SEMANTIC FEATURES AND SELECTION RESTRICTIONS

Elena V. Paducheva

Institute of scientific and technical information (VINITI)
Academy of Sciences of the USSR
125219 Moscow, Usievicha 20a
Telefax: (7.095) 9430060

ABSTRACT.One of the essential aspects is described of an expert system (called LEXICOGRAPHER), designed to supply the user with diverse information about Russian words, including bibliographic information concerning individual lexical entries. The lexical database of the system contains semantic information that cannot be elicited from the existing dictionaries. The priority is given to semantic features influencing lexical or grammatical co-occurrence restrictions. Possibilities are discussed of predicting selectional restrictions on the basis of semantic features of a word in the lexicon.

1. LEXICAL DATABASE OF THE SYSTEM

LEXICOGRAPHER is an expert system designed, in the first place, for the purposes of natural language processing. The work on the project is being conducted by a group of researchers including E.Belorusova, S.Krylov, E.Rakhilina, M.Filipenko e.a.; cf. Paducheva,Rakhilina 1989.

The system consists of two basic components:
— lexical database (LBD);
— bibliographical database (BBD).

LBD is a vocabulary presented in a machine readable form and consisting of several domaines, as in a usual relational database. The user may get information about morphology, syntactic features, semantic features, prosody and referential features of individual lexical items. Among the semantic features that are included or must be included in the database there are such features as: [+Speech act verb], [+Performative verb], [+Verb of motion], [+Kinship term], [+Part of the body], [+Person (as opposed to a physical body)], [+Parameter], etc., all in all — several dozen features.

Programs now existing give the following options:
— marking the vocabulary by a feature presented as a list of words;
— compiling lists of words possessing a common feature or a set of features named;
— compiling lists of documents containing information about the lexeme named, as well as about a given feature or a set of features;
— compiling alphabetical lists of words and features mentioned in the BBD; such lists may play, for the user, the role of a catalog representing the running state of the BBD.

In the BBD for every lexeme or semantic feature all the documents are mentioned that contain some lexicographically useful information about that lexeme or feature. In contradistinction to all the existing bibliographic catalogs, our BBD contains bibliographic information about individual lexemes, cf.Krylov 1989.

The vocabulary consists of some 12.500 words. Morphological information is taken from the dictionary Zalizniak 1977.

As for syntactic and semantic information, usually it cannot be found in existing dictionaries.

2. SEMANTIC FEATURE ACCORDING TO U.WEINREICH

Semantic features are the main subject of the present paper. The notion of semantic feature is associated, in the first place, with the name of U.Weinreich (1967), who proposed a useful distinction between a paradigmatic semantic feature (approximately as in componential analysis , cf.Bendix 1965) and a transfer feature. This distinction made it possible to use the notion of semantic feature in a broader sense than in transformational grammar (TG) where semantic features are strictly opposed to syntactic ones, namely, to selection features and to features of strict categorization: in TG semantic features do not take part in formulation of grammatical rules.

In Weinreich's conception semantic feature serves several different purposes:
1) it is regarded as a basis of semantic agreement (as in well known examples *pretty girl* vs. **pretty man; a year ago* vs. **a house ago; before breakfast* vs. **before John* etc.);

2) it explains deviant and metaphorical readings (as in *a grief ago, before the wall* etc.);

3) it adds provisional semantic contents to a potentially ambiguous word in order to impose semantic agreement where strictly speaking (i.e. under literal interpretation) there is none; thus, in example (1), from Вейнрейх 1981, p.159, the word house, with the inherent semantic feature [-Time], acquires in the given context feature [+Time] as a transfer feature imposed by the governing verb occur; as a result, the word house is interpreted as an event, e.g., as an event of somebody's perception of a house while passing by: (1) A red house occurred twice.

In example (2), from Вейнрейх 1981, p.159, the word craft acquires the feature [+Aircraft], as a transfer feature imposed by the verb to fly:

(2) They flew the craft.

Thus, following Weinreich, we divide semantic features into two groups. Categorial feature of a word (usually, of a noun) is understood as its own characteristics, possibly, as a common property of its referents; cf. such features as [+Person] or [+Place] of the words man and house correspondingly. Transitive feature of a word (usually, of a predicate) is a semantic condition imposed on one of the arguments - namely, on the semantic necessity of its presence in the utterance with the given word and on its categorial features. Thus, verbs of emotional state, such as to hope, possess the following transitive feature: their subject (necessarily existent) has a categorial feature [+Person]. Verbs of motion must have at least one argument with the categorial feature [+Place], etc.

3. SEMANTIC FEATURES IN SYSTEMS OF NATURAL LANGUAGE PROCESSING (NLP)

Semantic features belong to obviously significant NLP resources having no equivalents in existing dictionaries. The following problems of NLP may be listed where semantic features are constantly made use of:

1. Revealing predicate-argument relations in parsing algorithms: categorial features of the argument should agree with the transitive feature predicted for this argument by the predicate. As is pointed out in Апресян ed al. 1989, p.261, "in many cases adequate identification of a syntactic construction relics upon semantic agreement of words" Thus, semantic features can make a substantial contribution in syntactic parsing. 2. Disambiguation of a lexically homonymous predicate word: categorial feature of an argument may help in choosing the right lexical meaning of the predicate; cf. (3) a.Окна гостиницы выходят на юг;

b. Петя выходит на лужайку.

In (a) the word гостиница 'hotel' has a categorial feature [-Movable]; hence the stative meaning of the verb выходить 'go out'. In (b) мальчик 'boy' has the feature [+Movable] and the verb выходить has its usual meaning of a verb of motion.

3. Disambiguation of a lexically homonymous noun by addressing transfer features of the predicate. Thus, semantic features are usable for disambiguation of words in context.

4. Combinability of verbs with adverbials designating time, place, reason, purpose, instrument etc., always rely upon some sort of semantic concord, cf. Paducheva, Rakhilina 1990. E.g., the adverbial of purpose is only possible in the context of a verb denoting controlled action and, consequently, having an agent endowed with free will. If this condition is not fulfilled, the adverbial of purpose sounds deviant (cf. *Для оплаты проезда у водителя имеются в продаже абонементные книжечки: the adverbial of purpose is out of place here because имеются в продаже does not denote an action). The time adverbials denoting exact time (Падучева 1988а) are excluded, on semantic grounds, in the context of such non-action verbs as опоздать,отстать, затянуться <о докладе>, сохраниться etc.

5. In the course of analysis of coordinate constructions it is often necessary to carry out a transformation opposite to conjunction reduction, and semantic agreement is what gives a hint as to how this transformation is to be fulfilled.

6. Semantic features may be useful in the procedure of revealing anaphoric relations in the text, cf. example from Dahlgren, McDowell 1986: (4) The cat did not drink the milk. It spilled.

As the verb *to spill* presupposes a subject which is a liquid, the pronoun may be unambiguously associated with *the milk* and not with *the cat*.

7. Transfer semantic features may be used to distinguish texts allowing for literal interpretations from deviant or metaphoric (as in *the sea smiled*).

4. SEMANTIC FEATURES AND SELECTION RESTRICTIONS IN LEXICON AND GRAMMAR

In early 60-ies semantic features were almost unique theoretical instrument of semantic analysis. A progress in semantic theory achieved in the 70-ies and in the 80-ies (in the first place in works of Ju.Apresjan (1974) and A.Wierzbicka (1972), connected in the first place with semantic decomposition of lexical meanings, drew the notion of semantic feature aside, to a secondary and a more modest position. Semantic features were regarded at best as a subsidiary means in systems of NLP, cf. Апресян и др. 1989. Now I argue that the notion of semantic feature deserves a more prominent place, even in the context of modern intricate "garden variety" semantics.

The fact is that in many cases semantic features can be interpreted as a label for one or more semantic components in the semantic decomposition of a lexeme.

I am inclined to think that it is the semantic feature and not the syntactic one that plays the leading role in regulating selection restrictions in lexicon and grammar.

Anna Wierzbicka in her book "Semantics of grammar" takes an ambitious task - to present all selection restrictions in grammar as motivated by some semantic features of words and constructions: "grammatical distinctions are motivated <...>by semantic distinctions" (Wierzbicka 1988, p.3). While not wholly agreeing with this thesis, we can add some arguments in its favor. Here are several examples of selection restrictions that are usually considered to be purely syntactic, i.e. demanding lexical lists, and which can be proved to be semantically motivated, namely, motivated by some semantic component of a word or of a grammatical construction.

EXAMPLE 1. In Зализняк, Падучева 1987 a semantic characterization was proposed for the class of predicates allowing Neg-Raising. Such Neg-Raising predicates as *to believe* <*that*> possess two semantic features: [+Incompatibility of contraries] (*you cannot believe* that P and simultaneously *believe* that not-P, though, e.g., you can *assume* that P and simultaneously assume that not-P) and [+Excluded neutrality] (*I do not think that P* is out of place in the context when I never gave it a thought - whether P or not-P).

EXAMPLE 2. In Арутюнова 1988 it was shown that Russian conjunctions что 'that' and как 'as' obey the followingrule of semantic distribution: что is used after verbs with the semantic component 'know/believe' (cf, similar considerations about English that in Wierzbicka 1988) and как - after words with the component 'perceive', cf. Я помню, что мы там купались and Я помню, как мы там купались.

EXAMPLE 3. In Падучева 1988b the semantic invariant is revealed for the class of predicates capable of introducing indirect question or its equivalent - parameter word; cf. I *know* why he arrived; I *know* the reason of his arrival, on the one hand, and *I believe why he came, *I believe* the reason of his arrival - on the other (this problem was stated in Vendler 1981). It is the semantic component 'X knows' that is responsible for this semantic option.

5. ON SEMANTIC INVARIANT OF THE CLASS OF WORDS WITH GENITIVE SUBJECT

Our main object of attention in this paper is theconstruction with genitive subject in Russian : Ответа не пришло, Мороза не чувствуется, Катастрофы не произошло. Note that in some cases nominative is also possible: Ответ не пришел, Мороз не чувствовался.

In Apresjan 1985 it is claimed that the choice of the case of the subject in this construction is determined by a syntactic feature of a verb, and that this syntactic feature must be ascribed to the corresponding group of verbs (cf. произойти); verb forms - mainly, passive forms (cf. наблюдаться, чувствоваться) or predicatives (cf. видно, слышно) in the dictionary. The list is supposed tocontain more than two hundred items. These words, as Apresjan

believes, possess some semantic affinity, but this affinity is not sufficient for reliable prediction of the case of the subject: the list of words is supposed to be the only thing that is necessary and sufficient.

This thesis is demonstrated by the following differences in syntactic behavior of semantically cognate pairs of verbs:

(5) a. Старосты на собрании не было;

b. *Старосты на собрании не присутствовало.

(6) a. Поворота в постановке технической пропаганды не наступило;

b. *Поворота в постановке технической пропаганды не началось.

(7) a. Симптомов болезни не появилось;

b. *Симптомов болезни не исчезло.

We claim that different choice of the case of the subject in these examples has a semantic explanation. Verbs that can be used with genitive subject will be called **genitive verbs**. Now we claim that the set of genitive verbs (more precisely, the set of meanings these verbs have when used with a genitive subject) has a semantic invariant.

There are two semantic components, different but cognate, such that at least one of them is always present in every negative sentence with the genitive subject construction. Correspondingly, there are two semantic groups of genitive verbs. In **group I** genitive subject in a negative sentence is explained by the fact that the corresponding sentence without negation contains a semantic component 'X exists',where X stands for the referent of the subject NP (or 'X takes place' - if the subject NP does not denote any object but rather a process or an event). An important condition is that this essential component should have - in the semantic representation of a sentence - the status of an **assertion** or an **implication** (according to Karttunen 1973): it must not have the status of a presupposition. Now, if the semantic representation of the non-negative sentence contains a proposition of the form 'X exists' and if this proposition does not have the status of a presupposition of this sentence,then under negation (of the whole sentence) proposition 'X exists' will be negated. It is exactly this semantic component - negation of existence of X, - that is "responsible", in the 1-st group of genitive verbs, for the genitive subject.

If the meaning of the verb does not predict unambiguously whether the presupposition of existence must or must not be present in the semantic representation of a sentence then both genitive and nominaЛive subjects are possible: negative sentences with the nominative and with the genitive subject will have different meanings:

(8) a.](Ответ пришел) = Ответ не пришел (the existence of the answer is presupposed);

b.](Пришел ответ) = Ответа не пришло (the existence of the answer is not presupposed).

In examples (9)-(11), where only genitive subject is possible in a negative sentence, proposition 'X exists' cannot have the status of presupposition - it is always an implication; thus, the case of the subject is genitive:

(9) a. Осадка не выпало;

b. *Осадок не выпал.

(10) a. Затруднений не возникло;

b. *Затруднения не возникли.

(11) a. Разницы не усматривается;

b. *Разница не усматривается.

Nominative is only possible as an "aggressive" stylistic variant corresponding to a newly born norm, as in

(12) Договоренность не достигнута.

It is much better to say Договоренности не достигнуто.

In group II genitive subject is predicted by a component 'X is present in the field of vision of an observer'. When negated, this component has the following form: 'X is not present in the field of vision of the observer'. It is this component that is responsible for the genitive subject in the second group of genitive predicates.

On the contrary, Nominative case of the subject in the context of verb of group II expresses the presupposition that the object is present at the place mentioned, but is not accessible to perception:

(13) a.|(Деревня видна) = Деревня не видна.

б.|(Видна деревня) = Деревни не видно.

Semantic components responsible for the genitive subject in group I and in group II are cognate. In fact, proposition 'X is not present in the field of vision of the observer' often has a conversational implicature - 'And I doubt whether X exists at all'. In other words, the absence of the object in the field of vision casts doubt on the very fact of its existence. This implicature impends itself if the subject can only occupy the place that the speaker has in mind. Then if the subject is not perceived in this place it does not exist at all, as in Мороза не чувствуется. For persons who can occupy different places, the problem does not arise. Thus when stating

(14) Маши не видно <здесь>

the speaker does not call in question the existence of Masha.

Thus, our semantic invariant of the class of sentences with a genitive subject makes it possible to characterize semantically the class of genitive verbs; moreover, this invariant makes it possible to state conditions (on sentence structure) under which genitive subject is excluded, inspite of the fact that the verb belongs to the class of genitive verbs. Thus, we get explanation of the role of such factors (mentioned in Babby 1980) as

— animate vs. inanimate subject;

— referentiality vs. non-referentiality of the subject;

— topic-focus articulation of the implied non-negative sentence;

— presence vs. absence of the observer.

To recapitulate, our example shows that there is, though indirect, connection between selectional restrictions and semantic features of the word, i.e. semantic components of its semantic decomposition.

REFERENCES

Апресян Ю.Д. Лексическая семантика. М.: Наука, 1974.

Апресян Ю.Д., Богуславский И.М., Иомдин Л.Л. и др. Лингвистическое обеспечение системы ЭТАП-2. М.: Наука, 1989.

Арутюнова Н.Д. Факт, событие, оценка. М.: Наука, 1988.

Вейнрейх У. Опыт семантической теории. - В кн.: Новое в зарубежной лингвистике, вып. Ч. М.: Прогресс, 1981, с.50-176.

Зализняк А.А. Грамматический словарь русского языка. Словоизменение. М., 1977.

Зализняк Анна, Падучева Е.В. Под'ем отрицания: синтаксис, семантика или прагматика? - НТИ, сер.2, 1987.

Крылов С.А. О структуре документальной ИПС по русской лексикологии. - В: Тезисы конференции по машинному фонду русского языка. М., 1989.

Падучева Е.В. К семантической классификации временных детерминантов предложения. - В кн.: Язык: система и функционирование. М.: Наука, 1988а.

Падучева Е.В. Выводима ли способность подчинять косвенный вопрос из семантики слова. - В кн.: Знание и мнение. М.: Наука, 1988b, с.33-46.

Bendix E.H. Componential analysis of general vocabulary. The Hague, 1966.

Dahlgren K., McDowell Y. Kind types in knowledge representation. - In: 11-th International conference on Computational linguistics (COLING-1986). Proceedings. Bonn, University of Bonn, 1986, p. 216-221.

Karttunen L. La logique des constructions anglaises a complement predicatif. - Languages, 1973, N 30, p. 36-80.] Paducheva E.V., Rakhilina E.V. Predicting Co-Occurrence Restrictions by using Semantic Classifications in the Lexicon. - In: COLING-90. Papers presented to the 13-th International Conference on Computational linguistics. V. 3. Helsinki, 1990.

Vendler Z. Telling the facts. - In: Speech act theory and pragmatics. Dordrecht: Reidel, 1981.

Weinreich U. Explorations in semantic theory. - In: T.A. Sebeok (ed.). Current trends in linguistics, III. The Hague: Mouton, 1966, p. 395-477.

Wierzbicka A. Semantic primitives. - Frankfurt/M. Athenaum, 1972.

Wierzbicka A. The semantics of grammar: Amsterdam, etc.: John Benjamins, 1988.