

Session 9: SEMANTIC RESOLUTION

SEMANTIC CLASSIFICATION

Ariadne W. Lukjanow

C-E-I-R, Inc.

The ambiguity in the meaning of words is one of the most difficult problems in machine translation. First let us discuss the problem itself, and then consider possible ways and means for its solution. Actually, we have to contend with two kinds of ambiguity, namely, the multiple meaning of a word or words due to the subject matter or to the particular field of knowledge in which the word is used.

Example:

- bak = 1. "tank, cistern"  
2. "foredeck" (Naval term)
- stol = 1. "table"  
2. "board" (in the sense of food)  
3. "cuisine" (food in a restaurant or hotel)  
4. "department" or "office".

When words of this type are encountered singly without any unambiguous modifiers or other helpful identifiers, such as prepositions, adverbs, verbs, etc., the ambiguity of meaning cannot be solved within the immediate elements of the phrase or sentence. The information on the subject matter of the article or chapter translated can be used as a cue. That is, however, by no means an airtight answer to this type of ambiguity. The word in question can be used within the same paragraph both in its general (literal) meaning, as well as in relation to its specific meaning related to special subject matter.

Example:

- Ia zaplatila za stol = "I paid for the table" ,  
"I paid for the board"
- Matrosy krasili bak = "The sailors painted the foredeck",  
"The sailors painted the tank".

In these sentences there is no way of determining the meaning of the words stol and bak. We cannot give instructions which would permit us to solve the ambiguity of meaning within the boundary of a

Session 9: SEMANTIC RESOLUTION

single sentence. But the human translator has to face the same problem if he or she receives these sentences out of context. The cue to the meaning lies outside the sentence and might be contained elsewhere in the paragraph, or possibly even outside of it.

We are studying this fascinating problem at C-E-I-R, but we have as yet no answer to it and therefore expect a certain amount of inaccuracy in our translation. On the other hand, some of the subject matter ambiguity can be solved by examining the immediate environment of a given word where the microglossary signal can be successfully employed.

For example, when we encounter geograficheskaia karta = "geographical map", we do not speak of "playing cards"; or in kislotnoe obrazovanie = "acid formation", we are not referring to "education"; or in kamennaia baba = "stone image", we do not mean a "stone peasant woman".

We could classify the second type of semantic ambiguity as one of interrelated words, or the influence that one word or several words can have on another word or words within the same sentence.

Example:

Preposition-Noun: iz-za shuma, iz-za doshdia, iz-za buri, iz-za eë, iz-za nego, etc. , where the translation of the preposition is "because of", is not the same as iz-za ugla = "from behind the corner".

Adjective-Noun: universitetsnoe obrazovanie = "university education",  
kislotnoe obrazovanie = "acid formation".

What are the possible ways of resolving these ambiguities?

There are many:

(1) Compilation of special dictionaries with word combinations and instructions to the machine for a special lookup. This process would be time-consuming and expensive.

(2) Give instructions to the machine in the following form: If word iz-za occurs, search for words: shuma, doshdia, buri, etc. , i. e. , with indication of possible location, or number of items removed

## Session 9: SEMANTIC RESOLUTION

from the word under consideration. If the answer is "yes", take translation "x" for iz-za; if the answer is "no", take translation "y".

(3) Mark the words in the dictionary for a special search of identical or similar diacritics, or establish a numerical total of diacritics for positive and negative answers.

(4) Establish a numerical code which, when matched at the boundaries of the logical sequence or working area (phrase), would automatically select the appropriate equivalents by virtue of the equality of the code numbers.

Considering these four approaches, we note that despite their inherent differences, they have one thing in common, namely, they all strive for categorization or classification of words to a varying degree. Thus, we may say that in order to resolve the meaning problem, i. e. , semantic ambiguity for the purpose of MT, it is necessary to establish semantic or meaning classes of words and to determine their environment or the participant members of these classes.

The Unified Transfer System uses the fourth approach, i. e. , it is based on numerical codes and equality of codes for selection. We have established a classification system for the meaning of words in different environments. In other words, we have classified various parts of speech in different environments by their meaning. This has been done in a somewhat unorthodox fashion. For instance, in order to classify nominals, we first consider the possible environment and the other parts of speech that can influence nominals, or be influenced by them. We know that the meaning and translation of prepositions depends on the nominals used with them. There are innumerable nouns and their number would make the task of their classification in relation to prepositions extremely difficult, if not impossible. In contrast, the number of prepositions is rather small. Thus, the number of meaning concepts connected with them is also relatively small, and it is possible to establish the categories of nominals associated with those concepts. On the basis of these categories one can furnish nominal samples as a guide for classification.

Session 9: SEMANTIC RESOLUTION

Examples:

Prepositions:	Nominals:
<u>iz</u> = "of"	(a) <u>chlenov</u> , <u>uchitelej</u> , <u>zhenscin</u> , etc.
<u>mezhdu</u> = "among"	(b) <u>nas</u> , <u>tekh</u> , <u>ehtkh</u> , etc.
<u>sredi</u> = "among"	(c) <u>piati</u> , <u>semi</u> , etc.
<u>u</u> = "among"	

Thus, we can say that we have a class of animate plural nominals in the genitive case (pronouns and collective nouns) and numerals in the genitive, functioning as nouns. We call this group tentatively "class of selection" nominals.

By this method we have thus far established 349 such classes for various parts of speech. We have checked them against 25, 000 words of various texts and the 60, 000-word dictionary of Smirnitskij. In the near future we expect to publish these findings in organized form and to describe our procedure, as well as the actual classes.