Quantifying quantifier scope: a cross-methodological comparison

Oliver Bott Janina Radó

SFB 441 University of Tübingen

Linguistic Evidence 2006

Outline

The problem

- A methodological comparison
- 3 Pretesting the methods
- Oomparing the methods



Why speaker judgements?

- theoreticians want to identify properties of construction types
- but...
 - introspective data are not sufficient
 - corpus data can be scarce
- thus systematically collected judgements are important
 - from many speakers
 - several items representing a construction type

Syntactic judgements

judging grammaticality/acceptability

- how: e.g. questionnaire
- rating sentences one by one
- using a scale (e.g. 7 points) or magnitude estimation for subtle differences

Semantic judgements

judging which meaning/how many meanings a construction has

"Does the construction have meaning x ?"

what subjects have to do:

- compute meaning(s) for construction
- compute meaning for paraphrase
- compare the two meanings

difficult for naive speakers

 \rightarrow we need a better method

Outline



- 2 A methodological comparison
- Pretesting the methods
- Comparing the methods



Outline

1 The problem

- A methodological comparison
- Pretesting the methods
- Oomparing the methods



What is a suitable method?

- simple for subjects to work with
- does not introduce artificial bias
- can detect subtle differences

Quantifier scope

controversial judgements:

Everyone loves someone

- fully scope-ambiguous (e.g. May 1977, 1985; Hornstein 1984; Higginbotham 1985), or
- only wide scope universal (∀∃) reading (e.g. Reinhart 1976, 1983; Hornstein 1995; Beghelli & Stowell 1997)

Comparison procedure

materials:

- equally plausible under both scope readings
- maximally similar across experiments

two-step procedure:

- pretest: can the methods capture scope readings?
- main comparison: which methods deliver good results?

Methods under comparison

what subjects judge: sentence + disambiguation

form of disambiguation:

- disambiguating context (question)
- diagram
 - set diagram
 - natural-looking scenario

Outline

1 The problem

- A methodological comparison
- Pretesting the methods
- Omparing the methods

5 Summary

Pretesting the methods

- materials: unambiguous quantified sentences
- one universal and one existential quantifier
- factors:
 - order of quantifiers
 - disambiguation

Für genau einen Professor gilt, dass jede Studentin ihn verehrt hat. 'Exactly one professor is such that every student adores him.'

Für jede Studentin gilt, dass sie genau einen Professor verehrt hat. 'Every student is such that she adores exactly one professor.'

Experimental setup

- task: Does the sentence match the disambiguation?
- simple "yes, matches" "no, doesn't match" judgements
- 24 items, 36 fillers

Predictions





Kann man von genau einem Professor sagen, dass jede Studentin ihn verehrt hat? 'Can it be said of exactly one professor that every student adored him?'

Ja, stimmt. Für genau einen Professor gilt, dass jede Studentin ihn verehrt hat.

advantage:

- only linguistic material used
- method often used by linguists

disadvantage:

• word order - possible confound?

Kann man von genau einem Professor sagen, dass jede Studentin ihn verehrt hat? 'Can it be said of exactly one professor that every student adored him?'

Ja, stimmt. Für genau einen Professor gilt, dass jede Studentin ihn verehrt hat.

advantage:

- only linguistic material used
- method often used by linguists

disadvantage:

• word order - possible confound?

Kann man von genau einem Professor sagen, dass jede Studentin ihn verehrt hat? 'Can it be said of exactly one professor that every student adored him?'

Ja, stimmt. Für genau einen Professor gilt, dass jede Studentin ihn verehrt hat.

advantage:

- only linguistic material used
- method often used by linguists

disadvantage:

word order – possible confound?

Für genau einen Professor gilt, dass jede Studentin ihn verehrt hat.





 $\forall \exists \ diagram$

∃∀ diagram

advantage:

- easy to work with
- used in psycholinguistics (Gillen 1991; Jackson&Lewis 2005)

disadvantage:

● ∃∀ diagram somewhat complex for subjects

Für genau einen Professor gilt, dass jede Studentin ihn verehrt hat.





 $\forall \exists \text{ diagram}$

∃∀ diagram

advantage:

- easy to work with
- used in psycholinguistics (Gillen 1991; Jackson&Lewis 2005)

disadvantage:

● ∃∀ diagram somewhat complex for subjects

Für genau einen Professor gilt, dass jede Studentin ihn verehrt hat.





 $\forall \exists \text{ diagram}$

∃∀ diagram

advantage:

- easy to work with
- used in psycholinguistics (Gillen 1991; Jackson&Lewis 2005)

disadvantage:

• $\exists \forall$ diagram somewhat complex for subjects

Für genau ein Dreieck gilt, dass jedes Kind es in seiner Ecke hat. 'Exactly one triangle is such that it is in every child's corner.'





∀∃ diagram

ers corner

∃∀ diagram

advantage:

- relatively natural scenario
- no influence of lexical content or plausibility

disadvantage:

internal structure of diagram may influence visual search

Für genau ein Dreieck gilt, dass jedes Kind es in seiner Ecke hat. 'Exactly one triangle is such that it is in every child's corner.'





∀∃ diagram

∃∀ diagram

advantage:

- relatively natural scenario
- no influence of lexical content or plausibility

disadvantage:

internal structure of diagram may influence visual search

Für genau ein Dreieck gilt, dass jedes Kind es in seiner Ecke hat. 'Exactly one triangle is such that it is in every child's corner.'





∀∃ diagram

∃∀ diagram

advantage:

- relatively natural scenario
- no influence of lexical content or plausibility

disadvantage:

• internal structure of diagram may influence visual search

Pretest results – Summary



methods do quite well with unambiguous sentences

- but: differences in "false-yes" answers
- → how consistent are subjects within an experiment?

Pretest results – Summary



methods do quite well with unambiguous sentences

• but: differences in "false-yes" answers

• \rightarrow how consistent are subjects within an experiment?

Pretest results – Summary



- methods do quite well with unambiguous sentences
- but: differences in "false-yes" answers
- \rightarrow how consistent are subjects within an experiment?

Consistency across subjects

inter-rater reliability

- absolute intraclass correlations (ICCs)
- based on average scores per condition for each subject

results:

- question-answer pairs:
- set diagrams:
- scenarios:

ICC = 0.685 for single raters ICC = 0.847 for single raters ICC = 0.778 for single raters

conclusion:

- no significant differences, but
- subjects were least consistent using question-answer pairs

Consistency across subjects

inter-rater reliability

- absolute intraclass correlations (ICCs)
- based on average scores per condition for each subject
- results:
 - question-answer pairs:
 - set diagrams:
 - scenarios:

ICC = 0.685 for single raters ICC = 0.847 for single raters ICC = 0.778 for single raters

conclusion:

- no significant differences, but
- subjects were least consistent using question-answer pairs

Consistency across subjects

inter-rater reliability

- absolute intraclass correlations (ICCs)
- based on average scores per condition for each subject
- results:
 - question-answer pairs:
 - set diagrams:
 - scenarios:

ICC = 0.685 for single raters ICC = 0.847 for single raters ICC = 0.778 for single raters

conclusion:

- no significant differences, but
- subjects were least consistent using question-answer pairs

Outline

The problem

- A methodological comparison
- Pretesting the methods
- Comparing the methods



What we tested

- modified versions of the pretest materials:
- scope-ambiguous sentences with one universal and one existential quantifier (∀-subject, ∃-object)
- factors:
 - linear order of quantifiers
 - distributivity (alle vs. jede)

A sample item

Genau einen dieser Professoren haben alle Studentinnen verehrt. exactly one of-these professors have all fem.-students adored 'All students adored exactly one of these professors.'

Genau einen dieser Professoren hat jede Studentin verehrt. exactly one of-these professors has every fem.-student adored 'Every student adored exactly one of these professors.'

Alle Studentinnen haben genau einen dieser Professoren verehrt. 'All students adored exactly one of these professors.'

Jede Studentin hat genau einen dieser Professoren verehrt. 'Every student adored exactly one of these professors.'

Experimental setup

- all three methods used in the pretest
- judgements using magnitude estimation
- 24 items, 36 fillers

Criteria for evaluation

the judgements collected using our methods should

- correspond to consensus in the theoretical literature
- correspond to corpus data

Judgements in the literature

Pafel (2004)

- linear precedence increases potential for wide scope
- +distributive feature increases potential for wide scope

The problem A methodological comparison Pretesting the methods Comparing the methods Summary

Predictions according to Pafel (2004)



Bott, Radó Quantifying quantifier scope

Corpus data

corpus: Cosmas - all public written corpora of German

sentences maximally similar to experimental items, i.e.

- simple sentences, no embedding
- two quantifiers: ein as object and jede or alle as subject of a simple transitive verb
- one quantifier in sentence-initial position
- max. 4 words between determiners
- no indefinite use of ein
- scope clearly disambiguated by context

Corpus results

- 21 sentences with ein before jede
- 31 sentences with jede before ein
- 6 sentences with ein before alle
- 40 sentences with alle before ein

Corpus results



Kann man von genau einem Professor sagen, dass jede Studentin ihn verehrt hat? 'Can it be said of exactly one professor that every student adored him?'

Ja, stimmt. Genau einen Professor hat jede Studentin verehrt.

Question-answer pairs: Results



Genau einen Professor hat jede Studentin verehrt.





∀∃ diagram

 $\exists \forall \text{ diagram}$

Set diagrams: Results



Genau ein Dreieck hat jedes Kind in seiner Ecke. 'Every child has exactly one triangle in his corner.'



∀∃ diagram



∃∀ diagram

Scenarios: Results



Outline

The problem

- A methodological comparison
- 3 Pretesting the methods
- Oomparing the methods



What we set out to find

methods for collecting semantic judgements that

- subjects can use easily
- are sensitive to the semantic phenomenon
- capture subtle differences
- do not introduce task-related biases

Summary of results

- natural-looking scenarios attractive but unsatisfactory
- suitable methods: question-answer pairs, set diagrams
 - fine-grained distinctions in scope preferences across conditions
 - but: set diagrams more reliable
- \rightarrow importance of cross-methodological comparison

Implications and extensions

theoretical implications

• theories of scope must allow for a continuum of differences

possible extensions

- other quantifier interactions
- other factors influencing scope e.g. intonation
- modal operator negation interaction
- cumulative vs. distributive readings

Implications and extensions

theoretical implications

theories of scope must allow for a continuum of differences

possible extensions

- other quantifier interactions
- other factors influencing scope e.g. intonation
- modal operator negation interaction
- cumulative vs. distributive readings

- included in main experiment testing question-answer pairs
- definite NPs instead of quantifiers
- no effect of congruence (all F < 0)
- no effect of word order

potential sources of complexity:

- additional PP in seiner Ecke
- internal structure of diagram

Genau ein Dreieck hat jedes Kind in seiner Ecke. 'Every child has exactly one triangle in his corner.'



