Creating words by inflexion and derivation in RDFCFL graphs

Alena Lukasová, Martin Žáček
Department of Computers and Informatics
University of Ostrava
30. dubna 22, Ostrava
CZECH REPUBLIC
Alena.Lukasova@osu.cz http://www.osu.eu
Martin.Zacek@osu.cz http://www.osu.eu

Abstract: This paper presents an approach to express morphology and syntax rules of specifications of basic terminology of English language. The article continues the previous work of these authors in English Morphology in RDFCFL graphic metalanguage following the methodology of Andres Carstairs-McCarthy. This article aims are creating words by inflexion and derivation.

Key-Words: RDF model, Clausal Form Logic (CFL), RDFCFL formal system, morphological rules, Carstairs-McCarthy's, morphology.

1 Introduction

The authors of the paper completed the original RDF model by introducing general or existential statements (see [3]) as necessary conditions of formal deduction corresponding to the natural human mental activity. Moreover, they have extended the RDF model with "if – then" form of sentences following the Richard´s method known as "clausal form logics" (CFL).

The article continues the previous work (see [2]) of these authors in English Morphology in RDFCFL graphic metalanguage following the methodology of Andres Carstairs-McCarthy.

We present here an experiment to express ground terminology of the English language morphology and syntax on the base of first order logic's graphic tools RDFCFL. It seems that our meta-language representation of English grammatical rules taking into account predicate logic´s semantic principle with its high expressivity and easy-to-read graph-based form could generally bring new aspects into language theory.

The aim of this paper is modelling of linguistics at two levels

- syntactical: investigating lexical and functional categories of speech, in particular rules of their merging into higher units; here a graph-based tool of formal expressing became the RDFCFL metalanguage [4, 5],
- semantical: semantic-based morphology of a particular language which covers the importance of language morphemes; here it is better to use simple semantic (associative) networks together with some explaining by example.

All the life people learn to form sentences according to certain rules. Any natural language has its own rules, e.g. the English language form sentences using a simple formula SVOMPT (Subject, Verb, Objects, Manner, Place, Time) [7], while the Czech language has a structure that is much more complicated.

In order to model a natural language, it is necessary to find suitable means in the form of a formal system. As a default, the English language with its fairly simple morphology rules has been chosen for our experiment.

2 Drawing sentences about English linguistics by means of RDFCFL graph language or semantic network

T. Richards [4] proposed the Clausal Form Logic (CFL) built on the base of the FOPL and well corresponding with common using of the conditional „if – then "statement. Generally, a conditional statement (clause) says that the consequent composed as a disjunction of some predicate atoms follows from the antecedent composed as a conjunction of some predicate atoms.

The approach allows us to formulate clauses in the form

<antecedent> < implies> <consequent> (1)

Selecting a formal language for a knowledge representation is crucial. The formal basis should become here the first order predicate logic (FOPL) base for its high expressivity and a wide range of already developed formal deduction tools.

Knowledge Representation (originally those contained in Web resources), which are based on a domain ontology usually has been created in the framework of RDF (Resource Description Framework) model. An RDF model manipulates the semantic aspect of terms specified through URI references to resources in which their meanings are always elucidated by means of a certain position in a relevant ontology. The graphic RDF model in its form is easy and simple to understand even for the users who do not have experience with formal modelling. The idea is based on a simple statement concerning relations between items (resources) in the form of basic vector (Fig. 1):

Fig. 1. Basic vector.

The vector pattern corresponds with the SVO part of the general SVOMPT pattern of the English grammar.

The graph version of the vector representation uses notation of the Clausal Form Logic (CFL [6]).

Developments in the field of formal knowledge representation clearly show that the language of the FOPL and specifically its clausal form (in text or graph version) is an appropriate formal language that can virtually represent any assertion formulated in a natural language.

The graphic form of the CFL language [4] became the main idea of the RDFCFL graph language used here. Clauses use dashed lines in the cases of antecedent vectors and full lines for vectors of clause consequent.

By means of adding elements of Description logic [6] into the RDF model it has been possible to communicate with the web language OWL that also increases its expressiveness. Both languages in their text format are based on the XML syntax, making it easy for their machine processing.

Besides the RDFCFL representation of a clause we also use a simpler tool semantic network expressing interrelations of concepts in English linguistic known a long time before informatics ordered them among formal modelling tools. We use here the semantic network principle in the cases where the semantics of clauses is more important than syntax of their corresponding rules.

3 Words and morphemes

Andrew Carstairs-McCarthy writes in the book [1]:

“Two characteristics of morphemes, in the light of how the notion has been introduced. To allow the meanings of some complex words to be predictable, morphemes must

1. b e identifiable from one word to another (Fig. 2) and
2. contribute in some way to the meaning of the whole word (Fig. 3).”

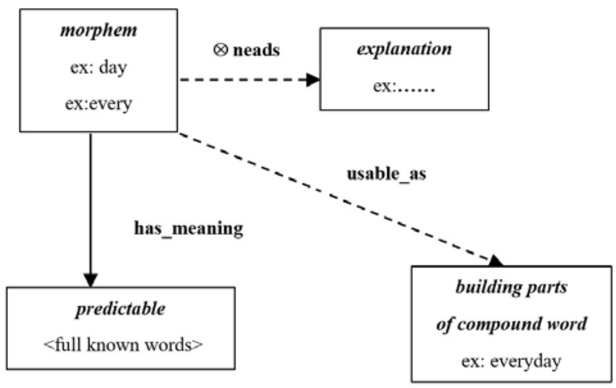

Fig. 2. Representation as a semantic network using the symbol of negation atom.

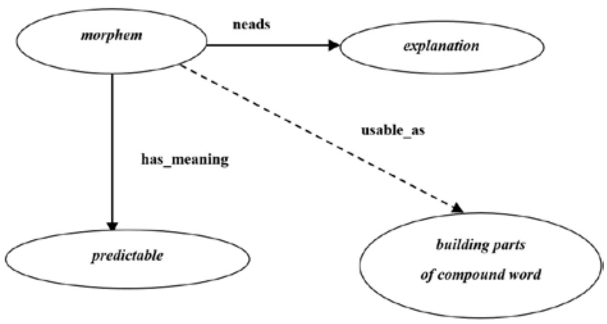

Fig. 3. Representation in the RDF CFL.

This two-aspect specification of the relation between concepts word and morpheme works with syntactic as well as semantic properties of both concepts. So it was natural to use a semantic network explaining relationships of concepts from the point of view of semantics in their corresponding relationship (fig.). Examples (ex:) are here only help-means of understanding. Usage of the RDFCFL representation form takes one contribute more: URIs referring towards properties of concepts “morpheme”, “predictability” or property of “heading” find a place in items hierarchy in ontology chosen as collaborating tool making the semantics of the represented sentences much more precise.

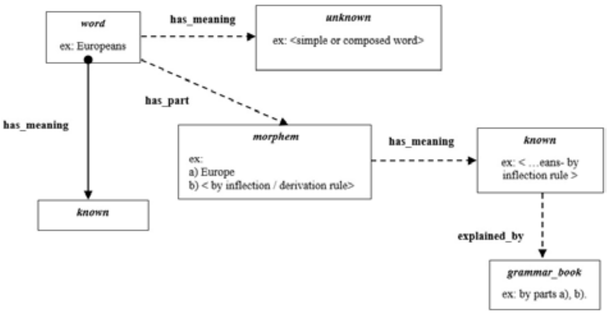

Fig. 4. Words and morphems in clausal representation as semantic network with examples.

The following figure shows the syntactic point of view of McCarthy's precise specification of interrelation between concepts word and morpheme. The meanings of both main concepts and predicates of their relationships define for each of them a reference URIref to ontology WORDNET chosen for this problem.

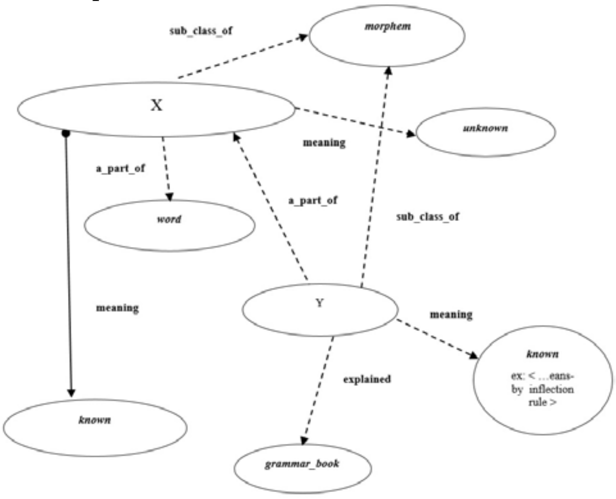

Fig. 5. Words and morphemes as RDF CFL network.

4 Creating words by inflexion and derivation

Some words (lexemes) have more than one word form, depending on the grammatical context or on choices that the grammar forces us to make (for example, in nouns, between singular and plural). This kind of word formation is called 'inflectional'. In so far as grammar affects all words alike, the existence of inflected word forms does not have to be noted in the dictionary; however, the word forms themselves must be listed if they are irregular.

4.1 Nouns: singular and plural forms

Most countable nouns in English have two word forms: a singular and a plural. Inflectionally, for any noun lexeme X, there are just two grammatical words, 'singular of X' and 'plural of X', contrasting in number.

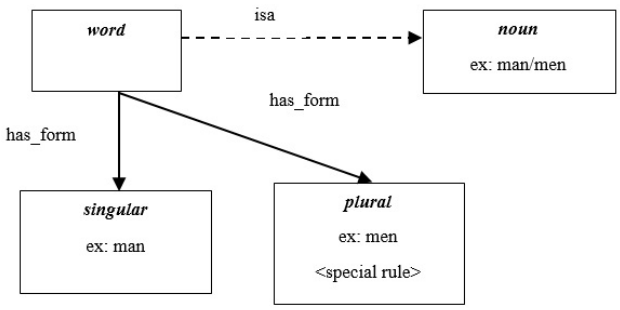

Fig. 6. Expresses only the facts of informational character.

To make the picture RDFCFL-like, it is necessary

- to use an elliptical shape of the node graphs,
- to add information that the statement holds for all the words that belong to the grammatical category of nouns,
- to write the completed URIs to all the items of the graph.

To fulfil the first point we can use of "Xword" character chain instead of the "word" one only.

$$\forall \text{word}(\text{isa}(\text{word, noun}) \rightarrow (\text{has\_form}(\text{Xword, singular}) \vee \text{has\_form}(\text{Xword, plular})))$$

Or in the clausal form

$$\text{isa}(\text{word, noun}) \rightarrow \text{has\_form}(\text{Xword, singular}), \text{has\_form}(\text{Xword, plular})$$

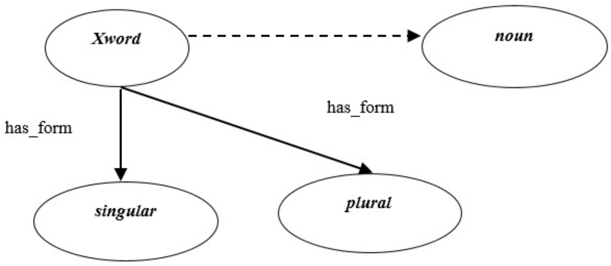

Fig. 7. Result previous statement.

4.2 Complex word creation by inflectional rule

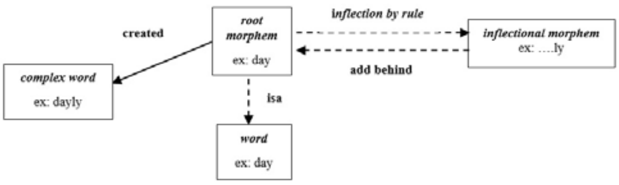

Fig. 8. Inflectional rule.

4.3 Complex word creation by derivational rule

Derivation - a word and its relatives: Derivation - a word and its relatives:

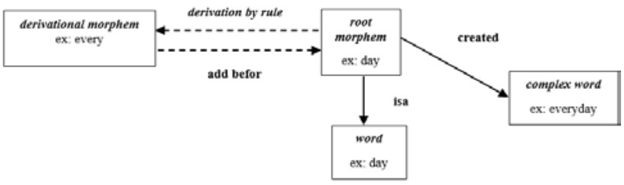

Fig. 9. Derivational rule.

4.4 Forms of verbs

In English, a verb lexeme has five distinct forms at the most, as illustrated here:

a. third person singular present tense: gives e.g. Mary gives a lecture every year.
b. past tense: gave e.g. Mary gave a lecture last week.
c. progressive participle: giving e.g. Mary is giving a lecture today.
d. perfect or passive participle: given e.g. Mary has given a lecture today. The lecture is always given by Mary.
e. basic form (used everywhere else): give e.g. Mary may give a lecture.

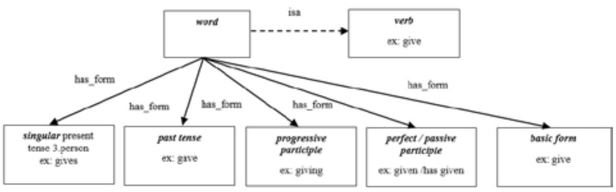

Fig. 10. Forms of verbs.

4.5 Forms of adjectives

Many English adjectives exhibit three forms, for example here:

- Grass is green.
- The grass is greener now than in winter.
- The grass is greenest in early summer.

The grammatical words that green, greener and greenest express are the positive, comparative and superlative, contrasting on the dimension of comparison.

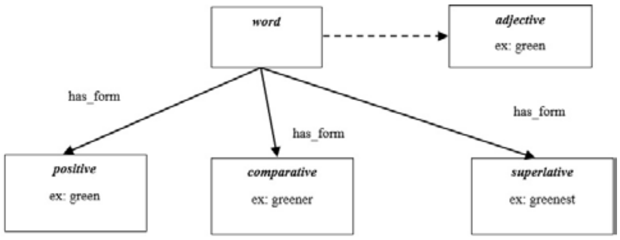

Fig. 11. Forms of adjectives.

5 Acknowledgments

The research described here has been financially supported by University of Ostrava grant SGS13/PŘF/16. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not reflect the views of the sponsors.

6 Conclusion

We see a usability of our RDFCFL approach in four directions:

1. As an easy-to-understand tool of representation of language grammatical rules. It is possible to use only a simple semantic network.
2. The approach based on logical principles could push authors of definitions or specifications of rules to hold within special grammar categories to specify carefully what are items of clauses, antecedents – prerequisites and consequents – conclusions of clauses represented truth-full general sentences.
3. As a part of the knowledge base of general rules in domain ontology for concrete natural language because it gives a possibility of rewriting also into OWL language.
4. As a general part of the knowledge base together with further clauses describing the represented world gives a possibility to obtain new conclusions by means of the RDFCFL formal system.

This article aims are creating words by inflexion and derivation, this goal we succeeded and we see the result in the article: Some words (lexemes) have more than one word form, depending on the grammatical context or on choices that the grammar forces us to make (for example, in nouns, between singular and plural). This kind of word formation is called ‘inflectional’. In so far as grammar affects all words alike, the existence of inflected word forms does not have to be noted in the dictionary; however, the word forms themselves must be listed if they are irregular.

References:

[1] Carstairs-McCarthy, A: An Introduction to English Morphology: Words and Their Structure. Edinburgh University Press 2002, ISBN 0748613269.

[2] Lukasová, A., Žáček, M., Vajgl. M. Carstairs-McCarthy´s morphological rules of English language in RDFCFL graphs. 1st International Conference on: Applied Physics, System Science and Computers.Dubrovnik, Croatia, 2016. ISSN: 1876-1100

[3] Lukasová, A., Vajgl, M., Žáček, M.: Reasoning in RDF graphic formal system with quantifiers.Proceedings of the International Multikonference on Computer Science and Information Technology. 2010. pp. 67-72.

[4] Lukasová, A., Žáček, M., Vajgl, M., Kotyrba, M.: Resolution Reasoning by RDF Clausal Form Logic. IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 3, No 1, May 2012. ISSN (Online): 1694-0814. www.IJCSI.org, 2012.

[5] Lukasová, A., Žáček, M., Vajgl, M.: Reasoning in Graph-based Clausal Form Logic. IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 1, No 3, 2012, pp. 37-43. ISSN (Online) 1694-0814.

[6] Richards, T.: Clausal Form Logic. An Introduction to the Logic of Computer Reasoning. Addison – Wesley, 1989.

[7] Miarka, R. a Žáček, M. Knowledge patterns for conversion of sentences in natural language into RDF graph language. In: Proceedings of the Federated Conference on Computer Science and Information Systems. USA: IEEE Computer Society Press, 2011. IEEE Computer Society Press, 2011. s. 63-68. ISBN 978-1-4577-0041-5.

[8] Miarka, R. Representation of Knowledge patterns for Semantic web. IJCSI International Journal of Computer Science Issues. 2013, č. 10, s. 201-207. ISSN 1694-0814.