Complexity of natural language quantifiers

Computational dichotomy between reciprocals

Jakub Szymanik

Institute for Logic, Language and Computation
Universiteit van Amsterdam

GLLC $14\frac{1}{2}$
September 28, 2007

ABSTRACT

- Study reciprocals, like “each other”.
- Define them as lifts over monadic GQs.
- Show computational dichotomy:
  - Strong r.l. over proportional quantifiers are NP-complete.
  - PTIME quantifiers are closed on intermediate and weak r.l.
- R.l. are frequent NP-complete constructions.
- Ask some general mathematical questions about r.l.

OUTLINE

Outline

1. Motivations
2. Previous works
3. Reciprocity in language
4. Reciprocals as lifts over GQs
5. Complexity of reciprocal lifts
   - Strong reciprocity
   - Intermediate and weak reciprocity
6. Further questions

- Link semantics and computational complexity.
- Evaluate complexity of semantic constructions in order to:
  - better understand our linguistic competence.
  - investigate into robustness of linguistic distinctions.
- Classify semantic constructions by their complexity.
- It will be valuable for cognitive science.
- Clarify concept of “meaning”.

Outline

GQs — a short reminder

Definition

A generalized quantifier Q of type $(n_1, \ldots, n_k)$ is a class of structures of the form $M = (U, R_1, \ldots, R_k)$, where R_i is a subset of U^{n_i}. Additionally, Q is closed under isomorphism.

Example

MOST $= \{(U, A^M, B^M) : card(A^M \cap B^M) > card(A^M - B^M)\}$.

Quantifiers and complexity

Definition

Let Q be of type $(n_1, \ldots, n_k)$. By complexity of Q we mean computational complexity of the corresponding class K_Q.

Our computational problem is to decide whether $M \in K_Q$. Equivalently, does $M \models Q[R_1, \ldots R_k]$?

Definition

We say that Q is NP-hard if K_Q is NP-hard.
Q is mighty if K_Q is NP and K_Q is NP-hard.

It was Blass and Gurevich 1986 who first studied those notions.

Mighty quantifiers — first example

Let us consider models of the form $M = (U, E^M)$, where E^M is an equivalence relation.

Definition

$M \models \mathrm{R_e} xy\ \varphi(x, y)$ means that there is a set $A \subseteq U$ such that $\forall a \in U\ \exists b \in A\ E(a, b)$ and for each $a, b \in A$ $M \models \varphi(a, b)$.

Theorem (Mostowski, Wojtyniak 2004)

$\mathrm{R_e}$ *is mighty.*

Mighty quantifiers — second example

Let us consider models of the form $M = (U, V^M, T^M)$, where V^M, T^M are subsets of U.

Definition

$M \models \mathrm{BMost}\, xy\ \varphi(x,y)$ means that there are sets $A \subseteq U$ and $B \subseteq U$ such that:

$$\mathrm{MOST}x\ (V(x), A(x)) \wedge \mathrm{MOST}y\ (T(y), B(y)) \wedge$$
$$\wedge \forall x \forall y (A(x) \wedge B(y) \Rightarrow \varphi(x,y)).$$

Theorem (Sevenster 2006)

BMost *is mighty.*

Motivation for previous results

Under branching interpretation the following sentences are NP-complete:

(1.) *Some relative of each villager and some relative of each townsman hate each other.*

(2.) *Most villagers and most townsmen hate each other.*

However, all these sentences are ambiguous and can be hardly found in the corpus of language.

Outline

Reciprocal expressions are common in English

(1.) *Andi, Jarmo and Jakub laughed at one another.*

(2.) *15 men are hitting one another.*

(3.) *Even number of the PMs refer to each other.*

(4.) *Most Boston pitchers sat alongside each other.*

(5.) *Some pirates were staring at each other in surprise.*

In BNC there are 10351 occurrences of “each other”.
Many sentences contain quantifiers in antecedents.

Various interpretations

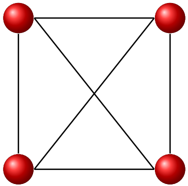
Dalrymple et al. 1998 classifies possible readings.
They explain variations in the meaning by:

Strong Meaning Hypothesis

Reading associated with the reciprocal in a given sentence is the strongest available reading which is consistent with relevant information supplied by the context.

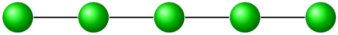
STRONG READING

(3.) *Even number of the PMs refer to each other.*



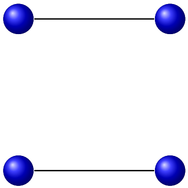
INTERMEDIATE READING

(4.) *Most Boston pitchers sat alongside each other.*



WEAK READING

(5.) *Some pirates were staring at each other in surprise.*



Outline

Strong reciprocal lift

Let Q be a monadic monotone increasing quantifier.

Definition

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

Example

(3.) *Even number of the PMs refer to each other indirectly.*

(3'.) Ram_S(EVEN)MP Refer.

Intermediate reciprocal lift

Definition

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

Example

(4.) *Most Boston pitchers sat alongside each other.*

(4'.) $\mathrm{Ram}_{\mathrm{I}}(\mathrm{MOST})$Pitcher Sit.

Weak reciprocal lift

Definition

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

Example

(5.) *Some pirates were staring at each other in surprise.*

(5'.) $\mathrm{Ram}_{\mathrm{W}}$(SOME)Pirate Staring.

Outline

Strong R.L. over counting quantifiers

Definition

$M \models \exists^{\geq k} y \varphi(y)[v] \iff card(\varphi^{(M,y,v)}) \geq v(k).$

Proposition

Quantifier $\mathrm{Ram}_S(\exists^{\geq k})$ *is mighty.*

Proof.

$M \models \mathrm{Ram}_S(\exists^{\geq k})AR$ if there is clique C s.t. $card(C) \geq v(k)$. □

Strong R.L. over proportional quantifiers

(6.) Most PMs refer to each other.

(7.) At least one third of the PMs refer to each other.

(8.) At least $q \times 100\%$ of the PMs refer to each other.

Definition

$M \models \mathrm{R_q} xy\ \varphi(x, y)$ iff there is $A \subseteq U$ s. t. for all $a, b \in A$ $M \models \varphi(a, b)$ and A is q-big, i.e. $\frac{card(A)}{card(U)} \geq q$.

Proposition

Let $q \in]0, 1[\cap \mathbb{Q}$, *then the quantifier* $\mathrm{R_q}$ *is mighty.*

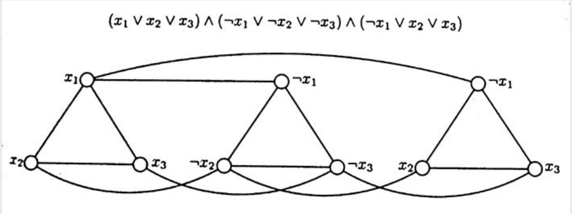
Proof of the Proposition 2

Corollary

q-big clique is NP-complete for $q \geq \frac{1}{k}$*, where* $k > 2$.

Proof.

It follows from the NP-completeness proof of INDEPENDENT SET. Consider graphs divided on complete disjoint k-agons.



Continuation of the proof

Lemma

For every $q \in]0,1[\cap \mathbb{Q}$ problem q-big clique is NP-complete.

Proof.

Let $G = (V, E)$ be s.t. $card(V) = ka$. In G exists $\frac{1}{k}$-big clique iff in G' exists $\frac{m}{k}$-big clique for $m < k$, where $G' = (V', E')$ is constructed as follows:

- $V' = V \cup U$, where U s.t. $card(U) = n = \lceil \frac{(m-1)ka}{k-m} \rceil$ and $U \cap V = \emptyset$;
- $E' = E \cup U \times (U \cup V)$.

It suffices to observe that $\frac{n+a}{n+ka} \geq \frac{m}{k} > \frac{n+(a-1)}{n+ka}$. □

Natural general question

Question

How can we describe the class of quantifiers for which Ram_S *results in NP-hard problem?*

Intermediate lift does not increase complexity

Proposition

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

Proof.

To check whether $M \in \mathrm{Ram}_{\mathrm{I}}(\mathrm{Q})$ use breadth-first search algorithm to compute all connected components of M. Their number is bounded by $card(U)$.Then check whether $\mathrm{Q}(C)$ holds for some connected component C. It can be done in polynomial time as Q is in PTIME. □

Weak lift is also weak

$\mathrm{Ram}_W(Q)$ says there is a subgraph bounded by Q without isolated vertices.

Proposition

If Q be PTIME quantifier, then also $\mathrm{Ram}_W(Q)$ *is in PTIME.*

Proof.

Check if sum of all connected components satisfies Q. □

OUTLINE

Complexity dichotomy between strong vs. intermediate and weak interpretations of reciprocal expressions.

- Does it influence our use of language?
- How to prove that $\mathrm{R_q}$ is NP-complete for irrational q?
- For which Q construction $\mathrm{Ram_S(Q)}$ is actually hard?
- Are those lifts interdefinable in $FO(L_{\infty\omega})$?
- How does it depend on quantifier in antecedent?
- ...

References

A. Blass and Y. Gurevich
Henkin quantifiers and complete problems,
APAL, 32(1986).

M. Dalrymple, M. Kanazawa, Y. Kim, S. Mchombo, S. Peters
Reciprocal expressions and the concept of reciprocity,
L&P, 21(1998).

M. Mostowski and D. Wojtyniak
Computational complexity of the semantics of some natural language constructions,
APAL, 127(2004).

M. Sevenster
Branches of imperfect information: logic, games, and computation,
PhD Thesis, ILLC 2006.