

QUANTIFICATION OF MEANING AND THE COMPUTER

Marie Těšitelová
Czech Language Institute
Czechoslovak Academy of Sciences
Prague
Č.S.S.R.

One of the major tasks of the quantitative semantic analysis is to disclose complex relations of sememes in communication, i.e. on the basis of their associations in the frame of syntactic structures. With the aid of computer it is possible to prepare a corpus of language material giving the possibility to quantify /1./ semantics of syntactic functions, /2./ lexical meanings, /3./ meanings of morphological categories, esp. those of parts of speech, and to create a new type of semantic frequency dictionary.

The assistance of computer in present-day quantitative linguistic studies makes possible to quantify not only language events, but also their relations, not only their static features, but also the dynamic tendencies of language. This refers, in the first place, to the phenomena of lexicon where the application of quantitative methods as well as the modern computer processing have been of a certain tradition; but the computer-aided research is applied on other language levels, too, e.g. in graphemics, phonemics, morphemics, and can be performed also in morphology and syntax. Thus, the contribution of L. Uhlířová a. collaborators, workers of the Dept. of Mathematical Linguistics and Phonetics of the Czech Language Institute, informed about the computer-aided quantification of patterns of simple, compound, and complex sentences and about their close relationship to the morphological and lexical levels.

The results achieved in quantitative analysis stimulate us to take a further step, i.e. to quantify the semantic aspects of language

phenomena. In this stage of analysis the computer is an indispensable instrument of research.

Let us remember in this place the thirties and fifties of this century when the first semantic frequency dictionaries of English were - manually - compiled [1], [2]. They were isolated achievements combining quantitative and qualitative approaches to obtain the frequency of individual lexical meanings.

In the Dept. of Mathematical Linguistics and Phonetics, we dispose with a statistical corpus of texts processed by computer the units of which /words/ have been encoded in context with regard to syntax, morphology and lexicon; the corpus provides an adequate basis for the quantification of the so-called context meaning, i.e. meaning in a certain context [3].

As it is known, a discourse is determined by external and internal linguistic conditions, in other words, it is a function of linguistic and extralinguistic variables. The complex relations between objects and other events of reality are reflected in language by syntagmatic and paradigmatic relations; to disclose their mutual conditioning is one of the major aims of the semantic analysis. The lexicon introducing the sememes within their proper systems is not capable to cope with their complex relations in the communication, and, thus, it is necessary to examine the linking of sememes in the frame of syntactic structure. This is a task which cannot be done without an assistance of computer, especially when the relations are to be quantified. And, at the same time, this is a task which ought to be performed by the quantitative linguistics as a first step to the quantification of meaning which is today at the centre of interest in linguistics, logic and other sciences.

A prerequisite for obtaining good results in the semantic research is a computationally processed corpus of linguistic units /sentence tokens/ in which we may quantify: /1./ semantics of syntactic functions, /2./ lexical meanings by means of which the semantics of sentence elements are expressed, /3./ meanings of morphological categories, especially those of parts of speech, when they prove to be of importance for semantics in points /1./ and /2./.

On this subject we add a few remarks. To point /1./: It is well

known that in spite of a considerable progress in sentence semantics in these last years [4], [5], there is a number of questions remaining unsolved or open. Nevertheless, we are able to quantify semantics on the sentence level applying the principles of L. Tesnière [6]. The approach consists in searching for semantic components corresponding to sentence members in different syntactic functions. This is certainly no easy task, but is possible under the condition that we dispose with a perfectly performed syntactic analysis of whole sentences /simple, compound, and complex/ and larger samples of running texts, as is the sampling used in the Dept. of Mathematical Linguistics for the mentioned corpus of texts. This sort of underlying language material allows us to study the semantics of language units even larger than a sentence - this, however, being a prospect for a future work.

Remarks to point /2./: The study of lexical meanings fixed in explanatory dictionaries shows some interesting aspects, too. The analysis based on their distribution in text and in relation to the respective semantics of syntactic functions gives a new picture of the frequency distribution of different context meanings associated with a word, and, in some cases, even corrects their position with regard to morphological categories.

Remarks to point /3./: Each part of speech, of course, "behaves" in a different way as far as the semantics of syntactic functions and the lexical meanings are concerned; sometimes, the meaning can be significantly influenced also by morphological categories. E.g. the animate and inanimate genders of masculines in Czech concern not only the semantics of the analyzed word, but also its "environment". So for example, the animateness of subject /e.g. agent/, or object /e.g. efficient/, is reflected in the verb /predicate/ and also is connected with certain lexical meanings; similarly, the verb categories /e.g. mood and tense/ often influence the use of lexical meanings as well as the constitution of new meanings.

These assertions will be now exemplified on two fundamental parts of speech, the verb and the noun, which in the quantitative semantic analysis manifest some specific features referring to their part-of-speech properties and syntactic functions.

E.g. the Czech verb in the predicative position represents together with its actants the semantic nucleus of sentence. The actants are formally expressed by the subject, object and/or adverbials as free complements. After the determination of meanings corresponding to the syntactic functions of verb as a further step we ascertain which of the lexical meanings functions most frequently as predicate. The semantics of predicate may be also influenced by the part-of-speech appurtenance of actants assigned to syntactic functions. assigned. It certainly makes a difference when a subject is expressed by a noun or a pronoun, or when it is not expressed at all. There are other important aspects, too, e.g. whether the function of subject /e.g. agent/ is expressed by an animate or an inanimate noun, whether it is a human being, animal or thing, whether the word stands in the singular or the plural, whether it is plurale tantum etc. In the relation of object to predicate we must take into account the lexical meaning of object and its morphological categories, especially case.

Between the verb, subject and object there exists a relation of semantic congruence, i.e. a classificatory agreement of noun and verb /i.e. of the predicate and its actants/. E.g. water, wine flows, not so the table, the earth; we chase the deer, not the people. This principle is rather close to the intention of verb in the sense of E. Pauliny [7].

With the noun in Czech we quantify the semantics corresponding especially to subject, object, attribute and adverbial. Compared to the previous remarks on the semantic quantification of verb we have to do for the most part with an inverse relation; as a point of departure we take now the noun in one of its syntactic functions /with different meanings/ and proceed to the verb syntactically functioning as a predicate. Further steps of analysis are practically the same as with the verb, i.e. the determination of a corresponding lexical meaning of the given noun, of its morphological categories and/or syntactic functions of its "environment." With the adverbial the attention is paid especially to whether it is obligatory or optional [5].

The semantic analysis of the remaining parts of speech in sentence, e.g. formal words /grammar words/, such as prepositions and conjunctions, have some aspects common with verbs or nouns and, in addition,

some special features connected with their status. In any case, the quantification of their semantics in the manner explained shows the functioning of words in text in a new light.

The quantitative linguistics can fulfill the task presented with the aid of computer which creates optimum conditions for the application of quantitative methods in semantics. Within the quantitative microanalysis of individual words in text /in context/ aimed at an identification of context meaning it is possible to obtain new facts concerning the theory of semantics in general and the quantitative semantics in special; the application of results may lead to a new type of semantic frequency dictionary. And this is at present, under my leadership, the principal task of the Dept. of Mathematical Linguistics and Phonetics, in the Czech Language Institute.

REFERENCES:

- [1] Lorge, I. - Thorndike, E.L., A Semantic Count of English Words. /New York 1938/.
- [2] West, M., A General Service List of English Words with Semantic References and a Supplementary Word-List for the Writing of Popular Science and Technology /London 1953/.
- [3] Hannappel, H. - Melenk, H., Alltagssprache. Semantische Grundbegriffe und Analysebeispiele /München 1979/.
- [4] Zimek, R., Sémantická výstavba věty /Praha 1980/.
- [5] Panevová, J., Formy a funkce ve stavbě české věty /Praha 1980/.
- [6] Tesnière, L., Eléments de syntaxe structurale /Paris 1959/.
- [7] Pauliny, E., Štruktúra slovenského slovesa /Bratislava 1943/.