Generalized Quantifiers

From Logic to Cognitive Science

Jakub Szymanik

Stockholm University

The 11th Szklarska Poręba Workshop

Outline

Outline

NL determiners

1. All poets have low self-esteem.
2. Some dean danced nude on the table.
3. At least 3 grad students prepared presentations.
4. An even number of the students saw a ghost.
5. Most of the students think they are smart.
6. Less than half of the students received good marks.

Quantifiers are useful

Everyone knows everyone here.

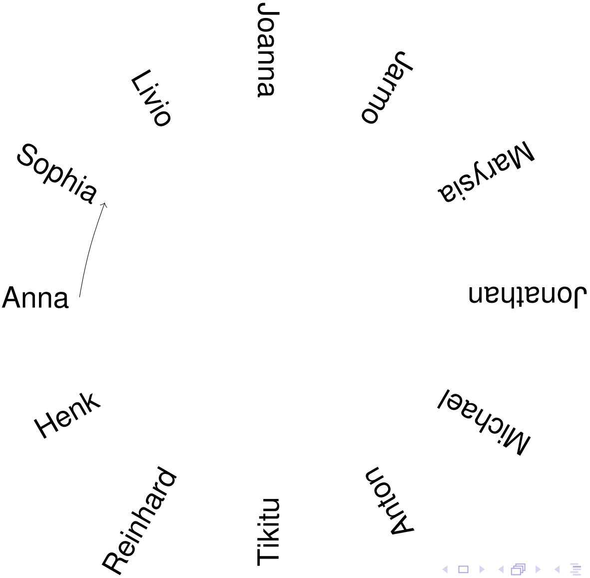
Quantifiers are useful

Everyone knows everyone here.

Joanna
Jarmo
Marysia
Jonathan
Michael
Anton
Tikitu
Reinhard
Henk
Anna
Sophia
Livio

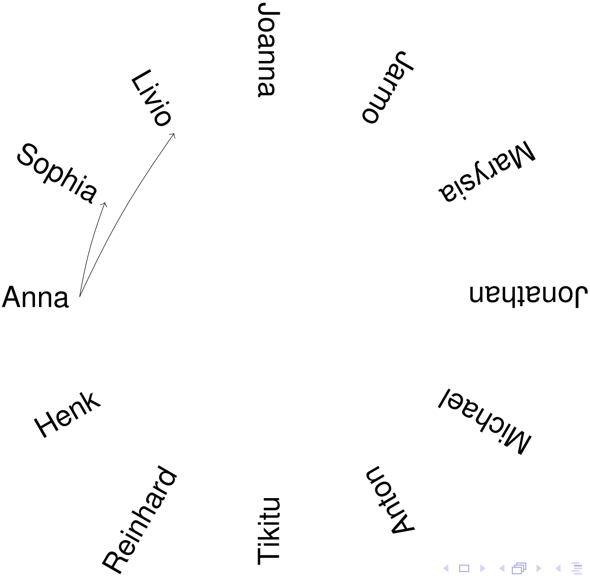
Quantifiers are useful

Everyone knows everyone here.



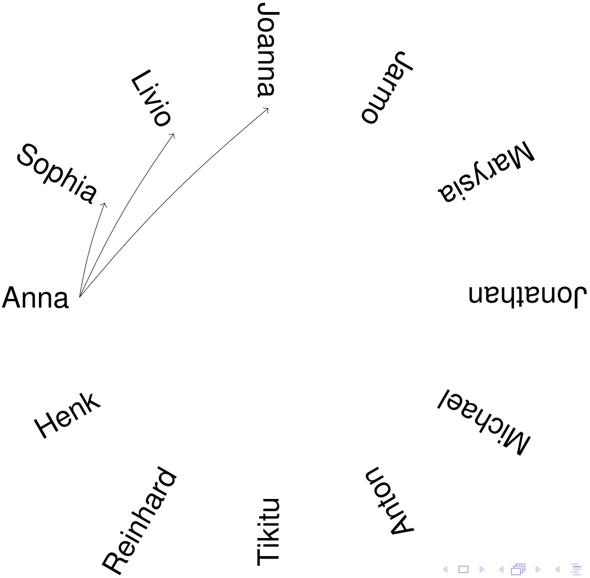
Quantifiers are useful

Everyone knows everyone here.



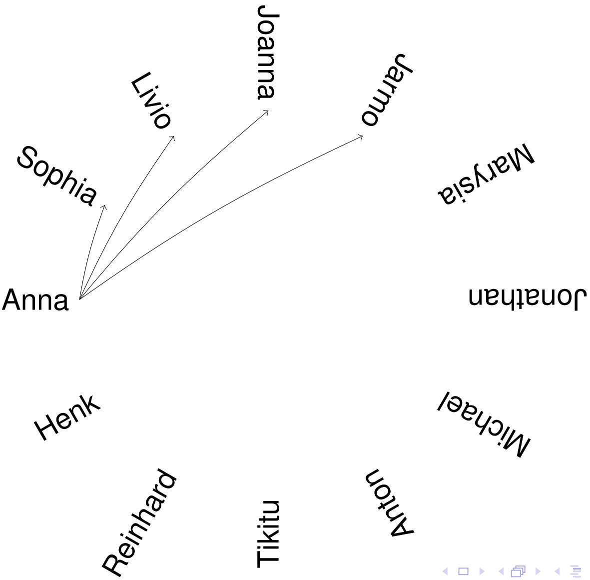
Quantifiers are useful

Everyone knows everyone here.



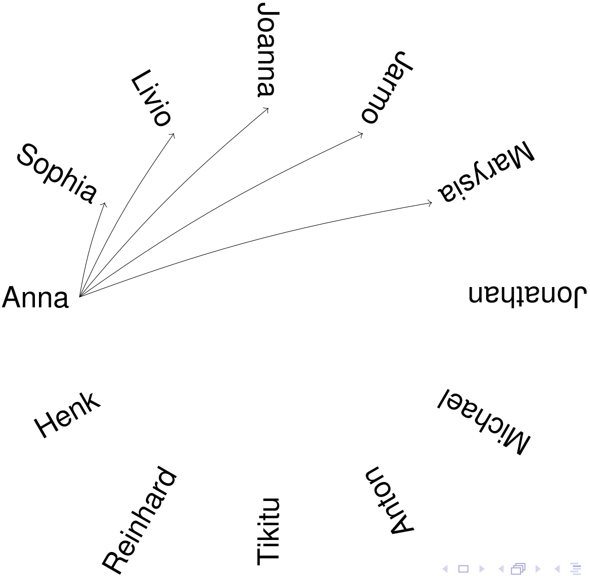
Quantifiers are useful

Everyone knows everyone here.



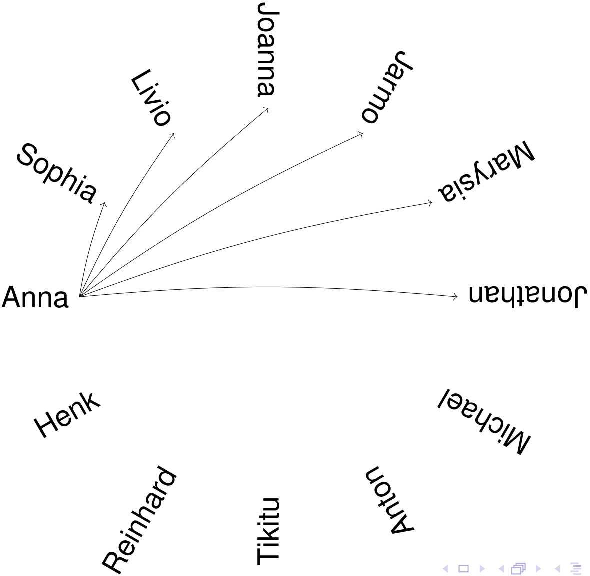
Quantifiers are useful

Everyone knows everyone here.



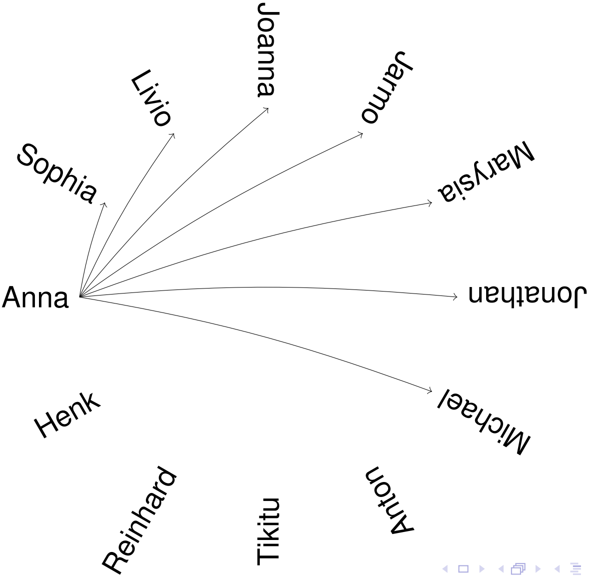
Quantifiers are useful

Everyone knows everyone here.



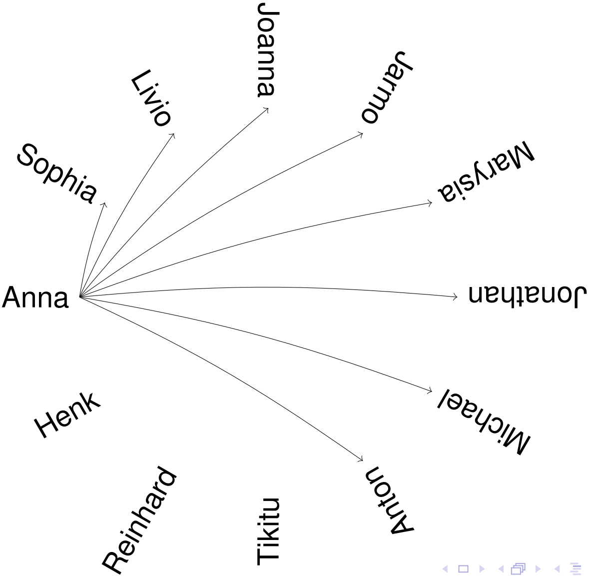
Quantifiers are useful

Everyone knows everyone here.



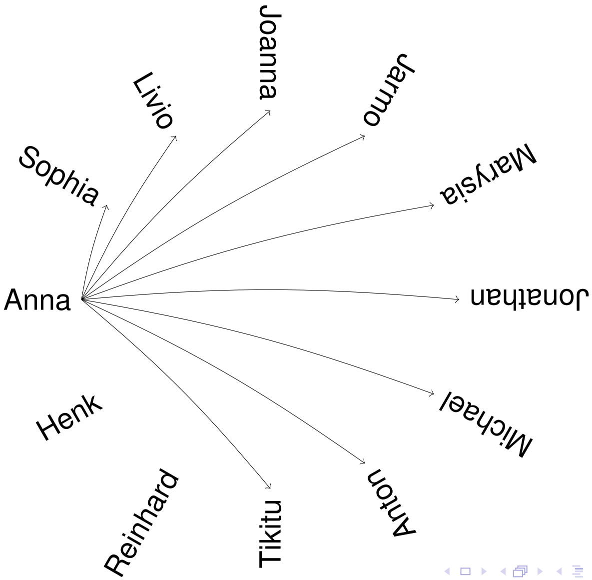
Quantifiers are useful

Everyone knows everyone here.



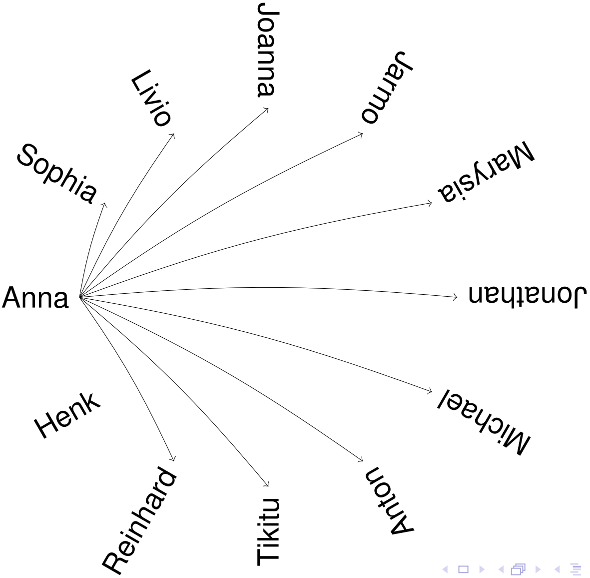
Quantifiers are useful

Everyone knows everyone here.



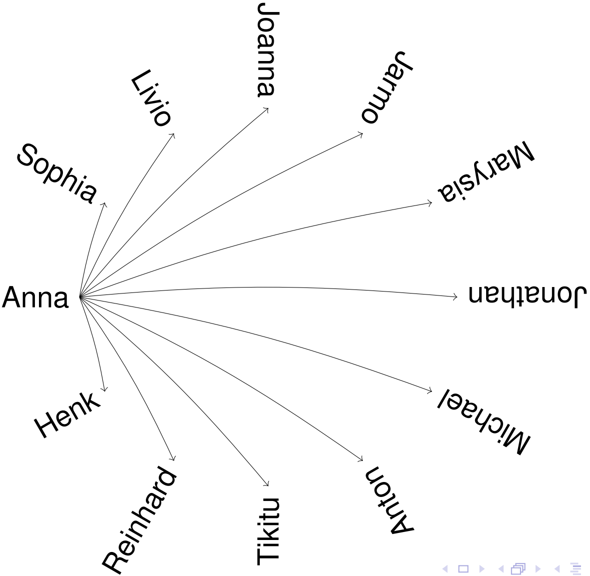
Quantifiers are useful

Everyone knows everyone here.



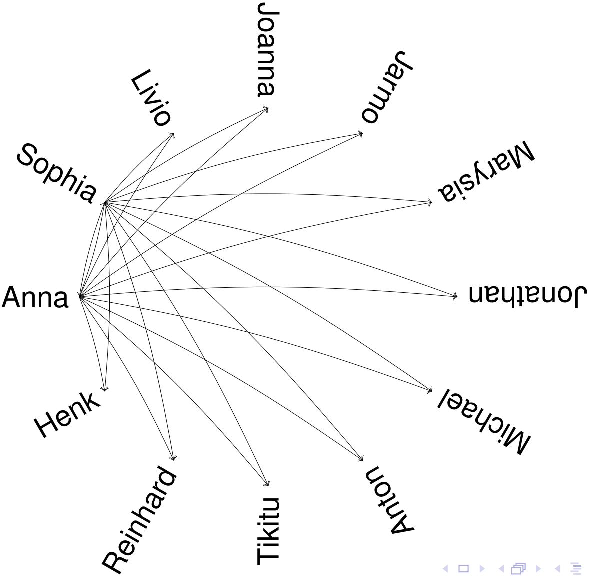
Quantifiers are useful

Everyone knows everyone here.



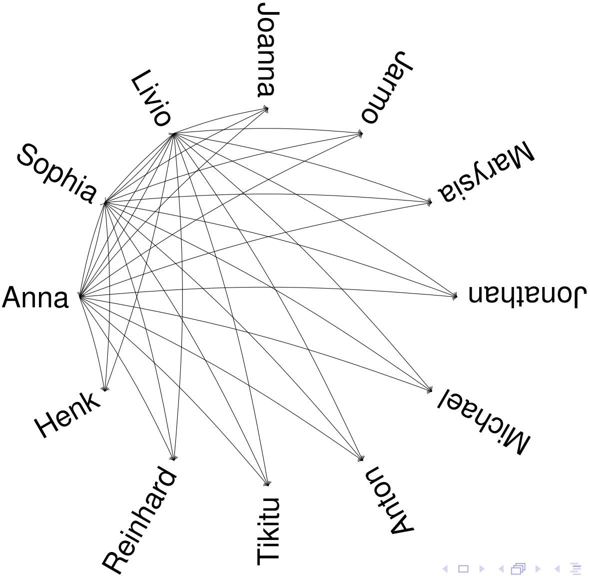
Quantifiers are useful

Everyone knows everyone here.



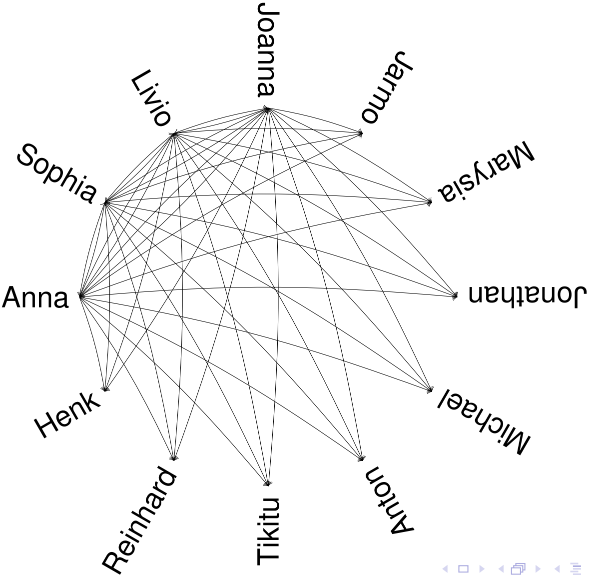
Quantifiers are useful

Everyone knows everyone here.



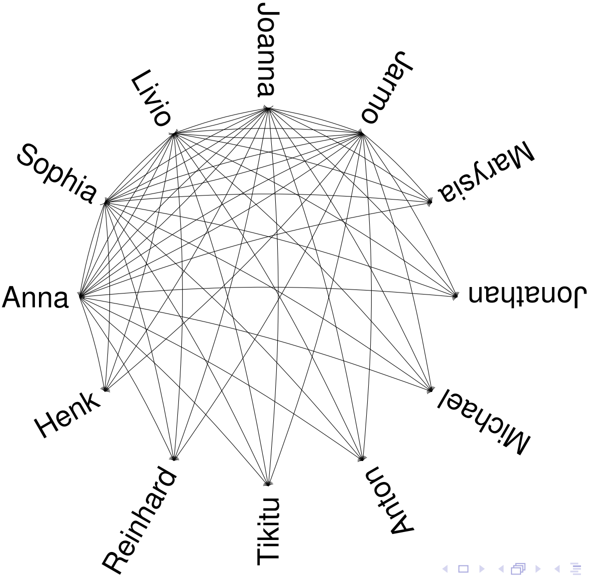
Quantifiers are useful

Everyone knows everyone here.



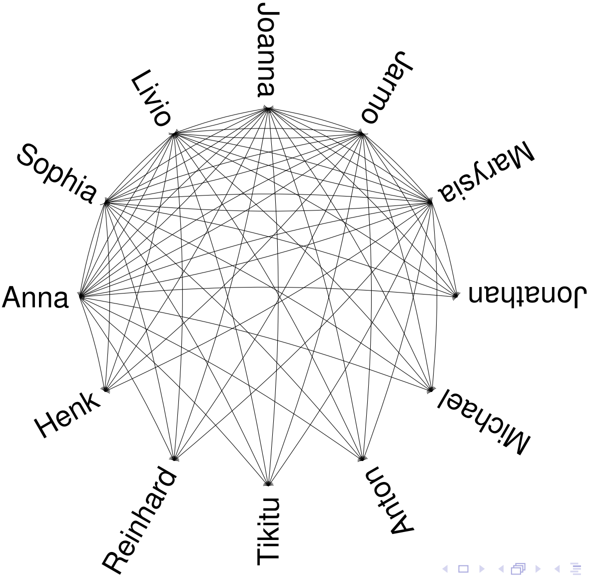
Quantifiers are useful

Everyone knows everyone here.



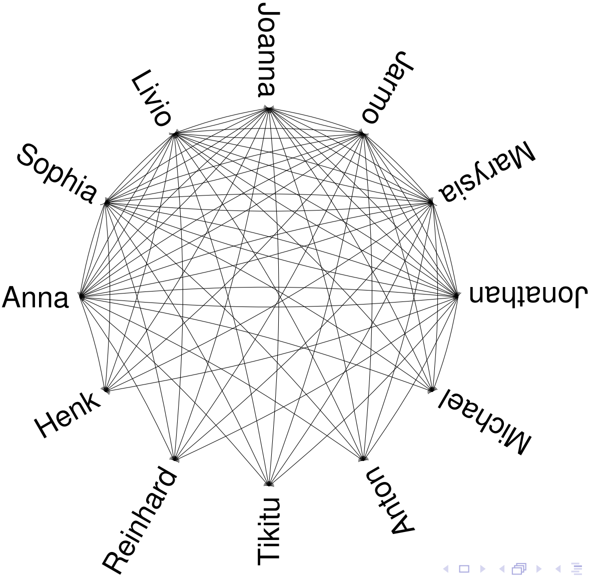
Quantifiers are useful

Everyone knows everyone here.



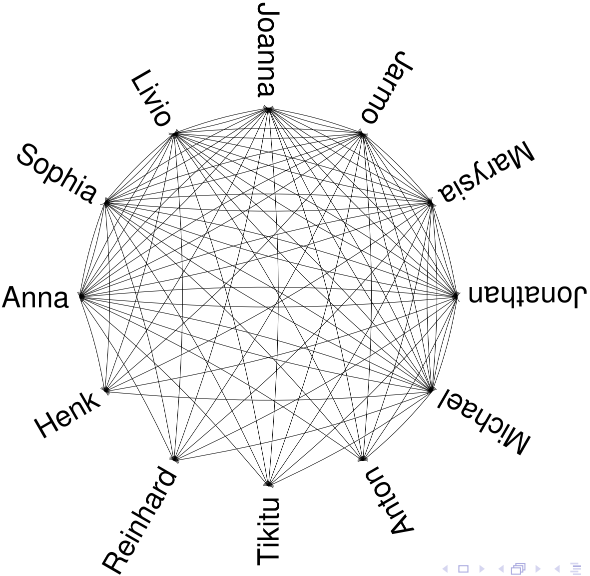
Quantifiers are useful

Everyone knows everyone here.



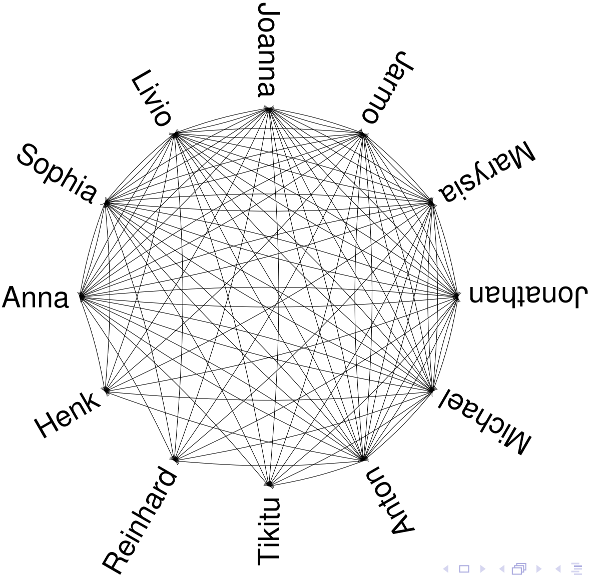
Quantifiers are useful

Everyone knows everyone here.



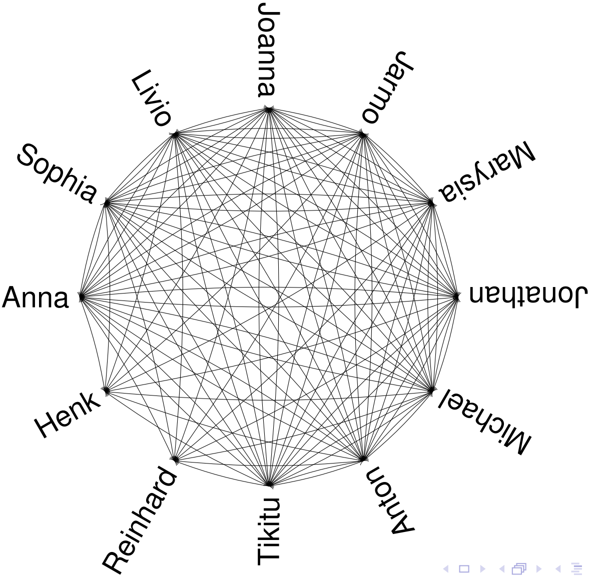
Quantifiers are useful

Everyone knows everyone here.



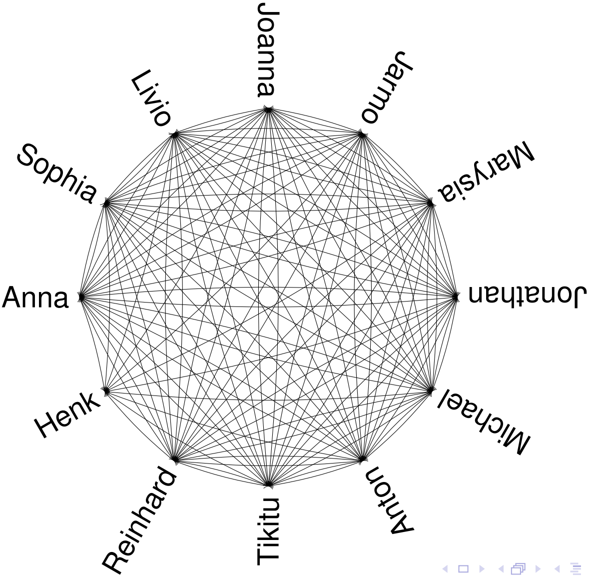
Quantifiers are useful

Everyone knows everyone here.



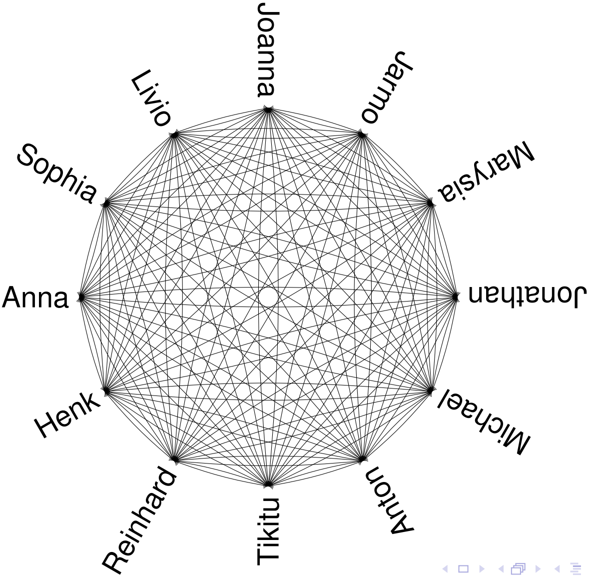
Quantifiers are useful

Everyone knows everyone here.

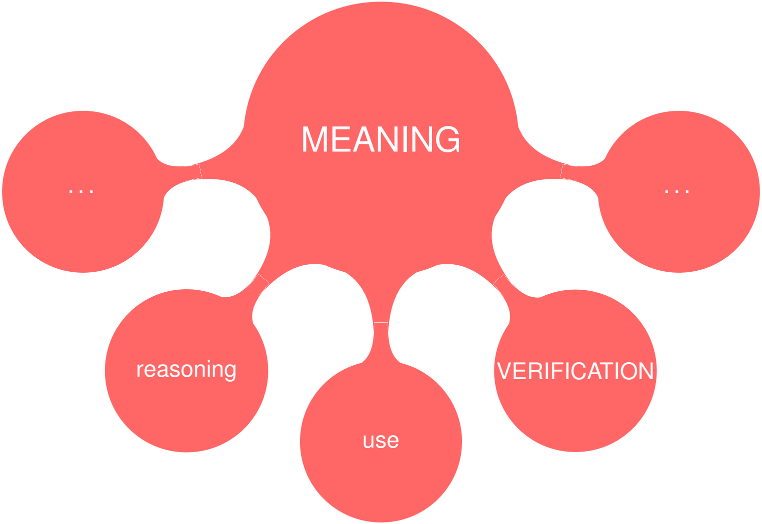


Quantifiers are useful

Everyone knows everyone here.



We understand quantifiers



Let’s focus on verification

More than half of the cars are yellow.

An example of a stimulus used in the sentence verification task

How are people doing it?

- They apply some strategies/procedures/algorithms.
- Those depend on:
  - quantifiers in question;
  - visual clues;
  - level of precision subjects want to achieve;
  - …

Meaning as algorithm

- Ability of understanding sentences.
- Capacity of recognizing their truth-values.
- Fregean tradition.
- Meaning is a procedure for finding extension in a model.
- Adopted often with psychological motivations.

Suppes, Variable-free semantics with remark on procedural extensions, 1982.

Lambalgen & Hamm,The Proper Treatment of Events, 2005.

Abstract task

From a computational perspective this is just model-checking:

Input: Qφ and M
Problem: $M \models Q\varphi$?
Answer: YES/NO

Immerman, Descriptive Complexity, Springer 1998.

A common question

Question

How complex are different quantifier fragments of NL?

A common question

Question

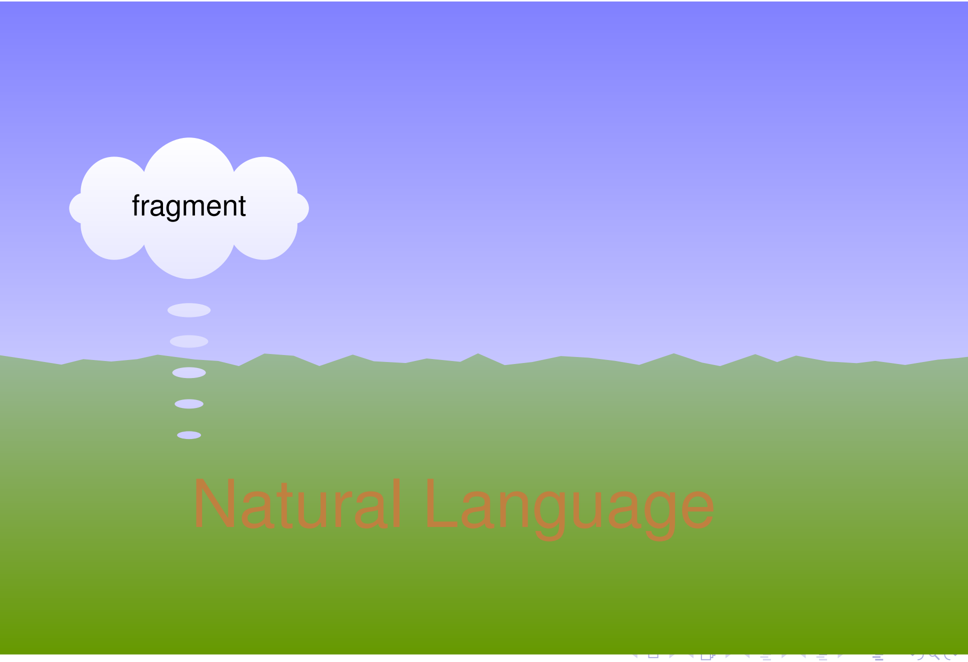
How complex are different quantifier fragments of NL?

1. Expressivity $\hookrightarrow$ controlled languages;
2. Difficulty $\hookrightarrow$ cognitive science;

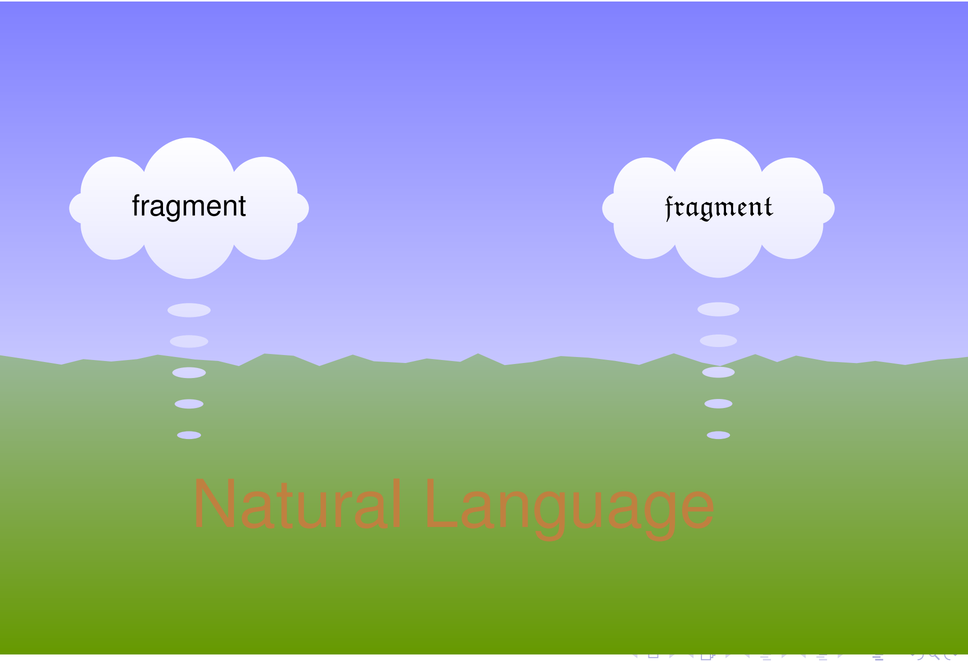
Pratt-Hartmann & Moss, Logics for the relational syllogistic, The Review of Symbolic Logic, 2009

Illustration

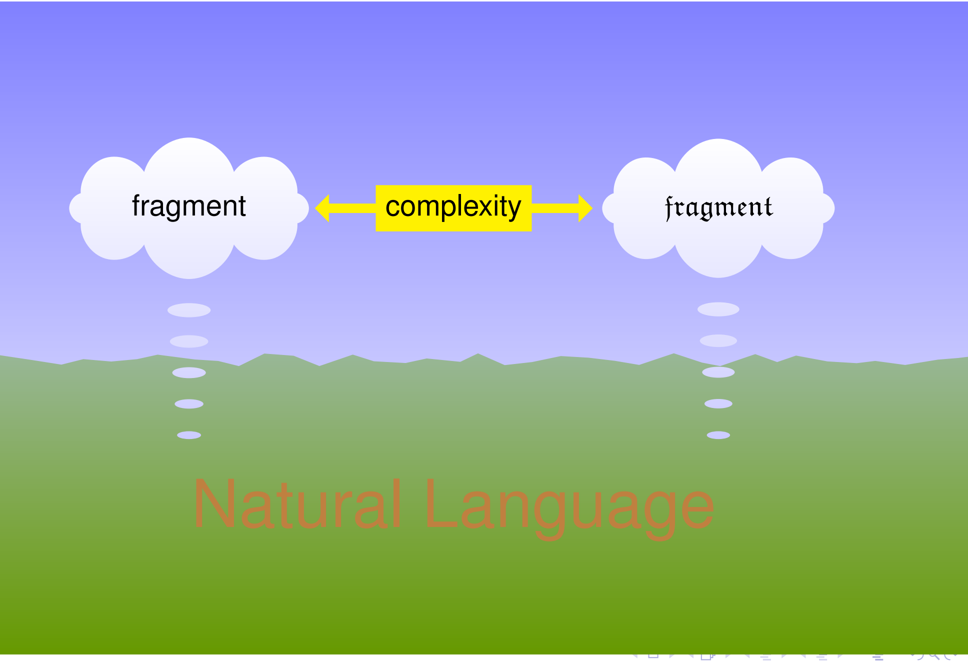
Illustration



Illustration



Illustration



Outline

Simplicity

Simple quantifiers can be computed by simple automata.

Simplicity

Simple quantifiers can be computed by simple automata.

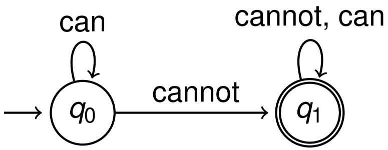
Question

What are the minimal automata for certain quantifier types?

van Benthem, Essays in logical semantics, 1986

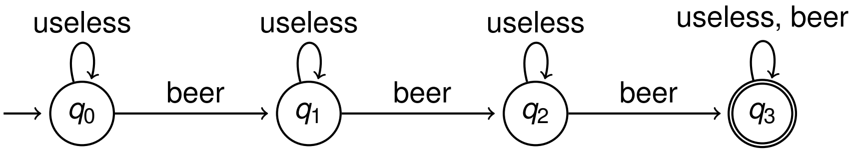
Example 1: Aristotelian quantifiers

- Someone cannot ski the black slope.



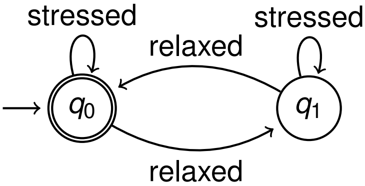
Example 2: Cardinal quantifiers

- There are at least 3 beers in that room.



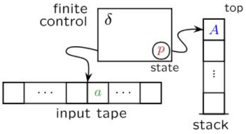
Example 3: Parity quantifiers

- An even number of us is relaxed.



Example 4: Proportional quantifiers

- “Most of us like Żubrówka.”



- Not computable by finite-automata.
- We need working memory.
- Simple push-down automata will do.

Does it say anything about processing?

Question

Do minimal automata predict differences in verification?

Does it say anything about processing?

Question

Do minimal automata predict differences in verification?

We'll try to convince you that the answer is positive!

Outline

Predictions

- RT will increase along with the computational resources.

Predictions

- RT will increase along with the computational resources.
- Aristotelian qua. < parity qua. < proportional qua.

Predictions

- RT will increase along with the computational resources.
- Aristotelian qua. $<$ parity qua. $<$ proportional qua.
- Aristotelian qua. $<$ cardinal qua. of high rank.

Participants

- 40 native Polish-speaking adults (21 female).
- Volunteers: undergraduates from the University of Warsaw.
- The mean age: 21.42 years (SD = 3.22).
- Each participant tested individually.

Materials

80 grammatically simple propositions in Polish, like:

1. Some cars are red.
2. More than 7 cars are blue.
3. An even number of cars is yellow.
4. Less than half of the cars are black.

Materials continued

More than half of the cars are yellow.

An example of a stimulus used in the first study

Procedure

- 8 different quantifiers divided into four groups.

Procedure

- 8 different quantifiers divided into four groups.
  - “all” and “some” (acyclic 2-state FA);

Procedure

- 8 different quantifiers divided into four groups.
  - “all” and “some” (acyclic 2-state FA);
  - “odd” and “even”(2-state FA);

Procedure

- 8 different quantifiers divided into four groups.
  - “all” and “some” (acyclic 2-state FA);
  - “odd” and “even”(2-state FA);
  - “less than 8” and “more than 7” (FA);

Procedure

- 8 different quantifiers divided into four groups.
  - “all” and “some” (acyclic 2-state FA);
  - “odd” and “even”(2-state FA);
  - “less than 8” and “more than 7” (FA);
  - “less than half” and “more than half”(PDA).

Procedure

- 8 different quantifiers divided into four groups.
  - “all” and “some” (acyclic 2-state FA);
  - “odd” and “even”(2-state FA);
  - “less than 8” and “more than 7” (FA);
  - “less than half” and “more than half”(PDA).
- Each quantifier was presented in 10 trials.

Procedure

- 8 different quantifiers divided into four groups.
  - “all” and “some” (acyclic 2-state FA);
  - “odd” and “even”(2-state FA);
  - “less than 8” and “more than 7” (FA);
  - “less than half” and “more than half”(PDA).
- Each quantifier was presented in 10 trials.
- The sentence true in the picture in half of the trials.

Procedure

- 8 different quantifiers divided into four groups.
  - “all” and “some” (acyclic 2-state FA);
  - “odd” and “even”(2-state FA);
  - “less than 8” and “more than 7” (FA);
  - “less than half” and “more than half”(PDA).
- Each quantifier was presented in 10 trials.
- The sentence true in the picture in half of the trials.
- Quantity of target items near the criterion of validation.

Procedure

- 8 different quantifiers divided into four groups.
  - “all” and “some” (acyclic 2-state FA);
  - “odd” and “even”(2-state FA);
  - “less than 8” and “more than 7” (FA);
  - “less than half” and “more than half”(PDA).
- Each quantifier was presented in 10 trials.
- The sentence true in the picture in half of the trials.
- Quantity of target items near the criterion of validation.
- Practice session followed by the experimental session.

Procedure

- 8 different quantifiers divided into four groups.
  - “all” and “some” (acyclic 2-state FA);
  - “odd” and “even”(2-state FA);
  - “less than 8” and “more than 7” (FA);
  - “less than half” and “more than half”(PDA).
- Each quantifier was presented in 10 trials.
- The sentence true in the picture in half of the trials.
- Quantity of target items near the criterion of validation.
- Practice session followed by the experimental session.
- Each quantifier problem was given one 15.5 s event.

Procedure

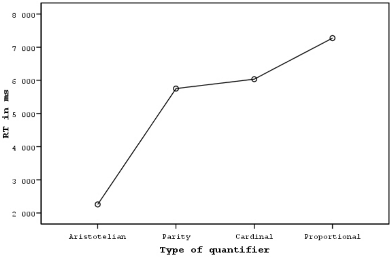
- 8 different quantifiers divided into four groups.
  - “all” and “some” (acyclic 2-state FA);
  - “odd” and “even”(2-state FA);
  - “less than 8” and “more than 7” (FA);
  - “less than half” and “more than half”(PDA).
- Each quantifier was presented in 10 trials.
- The sentence true in the picture in half of the trials.
- Quantity of target items near the criterion of validation.
- Practice session followed by the experimental session.
- Each quantifier problem was given one 15.5 s event.
- Subjects were asked to decide the truth-value.

Analysis of accuracy

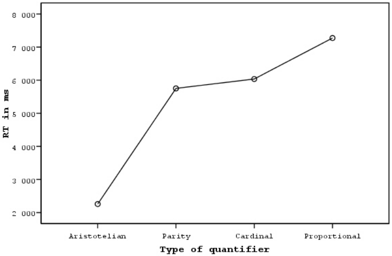
Quantifier group	Examples	Percent
Aristotelian FO	all, some	99
Parity	odd, even	91
Cardinal FO	less than 8, more than 7	92
Proportional	less than half, more than half	85

The percentage of correct answers

Analysis of RT

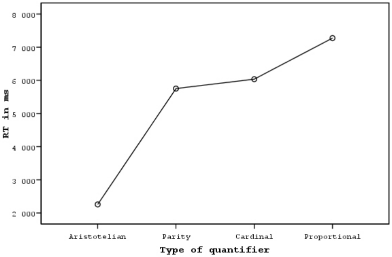


Analysis of RT



RT determined by quantifier type:

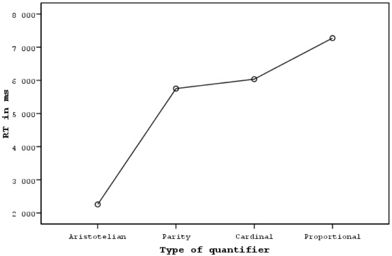
Analysis of RT



RT determined by quantifier type:

- All differences significant;

Analysis of RT



RT determined by quantifier type:

- All differences significant;
  - Aristotelian,
  - parity,
  - cardinal,
  - proportional.

Szymaniki & Zajenkowski, Comprehension of simple quantifiers. Empirical evaluation of a computational model, Cognitive Science, 2010

Outline

McMillan et al. fMRI studies

Differences in brain activity.

McMillan et al. fMRI studies

Differences in brain activity.

- All quantifiers are associated with numerosity: recruit right inferior parietal cortex;
- Only higher-order activate working-memory capacity: recruit right dorsolateral prefrontal cortex;

McMillan et al., Neural basis for generalized quantifiers comprehension, 2005

Szymanik, A Note on some neuroimaging study of natural language quantifiers comprehension, Neuropsychologia, 2007

Baddeley's model

WM unified system responsible for the performance in complex tasks.

Baddeley's model

WM unified system responsible for the performance in complex tasks.

- The model consists of:
  - temporary storage units:
  - a controlling system (central executive).

Baddeley, Working memory and language: an overview, 2003

Span test

Span test

- To asses the working memory construct.

Span test

- To asses the working memory construct.
- Subjects read sentences.

Span test

- To asses the working memory construct.
- Subjects read sentences.
- They are asked to:
  - remember the final words.
  - comprehend the story.

Span test

- To asses the working memory construct.
- Subjects read sentences.
- They are asked to:
  - remember the final words.
  - comprehend the story.
- What is:
  - the number of correctly memorized words?
  - the degree of understanding?

Span test

- To asses the working memory construct.
- Subjects read sentences.
- They are asked to:
  - remember the final words.
  - comprehend the story.
- What is:
  - the number of correctly memorized words?
  - the degree of understanding?
- Engagement of processing and storage functions.

Daneman and Carpenter, Individual differences in working memory, 1980

‘Computational’ theory of WM

Observation

A trade-off between processing and storage functions.

‘Computational’ theory of WM

Observation

A trade-off between processing and storage functions.

Hypothesis

One cognitive resource – competition for a limited capacity.

Daneman and Merikle, Working memory and language comprehension, 1996

Experimental setup

Question

How additional memory load influences quantifier verification?

Experimental setup

Question

How additional memory load influences quantifier verification?

Combined task:

- memorize sequences of digits;
- verify quantifier sentences;
- recall digits.

Predictions

Difficulty (RT and accuracy) should decrease as follows:

- proportional quantifiers,
- numerical quantifiers of high rank,
- parity quantifiers,
- numerical quantifiers of low rank.

Predictions

Difficulty (RT and accuracy) should decrease as follows:

- proportional quantifiers,
- numerical quantifiers of high rank,
- parity quantifiers,
- numerical quantifiers of low rank.

Additionally:

- processing of the PQs should influence storage functions;
- the effect should be stronger in more demanding situation.

Sentence verification

64 grammatically simple propositions in Polish, like:

1. More than 7 cars are blue.
2. An even number of cars is yellow.
3. Less than half of the cars are black.

Sentence verification

64 grammatically simple propositions in Polish, like:

1. More than 7 cars are blue.
2. An even number of cars is yellow.
3. Less than half of the cars are black.

- 8 different quantifiers divided into four groups.

Sentence verification

64 grammatically simple propositions in Polish, like:

1. More than 7 cars are blue.
2. An even number of cars is yellow.
3. Less than half of the cars are black.

- 8 different quantifiers divided into four groups.
    1. numerical quantifiers of relatively low rank, NQ4/5;

Sentence verification

64 grammatically simple propositions in Polish, like:

1. More than 7 cars are blue.
2. An even number of cars is yellow.
3. Less than half of the cars are black.

- 8 different quantifiers divided into four groups.
    1. numerical quantifiers of relatively low rank, NQ4/5;
    2. numerical quantifiers of relatively high rank, NQ7/8;

Sentence verification

64 grammatically simple propositions in Polish, like:

1. More than 7 cars are blue.
2. An even number of cars is yellow.
3. Less than half of the cars are black.

- 8 different quantifiers divided into four groups.
  1. numerical quantifiers of relatively low rank, NQ4/5;
  2. numerical quantifiers of relatively high rank, NQ7/8;
  3. parity quantifiers, DQ;

Sentence verification

64 grammatically simple propositions in Polish, like:

1. More than 7 cars are blue.
2. An even number of cars is yellow.
3. Less than half of the cars are black.

- 8 different quantifiers divided into four groups.
    1. numerical quantifiers of relatively low rank, NQ4/5;
    2. numerical quantifiers of relatively high rank, NQ7/8;
    3. parity quantifiers, DQ;
    4. proportional quantifiers, PQ.

Memory Task

- At the beginning of each trial a sequence of digits.

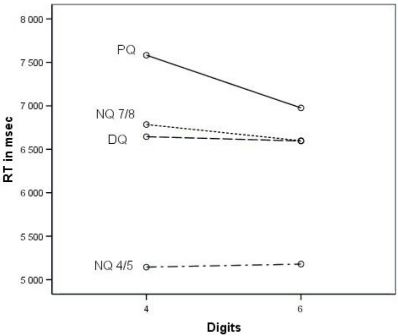
Memory Task

- At the beginning of each trial a sequence of digits.
- 2 experimental conditions:
  - 4 digits
  - 6 digits

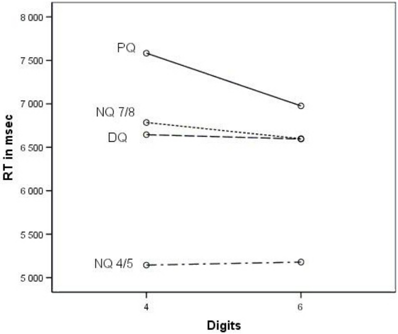
Memory Task

- At the beginning of each trial a sequence of digits.
- 2 experimental conditions:
  - 4 digits
  - 6 digits
- After verification task: recall the string.

RT in verification task

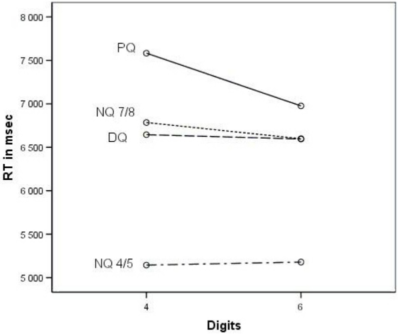


RT in verification task



RT determined by quantifier type in 4-digit:

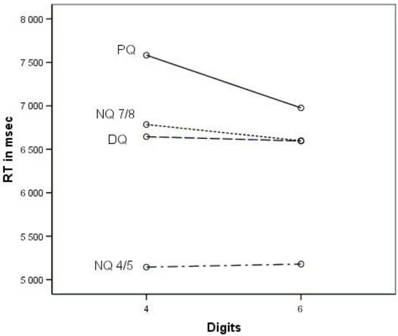
RT in verification task



RT determined by quantifier type in 4-digit:

- PQ solved longer than others;

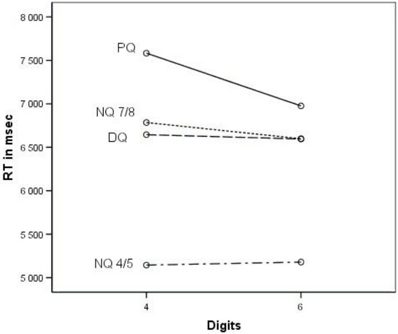
RT in verification task



RT determined by quantifier type in 4-digit:

- PQ solved longer than others;
- NQ 4/5 processed shorter than the rest;

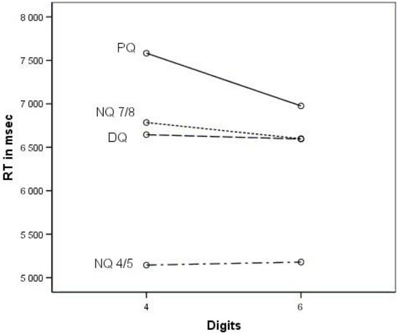
RT in verification task



RT determined by quantifier type in 4-digit:

- PQ solved longer than others;
- NQ 4/5 processed shorter than the rest;
- No difference between DQ and NQ 7/8.

RT in verification task

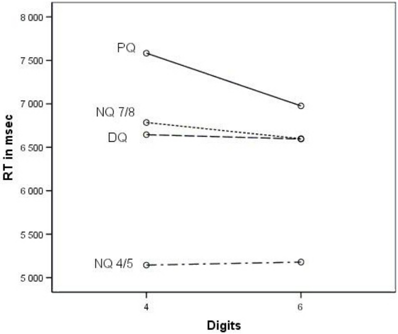


RT determined by quantifier type in 4-digit:

- PQ solved longer than others;
- NQ 4/5 processed shorter than the rest;
- No difference between DQ and NQ 7/8.

6-digit condition:

RT in verification task



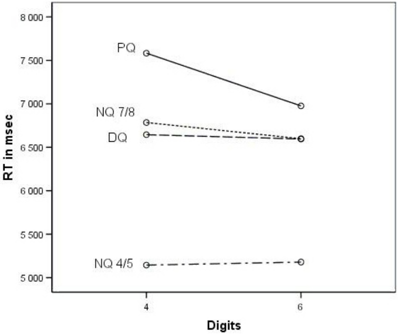
RT determined by quantifier type in 4-digit:

- PQ solved longer than others;
- NQ 4/5 processed shorter than the rest;
- No difference between DQ and NQ 7/8.

6-digit condition:

- NQ 4/5 had the shortest average RT.

RT in verification task



RT determined by quantifier type in 4-digit:

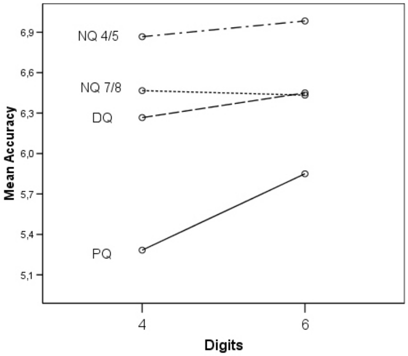
- PQ solved longer than others;
- NQ 4/5 processed shorter than the rest;
- No difference between DQ and NQ 7/8.

6-digit condition:

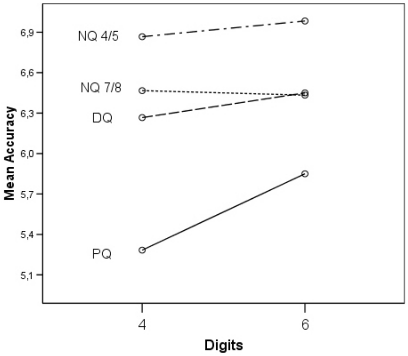
- NQ 4/5 had the shortest average RT.

Only PQ differed between memory load conditions.

Accuracy in verification task

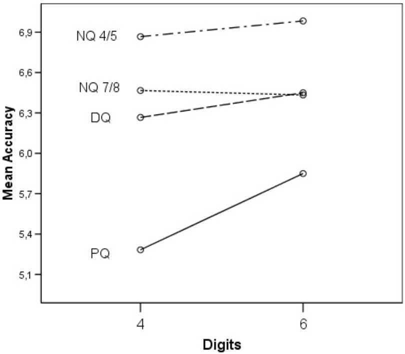


Accuracy in verification task



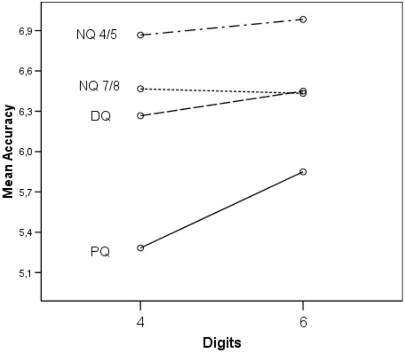
- All quantifiers differed significantly,
- besides DQ and NQ 7/8.

Accuracy in verification task



- All quantifiers differed significantly,
- besides DQ and NQ 7/8.
- Large effect for PQ!

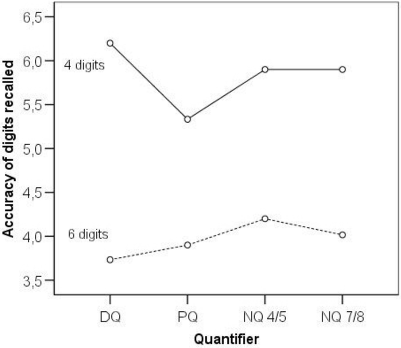
Accuracy in verification task



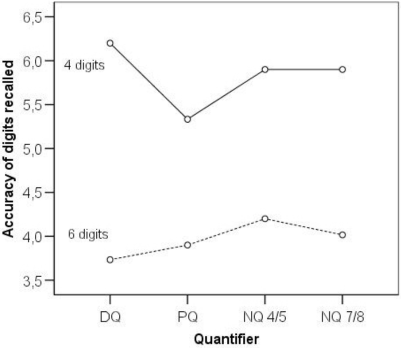
- All quantifiers differed significantly,
- besides DQ and NQ 7/8.
- Large effect for PQ!

In 4-digit condition all quantifiers were performed worse.

Memory task: recall accuracy

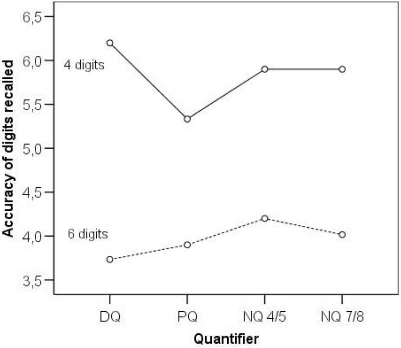


Memory task: recall accuracy



- In 4-digit with PQ: the worst;

Memory task: recall accuracy



- In 4-digit with PQ: the worst;
- In 6-digit: no differences.

Summary

- In 4-digit automata were good predictors of difficulty.

Summary

- In 4-digit automata were good predictors of difficulty.
- Discrepancy under two memory load conditions:
  - The real differences occurred only in 4-digit condition.
  - Holding six elements in memory was probably too difficult.
  - Trade-off between processing and storage.

Summary

- In 4-digit automata were good predictors of difficulty.
- Discrepancy under two memory load conditions:
  - The real differences occurred only in 4-digit condition.
  - Holding six elements in memory was probably too difficult.
  - Trade-off between processing and storage.
- Number of states is a good predictor of cognitive load.

Szymanik & Zajenkowski, Quantifiers and working memory, LNCS, 2010

Outline

Problem: Quantifier Verification

Computational Model

Reaction Time

Working Memory

Monotonicity

Outlook

A key property in logic and language

- Definability theory;
- Negative polarity items;
- Learnability theory;
- Reasoning;

Geurts, Reasoning with quantifiers, Cognition, 2003

Johnson-Laird, How we reason, OUP, 2008

Monotone quantifiers

Definition
Q is upward monotone if $X \subseteq Y$, then Q(X) entails Q(Y).

1. Every boy runs fast.
2. Every boy runs.

Definition
Q is downward monotone if $Y \subseteq X$, then Q(X) entails Q(Y).

1. No boy runs.
2. No boy runs fast.

Experiment

- 2 studies:
  - numerical quantifiers (“more than 7”, “less than 8”);
  - proportional quantifiers (“more than half”, “less than half”).
- upward monotone vs. downward monotone.

Average complexity and predictions

Assuming that people by default rather verify than falsify!

Average complexity and predictions

Assuming that people by default rather verify than falsify!

Cardinal quantifiers:

- "more than 7"
  - needs to check only $\left\lceil \frac{(n+7)}{2} \right\rceil$ on average.
- "less than 8"
  - always all n elements.

Average complexity and predictions

Assuming that people by default rather verify than falsify!

Cardinal quantifiers:

- "more than 7"
  - needs to check only $\left\lceil \frac{(n+7)}{2} \right\rceil$ on average.
- "less than 8"
  - always all n elements.
- RT will increase for the downward monotone quantifier!

Average complexity and predictions

Assuming that people by default rather verify than falsify!

Cardinal quantifiers:

- “more than 7”
  - needs to check only $\left\lceil \frac{(n+7)}{2} \right\rceil$ on average.
- “less than 8”
  - always all n elements.
- RT will increase for the downward monotone quantifier!

Proportional quantifiers:

- For both one has to go through all elements.
- No difference!

Results

Means (M) and standard deviations (SD) of RT.

Quantifier	M	SD
More than 7	5798.12	1130.15
Less than 8	6272.98	1117.43
More than half	7415.00	1735.60
Less than half	7131.92	1388.50

Discussion

1. Predictions were confirmed.
2. Effect sizes account around 45% of variance,
3. Before it was 90%.
4. Quantifier type explains more than monotonicity.

Outline

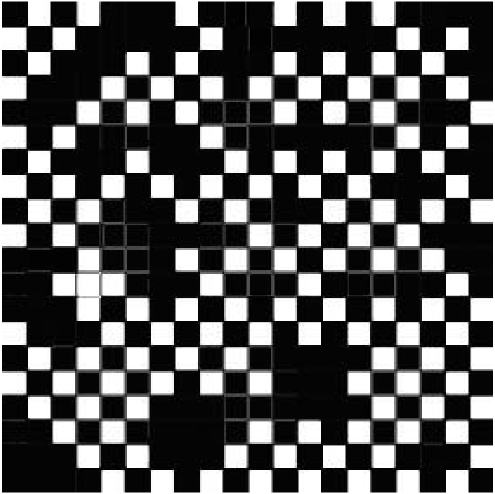
Bigger picture

- Enrich the model:
  1. Approximate Number System;
  2. Visual clues;

Dehaene, The number sense, OUP, 1999

Pietroski et al., The meaning of ‘most’, Mind & Language, 2009

Illustration



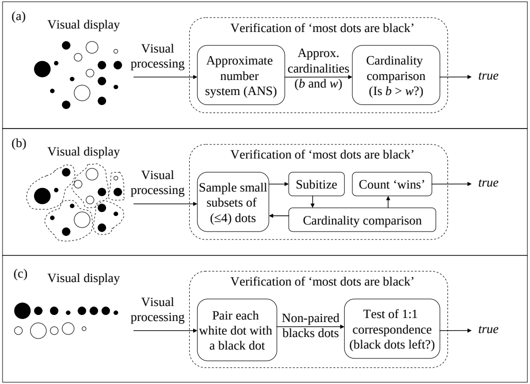
Neurocognitive computational modeling

- Mechanism selection;
- Translate to neurocognitive setting;
- fMRI experiments.

Hackl, On the grammar and processing of proportional quantifiers, Natural Language Semantics, 2009

Dehaene & Cohen, Cultural recycling of cortical maps, Neuron, 2007

Modeling example



Take home message

Take home message

All models are wrong but some are useful.

TH$\forall$NK YOU!