ lower, definitely not null.
- Our model in fact 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 mvt. of the DO indefinite.

Skolem strikes back

- But once we assume we have weights for LF rules that are reflected in RT magnitudes (because the heuristic values for the processor are derived from those weights), Skolem-function approaches (also Dependence Logic etc.) are back in the game.
- If covert LF operations are weighted, why not add weights/biases to the Skolem-arity specification procedure?
- Specify 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 the Skolem approach really needs the processor to enforce an ordering over the various scope configurations.
  [The LF approach provides the ordering on its own, the processor only specifies particular weights.]

Moral for the theoretical relevance of experimental data:

Being even minimally explicit about processing, i.e., the structure of the parser/interpreter, can have substantial consequences for the way we relate experimental data and semantic theories (grammars).

Summary / Conclusion

- 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 two experiments and their consequences for the two types of theories of scope.
- We outlined how probabilities for scope representations – and hence, for the LF construction rules that 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.

Acknowledgments

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 an SRG grant from the UCSC Committee on Research 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.

References I

AnderBois, Scott et al. (2012). “The Pragmatics of Quantifier Scope: A Corpus Study”. In: *Proceedings of Sinn und Bedeutung 16*. Ed. by A. Aguilar-Guevara et al. Cambridge, MA: MIT Working Papers in Linguistics, pp. 15–28.

Anderson, Catherine (2004). “The structure and real-time comprehension of quantifier scope ambiguity”. PhD thesis. Evanston, Illinois: Northwestern University.

Anderson, John R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Bos, Johan (1995). “Predicate Logic Unplugged”. In: *Proceedings of the 10th Amsterdam Colloquium*. Amsterdam.

Crain, Stephen and Mark Steedman (1985). “On not being led up the garden path: the use of context by the psychological syntax processor”. In: *Natural Language Parsing: Psychological, Computational and Theoretical Perspectives*. Ed. by L. Karttunen David Dowty and Arnold Zwicky. Cambridge: Cambridge University Press, pp. 320–358.

Ebert, Christian (2005). “Formal investigations of underspecified representations”. PhD thesis. London: King’s College.

References II

Filik, Ruth et al. (2004). “Processing doubly quantified sentences: Evidence from eye movements”. In: *Psychonomic Bulletin & Review* 11.5, pp. 953–959.

Fodor, Janet Dean (1982). “The mental representation of quantifiers”. In: *Processes, Beliefs and Questions*. Ed. by Stanley Peters and Esa Saarinen. Dordrecht: Reidel, pp. 129–164.

Frazier, Lyn and Janet Dean Fodor (1978). “The sausage machine: A new two-stage parsing model”. In: *Cognition* 6, pp. 291–325.

Frazier, Lyn and Keith Rayner (1982). “Making and Correcting Errors during Sentence Comprehension: Eye Movements in the Analysis of Structurally Ambiguous Sentences”. In: *Cognitive Psychology* 14, pp. 178–210.

Gibson, Edward (1991). “A computational theory of human linguistic processing: Memory limitations and processing breakdown”. PhD thesis. Pittsburgh, PA: Carnegie Mellon University.

Groenendijk, Jeroen and Martin Stokhof (1991). “Dynamic Predicate Logic”. In: *Linguistics and Philosophy* 14.1, pp. 39–100.

Hale, John (2001). “A Probabilistic Earley Parser as a Psycholinguistic Model”. In: *Proceedings of the 2nd Meeting of the North American Asssociation for Computational Linguistics*, pp. 159–166.

— (2011). “What a rational parser would do”. In: *Cognitive Science* 35, pp. 399–443.

References III

Ioup, Georgette (1975). “Some universals for quantifier scope”. In: *Syntax and Semantics 4*. Ed. by J. Kimball. New York: Academic Press, pp. 37–58.

Jackendoff, Ray (1972). *Semantic interpretation in generative grammar*. Cambridge, Massachusetts: MIT Press.

Johnson-Laird, P. N. et al. (1989). “Reasoning by Model: The Case of Multiple Quantification”. In: *Psychological Review* 96, pp. 658–673.

Jurafsky, Dan (1996). “A Probabilistic Model of Lexical and Syntactic Access and Disambiguation”. In: *Cognitive Science* 20, pp. 137–194.

Kamp, Hans (1981). “A Theory of Truth and Semantic Representation”. In: *Formal Methods in the Study of Language*. Ed. by Jeroen Groenendijk et al. Amsterdam: Mathematical Centre Tracts, pp. 277–322.

Kamp, Hans and Uwe Reyle (1993). *From Discourse to Logic. Introduction to Model theoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*. Dordrecht: Kluwer.

Kimball, John (1973). “Seven principles of surface structure parsing in natural language”. In: *Cognition* 2, pp. 15–47.

Kurtzman, Howard S. and Maryellen C. MacDonald (1993). “Resolution of Quantifier Scope Ambiguities”. In: *Cognition* 48, pp. 243–279.

Larson, Richard K. (1988). “On the Double Object Construction”. In: *Linguistic Inquiry* 19.3, pp. 335–391.

References IV

Levy, Roger (2008). “Expectation-based syntactic comprehension”. In: *Cognition* 106, pp. 1126–1177.

Marslen-Wilson, William (1973). “Linguistic Structure and Speech Shadowing at Very Short Latencies”. In: *Nature* 244, pp. 522–523.

Muskens, Reinhard (2001). “Talking about Trees and Truth-Conditions.” In: *Journal of Logic, Language and Information* 10, pp. 417–455.

Muskens, Reinhard A. (1999). “Underspecified Semantics”. In: *Reference and Anaphoric Relations*. Ed. by Klaus von Heusinger and Urs Egli. Dordrecht: Kluwer, pp. 311–338.

Newell, Alan and Herbert A. Simon (1972). *Human Problem Solving*. Englewood Cliffs, NJ: Prentice-Hall.

Plummer, Martyn (2013). *rjags: Bayesian graphical models using MCMC*. R package version 3-10. URL: `http://CRAN.R-project.org/package=rjags`.

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: `http://www.R-project.org/`.

Radó, Janina and Oliver Bott (2012). “Underspecified representations of quantifier scope?” In: *Logic, Language and Meaning: 18th Amsterdam Colloquium*. Ed. by Maria Aloni et al. The Netherlands: Springer.

References V

Reinhart, Tanya (1983). *Anaphora and Semantic Interpretation*. Chicago: University of Chicago Press.

— (2006). *Interface Strategies: Optimal and Costly Computations*. Cambridge, Massachusetts: MIT Press.

Reyle, Uwe (1993). “Dealing with Ambiguities by Underspecification: Construction, representation and deduction”. In: *Journal of Semantics* 10, pp. 123–179.

Sag, Ivan A. (1992). “Taking performance seriously”. In: *VII Congreso de Languajes Naturales y Lenguajes Formales*. Ed. by Carlos Martin-Vide. Barcelona. URL: `http://lingo.stanford.edu/sag/papers/vic-paper.pdf`.

Steedman, Mark (2001). *The Syntactic Process*. Cambridge, MA: MIT Press.

— (2012). *Taking Scope*. Cambridge, MA: MIT Press.

Sturt, Patrick et al. (1999). “Structural Change and Reanalysis Difficulty in Language Comprehension”. In: *Journal of Memory and Language* 40, pp. 136–150.

Szabolcsi, Anna (1997). *Ways of scope taking*. Dordrecht: Kluwer.

— (2011). *Quantification*. New York: Cambridge University Press.

Tanenhaus, M. K. et al. (1995). “Integration of visual and linguistic information in spoken language comprehension”. In: *Science* 268, pp. 1632–1634.

References VI

Tunstall, Susanne (1998). “The interpretation of quantifiers: Semantics and processing”. PhD thesis. Amherst: University of Massachusetts.