Witness Sets, Polarity Reversal and the Processing of Quantified Sentences

Oliver Bott¹, Udo Klein² & Fabian Schlotterbeck¹

¹Collaborative Research Centre 833
University of Tübingen

²Collaborative Research Centre 673
University of Bielefeld

26/11/2013, Konstanz
1 Introduction

2 Procedural Theories of Quantifier Verification: Automata

3 A Theory Based on Expansion

4 Predictions

5 Experiment 1: Simply Quantified Sentences

6 Experiment 2: Doubly Quantified Sentences

7 Conclusions
Inference:

Every nurse played against more than two foresters.
All foresters are socialists.

∴ Every nurse played against more than two socialists.

false: true: √
Inference:

Every nurse played against fewer than two foresters.  
All foresters are socialists.  

∴ Every nurse played against fewer than two socialists.  

false: √    true:
Inference:

At most three nurses played against fewer than two foresters. All foresters are socialists.

\[ \therefore \text{At most three nurses played against fewer than two socialists.} \]

false:  true: √
Inference:

- DE quantifiers are more difficult than UE quantifiers.
- Open question: Can semantic complexity be explained on the basis of individual quantifiers/sentences?
Verification:

(a) Superlative quantifiers (e.g. *at most* $n$) are more difficult than comparative quantifiers (e.g. *fewer than* $n$) (Geurts et al. 2010)

(b) UE quantifiers (e.g. *at most* $n$) are more difficult than DE quantifiers (e.g. *fewer than* $n$) (Geurts et al. 2010, Hackl 2009)

While there are theories that explain (a) (Geurts & Nouwen 2007, Cummins & Katsos 2010), (b) remains unexplained
Goal of this talk:

- Present a processing theory for quantifier verification that explains difficulty with DE quantifiers.
- Present an experiment that tested specific predictions derived from this theory.

A note on the the superlative/conservative- contrast:

- We will take it as given that superlative quantifiers pose processing difficulty.
- We use this as an additional factor in our experiments.
Procedural Theories of Quantifier Verification: Semantic Automata
What We Want

Input: A quantified sentence

1. **Interpretation**: Specify a procedure that can compute the truth value of the sentence in any model

2. **Verification**: Execute this procedure in order to evaluate the sentence in a specific model

(See Moschovakis 1994, Szymanik 2009, Pietroski et al. 2009)
Interlude: Monotonicity

Definition

A quantifier $Q$ is

- UE iff $Q_M(A, B) \land B \subseteq B' \Rightarrow Q_M(A, B')$
- DE iff $Q_M(A, B) \land B \supseteq B' \Rightarrow Q_M(A, B')$

More than two boys walk north. ⇒ More than two boys walk.

Fewer than two boys walk. ⇒ Fewer than two boys walk north.
Natural language quantifiers are CONS and EXT

This means that, in order to evaluate $Q(A, B)$, only $|A \cap B|$ and $|A \setminus B|$ are relevant.
If we encode models as strings, a quantifier $Q$ is a set of strings (language $L_Q$).

For natural language quantifiers it is sufficient to consider the alphabet $\Gamma = \{a_{\overline{AB}}, a_{AB}\}$

For instance: $L_{All} = \{\alpha \in \Gamma^* : \#a_{\overline{AB}}(\alpha) = 0\}$
Semantic Automata

- **No:**

- **An even number of:**

- **Fewer than 3 / At most 2:**

- **More than two 2 / At least 3:**
Proposition (van Benthem 1986)

For finite models, $L_Q$ is recognized by a finite automaton iff $Q$ is definable in divisibility logic.

- Proportional quantifiers need working memory.

Hypothesis

Proportional Qs are harder to process than Aristotelian Qs.

Hypothesis

The more states an automaton has, the harder it is to process.

The automaton model makes correct predictions about the least computational requirements a quantifier poses (in the worst case).

However, it is not a plausible model of quantifier verification because it does not capture:

- the superlative/comparative-contrast
- the difference between DE and UE quantifiers

**Fewer than 3 / At most 2:**

**More than two 2 / At least 3:**
Iterated Automata: The Languages

Bott, Klein & Schlotterbeck

The Processing of Quantified Sentences

(Steinert-Threlkeld & Icard, 2013)
(1) Every boy tickled some girl. No boy tickled no girl.

No difference between DE and UE quantifiers
A Theory Based on Expansion
Given: binary predicate $P \subseteq E^2$

Expansion: Add tuples $\langle x, Q_2 \rangle$ or $\langle Q_1, x \rangle$ to $P$

Truth: $Q_1 \ A \ P \ Q_2 \ B$ iff $\langle Q_1(A), Q_2(B) \rangle$ can be added to $P$

Similar for the $n$-ary case
Before Expansion

(1) Every boy tickled some girl.

\[
[tickle] = \{ <b_1, g_3>, <b_1, g_4>, <b_2, g_4>, <b_2, g_5>, <b_3, g_5> \}
\]
(1) Every boy tickled some girl.

\[
[tick\text{le}]^* = \{ < b_1, g_1, - >, < b_1, g_2, - >, < b_1, g_3, + >, < b_1, g_4, + >, \\
< b_1, g_5, - >, < b_2, g_1, - >, < b_2, g_2, - >, < b_2, g_3, - >, \\
< b_2, g_4, + >, < b_2, g_5, + >, < b_3, g_1, - >, < b_3, g_2, - >, \\
< b_3, g_3, - >, < b_3, g_4, - >, < b_3, g_5, + > \}
\]
(1) Every boy tickled some girl.

\[
[< b_1, g_3, + >]_{2}^{\text{tick}le}^* = \begin{cases}
< b_1, g_1, - > \\
< b_1, g_2, - > \\
< b_1, g_3, + > \\
< b_1, g_4, + > \\
< b_1, g_5, - > 
\end{cases} = \{ g_3, g_4 \}
\]
(1) Every boy tickled some girl.

If \[< b_1, g_3, + >][tickle]^* \] is witness set of some(girls), add \(< b_1, \text{some(girls)}, + >\) to \([\text{tickle}]^*\)
Every boy tickled some girl.

$\text{[tickle]}^* = \{< b_1, g_1, -, >, < b_1, g_2, -, >, < b_1, g_3, +, >, < b_1, g_4, +, >, < b_1, g_5, -, >, < b_2, g_1, -, >, < b_2, g_2, -, >, < b_2, g_3, -, >, < b_2, g_4, +, >, < b_2, g_5, +, >, < b_3, g_1, -, >, < b_3, g_2, -, >, < b_3, g_3, -, >, < b_3, g_4, -, >, < b_3, g_5, +, >, < b_1, \text{some(girl)}, +, >\}$
(1) Every boy tickled some girl.

\[
[tickle]^* = \{ <b_1, g_1, ->, <b_1, g_2, ->, <b_1, g_3, +>, <b_1, g_4, +>, <b_1, g_5, ->, <b_2, g_1, ->, <b_2, g_2, ->, <b_2, g_3, ->, <b_2, g_4, +>, <b_2, g_5, +>, <b_3, g_1, ->, <b_3, g_2, ->, <b_3, g_3, ->, <b_3, g_4, ->, <b_3, g_5, +>, <b_1, \text{some}(girl), +>, <b_2, \text{some}(girl), +>, <b_3, \text{some}(girl), +>, <\text{every}(boy), \text{some}(girl), +> \}
\]
Definition (w-function, w-quantifier, witness sets)

- w-function $q$:
  $$\mathcal{P}(E) \rightarrow \mathcal{P}(E) \times \mathcal{P}(\mathcal{P}(E)),$$
  $$A \mapsto \langle A, W \rangle$$
- $q(A) = \langle A, W \rangle$ is called a w-quantifier
- $W$ is the set of witness sets of $q$ at $A$. 

Bott, Klein & Schlotterbeck
The Processing of Quantified Sentences
**Definition (simple and complex $i$-expansion)**

Let $q$ be a w-function and $P$ be an $n$-ary predicate. Then the simple expansion of $P^*$ by $q(A)$ at position $i$, written as $\text{s-exp}_i(q(A), P^*)$ is the smallest set $Q$ such that $P^* \subseteq Q$ and clause 1 holds.

The complex expansion of $P^*$ by $q(A)$ at position $i$, written as $\text{c-exp}_i(q(A), P^*)$ is the smallest set $Q$ such that $P^* \subseteq Q$ and clauses 1-4 below hold:

1. $\sigma^+ \in P^* \land [\sigma^+]_{P^*, A} \in W \rightarrow \sigma^+[i/q(A)] \in Q$
2. $\sigma^+ \in P^* \land [\sigma^+]_{P^*, A} \notin W \rightarrow \sigma^+[i/q(A)] \in Q$
3. $\sigma^- \in P^* \land [\sigma^-]_{P^*, A} = A \land \emptyset \in W \rightarrow \sigma^+[i/q(A)] \in Q$
4. $\sigma^- \in P^* \land [\sigma^-]_{P^*, A} = A \land \emptyset \notin W \rightarrow \sigma^-[i/q(A)] \in Q$
Complex Expansion Necessary?

Proposition

If $q_1(A_1)$ and $q_1(A_1)$ are not empty-set quantifiers, then:

$$\langle q_1(A_1), q_2(A_2), + \rangle \in c\text{-}exp_1(q_1(A_1), c\text{-}exp_2(q_2(A_2), P^*)) \iff \langle q_1(A_1), q_2(A_2), + \rangle \in s\text{-}exp_1(q_1(A_1), s\text{-}exp_2(q_2(A_2), P^*)).$$
An Example

(1) Every boy tickled some girl.
(2) No boy tickled no girl.

For (1-a), **s-exp** suffices; for (1-b) **c-exp** is needed.
In contrast to the automata model, the present theory only allows us to evaluate sets like \( \{ x : R(x, a) \} \) if there is some \((x, a) \in R\).  

In the automata model, both positive and negative information is explicitly encoded in the language.
Predictions
Prediction 1: Comprehension (Exp. 2)

Sentences involving no empty-set \( Q \) should be read faster than sentences involving empty set \( Q \)'s, because the former require the specification of an algorithm involving \( s\text{-exp} \), whereas the latter require the complex expansion \( c\text{-exp} \).

Prediction 2: Verification (Exp. 1 and Exp. 2)

Since no UE quantifier is an empty-set-quantifier, whereas all DE \( Q \)'s are empty-set quantifiers, it follows that quantified statements involving only UE quantifiers can be evaluated by \( s\text{-exp} \) in all models/situations. On the other hand, the evaluation of statements involving DE quantifiers depends on the model: in some models \( s\text{-exp} \) suffices, in others \( c\text{-exp} \) is necessary.
Predictions

Prediction 1: Comprehension (Exp. 2)

Sentences involving no empty-set \( Q \) should be read faster than sentences involving empty set \( Q \)'s, because the former require the specification of an algorithm involving \( s\text{-}exp \), whereas the latter require the complex expansion \( c\text{-}exp \).

Prediction 2: Verification (Exp. 1 and Exp. 2)

Since no UE quantifier is an empty-set-quantifier, whereas all DE \( Q \)'s are empty-set quantifiers, it follows that quantified statements involving only UE quantifiers can be evaluated by \( s\text{-}exp \) in all models/situations. On the other hand, the evaluation of statements involving DE quantifiers depends on the model: in some models \( s\text{-}exp \) suffices, in others \( c\text{-}exp \) is necessary.
1-a) Genau ein Lehrer lobte jeden dieser Schüler voller Wohlwollen.
Exactly one teacher praised each of these pupils full of goodwill.

- Relative complexity of doubly quantified sentences may depend on their scopal reading(s)
- Pafel (2005): (1-a) should allow inverse scope
  - $QP_2$ is distributive and linked to discourse
- Bott & Schlotterbeck (2012); only linear scope: (1-a) = (1-b)

1-b) Für genau einen Lehrer gilt: er lobte jeden dieser Schüler voller Wohlwollen.
For exactly one teacher holds: he praised each of these pupils full of goodwill.
Which readings are available during online interpretation?

Step 1: present a disambiguating model

**linear:** $\exists ! \forall$

...praise full of goodwill...

[Diagram of linear model with arrows from teachers to pupils]

**inverse:** $\forall \exists !$

...praise full of goodwill...

[Diagram of inverse model with arrows from pupils to teachers]
Which readings are available during online interpretation?

Step 2: present a potentially ambiguous sentence (1-a) or a disambiguated control (1-b) in a stops-makes-sense task
Experiment 1
A) Mindestens ein Punkt ist blau. At least one dot is blue.

B) Höchstens ein Punkt ist blau. At most one dot is blue.

C) Weniger als zwei Punkte sind blau. Less than two dots are blue.

0-model: \[ \begin{array}{c}
  \vdots \\
  \vdots \\
  \vdots \\
  \vdots \\
  \vdots \\
  \vdots \\
 \end{array} \]

1-model: \[ \begin{array}{c}
  \vdots \\
  \vdots \\
  \vdots \\
  \vdots \\
  \vdots \\
  \vdots \\
 \end{array} \]

3-model: \[ \begin{array}{c}
  \vdots \\
  \vdots \\
  \vdots \\
  \vdots \\
  \vdots \\
  \vdots \\
 \end{array} \]

\( (MON \uparrow, \text{superl.}) \)

\( (MON \downarrow, \text{superl.}) \)

\( (MON \downarrow, \text{comp.}) \)
72 German participants read simply quantified sentences of type (A)–(C).

After pressing a button, the sentence disappeared and the picture was presented.

3 (quantifier) × 3 (model) within design.

Dependent variables:
- Judgment RTs
- Error rates

27 items, no fillers.

Three judgments per condition and participant.

Same participants as in Exp. 2 (Exp. 1 always after Exp. 2).
Exp. 1 – Judgment RTs

interaction quantifier × model ($p_{1/2} < .05$): 0-model hard for at most one and fewer than two but not for at least one!
interaction quantifier $\times$ model ($z = 2.42$): 0-models led to errors, but only for *at most one* and *fewer than two*
Only source of difficulty: evaluation of 0-models for DE Q’s
This difficulty is present both for at most and fewer than
Participants in fact seemed to search for positive information for empty set quantifiers

Relating the findings to our account

Processing difficulty is fully expected since this case requires the application of the most difficult clause of c-exp, namely clause 3
If the predicate is not empty, then there is an $a \in E$ with $\langle a, + \rangle \in P^*$, so clause 1 or 2 can be applied.
Experiment 2
Quantificational complexity of doubly quantified sentences

A-↑↑) Jeder | Junge | kitzelte | mehr als | drei | Mädchen.
Every | boy | tickled | more than | three | girls.

A-↑↓) Jeder | Junge | kitzelte | weniger als | drei | Mädchen.


S-↑↑) Mindestens ein | Junge | kitzelte | mehr als | drei | Mädchen.

S-↑↓) Mindestens ein | Junge | weniger als | drei | Mädchen.

S-↓↑) Höchstens ein | Junge | kitzelte | mehr als | drei | Mädchen.

S-↓↓) Höchstens ein | Junge | kitzelte | weniger als | drei | Mädchen.
Manipulated factors in the sentence materials:
- **type of Q₁** (Aristotelean $Q₁$ vs. superlative $Q₁$)
- **monotonicity of $Q₁$** (UE vs. DE)
- **monotonicity of $Q₂$** (UE vs. DE)

Three types of models: 0-model, 1-model and 3-model

Factorial within design with 24 conditions
Exp. 2 – Sentence-picture combinations

\{ Every \hspace{1cm} \text{No} \hspace{1cm} \text{At least one} \hspace{1cm} \text{At most one} \}

\{ \text{boy tickled} \hspace{1cm} \text{more than three} \hspace{1cm} \text{less than three} \hspace{1cm} \text{girls} \}

0-model

1-model

3/all-model

\begin{tabular}{|c|c|c|}
\hline
& 0-model & 1-model & 3/all-model \\
\hline
\multirow{2}{*}{UE Q₂} & \includegraphics[width=0.3\textwidth]{exp2 UE Q₂ 0-model.png} & \includegraphics[width=0.3\textwidth]{exp2 UE Q₂ 1-model.png} & \includegraphics[width=0.3\textwidth]{exp2 UE Q₂ 3-all-model.png} \\
& \includegraphics[width=0.3\textwidth]{exp2 UE Q₂ 0-model.png} & \includegraphics[width=0.3\textwidth]{exp2 UE Q₂ 1-model.png} & \includegraphics[width=0.3\textwidth]{exp2 UE Q₂ 3-all-model.png} \\
\hline
\multirow{2}{*}{DE Q₂} & \includegraphics[width=0.3\textwidth]{exp2 DE Q₂ 0-model.png} & \includegraphics[width=0.3\textwidth]{exp2 DE Q₂ 1-model.png} & \includegraphics[width=0.3\textwidth]{exp2 DE Q₂ 3-all-model.png} \\
& \includegraphics[width=0.3\textwidth]{exp2 DE Q₂ 0-model.png} & \includegraphics[width=0.3\textwidth]{exp2 DE Q₂ 1-model.png} & \includegraphics[width=0.3\textwidth]{exp2 DE Q₂ 3-all-model.png} \\
\hline
\end{tabular}
Two (independent) sources of difficulty during interpretation:

- Superlative $Q'$s are harder to comprehend than Aristotelean $Q'$s
  - due to a more complex logical form
  - due to (canceling) an ignorance implicature
  - ...

- Monotonicity: s-exp only possible in UE-UE sentences
Exp. 2 – Predictions II (expansion operation)

Same expansion of the object QP *(more than 3 vs. less than 3)*!

<table>
<thead>
<tr>
<th></th>
<th>0-model</th>
<th>1-model</th>
<th>3/all-model</th>
</tr>
</thead>
<tbody>
<tr>
<td>UE Q₂</td>
<td><img src="image1" alt="Diagram" /></td>
<td><img src="image2" alt="Diagram" /></td>
<td><img src="image3" alt="Diagram" /></td>
</tr>
<tr>
<td></td>
<td>$3 \times$ clause 2</td>
<td>$1 \times$ clause 1, $2 \times$ clause 2</td>
<td>$3 \times$ clause 1</td>
</tr>
<tr>
<td>DE Q₂</td>
<td><img src="image4" alt="Diagram" /></td>
<td><img src="image5" alt="Diagram" /></td>
<td><img src="image6" alt="Diagram" /></td>
</tr>
<tr>
<td></td>
<td>$3 \times$ clause 2</td>
<td>$1 \times$ clause 1, $2 \times$ clause 2</td>
<td>$3 \times$ clause 1</td>
</tr>
</tbody>
</table>
Exp. 2 – Predictions III (expansion operation)

We expected the following pattern during verification:

- Easiest case: apply \textit{s-exp} in all UE-UE conditions
- More difficult: \textit{c-exp} cases UE-DE, DE-UE and DE-DE
- Superlativity should independently add difficulty
- However, since the same clauses must be applied in the DE-UE and DE-DE conditions, we expected no differences
Exp. 2 – Methods

- 72 participants read doubly quantified sentences in a moving-window self-paced reading experiment.
- After reading the last segment the sentence disappeared and a set diagram appeared for which a truth-value judgment had to be provided.
- Dependent variables:
  - reading time (interpretation stage)
  - judgment RT (verification stage)
  - percent correct (verification stage)
- 72 items, 78 fillers, overall 50% true sentences, 24 lists in a latin square design.
- Always the same set of pictures in each sentence condition.
As predicted, interaction between $\textit{MON} \, QP_1$ and $\textit{MON} \, QP_2$ ($p_1 < 0.05; p_2 = 0.09$)
Exp. 2 – Judgment times

Main effects of *monotonicity* \(p_{1/2} < .01\), but no reliable interaction
Exp. 2 – Percent correct

- Participants had problems with *at most* in 0-models
- Substantially higher error rates than in Exp. 1 (**DE-DE** sign. below chance level)
Exp. 2 – Discussion

- Reading times: evidence for the proposed procedural theory of quantificational complexity
  - effects of superlativity and monotonicity added up
  - monotonicity effect took the form of an interaction

- Judgments: partial evidence for our model
  - evidence: 0-models difficult with superlative empty set quantifier *at most one*
  - counter-evidence: monotonicity effects of $Q_1$ and $Q_2$ added up even though UE and DE $Q_2$'s required the same expansion in the tested models
Exp. 2 – Discussion

- Reading times: evidence for the proposed procedural theory of quantificational complexity
  - effects of superlativity and monotonicity added up
  - monotonicity effect took the form of an interaction
- Judgments: partial evidence for our model
  - evidence: 0-models difficult with superlative empty set quantifier at most one
  - counter-evidence: monotonicity effects of $Q_1$ and $Q_2$ added up even though UE and DE $Q_2$'s required the same expansion in the tested models
Conclusions
Conclusions

Getting to the source of difficulty of DE quantifiers

The difference between s-exp and c-exp proved to be empirically important for the processing of UE and DE quantifiers

- We were able to account for differences in the online interpretation of UE-UE and UE-DE, DE-UE as well as DE-DE sentences
- We were able to identify the specific class of models that leads to difficulty during verification

What remains problematic, though, is

- how to account for the purely additive effects of monotonicity during verification
Conclusions

Getting to the source of difficulty of DE quantifiers

The difference between s-exp and c-exp proved to be empirically important for the processing of UE and DE quantifiers

- We were able to account for differences in the online interpretation of UE-UE and UE-DE, DE-UE as well as DE-DE sentences
- We were able to identify the specific class of models that leads to difficulty during verification

What remains problematic, though, is
- how to account for the purely additive effects of monotonicity during verification
Given Proposition 1 we would expect that all empty set quantifiers require $c$-exp. Therefore, our results should generalize to non-monotonic quantifiers:

1. Exactly three boys tickled more than two girls.
2. Fewer than three boys tickled more than two girls.
3. Exactly one boy or exactly three boys tickled more than two girls.
4. No boy or exactly three boys tickled more than two girls.
Thank you very much for your attention!!!

Thanks to Aysenur Sarcan for preparing the stimuli and running the experiments