

GENERALIZED SYNTACTIC RELATIONS AND SUBSTANTIONAL ATTRIBUTES

Arcady Borkowsky

Computing Center, Academy of Sciences, Moscow, USSR

The paper presents a conceptual framework for natural language analysis within which some experiments were held and some ideas had been developed.

The work concerns the means of translation of natural language text into its meaning representation in form of semantic network basing on frames formalism. The experiments used Russian as input language.

1. The analysis is essentially vocabulary driven. Semantic information is intensively used; indeed, the formalism does not make much difference between grammar and semantics. It could be regarded as a Generalized syntax. The approach leads to distribution of words into classes quite different from usual grammatical classes, but having obviously linguistic meaning.

The basic ideas are related with those of /1/, /2/; the earlier variant of vocabulary structure is given in /3/; the Lisp implementation of vocabulary and semantic network uses property lists with inheritance having much in common with PRL.

2. The language description consist of Semantic and Lexic vocabularies.

2.1. The entries of Semantic Vocabulary are notions, forming an abstract semantic network for meaning representation augmented with grammar information.

The vocabulary article contains the following:

a) a reference to supernotion; all information from supernotion is relevant to actual notion, if it isn't explicitly superceded. The "notion-supernotion" relation imposes hierarchical structure on the set of notions.

b) a list of the notion's attributes with corresponding Generalized Syntactic Relations (GSR). The set of all GSRs forms the grammar used in analysis. The GSR attached to attribute must hold for the words (or phrases) of the NL text, the first (master) referring an instance of the notion, the second (slave) referring to the attribute value. E.g. for Russian, the \$RECEPIENT attribute of \$GIVE would have a GSR demanding the slave to be an instance of \$PERSON and to have the form of Dative case. (The dollar sign is used to distinguish notions from words.)

The attributes inherited from supernotion may also have a specification of default or fixed value, which is immediately inserted into meaning representation.

2.2. The Lexic Vocabulary contains words. Words can be significant or auxiliary. Significant words are those, which name notions, attributes or attributes values. All other words are auxiliary; they are treated as components of analytic morphological forms and are processed by prescan. Therefore only significant words are present in the vocabulary.

Lexic Vocabulary article contains:

a) a reference to Semantic Vocabulary with indication of the role class (notion-name, attribute-name etc., see below) and, optionally, a lexical function (see /1/).

b) grammatical attributes of the word: grammar class, morphological pattern, fixed grammatical values (such as gender for nouns, aspect for verbs, etc.).

3. GSR is a logical function of master's and slave's attributes' values. These can be grammatical attributes from

Lexic Vocabulary or provided by prescan, attributes inherited by slave from supernotion or attributes values reflecting the meaning of the text. The most usual cases involve matching slave's grammatical attributes to some given values or to those of master, and demand for slave to refer a subnotion of a given notion. However some GSRs are more sophisticated.

3.1. The grammar part of GSRs has much in common with Surface Syntactical Relations (SSR) of /1/. Indeed the GSRs had been inspired by SSRs.

GSRs differ from SSRs in two aspects: first, they systematically use semantic information; second, GSRs usually deal with a deeper syntax level; e.g. if the grammar part of GSR postulates a "direct-object relation", its description may cover active and passive verbal, participial and nominal constructions: "to write a letter", "a letter is written", "the letter written by..." and "writting a letter".

3.2. However, the GSR technique allows different ways to describe a fragment of language; all depends on the attributes tested. Making the grammar part of GSRs trivial, one recieves a pure semantic grammar. On the other hand, if the set of notions is the set of grammatical classes, GSR-based analysis turns into traditional syntax analysis.

3.3. When NL text has been transformed into a set of property lists, fetched from the vocabularies and augmented by morphological prescan, there are many ways to order the application of the relevant GSRs, for, while each GSR is described procedurally, the description as a whole is declarative. In our experiments a simplification of the parsing algorithm described in /2/ was used.

4.1. The outlined approach demands a classification of words different from one based on grammatical classes.

Significant words are devided into classes depending on the role they play in naming corresponding notions. We disting-

uish four main classes - N, AV, A and SA.

Class N is the largest; it is comprised by words which name notions, instances of notions and values for some attributes. This class covers most nouns, verbs, verbal adjectives and numerals. It also contains a small but very important subclass of pronouns.

Class AV is formed by words, naming attribute together with its value. This class covers most adjectives and adverbs.

Class A consists of words naming attributes. Usually they are nouns.

Example:

The following words refer the same notion \$FLIGHT in different ways: "to fly", "flight", "flying" just name it and are of the class N; "speed" is an example of class A, it names an attribute of \$FLIGHT; "quick" and "quickly" refer to the same attribute, but provide a value (Magn) for it, these two are the member of class AV.

"Simple stories" use mostly N and AV. The words of class A are common in NL-access to data-bases.

4.2. The fourth class is formed by Substantial Attributes (SA). As a matter of fact, their discussion was the main motivation to write this paper. The separation of it from other classes makes it possible to process as "linguistic" some text relations usually treated as "semantic" and requiring some deductive system to process them.

Substantial Attributes combine the properties of the other three classes. They name attribute, provide its value but focusize on an object - the attribute's value. E.g. "capital" names an attribute of \$COUNTRY, but the focus is nor country neither the attribute-name but an instance of \$CITY.

an attribute has, so to say, its own substantiation
'ly separated from the master.

Indeed, unlike attributes of class A, Substantial Attributes may be used without explicit reference to its master: e.g. "the train goes to the capital". If the master is present, the master-slave relation need not to be expressed syntactically:

"Peter gave the son an apple."

Traditional syntax ignores the possessive link between subject ("Peter") and object ("son") in this phrase; nor does it consider the previous one incomplete.

Our GSR-grammar claims the presence of anaphora in these phrases. The first contains an unresolved reference to some country; in the second one the word "son" includes a reference to some "parent", the most probably - Peter (i.e. the denotat of "Peter").

Definition:

A word is Substantial Attribute if it refers an object by naming its relation to some other object or situation.

In text Substantial Attribute acts as if it were bound by possessive relation to a "virtual pronoun" of appropriate semantic class. For example, these two phrases can be viewed as

"The train goes to the capital [#]of-country"
and

"Peter gave his son an apple"

In the first phrase "[#]of-country" stands for such a "virtual pronoun". In the second one, "virtual pronoun" occasionally turned to be a real one; while in English its use is quite natural, in Russian the use of possessive pronoun would have an emphatic meaning.

The anaphora resolution for such "virtual pronouns" is done in the same way as for real (lexically expressed) pronouns. However, it is possible to take benefit of the fact,

that "virtual pronoun" refers an instance of specified notion, while for real pronoun only grammatical values are known.

Another example:

"President and wife came to capital"

(Articles and pronouns are dropped to reflect Russian),

This phrase is processed as

"President ^{of}-country with ^{his} wife came to capital ^{of}-country".

Indeed the wife is the wife of this president, and they came to the city, which is the capital of the country of which he is the president. To infer this from the phrase no extra-linguistic deduction is needed.

4.3. The experiments have shown, that treating SAs as two words: one referring an object and another a "virtual pronoun", is helpful in analysis oriented on extraction of meaning of the text. In fact it deals with a more general question of the limits between "language knowledge" and "knowledge of world".

References:

1. "A Theory of Linguistic Models MEANING == TEXT", Moscow, 1972 (in Russian).
2. M.C.McCord "Slot Grammars", Am. Journ. of Computational Linguistics, v.6, n.1, 1980, p. 31.
3. A.B.Borkowsky, G.V.Senin "Linguistical analysis of Queiries to Dialogue System", in "Interactive Systems" (the Proceedings of Soviet-Finnish Conference) Moscow, 1979.