The Pragmatics of Quantifier Scope: A Corpus Study

Scott AnderBois, Adrian Brasoveanu and Robert Henderson

UC Berkeley, Feb. 10 2012

Possible Readings

- semanticists have generally been concerned with accounting for the range of **possible scopes** for a given sentence
- the semanticist's aim is roughly what previous literature terms *scope generation*

Possible Readings

- semanticists have generally been concerned with accounting for the range of **possible scopes** for a given sentence
- the semanticist's aim is roughly what previous literature terms *scope generation*

Scope Prediction

- we as semanticists generally do not weigh in on the **actual patterns of usage** of a given possible reading
- that is, semantics is not concerned with the problem of *quantifier scope disambiguation* (QSD)

Pragmatics of Quantifier Scope

In order to develop a model for QSD, we examine the factors influencing quantifier scope in a controlled, but naturally occurring body of text: LSAT Logic Puzzles.

Goal

- today, our aim is to introduce the corpus and report the main findings

Pragmatics of Quantifier Scope

In order to develop a model for QSD, we examine the factors influencing quantifier scope in a controlled, but naturally occurring body of text: LSAT Logic Puzzles.

Goal

- today, our aim is to introduce the corpus and report the main findings
- in particular, we're interested in those aspects of a quantifier's usage that are correlated with its wide vs. narrow scope ...

In order to develop a model for QSD, we examine the factors influencing quantifier scope in a controlled, but naturally occurring body of text: LSAT Logic Puzzles.

Goal

- today, our aim is to introduce the corpus and report the main findings
- in particular, we're interested in those aspects of a quantifier's usage that are correlated with its wide vs. narrow scope . . .
- . . . e.g., its position (before or after the other quantifier/s), its grammatical function (S, O etc.), its lexical realization (each, all etc.)

Linguistically & Psychologically Plausible Predictors

We designed the tagging scheme to reflect the features that have been argued to bias QSD in (some of) the psychological and computational literature, which we summarize now.

Linguistically & Psychologically Plausible Predictors

We designed the tagging scheme to reflect the features that have been argued to bias QSD in (some of) the psychological and computational literature, which we summarize now.

Linear order/C-command

1. Every professor saw a student. *every* $\gg$ *a*
2. A student saw every professor. *a* $\gg$ *every*

Linguistically & Psychologically Plausible Predictors

We designed the tagging scheme to reflect the features that have been argued to bias QSD in (some of) the psychological and computational literature, which we summarize now.

Linear order/C-command

1. Every professor saw a student. *every* $\gg$ *a*
2. A student saw every professor. *a* $\gg$ *every*

Gillen (1991), Kutzman & McDonald (1993), Tunstall (1998), Anderson (2004)

Linguistically & Psychologically Plausible Predictors

We designed the tagging scheme to reflect the features that have been argued to bias QSD in (some of) the psychological and computational literature, which we summarize now.

Linear order/C-command

1. Every professor saw a student. *every* $\gg$ *a*
2. A student saw every professor. *a* $\gg$ *every*

Gillen (1991), Kutzman & McDonald (1993), Tunstall (1998), Anderson (2004)

- it is difficult in English to separate the effect of linear order from the next predictor, grammatical function

Linguistically & Psychologically Plausible Predictors

Grammatical function hierarchy

3. Joan told a child the story at every intersection. *every* $\gg$ *a*
4. Joan told everyone the story at an intersection. *a* $\gg$ *every*

Linguistically & Psychologically Plausible Predictors

Grammatical function hierarchy

3. Joan told a child the story at every intersection. *every* $\gg$ *a*
4. Joan told everyone the story at an intersection. *a* $\gg$ *every*

$$\text{S} \gg \text{Prep} \gg \text{IO} \gg \text{O}$$

Linguistically & Psychologically Plausible Predictors

Grammatical function hierarchy

3. Joan told a child the story at every intersection. *every* $\gg$ *a*
4. Joan told everyone the story at an intersection. *a* $\gg$ *every*

$$\text{S} \gg \text{Prep} \gg \text{IO} \gg \text{O}$$

Kutzman & McDonald (1993), Tunstall (1998), Micham et al. (1980)

Ioup's (1975) Quantifier Hierarchy

5. She knows a solution to every problem. *every* $\gg$ *a*
6. She knows a solution to all problems. *a* $\gg$ *all*

Linguistically & Psychologically Plausible Predictors

Ioup's (1975) Quantifier Hierarchy

5. She knows a solution to every problem. *every* $\gg$ *a*
6. She knows a solution to all problems. *a* $\gg$ *all*

each $\gg$ every $\gg$ all $\gg$ most $\gg$ many $\gg$ several $\gg$ some$_{pl}$ $\gg$ a few

Linguistically & Psychologically Plausible Predictors

Ioup's (1975) Quantifier Hierarchy

5. She knows a solution to every problem. *every* $\gg$ *a*
6. She knows a solution to all problems. *a* $\gg$ *all*

each $\gg$ every $\gg$ all $\gg$ most $\gg$ many $\gg$ several $\gg$ some$_{pl}$ $\gg$ a few

Tunstall (1998), Van Lehn (1978)

Numerical Typicality

Saba & Corriveau (2001) propose a formal model of the world knowledge used in QSD based on the number of restrictor entities that typically participate in the nuclear scope relation.

Numerical Typicality

Saba & Corriveau (2001) propose a formal model of the world knowledge used in QSD based on the number of restrictor entities that typically participate in the nuclear scope relation.

7. A doctor lives in every city.

- the narrow scope reading of *every* is dispreferred because it would require an individual to participate in the *living-in* relation with an atypically large number of cities.

Numerical Typicality (ctd.)

Srinivasan & Yates (2009) show that numerical typicality can be extracted from a large corpus and applied successfully to QSD.

Numerical Typicality (ctd.)

Srinivasan & Yates (2009) show that numerical typicality can be extracted from a large corpus and applied successfully to QSD.

- applied to a handpicked corpus of 46 items

Numerical Typicality (ctd.)

Srinivasan & Yates (2009) show that numerical typicality can be extracted from a large corpus and applied successfully to QSD.

- applied to a handpicked corpus of 46 items
- information about numerical typicality significantly improves prediction, especially for inverse scope

Higgins & Sadock (2003)

Higgins & Sadock (2003) build a scope corpus from the WSJ Penn Treebank with the following properties:

Higgins & Sadock (2003)

Higgins & Sadock (2003) build a scope corpus from the WSJ Penn Treebank with the following properties:

- exactly 2 scope taking elements

Higgins & Sadock (2003)

Higgins & Sadock (2003) build a scope corpus from the WSJ Penn Treebank with the following properties:

- exactly 2 scope taking elements
- scope taking elements include most NPs with a determiner, predeterminer or measure phrase, e.g., *more than half*

Higgins & Sadock (2003)

Higgins & Sadock (2003) build a scope corpus from the WSJ Penn Treebank with the following properties:

- exactly 2 scope taking elements
- scope taking elements include most NPs with a determiner, predeterminer or measure phrase, e.g., *more than half*
- the result was 893 sentences, coded for scope by 2 people

Higgins & Sadock (2003)

Higgins & Sadock (2003) build a scope corpus from the WSJ Penn Treebank with the following properties:

- exactly 2 scope taking elements
- scope taking elements include most NPs with a determiner, predeterminer or measure phrase, e.g., *more than half*
- the result was 893 sentences, coded for scope by 2 people

Corpus Worries

- leave out NPs headed by *a/an*
- do not separate conjoined or appositive clauses, so the two quantifiers do not interact in 61% of the corpus

Higgins & Sadock (2003) (ctd.)

Three models (Naive Bayes, Maximum Entropy, Single Layer Perceptron) were trained on a subset of the corpus and each had an accuracy of 70%-80% on the remaining corpus.

Higgins & Sadock (2003) (ctd.)

Three models (Naive Bayes, Maximum Entropy, Single Layer Perceptron) were trained on a subset of the corpus and each had an accuracy of 70%-80% on the remaining corpus.

Predictors:

- the first quantifier c-commands the second or the second quantifier c-commands the first

Higgins & Sadock (2003) (ctd.)

Three models (Naive Bayes, Maximum Entropy, Single Layer Perceptron) were trained on a subset of the corpus and each had an accuracy of 70%-80% on the remaining corpus.

Predictors:

- the first quantifier c-commands the second or the second quantifier c-commands the first
- the second quantifier is *each*, *every*, *all*, a superlative adverb or a numerical measure phrase

Higgins & Sadock (2003) (ctd.)

Three models (Naive Bayes, Maximum Entropy, Single Layer Perceptron) were trained on a subset of the corpus and each had an accuracy of 70%-80% on the remaining corpus.

Predictors:

- the first quantifier c-commands the second or the second quantifier c-commands the first
- the second quantifier is *each*, *every*, *all*, a superlative adverb or a numerical measure phrase
- there is an intervening S node

Higgins & Sadock (2003) (ctd.)

Three models (Naive Bayes, Maximum Entropy, Single Layer Perceptron) were trained on a subset of the corpus and each had an accuracy of 70%-80% on the remaining corpus.

Predictors:

- the first quantifier c-commands the second or the second quantifier c-commands the first
- the second quantifier is *each*, *every*, *all*, a superlative adverb or a numerical measure phrase
- there is an intervening S node

- conjoined or appositive clauses were not separated, so other important predictors were intervening comma or colon, intervening conjunct node, intervening quotation mark etc.

Summary of Predictors in Previous Literature

Scope predictors in the previous computational and (psycho)linguistic literature:

- Linear order/C-command
- Grammatical hierarchy
- Particular quantificational item
- Intervening clause boundaries
- World knowledge

LSATs

- the LSAT exam consists of several types of questions: reading comprehension, analytical reasoning etc.

LSATs

- the LSAT exam consists of several types of questions: reading comprehension, analytical reasoning etc.
- our corpus is drawn from one particular type of question: analytical reasoning questions, a.k.a. logic puzzles

LSATs

- the LSAT exam consists of several types of questions: reading comprehension, analytical reasoning etc.
- our corpus is drawn from one particular type of question: analytical reasoning questions, a.k.a. logic puzzles
- logic puzzles follow a particular format as follows

Structure of a Logic Puzzle

In the course of one month Garibaldi has exactly seven different meetings. Each of her meetings is with exactly one of five foreign dignitaries: Fuentes, Matsuba, Rhee, Soleimani, or Tbahi. The following constraints govern Garibaldi's meetings:

} Introduction

Structure of a Logic Puzzle

In the course of one month Garibaldi has exactly seven different meetings. Each of her meetings is with exactly one of five foreign dignitaries: Fuentes, Matsuba, Rhee, Soleimani, or Tbahi. The following constraints govern Garibaldi's meetings:

Introduction

She has exactly three meetings with Fuentes, and exactly one with each of the other dignitaries.
She does not have any meetings in a row with Fuentes.
Her meeting with Soleimani is the very next one after her meeting with Tbahi.
Neither the first nor last of her meetings is with Matsuba.

Laws

Structure of a Logic Puzzle

In the course of one month Garibaldi has exactly seven different meetings. Each of her meetings is with exactly one of five foreign dignitaries: Fuentes, Matsuba, Rhee, Soleimani, or Tbahi. The following constraints govern Garibaldi's meetings:

Introduction

She has exactly three meetings with Fuentes, and exactly one with each of the other dignitaries.
She does not have any meetings in a row with Fuentes.
Her meeting with Soleimani is the very next one after her meeting with Tbahi.
Neither the first nor last of her meetings is with Matsuba.

Laws

2. If Garibaldi's last meeting is with Rhee, then which one of the following could be true?

Question

Structure of a Logic Puzzle

In the course of one month Garibaldi has exactly seven different meetings. Each of her meetings is with exactly one of five foreign dignitaries: Fuentes, Matsuba, Rhee, Soleimani, or Tbahi. The following constraints govern Garibaldi's meetings:

} Introduction

She has exactly three meetings with Fuentes, and exactly one with each of the other dignitaries.
She does not have any meetings in a row with Fuentes.
Her meeting with Soleimani is the very next one after her meeting with Tbahi.
Neither the first nor last of her meetings is with Matsuba.

} Laws

2. If Garibaldi's last meeting is with Rhee, then which one of the following could be true?

} Question

(A) Garibaldi's second meeting is with Soleimani.
(B) Garibaldi's third meeting is with Matsuba.
(C) Garibaldi's fourth meeting is with Soleimani.
(D) Garibaldi's fifth meeting is with Matsuba.
(E) Garibaldi's sixth meeting is with Soleimani.

} Answers

Minimal ambiguity

- test takers are expected to select a single correct answer, so ambiguity must be minimal.

Why Logic Puzzles?

Minimal ambiguity

- test takers are expected to select a single correct answer, so ambiguity must be minimal.

Minimal world knowledge

- as an aptitude test, the LSAT explicitly states assumptions which might be left to world knowledge in ordinary conversation
- in essence, the entire discourse context is made linguistically explicit, allowing us to abstract away from world knowledge

Why Logic Puzzles?

Minimal ambiguity

- test takers are expected to select a single correct answer, so ambiguity must be minimal.

Minimal world knowledge

- as an aptitude test, the LSAT explicitly states assumptions which might be left to world knowledge in ordinary conversation
- in essence, the entire discourse context is made linguistically explicit, allowing us to abstract away from world knowledge

Multiple quantifiers frequent

- sentences with two or more quantifiers are (unsurprisingly) quite frequent in this register

Syntactic Constraints

- in Higgins & Sadock (2003), the sentence was taken as the domain for quantifier scope regardless of syntactic complexity

Syntactic Constraints

- in Higgins & Sadock (2003), the sentence was taken as the domain for quantifier scope regardless of syntactic complexity
- however, it is often clear that a sentence consists of multiple separate scopal domains

Syntactic Constraints

- in Higgins & Sadock (2003), the sentence was taken as the domain for quantifier scope regardless of syntactic complexity
- however, it is often clear that a sentence consists of multiple separate scopal domains
- e.g., if two quantifiers appear in a coordinate structure . . .

Syntactic Constraints

- in Higgins & Sadock (2003), the sentence was taken as the domain for quantifier scope regardless of syntactic complexity
- however, it is often clear that a sentence consists of multiple separate scopal domains
- e.g., if two quantifiers appear in a coordinate structure . . .

8. **[** Joe ate <u>three oranges</u> **]** and **[** Pam ate <u>two apples</u> **]**.

Syntactic Constraints

- in Higgins & Sadock (2003), the sentence was taken as the domain for quantifier scope regardless of syntactic complexity
- however, it is often clear that a sentence consists of multiple separate scopal domains
- e.g., if two quantifiers appear in a coordinate structure . . .

8. **[** Joe ate three oranges **]** and **[** Pam ate two apples **]**.

- . . . the example is best treated as two separate scopal domains, one per sentential conjunct

Scopal Domains

- quotations and parenthetical content like appositive relative clauses similarly involve multiple scopal domains

Scopal Domains

- quotations and parenthetical content like appositive relative clauses similarly involve multiple scopal domains
- therefore, we consider *scopal domains* with multiple quantifiers, rather than *sentences*

Scopal Domains

- quotations and parenthetical content like appositive relative clauses similarly involve multiple scopal domains
- therefore, we consider *scopal domains* with multiple quantifiers, rather than *sentences*
- this is consistent with our stated goal of studying the *pragmatics* of quantifier scope

Scopal Domains

- quotations and parenthetical content like appositive relative clauses similarly involve multiple scopal domains
- therefore, we consider *scopal domains* with multiple quantifiers, rather than *sentences*
- this is consistent with our stated goal of studying the *pragmatics* of quantifier scope
- the lack of relative scope between quantifiers in different conjuncts of a coordinate clause is largely an observation about the syntax/semantics of quantifiers, not their pragmatics

Method

- quantifier scope by its nature requires a trained linguist to tag

Method

- quantifier scope by its nature requires a trained linguist to tag
- first, we separated the data into individual sentences and then further into scopal domains

Method

- quantifier scope by its nature requires a trained linguist to tag
- first, we separated the data into individual sentences and then further into scopal domains
- second, we enlisted undergrads to identify sentences with multiple quantifiers and make a first attempt at tagging them

Method

- quantifier scope by its nature requires a trained linguist to tag
- first, we separated the data into individual sentences and then further into scopal domains
- second, we enlisted undergrads to identify sentences with multiple quantifiers and make a first attempt at tagging them
- finally, we individually went through the corpus by hand producing the final tags

Method

- quantifier scope by its nature requires a trained linguist to tag
- first, we separated the data into individual sentences and then further into scopal domains
- second, we enlisted undergrads to identify sentences with multiple quantifiers and make a first attempt at tagging them
- finally, we individually went through the corpus by hand producing the final tags

No effort was made to quantify inter-annotator agreement (this would require additional skilled coders).

Method

- quantifier scope by its nature requires a trained linguist to tag
- first, we separated the data into individual sentences and then further into scopal domains
- second, we enlisted undergrads to identify sentences with multiple quantifiers and make a first attempt at tagging them
- finally, we individually went through the corpus by hand producing the final tags

No effort was made to quantify inter-annotator agreement (this would require additional skilled coders).

- based on Higgins & Sadock's (2003) study, we would expect to find fairly high variability

Response variable

Scope: the relative scope of the 2 or more quantifiers in a scopal domain

Response variable

Scope: the relative scope of the 2 or more quantifiers in a scopal domain

Predictors:

1. Linear order
2. Grammatical function
3. Lexical identity of quantifier

Tagged Categories

The beginning of a tag is marked by **&**, the end of a tag is marked by #, subtags are separated by _.

Tagged Categories

The beginning of a tag is marked by **&**, the end of a tag is marked by #, subtags are separated by _.

Scope

- we coded scope numerically, with **1** corresponding to widest scope and other numbers indicating narrower scope

Tagged Categories

The beginning of a tag is marked by **&**, the end of a tag is marked by #, subtags are separated by _.

Scope

- we coded scope numerically, with **1** corresponding to widest scope and other numbers indicating narrower scope
- quantifiers with no relative scope (mainly cumulative readings) were 'co-tagged' with the same number

The beginning of a tag is marked by **&**, the end of a tag is marked by #, subtags are separated by _.

Scope

- we coded scope numerically, with **1** corresponding to widest scope and other numbers indicating narrower scope
- quantifiers with no relative scope (mainly cumulative readings) were 'co-tagged' with the same number
- this is merely a convenience for examples with 2 quantifiers (we could simply not tag them in such cases) . . .

Tagged Categories

The beginning of a tag is marked by **&**, the end of a tag is marked by #, subtags are separated by _.

Scope

- we coded scope numerically, with **1** corresponding to widest scope and other numbers indicating narrower scope
- quantifiers with no relative scope (mainly cumulative readings) were 'co-tagged' with the same number
- this is merely a convenience for examples with 2 quantifiers (we could simply not tag them in such cases) . . .

9. Exactly six**&1#** employees must be assigned to exactly three**&1#** committees.

Scope (ctd.)

- . . . but necessary for sentences with 3 or more quantifiers, where 2 quantifiers may take wide scope relative to a 3rd, but not relative to one another

Scope (ctd.)

- . . . but necessary for sentences with 3 or more quantifiers, where 2 quantifiers may take wide scope relative to a 3rd, but not relative to one another

10. Exactly six**&2#** of seven**&1#** jugglers are each**&3#** assigned to exactly one**&4#** of three**&1#** positions.

Scope (ctd.)

- . . . but necessary for sentences with 3 or more quantifiers, where 2 quantifiers may take wide scope relative to a 3rd, but not relative to one another

10. Exactly six**&2#** of seven**&1#** jugglers are each**&3#** assigned to exactly one**&4#** of three**&1#** positions.

In cases where no truth conditional difference was clear, we used the felicity of "such that" paraphrases as our ultimate criterion.

Linear order

- linear order was not explicitly tagged for since this information is implicit in the sequence of assigned tags

Tagged Categories

Linear order

- linear order was not explicitly tagged for since this information is implicit in the sequence of assigned tags

Grammatical function

- we distinguished four syntactic roles as follows: **S**ubject, **O**bject, **P**ivot, **A**djunct
- for prepositions, we tagged individual prepositions separately (today, we only analyze S and O)

Linear order

- linear order was not explicitly tagged for since this information is implicit in the sequence of assigned tags

Grammatical function

- we distinguished four syntactic roles as follows: **S**ubject, **O**bject, **P**ivot, **A**djunct
- for prepositions, we tagged individual prepositions separately (today, we only analyze S and O)

Lexical identity

- if the determiner / pre-restrictor material was complex, we tagged it as a unit – e.g., **more.than.two**, **a.different** etc.

11. Each**&1_S_each#** tape is to be assigned to a different**&2_to_a.different#** time slot, . . .
12. . . . and no**&1_S_no#** tape is longer than any**&2_than_any#** other tape.

11. Each**&1_S_each#** tape is to be assigned to a different**&2_to_a.different#** time slot, . . .
12. . . . and no**&1_S_no#** tape is longer than any**&2_than_any#** other tape.

13. Each**&1_S_each#** professor has one or**&2_O_or#** more specialities.

11. Each**&1_S_each#** tape is to be assigned to a different**&2_to_a.different#** time slot, . . .
12. . . . and no**&1_S_no#** tape is longer than any**&2_than_any#** other tape.

13. Each**&1_S_each#** professor has one or**&2_O_or#** more specialities.

14. The judge of the show awards exactly four**&1_O_exactly.four#** ribbons to four**&1_to_four#** of the dogs.

The Dataset

- we focus on sentences with 2 quantifiers only

The Dataset

- we focus on sentences with 2 quantifiers only
- we remove the cumulative sentences

The Dataset

- we focus on sentences with 2 quantifiers only
- we remove the cumulative sentences
- we focus on S and O only (we drop the other grammatical functions)

The Dataset

- we focus on sentences with 2 quantifiers only
- we remove the cumulative sentences
- we focus on S and O only (we drop the other grammatical functions)
- we are left with 497 observations

The Dataset

- we focus on sentences with 2 quantifiers only
- we remove the cumulative sentences
- we focus on S and O only (we drop the other grammatical functions)
- we are left with 497 observations

- we have double counting: some sentences have both an S and an O quantifier and the scope of one completely determines the scope of the other

The Dataset

- we focus on sentences with 2 quantifiers only
- we remove the cumulative sentences
- we focus on S and O only (we drop the other grammatical functions)
- we are left with 497 observations

- we have double counting: some sentences have both an S and an O quantifier and the scope of one completely determines the scope of the other
- 139 doubly counted sentences

The Dataset

- we focus on sentences with 2 quantifiers only
- we remove the cumulative sentences
- we focus on S and O only (we drop the other grammatical functions)
- we are left with 497 observations

- we have double counting: some sentences have both an S and an O quantifier and the scope of one completely determines the scope of the other
- 139 doubly counted sentences
- we randomly sample one quantifier from each of them

The Dataset

- final N(umber of observations): 358

- final N(umber of observations): 358

Response variable

- SCOPE: factor with 2 levels (narrow, wide); 'success' level: wide

- final N(umber of observations): 358

Response variable

- SCOPE: factor with 2 levels (narrow, wide); 'success' level: wide

Fixed effects

1. LIN.ORD: factor with 2 levels (first, last); reference level: first
2. GRAM.FUN: factor with 2 levels (S, O); reference level: S

- final N(umber of observations): 358

Response variable

- SCOPE: factor with 2 levels (narrow, wide); 'success' level: wide

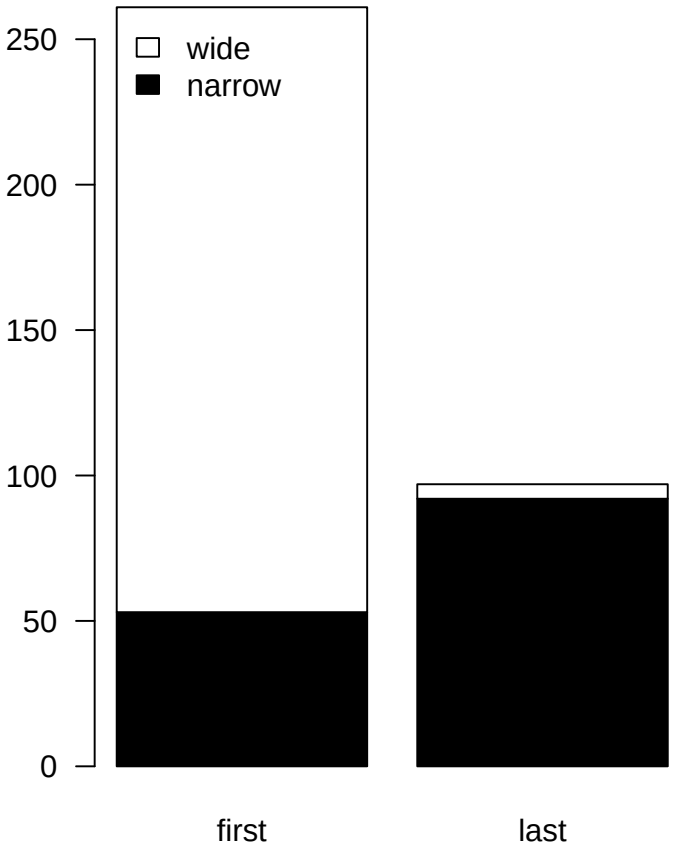
Fixed effects

1. LIN.ORD: factor with 2 levels (first, last); reference level: first
2. GRAM.FUN: factor with 2 levels (S, O); reference level: S

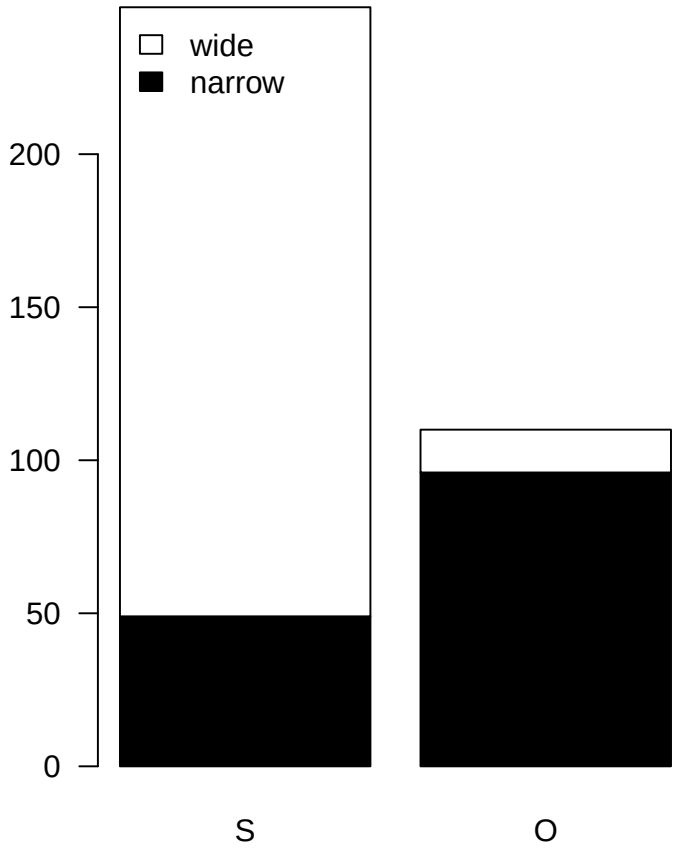
Random effects

1. LEX.REAL: factor with 17 levels (a, a.different, all, ...)
2. LEX.REAL.OTHER: factor with 19 levels (a, a.different, a.time, all, ...)

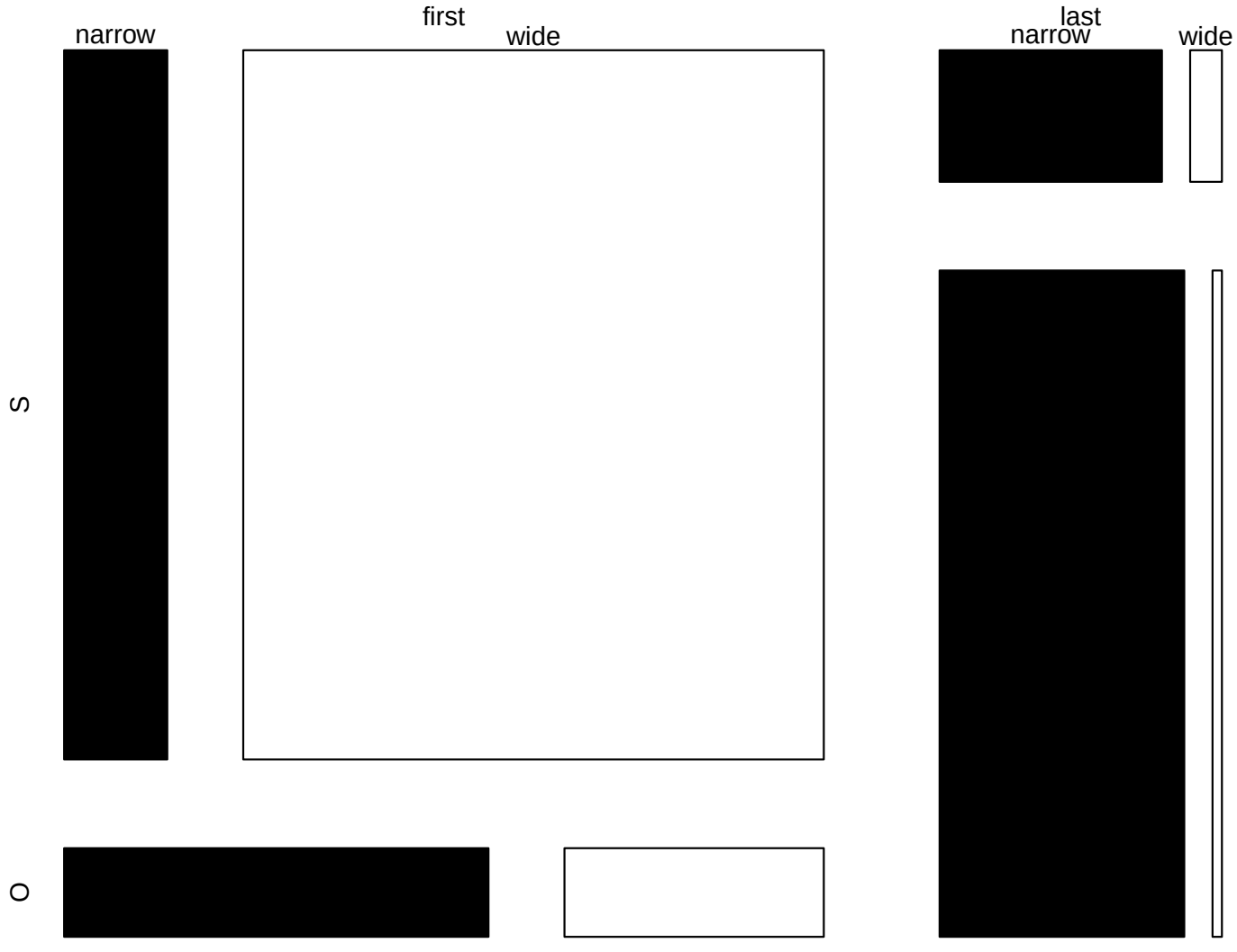
LIN.ORD by SCOPE



GRAM.FUN by SCOPE



The Dataset


LIN.ORD by GRAM.FUN by SCOPE
first
last
narrow
wide
narrow
wide
S
O

Examples for various conditions

GRAM.FUN=S, LIN.ORD=first

Examples for various conditions

GRAM.FUN=S, LIN.ORD=first

- SCOPE=wide:
  Each chair is occupied by exactly one representative.

GRAM.FUN=S, LIN.ORD=first

- SCOPE=wide:
  Each chair is occupied by exactly one representative.
- SCOPE=narrow:
  Exactly one child sits in each chair.

Examples for various conditions

GRAM.FUN=S, LIN.ORD=first

- SCOPE=wide:
  Each chair is occupied by exactly one representative.
- SCOPE=narrow:
  Exactly one child sits in each chair.

GRAM.FUN=S, LIN.ORD=last

Examples for various conditions

GRAM.FUN=S, LIN.ORD=first

- SCOPE=wide:
  Each chair is occupied by exactly one representative.
- SCOPE=narrow:
  Exactly one child sits in each chair.

GRAM.FUN=S, LIN.ORD=last

- SCOPE=wide:
  Every week **six crews** – A,B,C,D,E,F – were ranked from first (most productive) to sixth (least productive).

GRAM.FUN=S, LIN.ORD=first

- SCOPE=wide:
  Each chair is occupied by exactly one representative.
- SCOPE=narrow:
  Exactly one child sits in each chair.

GRAM.FUN=S, LIN.ORD=last

- SCOPE=wide:
  Every week **six crews** – A,B,C,D,E,F – were ranked from first (most productive) to sixth (least productive).
- SCOPE=narrow:
  On each day of other days of hiring, **exactly one** worker was hired.

Examples for various conditions

GRAM.FUN=0, LIN.ORD=first

Examples for various conditions

GRAM.FUN=O, LIN.ORD=first

- SCOPE=wide:
  He did not wash **any two of the objects** at the same time.

GRAM.FUN=O, LIN.ORD=first

- SCOPE=wide:
  He did not wash **any two of the objects** at the same time.
- SCOPE=narrow:
  The nine flowers used in the corsages must include **at least one flower** from each of the four types.

Examples for various conditions

GRAM.FUN=O, LIN.ORD=first

- SCOPE=wide:
  He did not wash **any two of the objects** at the same time.
- SCOPE=narrow:
  The nine flowers used in the corsages must include **at least one flower** from each of the four types.

GRAM.FUN=O, LIN.ORD=last

Examples for various conditions

GRAM.FUN=O, LIN.ORD=first

- SCOPE=wide:
  He did not wash **any two of the objects** at the same time.
- SCOPE=narrow:
  The nine flowers used in the corsages must include **at least one flower** from each of the four types.

GRAM.FUN=O, LIN.ORD=last

- SCOPE=wide:
  Exactly three girls perform **each dance**.

Examples for various conditions

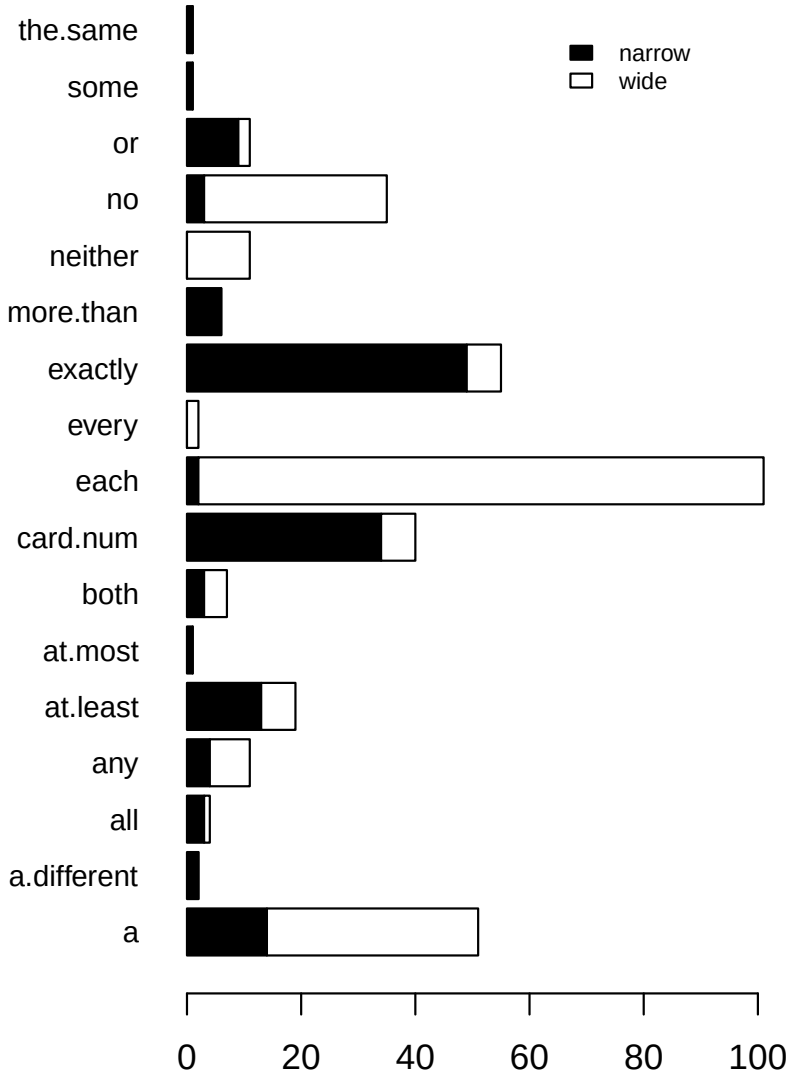
GRAM.FUN=0, LIN.ORD=first

- SCOPE=wide:
  He did not wash **any two of the objects** at the same time.
- SCOPE=narrow:
  The nine flowers used in the corsages must include **at least one flower** from each of the four types.

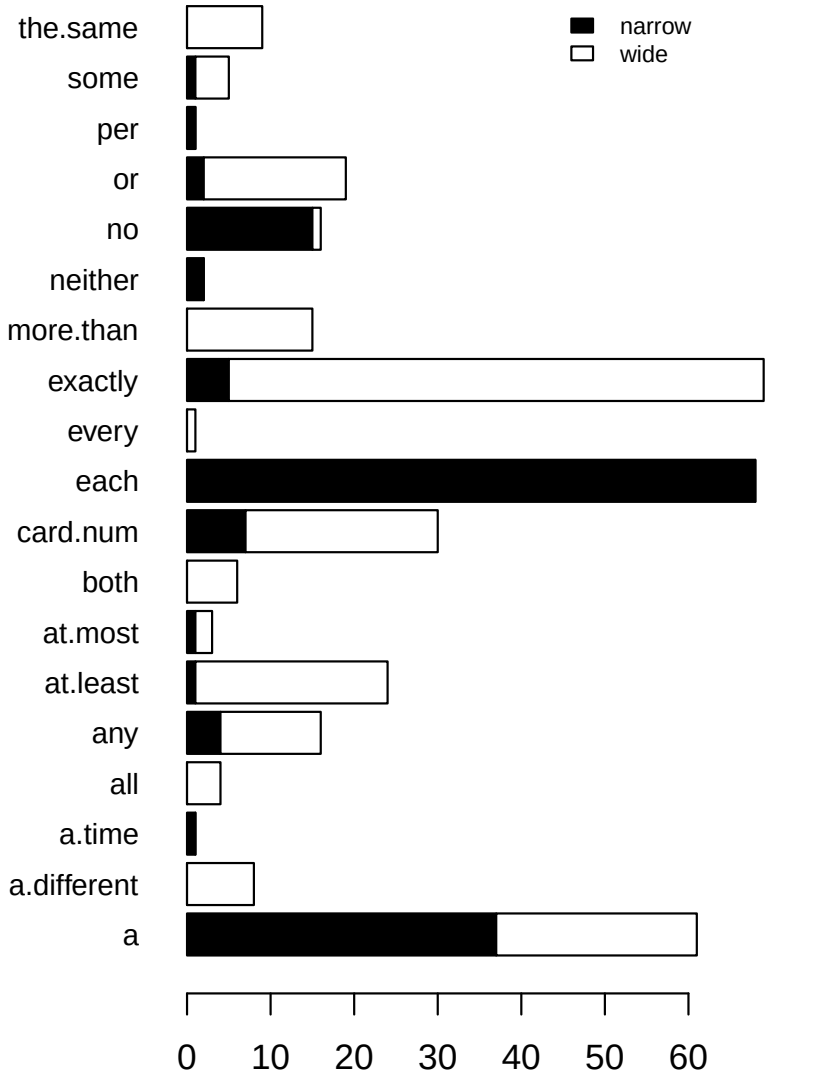
GRAM.FUN=0, LIN.ORD=last

- SCOPE=wide:
  Exactly three girls perform **each dance**.
- SCOPE=narrow:
  The official will also assign each runner to represent **a different charity**.

LEX.REAL by SCOPE



LEX.REAL.OTHER by SCOPE



Modeling and Resulting Generalizations

- we start with the full model for the fixed effects (the two main effects and their interaction) and intercept-only random effects for LEX.REAL and LEX.REAL.OTHER

Modeling and Resulting Generalizations

- we start with the full model for the fixed effects (the two main effects and their interaction) and intercept-only random effects for LEX.REAL and LEX.REAL.OTHER
- the interaction of LIN.ORD and GRAM.FUN is not significant (p=0.70), so we drop it

Modeling and Resulting Generalizations

- we start with the full model for the fixed effects (the two main effects and their interaction) and intercept-only random effects for LEX.REAL and LEX.REAL.OTHER
- the interaction of LIN.ORD and GRAM.FUN is not significant (p=0.70), so we drop it
- but adding GRAM.FUN to the model with LIN.ORD as the only fixed effect significantly reduces deviance (p=0.005)

Modeling and Resulting Generalizations

- we start with the full model for the fixed effects (the two main effects and their interaction) and intercept-only random effects for LEX.REAL and LEX.REAL.OTHER
- the interaction of LIN.ORD and GRAM.FUN is not significant (p=0.70), so we drop it
- but adding GRAM.FUN to the model with LIN.ORD as the only fixed effect significantly reduces deviance (p=0.005)
- similarly, adding LIN.ORD to the model with GRAM.FUN as the only fixed effect significantly reduces deviance (p=1.04e-07)

Modeling and Resulting Generalizations

- we start with the full model for the fixed effects (the two main effects and their interaction) and intercept-only random effects for LEX.REAL and LEX.REAL.OTHER
- the interaction of LIN.ORD and GRAM.FUN is not significant (p=0.70), so we drop it
- but adding GRAM.FUN to the model with LIN.ORD as the only fixed effect significantly reduces deviance (p=0.005)
- similarly, adding LIN.ORD to the model with GRAM.FUN as the only fixed effect significantly reduces deviance (p=1.04e-07)
- adding random effects for the LIN.ORD and / or GRAM.FUN slopes is not significant (when the MLEs of the resulting models can be estimated, which is not always possible)

Modeling and Resulting Generalizations

- we start with the full model for the fixed effects (the two main effects and their interaction) and intercept-only random effects for LEX.REAL and LEX.REAL.OTHER
- the interaction of LIN.ORD and GRAM.FUN is not significant (p=0.70), so we drop it
- but adding GRAM.FUN to the model with LIN.ORD as the only fixed effect significantly reduces deviance (p=0.005)
- similarly, adding LIN.ORD to the model with GRAM.FUN as the only fixed effect significantly reduces deviance (p=1.04e-07)
- adding random effects for the LIN.ORD and / or GRAM.FUN slopes is not significant (when the MLEs of the resulting models can be estimated, which is not always possible)
- but dropping the intercept random effects for LEX.REAL or LEX.REAL.OTHER significantly increases deviance (p= 3.21e-11 and p=2.08e-13, respectively)

Final Mixed-effects Logistic Regression Model

- intercept random effects for both LEX.REAL and LEX.REAL.OTHER
- fixed effects for LIN.ORD and GRAM.FUN (no interaction)

Final Mixed-effects Logistic Regression Model

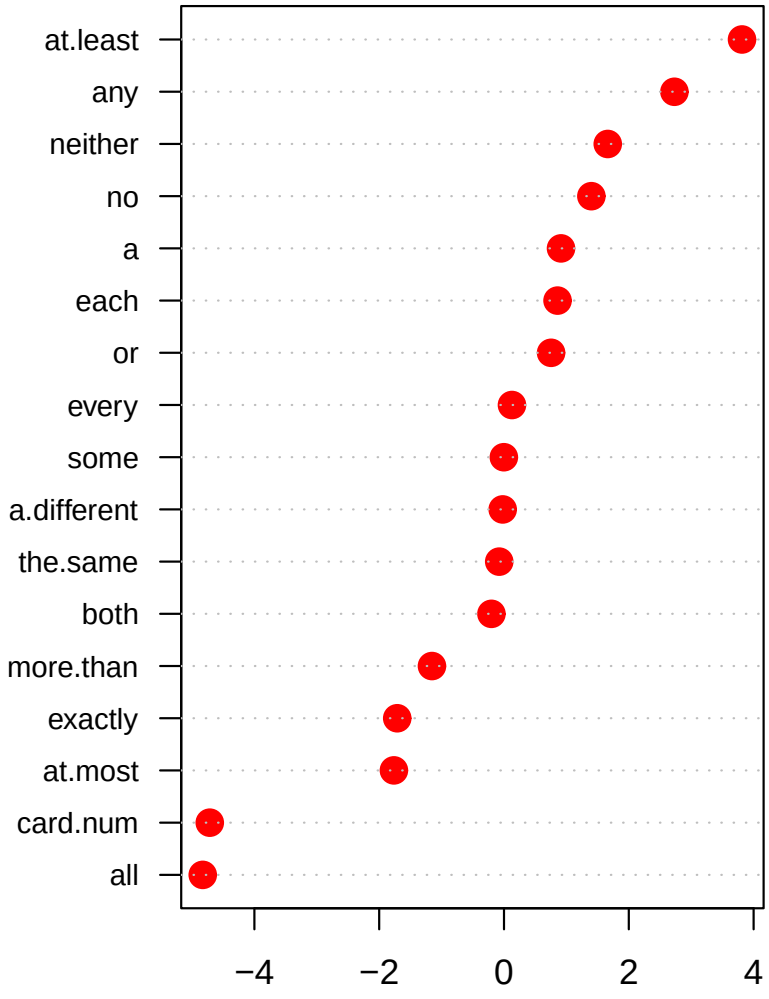
- intercept random effects for both LEX.REAL and LEX.REAL.OTHER
- fixed effects for LIN.ORD and GRAM.FUN (no interaction)

Maximum Likelihood Estimates (MLEs):

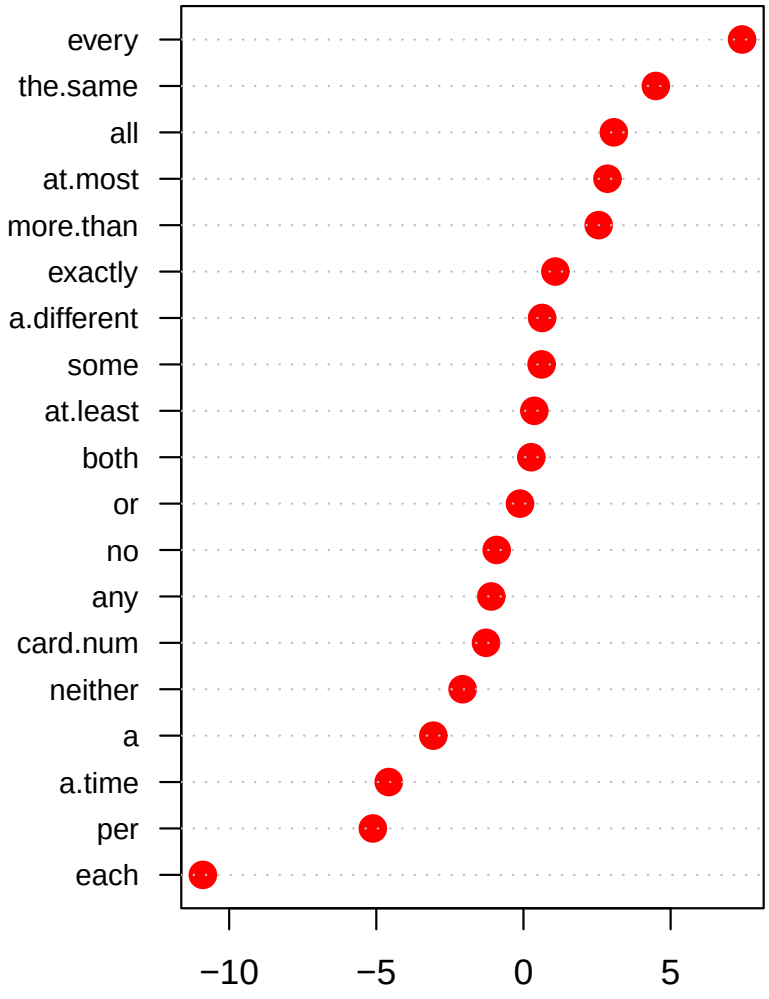
RANDOM EFFECTS		std.dev.		
	LEX.REAL	3.45		
	LEX.REAL.OTHER	5.55		
FIXED EFFECTS		**estimate**	**std.error**	**p-value**
	INTERCEPT	4.60	1.86	0.014
	LIN.ORD-LAST	-6.16	1.42	1e-05
	GRAM.FUN-O	-2.49	0.93	0.007

MLEs for Random Effects

LEX.REAL



LEX.REAL.OTHER



Main Results

- lexical effects on scoping preferences seem more important than linear order or grammatical function

Main Results

- lexical effects on scoping preferences seem more important than linear order or grammatical function
- in particular, the relational aspect of these lexical effects is important: LEX.REAL.OTHER seems to be at least as good a predictor of scope as LEX.REAL

Main Results

- lexical effects on scoping preferences seem more important than linear order or grammatical function
- in particular, the relational aspect of these lexical effects is important: LEX.REAL.OTHER seems to be at least as good a predictor of scope as LEX.REAL
- this provides a new kind of empirical support for *relational* theories of quantification that derive scopal behavior by focusing on the way in which one quantifier affects the context of interpretation for another quantifier, e.g., (in)dependence logic or dynamic plural logic

- the notion of interpretation context formalized in these logics is inherently relational because it focuses on context change
  - i.e., on the way in which an expression sets up the context of interpretation for a subsequent expression

Main Results

- the notion of interpretation context formalized in these logics is inherently relational because it focuses on context change
  - i.e., on the way in which an expression sets up the context of interpretation for a subsequent expression
- syntactic scoping mechanisms that focus on hierarchies of (classes of) quantifiers, e.g., Beghelli & Stowell (1997), are also supported

Semantics and processing

- identifying patterns of scoping behavior for quantifiers should ultimately enable us to group them into classes
- we might want our semantic theories to assign different kinds of semantic representations to these classes . . .
- . . . and / or we might want to hypothesize different processing strategies for these classes

Typology

- looking at the system of quantifiers in English opens the way towards examining cross-linguistic variation at system level, not only between individual (classes of) quantifiers

Typology

- looking at the system of quantifiers in English opens the way towards examining cross-linguistic variation at system level, not only between individual (classes of) quantifiers

Applied linguistics

- identifying and examining quantifier usage patterns is important for designing education and assessment materials in mathematics and sciences

Acknowledgments

First of all, we would like to thank the Law School Admission Council (LSAC) for access to practice test materials used in the analysis. We would also like to thank Pranav Anand, Donka Farkas, Matt Wagers and the participants in the UCSC Corpus Linguistics Group and the audiences at CUSP 3, MIT Syntax Square, ESSLLI 2011 PUQOL and SuB 2011 for helpful feedback.

This research was supported by an SRG (2009-2010) and an FRG grant (2010-2011) to Adrian Brasoveanu from Committee on Research, UC Santa Cruz.

References

Anderson, C. The Structure and Real-time Comprehension of Quantier Scope Ambiguity. PhD thesis, Department of Linguistics, Northwestern University.

Beghelli, F. & T. Stowell. (1997). Distributivity and Negation: The Syntax of Each and Every. In *Ways of Scope Taking*, A. Szabolcsi (ed.), Springer, 71-107.

Gillen, K. (1991) The Comprehension of Doubly Quantified Sentences. PhD Thesis, University of Durham.

Higgins, D. & Sadock, J. (2003). A Machine Learning Approach to Modeling Scope Preferences. *Computational Linguistics* 29(1): 73-96.

Ioup, G. (1975). Some Universals for Quantier Scope. In J. Kimball, editor, *Syntax and Semantics*, volume 4, pages 37-58. Academic Press, New York.

Kurtzman, H. & MacDonald, M. (1993). Resolution of quantifier scope ambiguities. *Cognition*, 48:243-279.

Micham, D., J. Catlin, N VanDerven & K. Loveland (1980). Lexical and Structural Cues in Quantifier Scope Relations. *Journal of Psycholinguistics Research*, 9: 367-377.

Van Lehn, K. (1978). Determining the scope of English quantifiers. Technical Report AI-TR-483, AI Lab, MIT.

Saba, W. & Corriveau, J-P (2001). Plausible reasoning and the resolution of quantier scope ambiguities. Studia Logica, 67: 271-289.

Srinivasan, P. & Yates, A. (2009). Quantier scope disambiguation using extracted pragmatic knowledge: Preliminary results. In Proceedings of the Conference on Empirical Methods in Natural Language Processing.

Tunstall, S. (1998) The interpretation of quantifiers: semantics and processing. PhD Thesis, University of Massachusetts Amherst.

Appendix: Model Fit

- C is an index of concordance between predicted probability and observed response
- Somers' Dxy is a rank correlation between predicted probabilities and observed responses related to C

Appendix: Model Fit

- C is an index of concordance between predicted probability and observed response
- Somers' Dxy is a rank correlation between predicted probabilities and observed responses related to C

The final mixed-effects logistic regression model

- C: 0.996
- Dxy: 0.992

Appendix: Model Fit

- C is an index of concordance between predicted probability and observed response
- Somers' Dxy is a rank correlation between predicted probabilities and observed responses related to C

The final mixed-effects logistic regression model

- C: 0.996
- Dxy: 0.992

Compare with the fixed-effects only model

- C: 0.859
- Dxy: 0.717

Appendix: Model Fit

- C is an index of concordance between predicted probability and observed response
- Somers' Dxy is a rank correlation between predicted probabilities and observed responses related to C

The final mixed-effects logistic regression model

- C: 0.996
- Dxy: 0.992

Compare with the fixed-effects only model

- C: 0.859
- Dxy: 0.717

Compare with the random-effects only model

- C: 0.982
- Dxy: 0.965

Appendix: Model Fit

- Nagelkerke R2: a common pseudo-R2 measure for (mixed-effects) logit models assessing how much of the 'variance' in the response is accounted for by the predictors

Appendix: Model Fit

- Nagelkerke R2: a common pseudo-R2 measure for (mixed-effects) logit models assessing how much of the 'variance' in the response is accounted for by the predictors
- it assesses the quality of a model with regard to the model with only the intercept

Appendix: Model Fit

- Nagelkerke R2: a common pseudo-R2 measure for (mixed-effects) logit models assessing how much of the 'variance' in the response is accounted for by the predictors
- it assesses the quality of a model with regard to the model with only the intercept

Nagelkerke R2

- our final model relative to the mixed intercept model: 0.404

Appendix: Model Fit

- Nagelkerke R2: a common pseudo-R2 measure for (mixed-effects) logit models assessing how much of the 'variance' in the response is accounted for by the predictors
- it assesses the quality of a model with regard to the model with only the intercept

Nagelkerke R2

- our final model relative to the mixed intercept model: 0.404
- our final model relative to the ordinary intercept model: 0.847

- Nagelkerke R2: a common pseudo-R2 measure for (mixed-effects) logit models assessing how much of the 'variance' in the response is accounted for by the predictors
- it assesses the quality of a model with regard to the model with only the intercept

Nagelkerke R2

- our final model relative to the mixed intercept model: 0.404
- our final model relative to the ordinary intercept model: 0.847
- the mixed intercept model relative to the ordinary intercept model: 0.743

Appendix: Bayesian Estimates

We can quantify these wide / narrow scope preferences more precisely based on the Bayesian estimates of their posterior distributions.

We can quantify these wide / narrow scope preferences more precisely based on the Bayesian estimates of their posterior distributions.

Priors

- fixed effects: the priors for the intercept and the non-reference levels of LIN.ORD and GRAM.FUN are all independent normals $N(0, 10^2)$

We can quantify these wide / narrow scope preferences more precisely based on the Bayesian estimates of their posterior distributions.

Priors

- fixed effects: the priors for the intercept and the non-reference levels of LIN.ORD and GRAM.FUN are all independent normals $N(0, 10^2)$
- random effects: we assume independent normal distributions $N(0, \sigma^2)$ and $N(0, \tau^2)$ for the LEX.REAL and LEX.REAL.OTHER random effects; the priors for the standard deviations σ and τ are independent uniforms $Unif(0, 30)$

We can quantify these wide / narrow scope preferences more precisely based on the Bayesian estimates of their posterior distributions.

Priors

- fixed effects: the priors for the intercept and the non-reference levels of LIN.ORD and GRAM.FUN are all independent normals $N(0, 10^2)$
- random effects: we assume independent normal distributions $N(0, \sigma^2)$ and $N(0, \tau^2)$ for the LEX.REAL and LEX.REAL.OTHER random effects; the priors for the standard deviations σ and τ are independent uniforms $Unif(0, 30)$

MCMC estimation: 3 chains, 3.5e+06 iterations per chain, 1e+06 burnin, 2500 thinning.

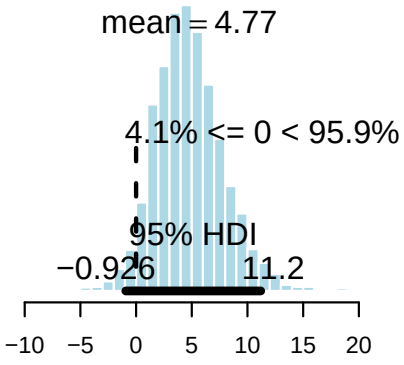
Summaries of the posterior distributions

The means and standard deviations of the posterior distributions for the random and fixed effects are fairly close to the MLEs:

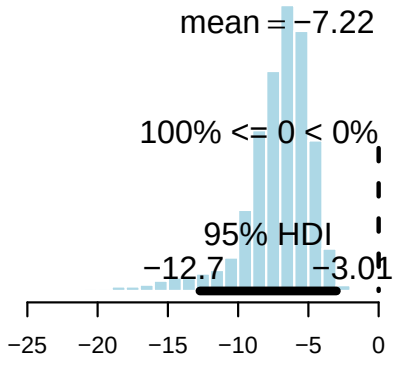
RANDOM EFFECTS		mean	std.dev.
	σ	5.37	2.48
	τ	9.65	4.32
FIXED EFFECTS		mean	std.dev.
	INTERCEPT	4.77	3.13
	LIN.ORD-LAST	-7.22	2.64
	GRAM.FUN-O	-2.67	1.07

Posteriors for Fixed Effects and Random Effect Std.Dev.s

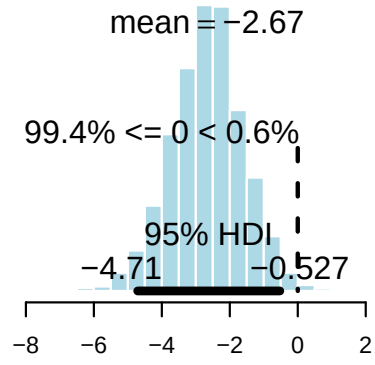
INTERCEPT



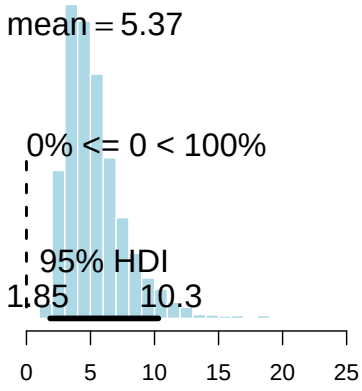
LIN.ORD-LAST



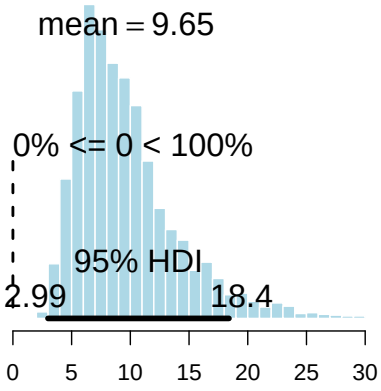
GRAM.FUN-O



LEX.REAL std.dev.

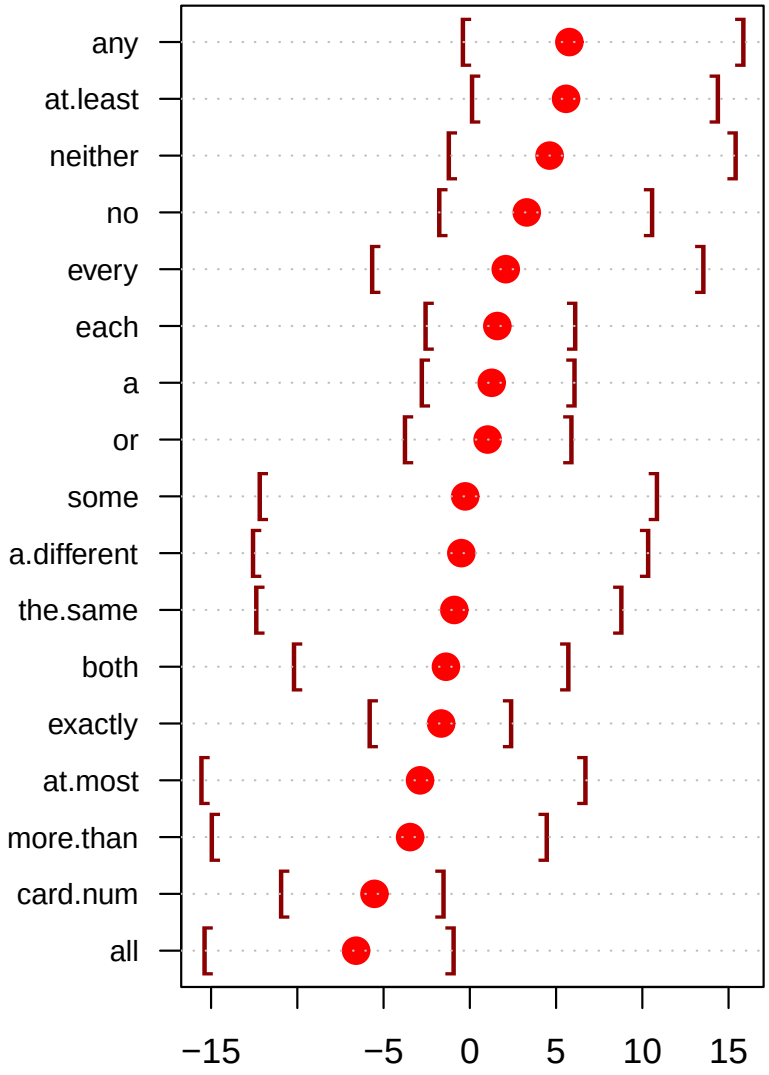


LEX.REAL.OTHER std.dev.

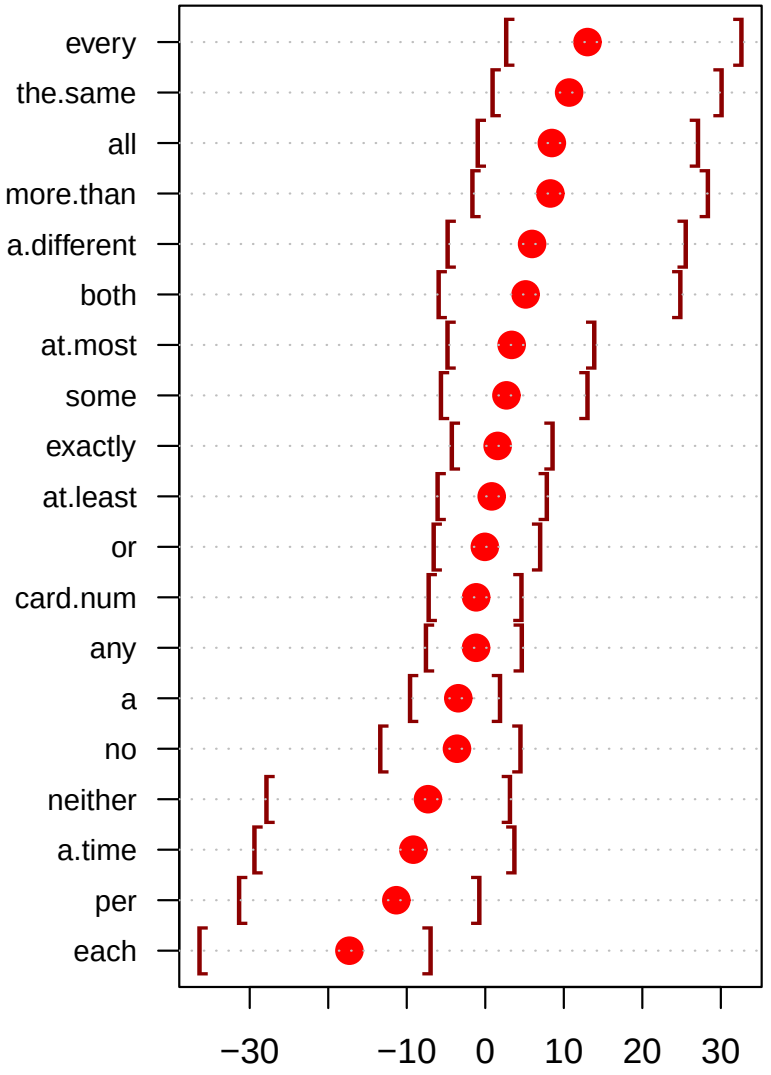


Random Effects (Means and 95% CRIs)

LEX.REAL

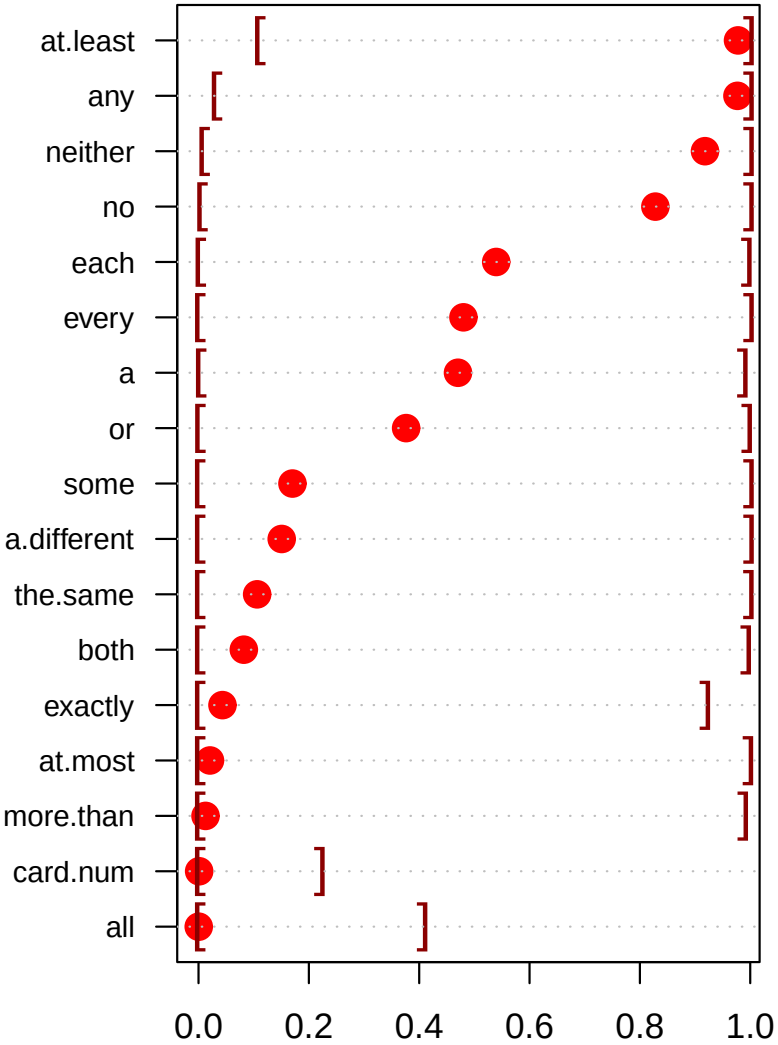


LEX.REAL.OTHER



Lex. Pref.s for Wide Scope (Median Prob.s and 95% CRIs)

LEX.REAL



LEX.REAL.OTHER