UDK 811'8
81'36
81'37
Review article
Accepted for publication 11. 10. 2002

Sanja Seljan, Kristina Vučković, Zdravko Dovedan
Filozofski fakultet, Zagreb

Sentence Representation in Context–Sensitive Grammars

Every language, whether it is natural or artificial, has its recognizable grammar that consists of allowed elements and rules for putting those elements together. The main aim of the formal grammar is to represent rules for generation of the artificial or natural languages. While artificial languages (such as note system, logic, mathematics, programming languages) are described by context–free formal grammar aiming to describe syntax, natural languages tend to be described by context–sensitive rules aiming to include, as much as possible, syntactic and semantic component. Among many formal grammars that tend to describe as much as possible the natural language sentences, in this paper two context–sensitive grammars will be presented: Lexical–Functional Grammar and Case Grammar that aim to include semantic roles (such as agent, theme, beneficiary, goal, location, etc.) in order to represent the natural language sentences.

1 Introduction

Every formal grammar aims at finding the most suitable way to represent sentence of the natural language, i. e. to represent the syntax and the meaning of the sentence. The search for bridging the gap between artificial and natural languages, bearing in mind specific problems such as ambiguity, world knowledge, context dependency, has motivated linguists and information scientists to develop different types of formal grammars and parsers in order to perform analysis of the natural language sentence. Every formal grammar has been trying to find the proper way to come as close to the natural language (NL) as possible, so different approaches have been developed: Transformational Grammar using transformations, generative semantics focusing on the semantic component, Unification–based Grammars using unification as main operation, Lexical–Functional Grammar using grammatical function represented in lexicon, Case Grammar using semantic cases, Systemic Grammar which is semiotics–driven, Semantic Grammar using semantic subcategories, Tree–Adjoining Grammar using trees, Head–driven Phrase Structure Grammar, etc.

Sentence representation is also one of the main problems. Grammars should enable representation of the NL sentence suitable for various types of languages, i. e. grammars should be suitable for highly configurational equally as for nonconfigurational types of languages.

In the paper, formal demands of context–sensitive grammar and two formal grammars: Lexical–Functional Grammar and Case Grammar will be presented. They both tend to analyze the NL sentences, representing them through different structures. Trying to perform morphological, lexical, syntactic and semantic analysis, each of the grammars is based on different basic principles in order to represent the NL sentence in the most suitable way.

2 Generative semantics

The *generative semantics* is usually considered separately from the Transformational Grammar of Noam Chomsky. The generative semantics focuses on the semantic component and on the deep structure. The semantic component is considered to be the center, which could generate syntactic structures, which is opposed to N. Chomsky's theory that focused in the beginning on the syntactic component (although in later works he added semantic component, which can apply also on the surface structure. This part of his theory is called *interpretive semantics,* by which he interprets the syntactic structure by adding semantic component. Still, it is the syntax that has the generative power, opposed to generative semantics.)

Somewhere between the interpretive and the generative semantics, the Case Grammar by Charles Fillmore (1962) has been developed. He considered that deep structure should be best characterized by deep cases (such as *agent, patient, instrument, source, destination, location and manner*).

This idea has been afterwards taken over in Lexical–Functional Grammar (Kaplan & Bresnan, 1982) using argument (a)–structure contained in the lexical unit and afterwards included in the functional (f)–structure. *Predicate–argument structure* in LFG model consists of predicator and a list of arguments that relate grammatical functions (subject, object, object2, etc.) and thematic roles (*agent, beneficiary, goal, instrument, theme, location*).

Θ*–theory* or *theory of thematic roles (agent, theme, etc.)* has been developed in the 60's and 70's, but has appeared in syntactic descriptions quite recently. Although numerous formal grammars use Θ*–theory*, there isn't complete list of all semantic or thematic roles and exact rules of assigning certain roles in a context, but some aspects are used more frequently.

3 Formal demands of context–sensitive grammar

The concept of context–sensitive grammars was introduced by Noam Chomsky in the 1950's, trying to formally describe type of formal languages suitable for description of NL (natural language) sentences. Although some parts of natural languages can be described using context–free languages or even regular expressions (e. g. lexicon), it is often the case that a word is described only in the suitable context. This type of formal grammar is especially needed in non–configurational languages where agreement or long–distance dependencies need to be formalized.

According to investigations of Laboratory for Linguistics and Computation at Brandeis University, typical NL constructions that require sensitive power are:

– *reduplication*, leading to languages of the form { ww | w is Sigma* }, where Sigma is equivalent to V, marking set of nonterminals–
– *multiple agreements,* corresponding to languages of the form { $a^nb^nc^n$ | n > 0 }, { $a^nb^nc^nd^n$ | n > 0 }, etc.
– *crossed agreements*, as modeled by { $a^nb^mc^nd^m$ | n > 0 },
– *Mildly context–sensitive grammars* have been proposed as capable of modeling the above–mentioned phenomena.

A context–sensitive grammar is a formal grammar G=(N, T, P, S) if all productions are of the form

$$\alpha A\beta \rightarrow \alpha\gamma\beta$$
$$\text{where } A\in N;\ \alpha,\beta\in N\cup T)^*;\ \gamma\in(N\cup T)$$

meaning that A in N is a single nonterminal; α and β in $(N\cup T)^*$ are strings of nonterminals and terminals and γ in $(N\cup T)+$ is a nonempty string of nonterminal and terminals. The quality of *context sensitive* is explained by α and β that form the context of A which can be replaced by γ only in the context α and β.

Rule of the form

$$S\rightarrow\varepsilon$$

with ε the empty string, is added only if S does not appear on the right side of any rule.

Another definition of context–sensitive grammars defines them as formal grammars with the restriction that for all rules

$$\alpha\rightarrow\beta$$
$$\text{where } |\alpha| \leq |\beta|$$

Such grammar is also called a *noncontracting grammar* because none of the rules decreases the size of the string that is being rewritten.

A context–sensitive language is a language generated by a context–sensitive grammar. Typical formal language generated by context–sensitive grammar is $a^nb^nc^n$.

4 Lexical–Functional Grammar

Lexical–Functional Grammar (LFG) is a formal model of generative non–transformational grammar, developed by R. Kaplan and J. Bresnan, 1982. As a Unification grammar, it uses unification as a principal operation.

In order to represent the NL sentence, the formal grammar has to be context–sensitive, which, besides syntax analysis, also performs certain semantic analysis (in quite a restricted way). Words are analyzed in the environment of the left and right context and not isolated one from another. Context–sensitive grammars are used to formally describe agreement, passive sentences, relative sentences, long–distance dependencies, etc.

4.1 Why LFG is suitable for different types of languages?

The formal model should be suitable for highly configurational type of languages (English, French, and German) and for nonconfigurational languages with relatively free word order providing simple morphological analyzer (J. Bresnan: »Morphology competes with syntax.«). Therefore, besides the representation suitable for the specific natural language (principle of variability), a more abstract representation is also needed (principle of universality). This means that the formal grammar should satisfy several conditions: to be context–sensitive, and to be suitable for highly structured ad well as for nonconfigurational languages.

Having in mind that the perfect formal model doesn't exist (because of ambiguity, complexity and unlimited possibilities of combinations in the language) and that formal descriptions are used for a certain subset of the language sentences or for a controlled language, there are several reasons for the LFG model to be chosen.

First, LFG model aims to unify *computational efficiency* and *linguistic theory*. It has been developed as a result of linguists and information specialists (J. Bresnan was a former student of N. Chomsky and R. Kaplan worked on ATN parsers) using knowledge from linguistics, informatics, logic and mathematics. LFG model is quite restricted in its generative power, but suitable for computer implementation.

Second, according to the principle of the Universal Grammar, which supposes that despite very different syntactic means of expression all natural languages have a verb and grammatical functions as subject and object, the crucial idea of the LFG model is to represent *grammatical functions in the lexicon,* assuming also that some linguistic phenomena like passivization are psychologically better explained in this way than using transformations.

Because of its possibility to represent highly structured language sentences with fix word order and sentences with relatively free word order and rich morphological system (simultaneous representation of c–and f–structures), the LFG model has been used for various types of languages and for different language phenomena: for English, French, German, Italian, Spanish, Irish, Dutch, Russian, Arabian, Japanese, Urdu, Walpiri, Bantu languages, Icelandic, Flemish, Norwegian, Malayalam, Moroccan, etc. Therefore, LFG tends to be suitable for different types of languages.

Third, in LFG model the sentence structure is represented in two basic forms that exist *simultaneously*

– As hierarchical *constituent or c–structure* which *varies* from one language to another, reflecting the surface structure
– As *functional or f–structure* as more abstract way of representation which tends to be *universal* for the same sentence across the languages although other levels were additionally added:
– *Lexical level* representing the lexical unit
– *Argument or a–structure* that was firstly included in the lexical level, and in 1989. Bresnan and Kanerva separated the a–structure as a distinct level of representation
– *Morphological or m–structure* which is lately subject of many discussions, representing morphological information.

Constituent structure is represented in the form of tree or in the form of context–free rules, enriched with functional annotations. C–structure encodes linear order, hierarchy and syntactic categories of constituents. Information form c–structure and lexicon participate in creating functional structure (except constraining equations). Using constraints, the principle of agreement between subject and the verb in person and number is added.

```
S   ──►   NP            VP
          (↑SUBJ)=↓     ↑=↓

VP  ──►   V                  (NP)
          ↑=↓                (↑OBJ)=↓
          (↑SUBJ NUM)=↓
          (↑SUBJ PRS)=↓

NP  ──►   (Det)  (Adj)   N
          ↑=↓    ↑=↓     ↑=↓
```

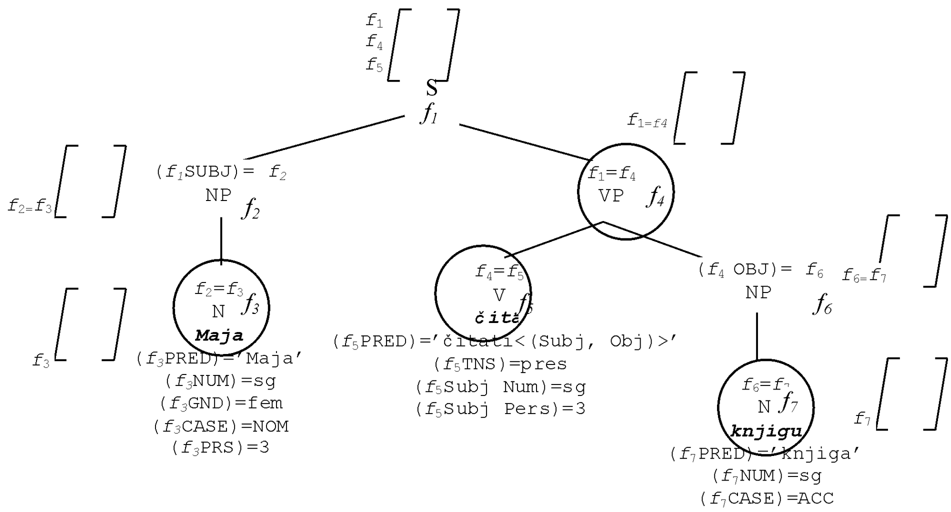

Figure 1. Annotated c–structure

Functional structure is represented in the form of matrix, which consist of attribute–value pairs. When a lexical item occupies the terminal node, it is inserted into the f–structure. In this way, functional structure integrates structural and lexical information. Functional structure does serve for more abstract presentation of grammatical structures, regardless of the position in the sentence.

```
        ┌                                               ┐
        │ PRED  'čitati <( f5Subj), (f5Obj)>'           │
        │ TNS    pres                                   │
        │                                               │
        │         ┌                    ┐                │
        │         │ SUBJ     'Maja'    │                │
        │   SUBJ  │ NUM       sg       │                │
        │         │ GND       fem      │                │
        │   f2    │ CASE      NOM      │                │
        │   f3    │ PRS       3        │                │
f1      │         └                    ┘                │
f4      │         ┌                    ┐                │
f5      │         │ PRED  'knjiga'     │                │
        │   OBJ   │ NUM    sg          │                │
        │   f6    │ CASE   ACC         │                │
        │   f7    └                    ┘                │
        └                                               ┘
```

Figure 2. Basic f–structure

Lexical structure includes information about meaning of the item, its argument structure and grammatical functions. Grammatical functions (such as subject, object, etc.) play an essential role in the LFG model and they mediate between lexicon and syntax.

Maja, N	*čita, V*	*knjigu, N*
(↑PRED)='Maja'	(↑PRED)='čitati<SUBJ,OBJ>'	(↑PRED)='knjiga'
(↑NUM)=sg	(↑TNS)=pres	(↑NUM)=sg
(↑GND)=fem	(↑NUM)=sg	(↑CASE)=ACC
(↑CASE)=NOM	(↑PRS)=3	
(↑PRS)=3		

Bresnan and Kanerva have added in 1989. *argument or a–structure* as separate structure from the lexical level, consisting of predicate and its arguments. According to *Function–Argument Biuniqueness*, every grammatical function (*subject, object*) has its thematic role (*agent, beneficiary, goal, instrument, theme, and locative*), so semantic meaning is described inside the' 'signs, where relations between functions and thematic roles are defined.

```
             (agent , theme)   ← thematic roles
'čitati      (SUBJ,    OBJ)'   ← grammatical functions
```

This idea to subcategorize grammatical functions and not syntactic categories, enable categorial independency, i. e. that different word categories (word types) can have the same grammatical function and the same category can have different functions.

Therefore, in the LFG model lexical, syntactic and semantic analyses are quite interrelated, represented through simultaneous c–structure and f–structure.

```
     ___S_____
    /         \
  NP         ___VP_____
   |        /          \
   N       V        ___Scomp___
   |       |       /           \
  Na1     V1    Comp          ___S_____
   |       |      |          /         \
 Marko   kaze    da        NP         __VP____
                            |        /        \
                            N       V        VBAR__
                            |       |       /      \
                           Na1    zeli     V       VBAR
                            |              |         |
                        Tomislav        nauciti      V
                                                     |
                                                   citati
```

```
PRED  Kazati<SUBJ,COMP>
  TNS   PRES
  SUBJ  PRED  Marko
        NUM  SG
        PRS  3
        GND  MASC
        CASE  NOM
  COMP  PRED  Zeljeti<SUBJ,XCOMP>
        TNS  PRES
        NUM  SG
        PRS  3
        SUBJ   <7108>
              PRED  Tomislav
              NUM  SG
              GND  MASC
              CASE  NOM
        XCOMP  PRED  Nauciti<SUBJ,XCOMP>
               MOOD  INF
               SUBJ   <7108>
               XCOMP  PRED  Citati<SUBJ>
                      MOOD  INF
                      SUBJ   <7108>
```

Figure 3. Constituent and functional structure for the sentence *Marko kaže da Tomislav želi naučiti čitati*

A. Frank, R. Kaplan and M. Butt give preference of introducing the new level of representation, i. e. *morphological (m)–structure* for representation of morphological properties, but without more important functional or semantic role. This structure would contain morphological features that are very important for description of the Croatian language, and also agreement properties.

Therefore, using m–structure the same sentence expressed in Croatian, English, German and French would have the identical f–structure (with main verb on the top in composite tenses), while all morphosyntactic differences would be presented in m–structure. The main verb is also the governing verb with PRED attribute and carrying information about tenses. M–structure would carry information about auxiliaries.

Therefore, all morphological differences in different languages for the same sentence would be described in m–structure, while the f–structure would be universal.

Eng. *The driver will turn the lever.*

Ger. *Der Fahrer wird den Heber drehen.*

Fr. *Le conducteur tournera le levier.*

Cro. *Vozač će pokrenuti ručicu (polugu).*

```
                              PRED  'turn/drehen/tourner/pokrenuti<(SUBJ, OBJ)>'
                              TNS   FUT
FIN  +
                                    [PRED  'driver/Fahrer/conducteur/vozač']
AUX  +                        SUBJ  [Gnd   masc                           ]
                                    [Num   sg                             ]
                                    [Spec  DEF                            ]
     [FIN     -   ]
DEP  [VFORM   BASE]                 [PRED  'lever/Hebel/levier/ručica']
                              OBJ   [Gnd   masc                      ]
                                    [Num   sg                        ]
                                    [Spec  DEF                       ]
```

Figure 4. Morphological and functional structure for one sentence in several languages

Fourth, in order to be well–formed, the functional structure has to satisfy several criteria:

– *consistency* also known as functional uniqueness, requiring that each attribute in the matrix has a unique value, i. e. it is not possible to have one attribute having two different values.

– *completeness* requiring that every function that is subcategorized by the predicate must have a value, e. g. if a verb subcategorizes subject and object, than each of these functions must have value. Otherwise, the sentence is incomplete (*Speaks.)

– *coherence* requiring for every grammatical function to have value mentioned in predicate argument structure.

Fifth, LFG model belongs to the group of Unification Grammars, using the operation of *unification*, which is a result of the work in ProLog. Unification of two functional structures A and B is a new structure A ∪ B that unifies all compatible features of the two structures. In the case of no compatible feature, unification fails.

Unification is especially important in agreement, e. g. it is not necessary to mention always every characteristic feature for determiner, adjective and noun, because all characteristic features are unified in the noun phrase.

Sixth, it is possible to *decompose* categories on *characteristic features* or to group categories having the same characteristic features (e. g. reflexive verbs having Refl=+, negatives having Neg=+, collective nouns Coll=+ etc... In this way, it is possible to *add new characteristic features*, which for example can be used for recognizing the cases having the same forms (Soc=+, Thg=+). Characteristic features are unified by the operation of *unification* in one syntactic group, incorporating in that way *contextual elements* that are in meta–language reflected as attribute–value pairs.

Seventh, syntactic rules are enriched by constraints, so LFG model is a type of *constraint–based* grammar defining set of constraints that need to be satisfied in order to produce the sentence. In that way, the accumulation of rules is avoided.

3.1. Practical use of LFG model

LFG model has been used as a linguistic theory of more recent Machine Translation systems:

The Knowledge Research Institute of the *Fuji Xerox Corporation* is doing research on knowledge processing systems based on Natural Language Processing technologies. They are writing computational grammar on the basis of LFG formalism in order to construct a high–quality parser and generator for Japanese. This should be basis for knowledge processing systems, such as content understanding systems and human–machine dialog systems. The Project aims at practical Japanese LFG system with broad coverage. Besides Machine Translation, other possible applications of the computational LFG model are: summarization, document collection management (detect and resolve redundant content, create composite documents), e–mail response, and human–machine dialog, question–answering and knowledge extraction.

This group is a member of *ParGram project, 1995. (Parallel Grammar Project)* where LFG computational grammars for English, French, German, Norwegian, Japanese and Urdu have been developed for multi–lingual applications. The major goal of ParGram project is to produce large LFG computational grammar for multiple languages, based on parallel principles, i. e. on common set of linguistic principles. Members of ParGram project are Palo Alto Research Center, University of Stuttgart, University of Bergen, and University of Konstanz.

Kant, 1991. is a knowledge–based machine translation system that produces semantically accurate translations, but requires massive knowledge acquisition for technical documentation. It works for document production in French, German and Japanese using LFG source grammar in order to produce *interlingua representation* for each sentence. The Kant system has two main restrictions: limited vocabulary (the general vocabulary has about 14.000 word senses) and level of syntactic complexity in order to avoid ambiguities in parsing, but still providing a subset of English sentences suitable for adequate expression.

1990. Gertjan van Noord has used Unification Grammars in the *MiMo2* system for translating international teletext news among English, German and Spanish.

In 1991. Zweigenbaum has used LFG model for the syntactic and semantic analysis of doctoral diagnosis (hospitalization, chronology of diseases) combining documentation approach and language analysis.

Another two MT systems using *transfer* (*Machine Translation Toolkit, Executive Communication System, Provo, Utah, 1985*) and *interlingua* (*PONS – Partial Translation between Closely Related Languages, University Bergen, 1990*) strategies. Both systems use as common denominators functional structure of the LFG formalism, tested on English–Norwegian translations of technical documentation.

5 Case Grammar

Case grammar (CG) was originally developed by Fillmore. It refers to a system of grammatical forms that expresses predicate–argument structure. It presents an alternative approach to grammar where cases represent roles in a sentence. Cases are not considered as classical cases in morphological or syntactic sense, but are considered to be semantic categories (agent, patient, instrument, source, destination, location and manner) that can have different forms in the syntactic realization. Every sentence has maximally one filer per role. Languages differ in terms of the number and kinds of cases they have and how they are expressed. Not all languages have elaborate case systems. English, for example, draws predicate–argument structure through word order and system of prepositions.

CG was introduced in order to avoid confusion about what argument plays what role in a predicate. It focuses more on developing a structure that more clearly marks aspects of the meaning. Fillmore suggested that syntactical analyses of sentences were insufficient, because this could suggest different meanings for a word depending on the syntactic role it plays in the sentence. Thus, he suggested that language be analyzed for meaning using case roles. Each word in a sentence plays a semantic role, which he refers to as the word's case role. The central notion of a case analysis is to translate sentence strings into a nested structure of case relations where each relation has a head term and an indefinite number of labeled arguments. An argument may itself be a case relation. Thus a sentence forms a tree of case relations.

Case grammar makes a sharp distinction between semantic relations and grammatical relations. Semantic relations and grammatical relations do not always bear a one–to–one relationship to each other. Even if the elements change grammatical location, the semantic roles, the actual events, do not change at all. Case grammar maintains the constancy of the semantic roles regardless of their grammatical location in a sentence.

English language has three cases: *nominative*, *possessive* or genitive, and *objective* or accusative for direct object and dative for indirect object. Fillmore argued that the traditional concept of case dealt only with the surface structure of the sentence and was not significant in any meaningful way. Thus, he suggested that a sentence is composed of the *proposition*, a tenseless *set of relationships* involving verbs and nouns, and *modality constituents*, negation, tense, mood and aspect with a case occurring only once in a given sentence.

CG describes language from a more semantically oriented perspective. It provides more than a description of the structural relationships within a sentence including notions about the functional relationships among the various phrases within a sentence, the part of syntax that conveys meaning. The verbal elements of the sentence are the major source of the structure: the main verb in the proposition is the focus around which the other phrases, or cases, revolve and the auxiliary verbs contain much of the information about modality.

Fillmore's grammar suggests that underlying all (surface) sentences is a deep structure, which consists basically of a proposition plus a modality. The modality component contains elements that modify the proposition as a whole. The proposition »consists of a verb and one or more noun phrases, each associated with the verb in a particular case relationship« (Fillmore, 1968). The case is one of the underlying syntactic – semantic relationships in a language which make up a universal set of innate concepts that explain judgments about notions such as »ho did what to whom« (Palmatier, 1972). Fillmore gives his list of case relationships that can appear in a sentence: agentive, instrumental, dative, factitive, locative and objective. Each verb is associated with a particular selection from this list, and there are rules which determine how the various underlying cases are realized in terms of subject, direct object, indirect object, and so on.

Description of case grammar can be seen through the set of rules:

$S \rightarrow M + P$

$M \rightarrow$ tense, aspect, form, mood, essence, modal, manner, time

$P \rightarrow V + C1... Cn$

$V \rightarrow$ existing verbs in the vocabulary

$C_i \rightarrow K + NP$

$K \rightarrow$ *(null, PREP).*

Where *tense* in modality indicates the basic time of an action (present, past, future); *aspect* says whether the action is continuing (perfect) or completed (imperfect); *form* determines whether the action is simple, emphatic or progressive; *mood* relates to the order of the elements of the sentence whether it is declarative, interrogative or imperative; *essence* shows whether the action is positive, negative or indeterminate; *modals* are the helping verbs (may, can and must); and both *manner* and *time* are indicated by the adverbial parts of a sentence. Since in case grammar the verbal elements present the focus of a sentence, it is of a great importance to determine its modality.

Mapping of the cases indicated in a case frame onto the actual surface elements of a sentence can be done in two ways. The first is *sentence generation* where the concepts, which need to be represented, are mapped onto the case frame. In every case frame, cases can be optional, obligatory or not allowed. Each of these states has to be marked appropriately. Than, the three rules are applied in the given order: creating the subject, creating the object and finally, specifying prepositional phrases.

The second is *sentence analysis* where the process is reversed, i. e., the cases in the case frame are mapped onto the phrases found in the sentence. The analysis is started with the process of finding the verb followed by prepositional phrases so that a preliminary match can be made between founded phrases and the responding cases. When all of this is completed, rules for finding subject and object are applied.

However, given rules do not handle all possible situations. Additional semantic information about objects is needed so that the correct decision could be made. Case grammar does not handle modality adequately. Thus, additional considerations must be added in order to deal with more complicated sentence types.

Case grammar gives graphical representation of its general case sentences. Specific graphs are built during the process of synthesis or analysis, as parts of a sentence are discovered. During synthesis, the nodes are created from the given information about a sentence. This data is stored until the surface structure is generated from the graph. During analysis, the nodes of the graph are added to the list as they are found and filled in, as parts of a sentence are interpreted.

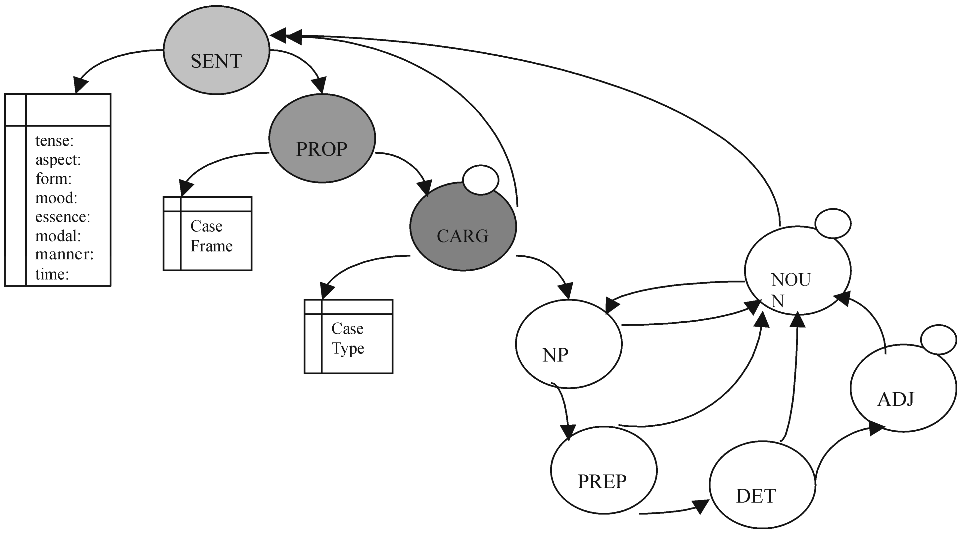

Figure 5. Graph representation

Case grammar graphs have a very good application in answering questions about the information stored in them. The question can be answered by finding at least one case argument in the sentence.

Although Fillmore's grammar has the possibility of a closer union between linguistics and psychology in regard to the understanding of the language–thought relationship, Segalowitz criticizes Fillmore's approach saying that although radical in using »explicitly psychological aspects of language«, it doesn't go far enough – it is »still subordinated to the task of relating surface structures to each other« (Segalowitz, 1970). Indeed, Fillmore gives quite a little attention to the non–linguistic analysis of his case categories but rather stresses their relationship to surface structure. Even though the definitions of these categories are all predicated on something called *the verb, the verb* itself is left undefined in terms of cognitive structure. The definitions, moreover, leave something to be desired in terms of their relationship to the overall cognitive structure. Another discrepancy is that determiners are introduced into the deep–structure phrase markers without any statement as to their origin or semantic correspondence.

Still, case grammar, combined with phrase structure grammar, has found its implementation in solving word problems in Physics (Chatterjee, Barboun, 1999). Also, speech and language clinicians use it for explaining the basic sentence in their work with clinical language teaching. The system of deep case has become one of the modules of generative Government Binding theory under Theta theory (Θ–theory) or the theory of thematic roles (Chomsky, 1981).

6 Conclusion

It can be seen from number of congresses discussing its theoretical aspects and practical use that formal description of natural languages is an interesting research area.

LFG model does represent one step forward in language description ensuring flexible tool for highly structured and nonconfigurational languages, but still many linguistic phenomena are not sufficiently described (non–agreement in coordinated structures, too long dependencies, some relative and causative constructions, composed tenses, etc.). Besides problems valid for all natural languages, the Croatian language has some specific demands. Still, in spite of imperfections of formal model, its formal analysis and teamwork of linguists and information specialists would give certain result.

Case Grammar focuses mostly on the semantics of a sentence using the case relationships that exist between the main verb and other sentence elements. In spite of the fact that it still lucks the rules to handle all possible sentence situations in all types of languages and that it puts all its attention to the surface structure, it has shown very good results in many projects concerned with the language processing. Further work is needed to broaden its implementation which might prove useful in eve wider selection of language concerned problems.

References

1. Bresnan, Joan. *Lexical–Functional Syntax.* Blackwell Publishers, 2001.
2. Butt, Miriam. *The Treatment of Tense*. // Proceedings of the LFG01 Conference. CSLI Online Publications. http://csli–publications. stanford.edu/LFG/6/lfg01butt.pdf
3. Chatterjee, A., Barboun, L. *A Natural Language Processor for Physics Word* Problems, Proceedings of the 35th International Linguistics Colloquium, Germersheim, Germany, 1999. http://www.hyperdigm.com/publications/Germersheim1999.doc
4. Chomsky, Noam. *Aspects of the Theory of Syntax. Cambridge:* MIT Press, 1965.
5. Chomsky, Noam. *Syntactic Structures. Paris:* Mouton, 1972.
6. Chomsky, Noam. *Lectures in Government and Binding.* Studies in generative grammar 9. Dordrecht, Foris, 1981.
7. Fillmore, Charles J. *The Case for Case.* // Universals in linguistic theory, edited by Emmon Bach, Robert T. Harms, Holt, Rinehart and Winston, New York, 1968
8. Frank, Anette. *Syntax and Morphology of Tense in LFG.* // Proceedings of the LFG00 Conference. http://www.xrce.xerox.com/people/frank/papers.html

9. Eiselt, K; Holbrook, J.: *Universal Grammar Theory,* http://www.cc.gatech.edu/computing/classes/cs4344_98_winter/lec06.htm l

10. Hancox, Peter. *Annual Research Review*, 1997. The University of Birmingham. http://www.cs.bham.ac.uk/research/booklet_97/arr/projects/node8.html

11. Harris, M. D.: *Introduction to Natural Language Processing,* Reston Publishing Company, Inc., Reston, Virginia, 1985.

12. Neidle, Carol. *Lexical Functional Grammar.* // Encyclopedia of Language and Linguistics. New York: Pergamon Press, 1994. Str. 2147–2153. http://www.bu.edu/asllrp/neidle–lfg.pdf

13. JLTT: Japanese Linguistic Theory and Technology. Knowledge Research Institute, Corporate Research Center, Fuji Xerox Co, Ltd. http://www.fujixerox.co.jp/crc/rf–kndg/linguistic/linguistic–e.html

14. Palmatier, Robert A. *A Glossary for English Transformational Grammar.* Appleton–Century–Crofts, New York, 1972.

15. Segalowitz, Norman. *Approaches to linguistic theory.* Informally circulated paper, originally presented to the Psycholinguistics Seminar, Oxford, 1970.

16. Steels, L.: *Computer simulations of the origins of case grammar,* http://www.ling.ed.ac.uk/evolang2002/ABSTRACTS/steels.txt

17. Thunes, Martha. *Machine Translation Strategies:* A Comparison of F–Structure Transfer and Semantically Based Interlingua. Eklund, R. (ed.) Proceedings of the 9th Nordic Conference on Computational Linguistics (NODALIDA93). http://www.nodali.sics.se/bibliotek/nodalida/1993_sthlm/NODA93–26/NODA93–26.ht

18. van Noord, Gertjan. *Reversible Unification Based Machine Translation.* Proceedings of the 13th International Conference on Computational Linguistics (COLING1990). http://grid.let.rug.nl/~vannoord/papers/coling90/coling90.html

19. Zweigenbaum, Pierre. Soutenance d'habilitation à diriger des recherches: Traitement automatique de la langue médicale. http://www.biomath.jussieu.fr/~pz/hdr/

20. Wong, Sylvia. *Machine Translation Techniques.* http://www.fi.muni.cz/usr/wong/teaching/mt/notes/index.html.iso–8859–1

21. Yehuda, N. Falk. *Lexical–Functional Grammar:* An Introduction to Parallel Constraint–Based Syntax. Stanford: CSLI Publications, 2001.

Prikaz rečenica u kontekstnim gramatikama

Svaki jezik, bilo prirodan ili umjetan, sadrži gramatiku kojom se određuju dozvoljeni elementi jezika i pravila kojima se određuje grupiranje i redoslijed dozvoljenih elemenata. Glavni cilj formalnih jezika jest prikazati pravila za izgradnju umjetnih ili prirodnih jezika. Dok se umjetno stvoreni jezici (npr. notni sustav, jezik logike, matematike, jezici za programiranje) opisuju beskontekstnim pravilima usmjerenim isključivo na sintaksu, prirodni se jezici opisuju kontekstnim pravilima koja nastoje uključiti, koliko je to moguće, sintaktičku i semantičku komponentu. Između brojnih formalnih gramatika, od kojih se svaka nastoji što je moguće više približiti opisu prirodnih jezika, u ovom će se radu prikazati dvije formalne kontekstne gramatike koje nastoje uključiti i semantičke uloge (kao agens, tema, dobročinitelj, cilj, mjesto, itd.) kako bi se što bolje opisale rečenice prirodnoga jezika.

Key words: formal grammar, natural language, context–sensitive grammar, lexical–functional grammar, case grammar

Ključne riječi: formalna gramatika, prirodni jezik, kontekstna (okolinska) gramatika, leksičko–funkcionalna gramatika, padežna gramatika