An Investigation into the Computational Complexity of Quantifying Expressions

Fabian Schlotterbeck & Oliver Bott

Collaborative Research Centre 833
University of Tübingen

18/03/2010

CUNY 2010, NYU
1. Semantic Complexity and Measures of Complexity

2. Experiment 1: Comparing Different Quantifiers

3. Experiment 2: Controlling the Reading
   - Experiment 2a: Monotonicity
   - Experiment 2b: Cardinality

4. Conclusions
Outlook

1 Semantic Complexity and Measures of Complexity

2 Experiment 1: Comparing Different Quantifiers

3 Experiment 2: Controlling the Reading
   • Experiment 2a: Monotonicity
   • Experiment 2b: Cardinality

4 Conclusions
Natural language determiners like *every, some, exactly two, at most three*

From semantic theory and cognitive psychology it is known that quantifiers are semantically heterogenous (anaphora resolution, NPI/PPI licensing, reasoning, verification)

Are there differences of complexity in comprehension?
How Difficult is a Quantifier? Relatively Easy

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

∴ Every nurse played against more than two socialists.

false: □   true: ✓
Every nurse played against fewer than two foresters.
All foresters are socialists.

\[
\therefore \text{Every nurse played against fewer than two socialists.}
\]

false: ☐  true: ☑
At most three nurses played against fewer than two foresters. All foresters are socialists.

∴ At most three nurses played against fewer than two socialists.

false: □   true: ✓
How Difficult is a Quantifier?

- Geurts & van der Slik (2005) have shown that in this reasoning task certain quantifiers lead to more errors.

<table>
<thead>
<tr>
<th>Quantifiers</th>
<th>Correct Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Every + more than two</td>
<td>91%</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Every + fewer than two</td>
<td>71%</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>At most three + fewer than two</td>
<td>36%</td>
</tr>
</tbody>
</table>

from Geurts & van der Slik (2005)
Why Should Some Quantifiers be More Difficult than Others?

- Formal properties may influence processing difficulty

Three candidates:

H1  Complexity of Model Checking
H2  Monotonicity
H3  Superlative vs. Comparative
Type (1,1) quantifiers are

1. classes of models $\mathcal{M} = (M, A, B)$,

   e.g. $\text{Every} = \{(M, A, B) : A \subseteq B\}$

or equivalently

2. relations between relations $A, B$ on a set $M$,

   e.g. $\text{Every}_M = \{(A, B) : A \subseteq B\}$
Notation in Generalized Quantifier Theory

Natural language sentences are translated into a logic $\mathfrak{L}$

\[
\text{Every} \quad \text{boy} \quad \text{runs.} \\
\text{Quantifier} \quad \text{Restriction} \quad \text{Nuclear Scope} \\
Q \quad A \quad B
\]

\[\leadsto \text{Every}_x(\text{Boy}(x), \text{Run}(x))\]

Sentences of $\mathfrak{L}$ may be satisfied by a Model

\[M \models \text{Every}_x(\text{Boy}(x), \text{Run}(x)) \iff (M, \text{Boy}^M, \text{Run}^M) \in \text{Every}\]
Notation in Generalized Quantifier Theory

Some Examples:

\[
\text{No} \quad = \quad \{(M, A, B) : |A \cap B| = 0\}
\]
\[
\text{Exactly one} \quad = \quad \{(M, A, B) : |A \cap B| = 1\}
\]
\[
\text{More than two} \quad = \quad \{(M, A, B) : |A \cap B| > 2\}
\]
\[
\text{Fewer than two} \quad = \quad \{(M, A, B) : |A \cap B| < 2\}
\]
\[
\text{Every} \quad = \quad \{(M, A, B) : |A \setminus B| = 0\}
\]
H1 Complexity of Model Checking

- Encoding 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 $\Gamma = \{a_{AB}, a_{AB}\}$
- For instance: $L_{\text{All}} = \{\alpha \in \Gamma^* : \#a_{AB}(\alpha) = 0\}$
**H1 Complexity of Model Checking**

- Encoding 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 $\Gamma = \{a_{AB}, a_{AB}\}$

For instance: $L_{All} = \{\alpha \in \Gamma^* : \#a_{AB}(\alpha) = 0\}$
H1 Semantic Automata

An automaton for \textit{All}:

\[ q_0 \xrightarrow{a_{AB}} q_1 \]

Example models:

\[ \alpha_1 = a_{AB} a_{AB} a_{AB} a_{AB} \in L_{\text{All}} \]
\[ \alpha_2 = a_{AB} a_{AB} a_{\overline{AB}} a_{AB} \notin L_{\text{All}} \]
H1 Semantic Automata

- **No:**

- **Exactly one:**

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

- **More than two / At least 3:**
For finite models

$L_Q$ is recognized by a finite acyclic automaton iff $Q$ is definable in First Order Logic.\hfill van Benthem (1986)

Possible complexity measure within this class: Number of states in the (smallest) automaton
**H1 (Semantic Automata)**

The more states an automaton (first order quantifier) has, the harder it is to process.

Szymanik & Zajenkowski (2010)

- Evidence for H1 from picture verification
- No evidence from comprehension
**H1 Semantic Automata**

- Prediction of H1: no < exactly one < \{ at most 2, at least 3, fewer than 3, more than 2 \}

- Problem: comprehension ≠ verification
**H2 Monotonicity – Inference From Sets to Supersets**

A quantifier $Q$ is $\text{MON} \uparrow$ if

$$Q_M(A, B) \land B \subseteq B' \Rightarrow Q_M(A, B')$$

More than two boys walk north. $\Rightarrow$ More than two boys walk.
A quantifier $Q$ is $\text{MON} \downarrow$

$$Q_M(A, B) \land B \supseteq B' \Rightarrow Q_M(A, B')$$

Fewer than two boys walk. $\Rightarrow$ Fewer than two boys walk north.
H2 (Monotonicity)

Monotonically decreasing quantifiers are harder to process than monotonically increasing quantifiers


- Evidence for H2 from picture verification, reasoning, acquisition
- No conclusive evidence from comprehension
H3 Superlative vs. Comparative Scalar Quantifiers

H3 (Superlative vs. Comparative)
Superlative quantifiers are harder to process than comparative quantifiers

Musolino (2004), Geurts & van der Slik (2007), Geurts et al. (2010)

- Evidence for H3 from picture verification, reasoning, acquisition
- No conclusive evidence from comprehension
## A Previous Study

<table>
<thead>
<tr>
<th>Sentence</th>
<th>Ms</th>
<th>Picture</th>
<th>Ms</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) There are <strong>at most two</strong> As</td>
<td>1970</td>
<td>AAA</td>
<td>AA</td>
</tr>
<tr>
<td>(2) There are <strong>at least three</strong> As</td>
<td>1921</td>
<td>AAA</td>
<td>AA</td>
</tr>
<tr>
<td>(3) There are <strong>fewer than three</strong> As</td>
<td>1940</td>
<td>AAA</td>
<td>AA</td>
</tr>
<tr>
<td>(4) There are <strong>more than two</strong> As</td>
<td>1886</td>
<td>AAA</td>
<td>AA</td>
</tr>
<tr>
<td>(5) There are <strong>exactly three</strong> As</td>
<td>1580</td>
<td>AAA</td>
<td>AA</td>
</tr>
</tbody>
</table>

from Geurts et al. (2010)

- No evidence for H2/3 in RT of complete sentences
- Evidence for H2/H3 in judgment times
- Do quantifiers remain underspecified during reading?
- possible confound: lexical differences!
How to Measure QP Complexity?

Linear scope interpretation of (1) proceeds as follows:

(1) **Exactly one** teacher praised every student.

1. Compose second quantifier and verb
   - Set of individuals praising every student

2. Evaluate first quantifier
   - There is exactly one teacher in this set
   - Comprehending second quantifier indicates complexity of the first.
How to Measure QP Complexity?

Linear scope interpretation of (1) proceeds as follows:

(1) Exactly one teacher praised every student.

1. Compose second quantifier and verb
   - Set of individuals praising every student

2. Evaluate first quantifier
   - There is exactly one teacher in this set
   - Comprehending second quantifier indicates complexity of the first.
How to Measure QP Complexity?

Linear scope interpretation of (1) proceeds as follows:

(1) Exactly one teacher praised every student.

1. Compose second quantifier and verb
   - Set of individuals praising every student
2. Evaluate first quantifier
   - There is exactly one teacher in this set
   - Comprehending second quantifier indicates complexity of the first.
How to Measure QP Complexity?

Linear scope interpretation of (1) proceeds as follows:

(1) Exactly one teacher praised every student.

1. Compose second quantifier and verb
   - Set of individuals praising every student
2. Evaluate first quantifier
   - There is exactly one teacher in this set
   - Comprehending second quantifier indicates complexity of the first.
How to Measure QP Complexity?

Linear scope interpretation of (1) proceeds as follows:

(1) Exactly one teacher praised every student.

1. Compose second quantifier and verb
   - Set of individuals praising every student
2. Evaluate first quantifier
   - There is exactly one teacher in this set
   - Comprehending second quantifier indicates complexity of the first.
How to Measure QP Complexity?

Linear scope interpretation of (1) proceeds as follows:

(1) Exactly one teacher praised every student.

1. Compose second quantifier and verb
   - Set of individuals praising every student
2. Evaluate first quantifier
   - There is exactly one teacher in this set
   - Comprehending second quantifier indicates complexity of the first.
How to Measure QP Complexity?

Linear scope interpretation of (1) proceeds as follows:

(1) Exactly one teacher praised every student.

1. Compose second quantifier and verb
   - Set of individuals praising every student
2. Evaluate first quantifier
   - There is exactly one teacher in this set
   - Comprehending second quantifier indicates complexity of the first.
Recipe for the Following Experiments

1. Ensure linear interpretation (SO word order, binding)
2. Measure on second quantifier – same across conditions – to exclude confounding factors (frequency, morphology, syntax, etc.)
3. Confounds can only be excluded if there are no differences on the pre-quantifier segment.
Outlook

1. Semantic Complexity and Measures of Complexity

2. Experiment 1: Comparing Different Quantifiers

3. Experiment 2: Controlling the Reading
   - Experiment 2a: Monotonicity
   - Experiment 2b: Cardinality

4. Conclusions
Design – Numeral Quantifiers

1. **At least** four (six/eight) teachers praised one of their pupils.
2. **At most** three (five/seven) teachers praised one...
3. **More than** three (five/seven) teachers praised one...
4. **Less than** four (six/eight) teachers praised one...
Experimental Manipulations

- **Within factors:**
  - **Monotonicity** \((\text{MON} \uparrow \text{(at least, more than)} \text{ vs. } \text{MON} \downarrow \text{(at most, less than)})\)
  - **Superlative** \((\text{at least, at most}) \text{ vs. comparative} \text{ (more than, less than)}\) quantifiers

- **Between factor:**
  - **Cardinality** – number of states in the according acyclic finite automaton (three levels: five vs. seven vs. nine states)
Experimental Manipulations

- **Within factors:**
  - Monotonicity \((\text{MON} \uparrow \text{at least, more than}) \text{ vs. } \text{MON} \downarrow \text{at most, less than})\)
  - Superlative \((\text{at least, at most}) \text{ vs. comparative} \text{ (more than, less than)}\) quantifiers

- **Between factor:**
  - Cardinality – number of states in the according acyclic finite automaton (three levels: five vs. seven vs. nine states)
Design – Additional Quantifiers

(5) **Exactly** one|teacher|praised|one of his pupils.
(6) **No**|teacher|praised|one of his pupils.
(7) **The**|teacher|praised|one of his pupils.

- Within manipulation:
  - *exactly one* (three states, neither $MON \uparrow$ nor $MON \downarrow$)
  - *no* (two states, $MON \downarrow$)
  - *the* (control)
(1) **At least** four (six/eight) teachers praised each/one of their pupils.
(2) **At most** three (five/seven) teachers praised each/one...
(3) **More than** three (five/seven) teachers praised each/one...
(4) **Less than** four (six/eight) teachers praised each/one...
(5) **Exactly** one teacher praised each/one of his pupils.
(6) **No** teacher praised each/one of his pupils.
(7) **The** teacher praised each/one of his pupils.

Quantifiers in object position:
- Half of the items: *einen* (one); other half: *jeden* (each)
- Bound variable *of his/their* guarantees linear scope
Methods

- Self-paced reading using moving window presentation
- Yes-no-question after a third of the trials, all querying quantities, eg.

  **Sentence:**  *At most eight professors praised all of their students.*

  **Question:**  *Is this consistent with a situation in which nine professors praised all of their students?*

- 42 participants
- 42 items
- 80 fillers
- Latin square design
RTs of 2nd QP: the < exactly one, no < numeral QPs
Results – RTs of Second Quantifier Region

QP1 verb QP2

RTs 2nd QP in ms (+95% CIs)

- at least
- at most
- more than
- less than
- no
- exactly one
- the
Results – Cardinality Effect at 2nd QP

- 4/3/3/4 vs. 8/7/7/8: only sign. main effect of **Cardinality**
- The higher the num. value of QP1, the higher RT of QP2
Interim Summary

Semantic Automata (H1): supported
The more states a quantifier requires, the more difficult it is to comprehend.

Monotonicity (H2): not supported
Monotonically decreasing quantifiers are harder to comprehend than monotonically increasing quantifiers.

Superlative vs. comparative QP (H3): not supported
Superlative quantifiers are harder to comprehend than comparative quantifiers.
A Concern – RTs Already Differed at the Verb!

Why do we find a cardinality effect on this region? This clearly goes against the recipe from the introduction...
An Alternative Explanation of the Cardinality Effect

- Readers had to parse two quantificational sentences (target sentence plus question) both containing numeral expressions
- This may have forced them to memorize the exact value of QP 1
- Small numbers (3/4) may be easier to memorize than greater numbers (8/9)

▶ The cardinality effect may be a mere artefact of the method!
An Alternative Explanation of the Cardinality Effect

- Readers had to parse two quantificational sentences (target sentence plus question) both containing numeral expressions
- This may have forced them to memorize the exact value of QP 1
- Small numbers (3/4) may be easier to memorize than greater numbers (8/9)

▶ The cardinality effect may be a mere artefact of the method!
Readers had to parse two quantificational sentences (target sentence plus question) both containing numeral expressions

This may have forced them to memorize the exact value of QP 1

Small numbers (3/4) may be easier to memorize than greater numbers (8/9)

The cardinality effect may be a mere artefact of the method!
How Did Readers Interpret \textit{MON} $\downarrow$ Quantifiers?

- We had rather strong intuitions about the complexity of \textit{at most} and \textit{less than}.
- Do readers represent monotone decreasing quantifiers in the right way?
- Comprehension questions following \textit{at most} were answered most accurately in approx. 90\% ($no \approx 70\%$).

- We have to control the interpretation comprehenders actually compute!
How Did Readers Interpret $MON \downarrow$ Quantifiers?

- We had rather strong intuitions about the complexity of *at most* and *less than*
- Do readers represent monotone decreasing quantifiers in the right way?
- Comprehension questions following *at most* were answered most accurately in approx. 90% ($no \approx 70\%$)!

- We have to control the interpretation comprehenders actually compute!
How Did Readers Interpret $\text{MON} \downarrow$ Quantifiers?

- We had rather strong intuitions about the complexity of *at most* and *less than*
- Do readers represent monotone decreasing quantifiers in the right way?
- Comprehension questions following *at most* were answered most accurately in approx. 90% (*no* ≈ 70%)!

- We have to control the interpretation comprehenders actually compute!

Fabian Schlotterbeck & Oliver Bott
Computational Complexity of Quantifiers
Reconsidering Geurts et al.’s (2010) acquisition data, 11 year olds’ representations of *at most two* were as follows:
Outlook

1. Semantic Complexity and Measures of Complexity

2. Experiment 1: Comparing Different Quantifiers

3. Experiment 2: Controlling the Reading
   - Experiment 2a: Monotonicity
   - Experiment 2b: Cardinality

4. Conclusions
Do Comprehenders Interpret *At Most* Correctly?

**At most** At most one | professor | called | fewer than three | students | during | the evening.

- *At most one*: MON ↓, automaton with 3 states
- *At most*-sentence followed by one of three picture types

**true 0:**

**true 1:**

**false 2:**
No)

No | professor | called | fewer than three | students | during | the evening.

- No: MON ↓, automaton with 2 states
- No-sentence followed by either true or false picture

true:

false:
Each) Each professor called fewer than three students during the evening.

- Each: $MON \uparrow$, automaton with 2 states, control
- Each-sentence followed by either true or false picture

**true:**

- ... call yesterday evening...
- professors
- students

**false:**

- ... call yesterday evening...
- professors
- students
The Complete Design – Monotonicity

At most) At most one | professor | called | fewer than three | students | during | the evening.

No) No | professor | called | fewer than three | students | during | the evening.

Each) Each | professor | called | fewer than three | students | during | the evening.

▷ Seven conditions

▷ *At most* and *No* are **MON ↓**, *each* is **MON ↑**

▷ *Each* and *no* correspond to automata with two states, *at most* requires three states
Methods

- Self-paced reading using moving window presentation
- Picture verification task after each sentence

- 35 participants
- 42 items
- 66 fillers (32 false)
- Latin square design
Results – Picture Verification Task

- *Each* and *no* interpreted correctly
- *At most* at chance level in *true 0*-pictures
- *At most one* isn’t interpreted *MON \downarrow!*

![Bar chart showing percent correct for different conditions: each true, each false, no true, no false, at most true 0, at most true 1, at most false.](chart.png)
Results – Reading Times

At most starts slow and ends up fast
Results – Reading Times 2nd QP

- **Experiment 2a: Monotonicity**
- **Experiment 2b: Cardinality**

**No** is more difficult to integrate with 2nd QP than is **each**!
Monotonicity Affects RT

Monotonicity (H2): supported
Monotonically decreasing quantifiers are harder to comprehend than monotonically increasing quantifiers.
**Low** At least three students read one of the books during the spring break.

- Low cardinality sentence followed by either true or false picture
Design – Cardinality

**Low**) At least three | students | read | one of the | books...

**High**) At least eight | students | read | one of the | books | during | the spring break.

- High cardinality sentence followed by either true or false picture
Methods

- Same as in Exp. 2a
- 36 participants
- 12 items
- 96 fillers (47 false)
- Latin square design
Results – RTs of Low vs. High Number of States

- Small numbers of states numerically even slower than high number of states!

Fabian Schlotterbeck & Oliver Bott
Computational Complexity of Quantifiers
No Influence of Cardinality

Semantic Automata (H1): not supported

The more states a quantifier requires, the more difficult it is to comprehend.
Outlook

1. Semantic Complexity and Measures of Complexity

2. Experiment 1: Comparing Different Quantifiers

3. Experiment 2: Controlling the Reading
   - Experiment 2a: Monotonicity
   - Experiment 2b: Cardinality

4. Conclusions
Experiments 1 and 2a/b provide conflicting evidence, but taken together they demonstrate that...

- The task has an important influence on the comprehension of quantifiers

- Readers may choose a wrong but easier interpretation altering their semantic properties (e.g. monotonicity)

- We have to carefully control both processing complexity and the final interpretation

- During online comprehension, quantifiers differ in complexity
Conclusions

Experiments 1 and 2a/b provide conflicting evidence, but taken together they demonstrate that...

- The task has an important influence on the comprehension of quantifiers

- Readers may choose a wrong but easier interpretation altering their semantic properties (e.g. monotonicity)

- We have to carefully control both processing complexity and the final interpretation

- During online comprehension, quantifiers differ in complexity
Conclusions

Experiments 1 and 2a/b provide conflicting evidence, but taken together they demonstrate that...

- The task has an important influence on the comprehension of quantifiers
- Readers may choose a wrong but easier interpretation altering their semantic properties (e.g. monotonicity)
- We have to carefully control both processing complexity and the final interpretation
- During online comprehension, quantifiers differ in complexity
Conclusions

Experiments 1 and 2a/b provide conflicting evidence, but taken together they demonstrate that...

- The task has an important influence on the comprehension of quantifiers

- Readers may choose a wrong but easier interpretation altering their semantic properties (eg. monotonicity)

- We have to carefully control both processing complexity and the final interpretation

- During online comprehension, quantifiers differ in complexity
Thank you for your attention!