Outline of Semantic Research
for Future MT Development

MICHAEL
ZARECHNAK

Georgetown University.
Washington, D. C.. USA

The underlying intention of the paper is to call for concerted and multifaceted efforts for further improvements in the field of MT. The decisive components in this improvement should and must include various semantically codable layers within the hierarchically-designed overall semantic system. Different existing theories of semantic classification are discussed from the point of view of their application to MT.

The purpose of this research is to survey existing semantic classification theories in order to evaluate them in terms of their applicability to automatic assignment of features and properties for machine dictionary entries. These features are such morpho-syntactic and semantic categories whose values are context-free, while the properties constitute the corresponding context-sensitive values.

In our specific case, in dealing with the Russian input sentence, prior to its English synthesis, we have to automatically parse the Russian sentence. The dictionary will provide the codes for the features and properties as defined above. The set of such Russian categories as gender, number, case, voice, tense, person, mood, and degrees of comparison constitute the area of Russian morpho-syntactic layers. The set of semantic categories such as animate vs. inanimate, solid, liquid, gaseous, fluid (objects); qualifiers, quantifiers (attributes); relators (free form to a free form, bound form to a free form, bound form to a bound form; simple to complex sentences; parts of a sentence to another part of a sentence; anaphoric connectors within the paragraph, including the zero connector) represent the endo- and exostructures with corresponding ranges for specific transformational capabilities as a component within the Russian parser.

Idiomatic expressions are one of the manifestations of the semantic exocentric structures. So are case relations, prepositional phrases, and kernel sentences. All things being equal, a survey of existing semantic classification theories should provide as its final result a set of eclectically selected fragments from various theories (due to the fact that no single theory was designed for MT purposes) with the intention of the disambiguation of the deficient parsing output.

Since there is no complete and satisfactorily developed classificatory semantic theory for machine translation dictionaries and grammar in translating from the Russian scientific text into its English adequate information equivalent, the researcher should develop some general criteria which he could use in classifying the existing theories in terms of their applicability for disambiguation of the 'unwanted' results of the existing Russian MT parsers.

BACKGROUND FOR EVALUATION OF EXISTING SEMANTIC CLASSIFICATION THEORIES

Let us take a single scientific sentence in Russian (with its English translation):

1. V VOZNIKNOVENII ZABOLEVANIJA OCHEVIDNUJU ROL' IGRAL KONTAKT LJUDEJ S GRYZUNAMI.

English gloss: In the emergence of the disease, an apparent role was played by contact of people with rodents.

Let us paraphrase this Russian sentence in various ways:

1.1 Zabolevanija ochevidno *voznikali* ot kontakta ljudej s gryzunami.

Diseases obviously cropped up from contact of people with rodents.

1.2 Zabolevanija ochevidno *vyzyvalis'* kontaktom ljudej s gryzunami.

Diseases obviously were caused by people's contact with rodents.

1.3 Ljudi *zabolevali* ochevidno ot kontakta s gryzunami.

People become sick, obviously, from contact with rodents.

1.4 Gryzuny ochevidno *peredavali z*abolevanija ljudjam pri kontakte.

Rodents obviously transmitted diseases to people through contact.

1.5 Ljudskoj kontakt s gryzunami ochevidno *vyzyvaet* 3ti zabolevanija.

Human contact with rodents evidently has caused these diseases.

1.6 Rol' kontakta s gryzunami pri zabolevanijakh ljudej *ochevidna.*

The role of contact with rodents in human diseases is obvious.

1.7 Vozniknovenie zabolevanij *svjazano s* kontakitom mezdu ljud'mi i gryzunami

Emergence of diseases is connected with contact between people and rodents.

1.8 Gryzuny *prinosjat* ljud'am zabolevanija, soprika-sajas' s nimi.

Rodents bring people diseases by contact with them.

1.9 Soprikosnovenie ljudej s gryzunami *vyzyvaet* zabolevanija u ljudej.

Contact of people with rodents causes human diseases.

1.10 Ljudskije zabolevanija, v chastnosti, *vyzyvajutsja* kontaktom s gryzunami.

Human diseases, in particular, are being caused by contact with rodents.

Denotative Meaning. If one takes the kernel sentence as the basic structure in a language and expresses it in a functional way, then the predicate will be considered as the function proper. This function depends on several supportive, factual statements. Each kernel sentence reflects some extralinguistic situation. Sentence number 1 reflects a situational configuration consisting of human beings, rodents, disease, and the transmission of disease from rodents to human beings. All the above sentence paraphrases have the same reflection capability. This function of the sentence, to reflect some objective, extralinguistic situation, constitutes the denotative meaning of the sentence.

Significative Meaning. Given the specific, unchanging situation, the observer may express himself regarding the observed situation, in varying manners, depending on the facet of the entire situation to which the observer attaches some significance, focus, importance, or emphasis, without changing the identity of the situation as a whole. Thus, if one considers the expressive means for the function proper, i. e. the predicate of the paraphrases of sentence number one, they would include: IGRAT ROL', VOZNIKAT, VYZYVAT', VYZYVAT'SJA, ZABOLEVAT', PEREDAVAT', BYT' OCHEVIDNYM, BYT SVJAZANYM, PRINOSIT', PEREDAVAT', JAVIT'-SJA (PRICHINOJ). Each of these predicate fillers is a two-place predicate structure, calling for certain co-actants within the sentence structure to express the particular focus of attention, and yet preserve the extralinguistic situation as unchanged.

This capacity of the observer to report the event with a focus of his own constitutes the significative meaning of the sentence. Note the following examples:

1.1.1 3ti zabolevanija peredajutsja ljud'am gryzunami.

These diseases are being transmitted to people by rodents.

1.12 Gryzunno-ljudskie kontakty igrajut ochevidnuju rol' v 3tikh zabolevanijakh.

Rodent-human contact plays an evident role in these diseases.

The human translator rendered the Russian sentence under (1) in the following manner:

2. CONTACT BETWEEN PEOPLE AND RODENTS WAS EVIDENTLY RESPONSIBLE FOR THE DISEASE which, if translated back into Russian, would produce one more paraphrase:

2.1 Kontakt mezhdu ljud'mi i gryzunami *javilsja* ochevidnoj *prichinoj* 3tikh zabolevanij.

The contact between people and rodents turned out to be an obvious cause of these diseases.

One could think of many more paraphrases of this sentence.

Syntactic Meaning. An observer can express the same focus via varying morpho-syntactic representations.

Let us take the single focus, VYZYVAT'SJA, i. e., the passive transformation of VYZYVAT' produces the paraphrases 1.2 and 1.10. If we replace the words in these sentences with their morpho-syntactic symbols, we shall see that while the message is the same, the distribution pattern is not the same. Sentence 1.2 will appear as:

$$N_1 \quad D_1 \quad VYZYVALIS'_1 \quad N_2 \quad N_3 \quad P_1 \quad N_4.$$
$$c1 \quad\quad Vcd5 \quad\quad\quad c5 \quad c2 \quad cd5 \quad c5$$

where c — case, c1 — nominative case, c5 — instrumental case, and cd — case determiner.

Sentence 1.10 will appear as:

$$A \quad N, \quad P \quad N, \quad VYZYVAJUTSJA \quad N \quad P \quad N.$$
$$c1 \quad c1 \quad cd6 \quad c6 \quad Vcd5 \quad\quad\quad c5 \quad cd5 \quad c5$$

where Ac1 is a transformation of Nc2 from 1.2.

This ability of the speaker to use both endocentric and exocentric morpho-syntactic tools for depicting an identical focus of the same situation is called the syntactic (structural) meaning of the sentence.

Pragmatic Meaning. In the communicative process, in addition to the message to be transmitted, there are participants (speaker/listener), and the sentence structure might contain the participants' evaluation of either the message itself or the modal attitude toward the participants, including self-evaluation. In the above paraphrases where the word 'ochevidno' 'v chastnosti' or its derivative is used, the pragmatic meaning is present. In 1.10 there is an existential modifier ('v chastnosti'), and in 1.2, assertion of truth ('ochevidno').

This ability of the observer to add his evaluation constitutes the pragmatic meaning of the sentence. In a diagram, these four meanings could be represented by two triangles with a common vertex:
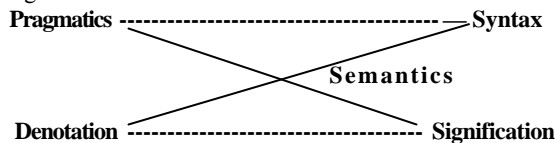
DIRECTION AND SCOPE
OF THE SEMANTIC THEORIES
TO BE EXAMINED

If the above four kinds of meanings are to be efficiently coded in the dictionary for automatic parsing, then the researcher has to seek the corresponding statements in each theory under examination, and select the appropriate fragments from it. The selection could be done against the historical semantic development with the emphasis on the theories containing, in addition to the *naming* semantics (with its various problems), the *combinatory* semantics (both within the system and the structure axes of the sentence, and each component of the sentence).

Classical semantics concentrated essentially on the individual word and the predicate word in the sentence, Here belong the works [1—9].

The main focus of classical semantics was on the processes developing within the individual word or the synonymic series (narrowing, widening, generalisation, specification, improvement, degrading, metaphoric, meto-

13

nymic, radial, chain-like development of meanings of words), and later also on such processes as polysemy, homonymy, synonymy vs. antonymy, meaning and use, usual (regular) and figurative uses, and the like.

Modern semantics concentrates on the sense of the whole sentence(s). Here belong the works [10—16]. In a more technical sense, the modern varieties of the semantic theories are within the orbit of the predicate calculus of the first and second order. In the first, the attention is paid to the predicate function and the obligatory arguments going with it as optimal fillers for the positions within the sentence. Here belong the worl6. [17—20] and others. The main reason for the emergence of modern semantics was the stimuli provided by the logical foundations of mathematics, and the use of computers for machine translation purposes. There developed two main streams of modern semantics: (a) 'Let us think harder', with its resulting work in creation of mathematical linguistics, and (b) 'Let us work harder', with the resulting artificial intelligence.

Language models (Revzin, Markus) and similar works elsewhere try to create the automatic facility for both generating and understanding the sentence, concentrating essentially on the written variety of natural language. Basically, three semantic problems are being considered in these works:

1. The ability of the speaker to express the same thought in many ways;

2. The ability of the listener to recognize varying morpho-syntactic structures as expressions of the same thought; and

3. The ability of the participants in the communication process to draw inferences of a semantic nature, on the basis of the information already perceived.

Those three problems indicate a possible universal semantic language to be discovered, and used for coding of the dictionary entries and the rules for establishing the relations between the elements of the sentence.

OPERATIONAL SEMANTIC MODELS

After the survey of semantic theories is carried out, one has to build the operational semantic model to test whether or not the existing deficiencies in the Russian parser could be disambiguated. Here, one would like to mention that the inclination based on our own experience leads us to believe that the coding system based on Nida's ideas in the area of semantic translation, coupled with the formalisation apparatus parallel with that of W. J. Hutchins [21] and Simmons [22] might secure the further improvement of the Russian-to-English machine translation, information transfer and the naturalness of the English synthesis.

REFERENCES

1. Bréal, M. *Essai de sémantique.* Paris: Hachette, 1911.
2. Darmesteter, A. *La vie des mots étudiée dans leurs significations.* Paris: Delagrave, 1893.
3. Pokrovsky, M. M. *Semasiological studies in the field of ancient languages.* Moscow, .1896 (in Russian).
4. Paul, G. *Principles of language history.* Moscow: Foreign Literature Publishers, 1960 (in Russian).
5. Meillet, A. *Linguistique historique et linguistique générale.* Paris: Champion, 1921.
6 Firth, J. The techniques of semantics. — In: *Papers in linguistics, 1934—1951.* London, 1957.
7. Baldinger, K. *Die Semasiologie. Versuch eines Überblicks.* Berlin: Akademie-Verlag, 1957.
8. Putnam, H. *Some issues in the theory of grammar. Structure of language and its mathematical aspects* New York, 1961.
9. Stern, G. *Meaning and change of meaning.* Bloomington: Indiana Univ. Press, 1964.
10. Masterman, M. Investigation into text semantic structures for the purposes of machine translation with the aid of an intermediate language. — In: *Mathematical linguistics.* Moscow: Mir Publishers, 1964 (in Russian).,
11. Mathesius. V. *Obsahový rozbor současne angličtiny na základě obecně lingvistickém.* Praha: Československá akad. věd, 1961.
12. Zholkovsky, A. et al. On principal use of sense in machine translation. — In: *Machine translation* (Moscow), 1961, No. 2 (in Russian).
13. *Machine translation and applied linguistics (Moscow),* 1964, No. 8.
14. Katz, J. J. and Fodor, J. The structure of a semantic theory. *Language,* 1963, 39, No. 1.
15. Chafe, W. *Meaning and the structure of language.* Chicago — London: The Univ. of Chicago Press, 1971.
16. Leech, G. N. *Towards a semantic description of English.* London: Harlow, 1969.
17. Bierwisch, M. *'Semantics' in Lyons.* 1970.
18. Tesnière, L. *Eléments de syntaxe structurale.* Paris: Klincksieck, 1959.
19. Kholodovich, A. A. On typology of word order. *Philological Sciences,* 1966, No. 3, 3—13 (in Russian).
20. Fillmore, Ch. J. The case for case. — In: Bach, E. and Harms, R. T. (eds). *Universals in linguistic theory.* London: Holt, Rinehart and Winston, 1968.
21. Hutchins, W. J. *The generation of syntactic structures from a semantic base.* Amsterdam — London, 1971.
22. Simmons, R. F. *Semantic networks: their computation and use for understanding English sentences* p. 63—117.