Knowledge Representation Using Predicate-Argument Structures with Nonclassical Quantifiers

Zygmunt Vetulani

Adam Mickiewicz University in Poznań
Department of Computer Linguistics and Artificial Intelligence
ul. Umultowska 87,
61614 Poznań
Poland
vetulani@amu.edu.pl

Abstract

The paper aims at contributing to the problem of translating natural (ethnic) language into the framework of formal logic in a structure-preserving way. There are several problems with encoding knowledge in logic-derived formalisms. Among the most difficult are those connected with natural language quantifiers. Even for relatively simple natural language sentences there are readings hard to encode in the classical formal logic. The solution we propose is to expand the language of the classical first-order logic with a number of new quantifier symbols (corresponding to the natural language quantifiers possibly co-occurring in a sentence) with a nonclassical semantics/interpretation similar in a way to the Mostowski's generalized quantifier, i.e. as a family of subsets of the Cartesian product of variable ranges. We claim that the logic-based formalism involving such nonclassical quantifiers may be considered a good knowledge representation tool closer to the natural language then the classical logic.

Introduction

One of the main functions of the human language is information exchange. Execution of this function is very complex and based on several pre-requirements. One of them is sharing some knowledge about the world by information exchange parties and, what is even more important - sharing conceptualization (at least partially). Sometimes, despite some incompatibility between the knowledge/conceptualization of communicating agents, communication is still possible, but requires additional effort to reduce incompatibility. In such cases one observes usually additional speech acts performed in order to clear the problem (e.g. an explanatory subdialogue). Colloquially speaking, in order to communicate the agents must use *the same language* and the same *logic.* Common language used by agents reflects the *common understanding* of the world. In the AI terminology, the human language is called "natural". This term emphasize the fact that human language is a product of evolution. We may suppose that this evolution goes in the direction of making communication easier. Our working hypothesis is that linguistic inspiration when looking for "good" knowledge representation (KR) is sound.

Our claim is that "NL-based KR" means approximately the same as "logic-based KR" because logic tries to capture the most general phenomena of "knowledge management" in the humans' brains and the knowledge structures are reflected in the natural language. This was the idea behind the fundamental observation by Richard Montague e.g. in his classical publication "English as formal language" that natural language (English) may be thought as a formal language. By extension the language structures are close to the logical structures. This idea is enhanced by the fact that indeed, simple sentences about named individuals may be considered as having the logical structure similar to the atomic formulas in the predicate calculus (predicate logic). Following this reasoning, we conclude that a good, well motivated candidate to the role of knowledge representation system is logic.

This idea is by no means new in AI, as it was explored already at the time of early logic programming (cf. Veronica Dahl: *Translating Spanish into Logic through Logic* (Dahl, 1981) or early system ORBIS by Colmerauer and Kittredge, early 80s). The emergence of logic programming (LP) with PROLOG as very high order programming language made it possible to use the LP as a uniform paradigm to cover the whole of language processing, e.g. in man-machine communication. Within this approach, knowledge is typically represented using a fragment of predicate logic limited to facts (atomic formulas) and rules (Horn clauses), whereas the sentence processing (parsing) is performed by the inference engine of the logic programming system. The elegance of this approach is however negatively counterbalanced by low efficiency of the standard processing techniques (combinatorial explosion)[1] and the limited expressiveness of the so limited logic. It is to notice here that even the full classical logic (e.g. the classical predicate calculus or any equivalent) does not provide a satisfactory solution in what concerns expressiveness. Indeed, every student of mathematics knows how awkward are sometimes mathematical statements when translated into logic. This is so because the classical formal logic is intentionally poor in formal devices (as e.g. quantifiers) in order to make relatively simple its inference system.

Problems with quantifiers

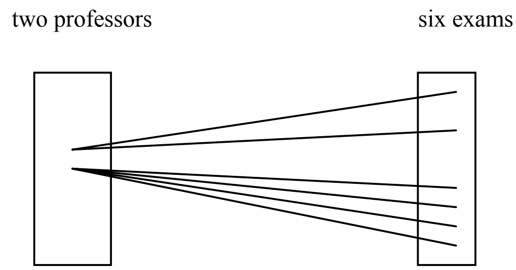
One of the major problems, well identified since long time ago, is how to formally represent knowledge expressed in natural language with NL quantifiers. It is to notice here that already Andrzej Mostowski in (Mostowski, 1957) extended the Tarski's semantics in a way to make possible interpretation of various nonstandard quantifiers (covered by a generic term *generalized quantifier*). This contribution was a side effect of Mostowski's works in metamathematics and has remained practically unexplored until early 80-ties. At that time Barwise and Cooper published their very influential paper on *Generalized quantifiers and natural language* (Barwise & Cooper, 1981) where NL-quantifiers and their mathematical properties were systematically studied on the ground of Mostowski's earlier works. (By NL-quantifier we mean, after Barwise and Cooper, expressions used in natural language to represent quantitative aspects of relationships.) These may be traditional quantifiers (every, some...), numerals (one, two,...), articles and other (many,...). One of the most important results due to Barwise and Cooper is the proof that some of NL-quantifiers may not be defined on the ground of classical logic. We may conclude from this observation that if we wish the logic-based knowledge representation system to be possibly close to natural language, then we should consider such undefinable quantifiers and define their formal semantics. The existence of undefinable quantifiers may be considered surprising because we have tendency to believe that logic is a common and universal background to all knowledge. In fact, the traditional formal logic (e.g. in form of predicate calculus) was created in order to formalize mathematics in the form of a formal axiomatic system (as it was described e.g. in *Principia Mathematica*). For this purpose considering the two classical quantifiers was enough.

[1] Our works on question answering (system POLINT implemented for the Polish language) show how to cope with this problem (successfully) (Vetulani, 2002).

Unfortunately, extending logic by introduction of "new" quantifiers and giving them semantics following the ideas of Mostowski, Barwise and Cooper does not solve all problems connected with quantifiers. Below, we will focus on the case of ambiguous sentences in which co-existence of different NL-quantifiers allows several readings, where some these readings can not be expressed in logic, even extended by the respective quantifiers. As example we will consider the sentence already investigated by several authors (Kempson, Cormarck, Bellert):

(S) *Two professors marked six exams.*

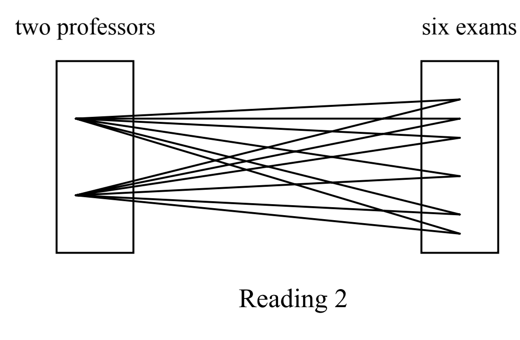
Considering "two" and "six" as quantifiers causes problems when trying to extend the traditional interpretation to cover all possible readings practiced by humans. The reading that makes trouble may be graphically represented as follows:



Reading 1

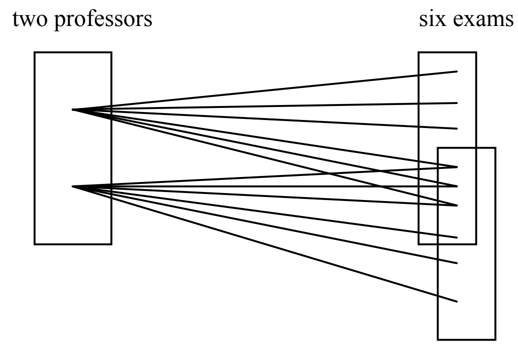
In Reading 1, both NL-quantifiers ("two" and "six") are used in the "+absolute" way, i.e. there exist exactly two professors and exactly six exams involved in the situation. (Cf. below in the section “Context/reading parameter”.) This reading leaves open the problem what links between professors and exams are referred to by the predicative word "marked" (which professor is "connected" with which exam). We say that within this reading both NL-quantifiers considered are "-distributive" one with respect to the other.

The remaining readings may be grouped into three classes as illustrated by the corresponding figures.



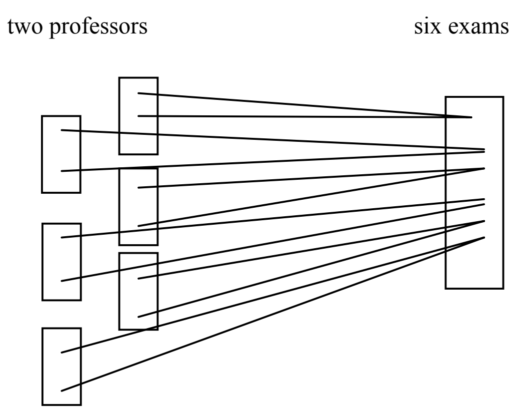
Reading 2

In Reading 2 both NL-quantifiers ("two" and "six") are used in the "+absolute" way (as in Reading 1), but are respectively "+distributive". The relation distributes in the "each-to-each" way. The remaining two readings corresponds to situations where only one of quantifiers is used in the "+absolute" way.



Reading 3

In Reading 3 each element form the domain of the "+absolute" quantifier is connected with exactly six elements of the other domain (we say that "two" is +distributive with respect to "six") but this does not go in the other direction (in this reading "six" is "-distributive" with respect to "two"). (Formaly "six" is in the scope of "two".) The following scheme presents the symmetrical situation of Reading 4.



Reading 4

Problems with formal representation of the first considered reading (Reading 1) are due to the fact that application of the extension of the Tarski's truth definition proposed by Mostowski requires one quantifiers to be *in the scope of the other*.

The Mostowski truth definition for quantifiers is as follows:

$$(*)\quad \mathsf{M} \models Qx\, P(x) \text{ iff } \{x \in U: \mathsf{M} \models Qx\, P(x)\} \in \|Q\|$$

where $\|Q\|$ is a family of (some) subsets of the universe U of the structure M.[2] Clearly, the Reading 1 can not be captured under the assumption of the above mentioned scope dependency.

The family of sets $\|Q\|$ is considered as the "interpretation" (or "meaning") of the NL-quantifier Q with respect to M. (The interpretation of the quantifier "two" will be the family of all two-element subsets of U.) It is clear that Tarskis's semantics extended in this way covers only the last three readings of our example.

The solution proposed

The solution we propose consists in considering the whole configuration of NL-quantifiers as a single (although complex) one (i.e. as a kind of single generalized quantifier).

Nonclassical quantifiers

We propose to represent simple NL sentences involving several quantifiers as composed of a predicate-argument structure and a configuration of quantifiers[3] corresponding to the argument positions. We propose the following notation:

$$(Q_1x_1, Q_2x_2, \ldots, Q_nx_n)\ \ P(x_1, x_2, \ldots, x_n)$$

where $(Q_1x_1, Q_2x_2, \ldots, Q_nx_n)$ will be called *nonclassical quantifier*.

In the case we investigated above the corresponding formula is:

(Two professors, Six exams) marked(professors, exams)

Contrary to the classical definition (*) provided above, the semantics of this complex quantifier will be defined for *particular reading* depending of the *context*. By *context* we mean *this factor which differentiates readings*. The nature of this factor is (usually) pragmatic, but sometimes the reading may also be indicated by formal features, as e.g. word order.

In the following, the context will be represented by the parameter C (we will say *contex/reading parameter*). As in Mostowski's truth condition in (*), the interpretation $\|(Q_1x_1, Q_2x_2, \ldots, Q_nx_n)\|_C$ will be some family of subsets of the universe and more precisely a family of subsets of the cartesian product $|U_1| \times |U_2| \times \ldots \times |U_n|$ of domains corresponding to arguments, i.e.

---

[2] The Tarski's truth relation $\models$ is relativised to some relational structure M. In order to simplify notation, further in this text we will omit the reference to M; $\|Q\|$ is also called *Mostowskian quantifier*.

[3] The concept of *configuration of quantifiers* was discussed already in (Vetulani, 1987).

$$\| (Q_1x_1,Q_2x_2,...,Q_nx_n) \|_C \subset P(|U_1| \times |U_2| \times ... \times |U_n|)$$

The definition is as follows:

$\models_C (Q_1x_1,...,Q_nx_n)\ P(x_1,...,x_n)$ iff

$\{(x_1,...,x_n): \models_C P(x_1,...,x_n)\} \in \| (Q_1x_1,...,Q_nx_n) \|_C,\ x_i \in |U_i|\}$

Now, what remains to do is to define the interpretations $\|Q\|_C$ for each configuration of quantifiers Q and each context (reading) C.

Let us remark that the formula $(Q_1x_1,Q_2x_2,...,Q_nx_n)$ $P(x_1,x_2,...,x_n)$, makes abstraction of the surface linear ordering of quantifiers in the sentence and also of the scope dependencies between quantifiers reflected in the traditional notation. This effect is intended as in many free word order languages (e.g. Slavonic languages, Latin) the surface succession of quantifiers do not necessarily correspond to the scope relations marked rather by inflection cases. Consistently, we prefer to consider this information as a part of context/reading parameter. This means that the predicate logic formulas $(\forall x(\exists y P(x,y))$ and $(\exists y(\forall x P(x,y))$ are both represented by one and the same formula $(\forall x,\exists y)\ P(x,y)$ and the scope relationship between $\forall$ and $\exists$ is to be encoded by the context/reading parameter.

The interpretation $\|(\forall x,\exists y)\|_{C1}$ for C1 corresponding to the classical reading of the formula $(\forall x(\exists y P(x,y))$ is as follows:

$$\|(\forall x,\exists y)\|_{C1} = \{R \subset |x| \times |y| :\ Dom(R) = |x|\}$$

For C2 corresponding to $(\exists y(\forall x P(x,y))$ the interpretation is different:

$$\|(\forall x,\exists y)\|_{C2} = \{R \subset |x| \times |y| : Dom(R)=|x| \wedge \cap\{\{y: R(x,y)\}: x \in X\} \neq \varnothing\}$$

For the sentence *Two professors marked six exams* considered in the context C corresponding to the Reading 2 in the case where professors and exams are unambiguously identified as respectively p_1, p_2 and $e_1,...,e_6$, we obtain the following interpretation:

$$\|(\text{Two professors, Six exams})\|_C = \{\{(p_i, e_j): i=1,2, j=1,...,6\}\}$$

Let C' be the context corresponding to the Reading 2 but without the assumption that the involved entities are identified. In this case the interpretation is more complex and consists of a family of sets, each composed of 12 ordered pairs (linking exams to professors).

$\|(\text{Two professors, Six exams})\|_{C'} =$
$\{\{(x_i, y_j): i=1,2, j=1,...,6\}: x_1 \neq x_2\}, y_m=y_n \rightarrow m=n,$
for $m,n=1,2,\ldots,6$, where x_i are professors, y_j are exams}

We leave the illustration of the remaining reading as an exercise for the Reader.

It follows from the above definitions and the examples that truth conditions *for a given reading* of a sentence with n quantifiers may be articulated as properties of some n-ary relation.

Context/reading parameter

From both the practical and theoretical points of view the question concerning the nature of the context/reading parameter is of fundamental importance. In this chapter we consider what kind of information should be encoded in the context/reading parameter and how this information may be represented.

One of possible answers to this problems may be derived from Bellert's monograph on NL quantifiers (Bellert, 1989). Among various issues addressed there, the one about possible readings of the NL-quantifiers co-occurring in the sentence is of special interest for us. The author formulates and discusses the idea of characterization of all possible readings in terms of the two features associated to NL-quantifiers in the sentence. These are +/- absoluteness and +/- distributiveness face to the co-occurring NL-quantifiers. Informally, "the value plus of absoluteness indicates that a given N″ may represent exactly one reference class in any particular context of use (e.g. as in the case of 'these girls')", similarly "the value minus of absoluteness indicates that the N″ in question may represent more then one reference class".[4] The value plus of distributiveness of a quantifier with respect to a co-occurring quantifier indicates that the relation expressed by the predicate is distributed among the all members of some reference class of this co-occurring quantifier (or *the* reference class if this co-occurring quantifier appears to be +absolute). Minus of distributiveness indicates that the relation is "is distributed among *some,* not necessarily all, members of the reference classes of the corresponding N″ (as in "Five students were sitting in these two rooms")".[5] In our example

(S) *Two professors marked six exams.*

The Reading 1 is characterized by the feature values attributed as follows:

"two professors":
- +absolute
- -distributive with respect to "six exams"

"six exams":

[4] In Chomsky's X-bar notation N″ stands for "full noun phrase".

[5] Although formally similar to our example "Two professors marked six exams", the sentence "Five students were sitting in these two rooms" has only one reading, the other being excluded by the selection of the lexical material.

+absolute
-distributive with respect to “two professors”,

whereas the Reading 3 by:

“two professors”:
+absolute
+distributive with respect to “six exams”

“six exams”:
-absolute
-distributive with respect to “two professors”

The main observation of the Bellert’s monograph we refer to is that not all of the possible combinations of feature values are linguistically admissible. The rules are proposed in application to pairwise co-occurring quantifiers in order to determine the linguistically admissible readings of quantifiers.[6] Thus the configurations of feature values as admitted by these rules may be considered as candidates for context/reading parameters to be used in the definition of semantics for the *nonclassical quantifier* as provided above.

Remark. The complexity of the problem we try to manage in this paper (i.e. problems of interpretation of sentences with many quantifiers) may be illustrated by the example communicated to us by the Anonymous Reviewer of this paper who objected that our approach does not capture the case where “Professor A graded all 6 of Problem 1’s and Professor B graded all 6 of the Problem 2’s”. A possible answer is that the sentence (S) with such an intended meaning is not about *exams*, but about *<exam, problem>* pairs, and therefore the Reading 3 perfectly applies in this case.

Another solution?

The representation problems discussed in Chapter 2 may also find solution on the ground of higher order logic, i.e. logic allowing higher order quantifiers. Let us then try to describe readings with the help of second order NL-quantifiers "The Two" and "The Six" applied respectively to professors and to exams.

On this ground the particular readings of the sentence

Two professors marked six exams

may be characterized by the corresponding second order formulas as follows:

- READING 1

For each X of The Two professors,
There exists Some Y
among The Six exams that:
marked(X,Y)
and
For each Y of The Six exams,
There exists Some X
among The Two professors that: marked(X,Y)

- READING 2

For each X of The Two professors
and
For each Y of The Six exams: marked(X,Y)

- READING 3

For each X of The Two professors,
There exists The Six exams
and
There exists Y among The Six exams
that: marked(X,Y)

- READING 4

For each Y of The Six exams,
There exists The Two professors
and
There exists X among The Two professors
that: marked(X,Y)

Final remarks

From the two solutions presented in this paper we claim that the first one better satisfies the postulate to maintain a possibly close formal relationship between the Natural Language structures and the corresponding formal Knowledge Representation structures. Although the other solution presented above seems simple and elegant, it has a serious drawback as it requires involvement of the very strong second order formalism which is computationally inconvenient. In any case further investigation is necessary to evaluate the computational aspects of the proposed solutions. Another class of problems that appear as a result of the introduction of the *nonclassical quantifier* as defined above concerns the mathematical properties of logic systems allowing such quantifiers (possible

[6] Three such rules were proposed:
— (absoluteness) If Q_i is –absolute then there is an +absolute co-occuring quantifier Q_k such that Q_k is + distributive with respect to Q_i and Q_i is -distributive with respect to Q_k.
— (weak-symmetry) If If Q_k is +absolute and Q_i is +distributive with respect to Q_k,, then Q_i is +distributive with respect to Q_i.
— (transitivity) If Q_i is -distributive with respect to Q_k and Q_k is -distributive with respect to Q_m, then Q_i is -distributive with respect to Q_m.

axiomatisations and their properties: completeness, compactness, Löwenheim property,…). Another problem of interest is how frequently the proposed techniques may appear necessary in practical considerations. We intend to tackle this problem in the future.

Acknowledgements

Thanks are due to Irena Bellert whose enthusiasm for quantifier related issues helped me to better understand the importance of quantifier studies. The Author wishes to thank the four Anonymous Reviewers for constructive remarks which encouraged him to continue working on the problems presented in this paper.

References

Barwise, J., and Cooper, R. 1981. Generalized Quantifiers and Natural Language. *Linguistic and Philosophy* 4: 159–219.

Bellert, I. 1989. *Feature System for Quantification Structures in Natural Language*. Dordrecht-Holland / Providence RI-USA: Foris Publications.

Dahl, V. 1981. Translating Spanish into logic through logic. *Computational Linguistic.* Volume 7, Issue 3: 149–164.

Mostowski, A. 1957. On a Generalization of Quantifiers. *Fundamenta Mathematicae* 44: 12-36.

Vetulani, Z. 1987. On Bellert's approach to quantificational universals. *Studia Logica* XLVI, 4: 311–320.

Vetulani, Z. 2002. *Question Answering system for Polish (POLINT) and its language resources.* In: M. Maybury (ed.). *Proceedings of the Workshop on Question Answering: Strategy and Resources, LREC 2002.* 51–55. Paris: ELRA.