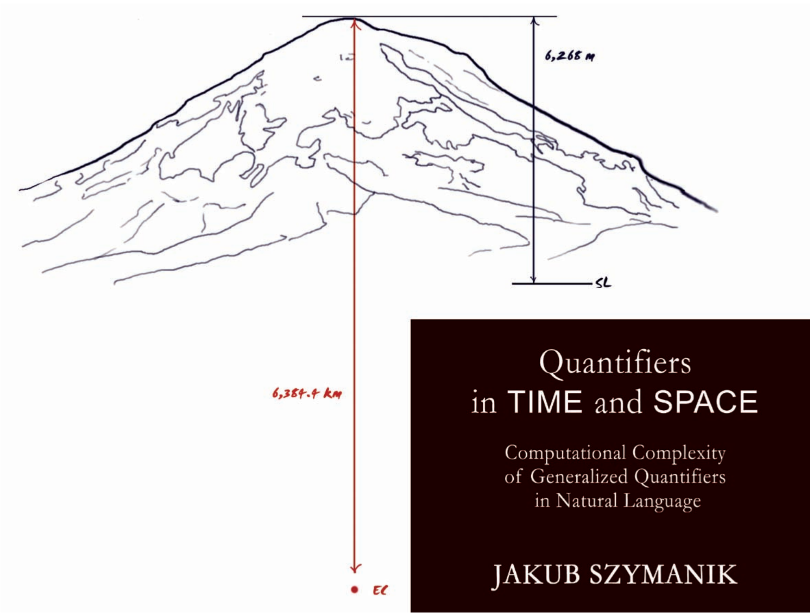

6,268 m
SL
6,384.4 km
EC
Quantifiers
in TIME and SPACE
Computational Complexity
of Generalized Quantifiers
in Natural Language
JAKUB SZYMANIK

Data complexity of the quantifier fragments in natural language

Jakub Szymanik

GLoRiClass Farewell

Introduction

Simple quantifiers

Complex constructions

Complex quantifiers

- Some happen to be intractable
- but most of them are tractable

Collective quantification

Outline

A very natural question

Question

How complex are different fragments of NL?

A very natural question

Question

How complex are different fragments of NL?

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

A very natural question

Question

How complex are different fragments of NL?

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

Pratt-Hartman, Moss: SAT-problems for various fragments.

Descriptive complexity of quantifiers

- How much logic is needed to formalize NL?
- What is the hierarchy of various fragments?
  - "Some", "At least 5", and "More than 7" are FO.
  - But "An even number of" and "Most" are not.
- It doesn't say much about processing.

Computational complexity of quantifiers

- Amount of resources needed for processing.
- W.r.t. to model size.
- Restriction to finite models.
- How difficult is it to verify quantifiers?

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.

NL determiners...

- ...mostly correspond to GQs of type (1,1).

Definition

A monadic generalized quantifier of type (1,1) is a class Q of structures of the form $M = (U, A_1, A_2)$, where $A_1, A_2 \subseteq U$. Additionally, Q is closed under isomorphism.

NL determiners...

- ...mostly correspond to GQs of type (1,1).

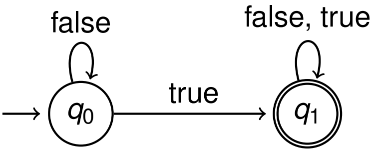
Definition

A monadic generalized quantifier of type (1,1) is a class Q of structures of the form $M = (U, A_1, A_2)$, where $A_1, A_2 \subseteq U$. Additionally, Q is closed under isomorphism.

They can be easily computed by van Benthem's automata.

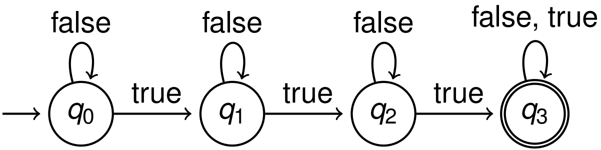
Example 1: Aristotelian quantifiers

- Some sentences in my thesis are true.



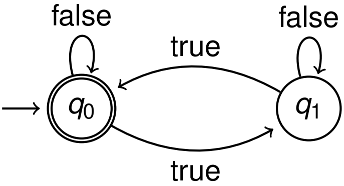
Example 2: cardinal quantifiers

- At least 3 sentences in my thesis are true.



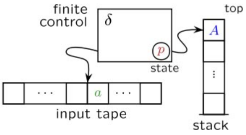
Example 3: parity quantifiers

- An even number of sentences in my thesis is true.



Example 4: proportional quantifiers

- "Most of the sentences in my thesis are true."
- Not computable by finite-automata.
- We need working memory.
- Simple push-down automata will do.



Does it say anything about processing?

Question

Does minimal automata predict differences in verification?

Observation

Empirical evidence converges to "YES".

- Szymanik, A Note on some Neuroimaging Study of NL Quantifiers Comprehension, Neuropsychologia 2007
- Szymanik and Zajenkowski, Comprehension of Simple Quantifiers, Cognitive Science 2009
- Szymanik and Zajenkowski, Working Memory and Quantifiers, LNCS 2010

Outline

What next?

- What about more complex quantifiers?
- What about the border tractable/intractable?

Outline

Polyadic GQs

Definition

Let $t = (n_1, \ldots, n_k)$ be a k-tuple of positive integers. *A generalized quantifier* of type t is a class Q of models of a vocabulary $\tau_t = \{R_1, \ldots, R_k\}$, such that R_i is n_i-ary for $1 \leq i \leq k$, and Q is closed under isomorphisms.

Polyadic GQs

Definition

Let $t = (n_1, \ldots, n_k)$ be a k-tuple of positive integers. *A generalized quantifier* of type t is a class Q of models of a vocabulary $\tau_t = \{R_1, \ldots, R_k\}$, such that R_i is n_i-ary for $1 \leq i \leq k$, and Q is closed under isomorphisms.

Definition

If in the above definition for all i: $n_i \leq 1$, then we say that a quantifier is *monadic*, otherwise we call it *polyadic*.

Outline

Introduction

Simple quantifiers

Complex constructions

Complex quantifiers

- Some happen to be intractable
- but most of them are tractable

Collective quantification

Possibly branching sentences

1. Most villagers and most townsmen hate each other.
2. One third of villagers and half of townsmen hate each other.
3. 5 villagers and 7 townsmen hate each other.

Branching reading

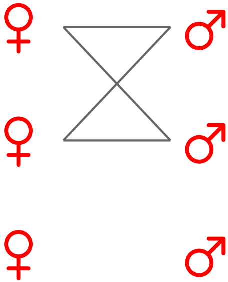
Most girls and most boys hate each other.

Branching: $\begin{array}{l} \text{most } x : G(x) \\ \text{most } y : B(y) \end{array} H(x, y).$

$$\exists A \exists A' [\text{most}(G, A) \wedge \text{most}(B, A') \wedge \forall x \in A \forall y \in A' H(x, y)].$$

Illustration

Most girls and most boys hate each other.



Branching readings are intractable

Theorem

Proportional branching sentences are NP-complete.

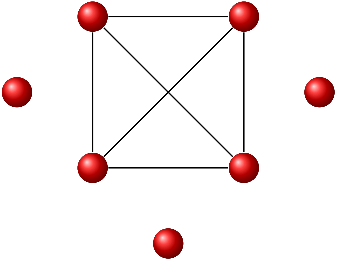
Sevenster, Branches of imperfect information, PhD 2006

Potentially strong reciprocal sentences

1. Andi, Jarmo and Jakub laughed at one another.
2. 15 men are hitting one another.
3. Most of the PMs refer to each other.

Strong reading

Most of the PMs refer to each other.



Strong reciprocal lift

Let Q be a monadic monotone increasing quantifier.

Definition

$\text{Ram}_S(\text{Q})(A, R) \iff$

$$\exists X \subseteq A[\text{Q}(X) \wedge \forall x, y \in X(x \neq y \Rightarrow R(x, y))].$$

Example

- Most of the PMs refer to each other indirectly.
- $\text{Ram}_S(\text{most})[\text{PMs}, \text{Refer}]$.

Strong reciprocity is intractable

Theorem

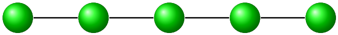
Model-checking for strong reciprocal sentences with proportional quantifiers is NP-complete.

Szymanik, The Computational Complexity of Quantified Reciprocals, LNCS 2007

Outline

Intermediate reading

Most Boston pitchers sat alongside each other.



Intermediate reciprocal lift

Definition

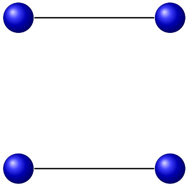
$$\mathrm{Ram_I(Q)}(A, R) \iff \exists X \subseteq A[\mathrm{Q}(X) \wedge \forall x, y \in X$$
$$(x \neq y \Rightarrow \exists \text{ sequence } z_1, \ldots, z_\ell \in X \text{ such that}$$
$$(z_1 = x \wedge R(z_1, z_2) \wedge \ldots \wedge R(z_{\ell-1}, z_\ell) \wedge z_\ell = y)].$$

Example

- Most Boston pitchers sat alongside each other.
- $\mathrm{Ram_I(most)[Pitcher,\ Sit]}$.

Weak reading

Some pirates were staring at each other in surprise.



Weak reciprocal lift

Definition

$\mathrm{Ram}_{\mathrm{W}}(\mathrm{Q})(A, R) \iff$

$$\exists X \subseteq A[\mathrm{Q}(X) \wedge \forall x \in X \exists y \in X(x \neq y \wedge R(x, y))].$$

Example

- Some pirates were staring at each other in surprise.
- $\mathrm{Ram}_{\mathrm{W}}(\text{some})[\text{Pirate}, \text{ Staring}]$.

Complexity dichotomy

Theorem

If Q *is PTIME, then also* $\mathrm{Ram_I(Q)}$ *and* $\mathrm{Ram_W(Q)}$ *are in PTIME.*

Complexity dichotomy

Theorem

If Q *is PTIME, then also* $\mathrm{Ram}_{\mathrm{I}}(Q)$ *and* $\mathrm{Ram}_{\mathrm{W}}(Q)$ *are in PTIME.*

Theorem

If Q *is PTIME and bounded, then also* $\mathrm{Ram}_{\mathrm{S}}(Q)$ *is in PTIME.*

Question

Is there a full duality: $\mathrm{Ram}_{\mathrm{S}}(Q)$ *is either PTIME or NP-hard?*

Szymanik, The Computational Complexity of Quantified Reciprocals, LNCS 2007

Iteration

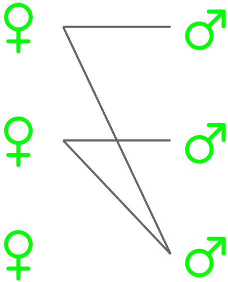
1. Most logicians criticized some papers.
2. It(most, some)[Logicians, Papers, Criticized].

Definition

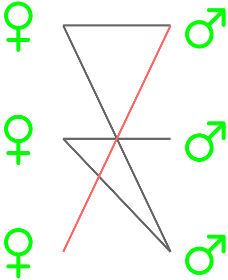
$\mathrm{It}(\mathrm{Q}, \mathrm{Q}')[A, B, R] \iff \mathrm{Q}[A, \{a \mid \mathrm{Q}'(B, R_{(a)})\}]$, where $R_{(a)} = \{b \mid R(a, b)\}$.

Iteration

Most girls and most boys hate each other.



Rather two-side-iteration than branching



Gierasimczuk and Szymanik, Branching Quantification vs. Two-way Quantification, Journal of Semantics 2009

Cumulation

1. Eighty professors taught sixty courses at ESSLLI'08.

Definition

$\mathrm{Cum}(\mathrm{Q}, \mathrm{Q}')[A, B, R] \iff$

$$\mathrm{It}(\mathrm{Q}, \mathrm{some})[A, B, R] \wedge \mathrm{It}(\mathrm{Q}', \mathrm{some})[B, A, R^{-1}]$$

Basic operations are tractable

Theorem

Let Q *and* Q′ *be generalized quantifiers computable in PTIME with respect to the size of a universe. Then the following quantifiers are also in PTIME:*

1. $\neg Q$;
2. $Q\neg$;
3. $Q \wedge Q'$;
4. $\mathrm{It}(Q, Q')$;
5. $\mathrm{Cum}(Q, Q')$;
6. $\mathrm{Res}(Q)$.

Szymanik, Almost All Complex Quantifiers are Simple, LNCS 2010

Outline

Introduction

Simple quantifiers

Complex constructions

Complex quantifiers

- Some happen to be intractable
- but most of them are tractable

Collective quantification

Two examples

1. Tikitu and Samson lifted the poker table together.
2. Most groups of students played Hold’em together.

Shifting strategy

1. Five people lifted the table.

1'. $\exists^{=5}x[\text{People}(x) \wedge \text{Lift}(x)]$.

1". $\exists X[\text{Card}(X) = 5 \wedge X \subseteq \text{People} \wedge \text{Lift}(X)]$.

2. Some students played poker together.

2'. $\exists X[X \subseteq \text{Students} \wedge \text{Play}(X)]$.

Collective "Most"

1. Most groups of students played Hold'em together.

1'. $\textsc{Most}\, X, Y[\text{Students}(X), \text{Play}(Y)]$.

Definition

$\textsc{Most} = \{P, S \subseteq \mathcal{P}(M) \text{ and } \text{card}(P \cap S) > \text{card}(P \setminus S)\}$.

Undefinability result

Theorem (with J. Kontinen)

The quantifier MOST *is not definable in second-order logic.*

Kontinen and Szymanik, A Remark on Collective Quantification, JoLLI 2008

Summary

1. Data complexity of model-checking for NL-quantifiers.
2. Systematically: from simple to complex constructions.
3. Its interplay with linguistics and cognitive science.
4. Computational complexity in cognitive science.
5. Algorithmic theory of meaning.

THANK YOU!