

**PREDICTING CO-OCCURRENCE RESTRICTIONS
BY USING SEMANTIC CLASSIFICATIONS IN THE LEXICON**

Elena V. PADUCHEVA and Ekaterina V. RAKHILINA

Institute of Scientific
and Technical Information,
Academy of Sciences of the USSR
Usievicha 20a 125315 Moscow

A b s t r a c t

In this paper we investigate general principles of constructing semantic classifications that yield useful predictions concerning combinatory options of words. Several semantic classes of Russian words are discussed, implemented in an expert system named "Lexicographer", the

"Lexicographer" is supposed to provide its users with all kind of information concerning some 15.000 most common Russian words. Along with morphological, syntactic and semantic information usually stored in conventional dictionaries, the system should contain information about referential characteristics of words and about restraints in combinability with other words in syntactic constructions of different types. In its final version "Lexicographer" should provide the users with all sorts of bibliographical information

system being conceived as an aid both in the area of natural language processing and in traditional lexicography. Semantic features proposed regulate co-occurrence of verbs with their non-obligatory dependents - such as Modifiers of place or time; Instrumental and Benefactive objects and the like.

(concerning both individual words and semantic classes of words) and with concordances made on the basis of a sufficiently representative corpus of Russian texts.

One of the basic components of the system is its lexicon; the lexicon contains information not only about individual lexemes, but also about semantic and syntactic classes of lexemes. Thus, for nominal lexemes such features are given as: "NATURAL CLASS", "ARTEFACT", "MASS TERM", "SET", "BODY PART" and the like.

For predicates the lexicon gives the values of such

features as Vendlerian classes (STATE, ACTIVITY, ACCOMPLISHMENT, ACHIEVEMENT), cf. [1]; CONTROLLED vs. NONCONTROLLED action [2, p. 32]; possibility of having a subordinate indirect question [9], etc.

It is argued that in many cases combinability of a word can be predicted on the basis of the semantic class(es) it belongs to. Thus, semantic classifications will be the main topic of our exposition.

It turned out that altogether different approaches are required for semantic classifications of predicates and for lexemes with non predicative meaning. In this paper we give central attention to the predicative zone of the lexicon.

Semantics of verbs has been studied intensely in recent years, cf. works by Ch. Fillmore [3], Z. Vendler [1], L. Karttunen [3]. It was a long tradition for linguists to lay the strongest emphasis on the individual semantic properties of separate verbs. Our goal in this paper will be to attract attention to those semantic properties of verbs that affect their combinability and still are common to more or less numerous verb classes.

Thus, combinability of verbs with subordinated noun phrases or adverbials designating place, time, reason, purpose,

instrument and other parameters of the situation denoted by a predicate (taken together with its arguments) was usually assumed to be non-restricted. This is definitely not the case, and the problem deserves special attention. There are some well known important observations made in this area. Thus, as is known, one of the Vendlerian classes - accomplishment verbs - is singled out by the fact that some verbs combine with a special kind of time adverbials, such as *in two hours*; these adverbials designate a period of time during which the process went on and finally reached its inherent limit.

In what follows we shall demonstrate several semantic classes of Russian verbs with the corresponding rules of semantic co-occurrence restraints. We argue that the restraints discussed are semantic in their nature; thus, it is natural to suppose, that some analogous or cognate restraints would hold for corresponding semantic classes of verbs in other languages. Cf. the look by A. Wierzhbicka [8] where the author strives for semantic explanations in grammar in a broad sense of the word, namely, tries to find semantic motivation for every co-occurrence restriction in the choice of morphemes or grammatical constructions.

1. There are classes of

predicates for which neither location in time nor in space is possible.

One of such classes is constituted by the so-called **stable states**, cf. *весить* 'to weigh', *владеть* 'to possess', *знать* 'to know', *надеяться* 'to hope', *помнить* 'to remember' etc. (Stable states are opposed to the so-called **temporary states** cf. such predicates as *веселиться* 'to make fun', *быть голодным* 'to be hungry', *пьяным* 'drunk' etc for which localization in time is possible. Another class with the same combinatorial properties is constituted by causative verbs denoting situations that give states as their results, cf. *радовать* 'to cause joy', *огорчать* 'to grieve', *рекомендовать* 'to recommend', *злить* 'to make angry' etc. Note especially verbs denoting changes in social state such as *овдоветь* 'to become a widow'.

Verbs of these semantic classes do not allow of place or time adverbials (of all time adverbials only those cooccur with such verbs that denote super-long intervals, such as *раньше* 'before', *в молодости* 'in the youth', *когда-то* 'long ago', cf. [5]).

In order to give semantic explanation of co-occurrence restraints of a verb it is sometimes necessary to use a more exact classification of time modifiers. Thus, the verb

опоздать 'to be late' does not co-occur with an adverbial expressing exact time (**Он опоздал в 14.30*), though for other Perfective verbs in Russian such modification usually is not precluded (*Он пришел в 14.30*), cf. differentiation of exact time and embracing time in [5].

The verb *опоздать* is not unique in its semantic class, cf. such verbs as *успеть* 'to be in time', *отстать*, *затянуться* (*Доклад затянулся*), *задержаться*, *сохраниться* and some others.

2. There is an interesting distinction that has remained unnoticed up till now, namely, a distinction between concrete and non-concrete verbs. A verb is **concrete** if it designates (in a proper context) a situation that is supposed to go through a series of phases in its natural development each of which can be given a detailed description. Indeed, for a situation described by a verb *бежать* 'to run' or *резать* 'to cut' (both of them concrete) it is always possible to say, for every fixed moment, what the subject involved in the situation is doing at this moment. On the other hand, there are verbs that present the situation in such a way that they emphasize the final result of the action described while the means used in order to achieve this result remain in the shadow; and this semantic feature characteristic of their use and understanding

must be reflected in their lexicographic representation. We call such verbs **non-concrete**. Thus, *ОТМСТИТЬ* means 'to revenge oneself for an evil committed earlier'; but the means used in order to attain this goal are irrelevant. As a consequence, no concrete phases can be discerned in the action denoted by this verb: the verb *ОТМСТИТЬ* does not express "an unfolding action", even when used in the Imperfective. The same is true for the verb *ИСПОРТИТЬ* 'to spoil', which is also non-concrete: *ИСПОРТИТЬ* means 'to deprive of the possibility of normal functioning'; but the way in which this deprivation is fulfilled is irrelevant for the meaning of the word, so it has no phases, it does not denote a developing process.

Non-concrete verbs very often are aspectually defective; namely, they either have no Imperfective or, being used in the Imperfective, they lack Progressive meaning. Thus, among the verbs that, according to Ju. Апресян [7] belong to the class of the so-called momentary verbs (achievements in Vendler's terminology) and thus lack a Progressive meaning there is a major group of verbs that owe their momentariness to non-concreteness; cf. *побеждать* 'to win', 'to succeed'; *застигать*, *заставать*; *изменять* <родине>; *нарушать правило* 'to violate the rule'; *обходить закон* 'to

violate the law'; *предавать* <товарищам> 'to fail'; *разглашать* <тайну>; *следовать* <совету> 'to follow the advice'; *совершать грех* 'to sin', etc.

(1) а. *бежал с факелом* 'ran with a torch <in his hand>;

б. **Мстил с портфелем* <в руке> 'revenged himself with a bag <in his hand>

(2) а. *разбил часы молотком* 'broke the watch with the hammer'; *ударил палкой* 'hit with a stick';

б. **Испортил* <часы> *молотком* 'spoiled the watch with a hammer'; **оскорбил палкой* 'revenged himself with a stick'.

3. Not every predicate can co-occur with such a common type of verb-complement as Benefactive (expressed in Russian with the help of the proposition *для* 'for'). Thus, it is impossible to say **бояться для кого-либо* 'to be afraid for smb', **клясться для кого-либо* 'to swear for smb' etc. These restrictions are easily deducible from semantic decomposition of the Benefactive construction: *X P для Y* = 'X did P having in mind to provide Y with the result of P'. It follows from this decomposition that a predicate for which this construction is possible must denote an action that is both a controlled one and an action having some definite result.

This is why Benefactive complement cannot co-occur with the verb *бояться*, denoting non-controlled state, or with the verb *клясться*, denoting an action with no result.

It is a well-known fact that verbs denoting non-controlled actions normally do not co-occur with a modifier of purpose; thus, such question as **Зачем идет снег?* **Зачем море шумит?* are deviant (questions about the cause of such states are normal: *Почему идет снег?* *Почему море шумит?*). Thus, if the dictionary contains information about the feature CONTROL, then the

behavior of a verb with respect to a modifier of purpose becomes predictable.

(3) **Почему он знает?*

**Почему он болеет?*

**Почему он молодой?*

(4) *Почему ты стоишь? Почему ты молчишь?*

Predicates in (3) denote states that are called **determinate** in [6]: these states develop in time according to their own laws, and it is impossible for the subject to change his state of his own accord. And this semantic feature explains the difference between (3) and (4).

S u m m a r y

It is customary to think that co-occurrence possibilities are determined only by individual properties of words: dependencies that do not correspond to predicate-argument relations are not semantically

motivated. Our research shows that there are no syntactic relationships without semantic backgrounding, though semantic features of words that regulate co-occurrence are not always easy to discover.

R e f e r e n c e s

1. Vendler Z. Verbs and times. - Philosophical reviews, 1957, v.64, p. 143-160.
2. Dik S. Functional grammar. - Amsterdam etc: NorthHolland P. C., 1978.
3. Karttunen L. La logique des constructions anglaises a complement predicatif. - Languages, 1973, N 30, p. 36-80.
4. Fillmore Ch. How to know whether you're coming or going. - In: Essays on deixis. / Ed. by Rauh G. Tubingen: Narr, 1983. - p.219-227.
5. Wierzbicka A. The semantics of grammar: Amsterdam, etc.: John Benjamins, 1988.
6. Падучева Е. В. К семантической классификации временных детерминантов предложения. - В кн.: Язык: система и функционирование. М.: Наука, 1988.

7. Падучева Е. В. Выводима ли способность подчинять косвенный вопрос из семантики слова? - В кн.: Знание и мнение. М.: Наука, 1988, с. 33-46.
8. Рахилина Е. В. Цель и причина. Вопросы языкознания, 1989, N 6.
9. Апресян Ю. Д. Глаголы моментального действия и перформативы в русском языке. - В кн.: Русистика сегодня. М.: Наука, 1988.