БЪЛГАРСКА АКАДЕМИЯ НА НАУКИТЕ . BULGARIAN ACADEMY OF SCIENCES

ПРОБЛЕМИ НА ТЕХНИЧЕСКАТА КИБЕРНЕТИКА И РОБОТИКАТА, 50
PROBLEMS OF ENGINEERING CYBERNETICS AND ROBOTICS, 50

София . 2000 . Sofia

The Association-Cup Model – an Outline

Velina Slavova, Theodor Kujumdjieff

New Bulgarian university, 1635 Sofia

1. Goal

The goal of this project is to create a framework for natural language understanding by means of a model of word associations. In order to achieve speech comprehension in a certain context of conversation we model the flow of verbal information and the "stream of thoughts". Because of the high complexity-degree of such an endeavour an overall description of the relations among conceptual objects (concept) in a sentence on the basis of lexical, morphological, syntactical, semantic and pragmatic information has been targeted.

It has been realised that a sole description of language or thought phenomena can hardly give a sound basis for modelling the processes in the mind, which are taking place in parallel and to a certain extent in competition for the same mental resources. That is why we have chosen a global syncretic approach unifying wherever feasible non-contradicting ideas that seemed to support our goal. Despite the complexity of the model we strove for the utmost simplicity possible.

The proposed framework reflects the structure of two mutually dependent psychological spaces: the space of words (word forms, lexemes) and the space of concepts (ideas, conceptions). The framework models the processes taking place in each of the spaces as well as the connection and influence exchange between them

2. Preliminaries – basic ideas and hypotheses

In developing the framework and the model we have presumed that on one hand language and thought are highly parallel in their functioning. On the other hand the different levels (views)of structuralistic linguistics reflect real interacting processes in the mind. These processes are related with one another which fact is reflected into

the division among the concepts of lexis, morphology, syntax, semantics and pragmatics, and the lack of their precise separation.

Comprehension takes place within a system of knowledge about the world and the language, which, although intuitive by its nature, can be described by means of fairly formalised rules. The context of a communicative situation as well as the listener's personal tendency to create lexical associations - his "associalect" - guides the set of concepts that emerges in his mind during the conversation. This leads to a certain stream of thoughts that reflects the context in which the listeners understands the message. It is natural to expect that the associations as well as the stream of thoughts change their direction in due course and concentrate around some thematic (semantic) focus.

3. Structure of the framework

The structure of the framework is presented on Fig. 1. It consists of four modules: *a graph of word forms*, *a semantic network*, *a parser* and *a semantic analyser*. All these modules are connected through appropriate data structures in order to provide the probabilistic imitation of the comprehension processes modelled by the framework. The functioning of the framework will be elaborated on in the next section. Now let us offer a more detailed description of the main elements of the network.

Graph of word form. The nodes of this graph are the word forms and the edges are the so called connection channels. Every word form is provided with its own identification tag. The graph of word forms is a two way multigraph that has basically to perform several tasks as to:

– connect all the word forms participating in the language description in a structure;

– introduce the associative connections between the segments of the message and some of the other word forms;

– establish the associative expectations of the different word forms in the association cups during the process of comprehension;

– create the stream of thoughts through the association cups;

– activate the semantic network;

– provide the actual message to the syntactic parser.

Semantic network. The (*augmented)* semantic network (Fig. 2) is in principle a slightly modified model of Rummelhard. The nodes are concepts accompanied by their attributes (features). The connections between the concepts reflects partly the description of the world. The main tasks of the semantic network are to:

– establish a hierarchy among the different concepts and their features;

– establish a relation between actions and concepts;

– create a feed-back connection between a concept and the corresponding word forms (the lexeme);

– create secondary associations in the graph of word forms;

– serve as a basis for identification of the thematic (semantic) focus;

– decide upon the semantic roles of the sentence's constituents.

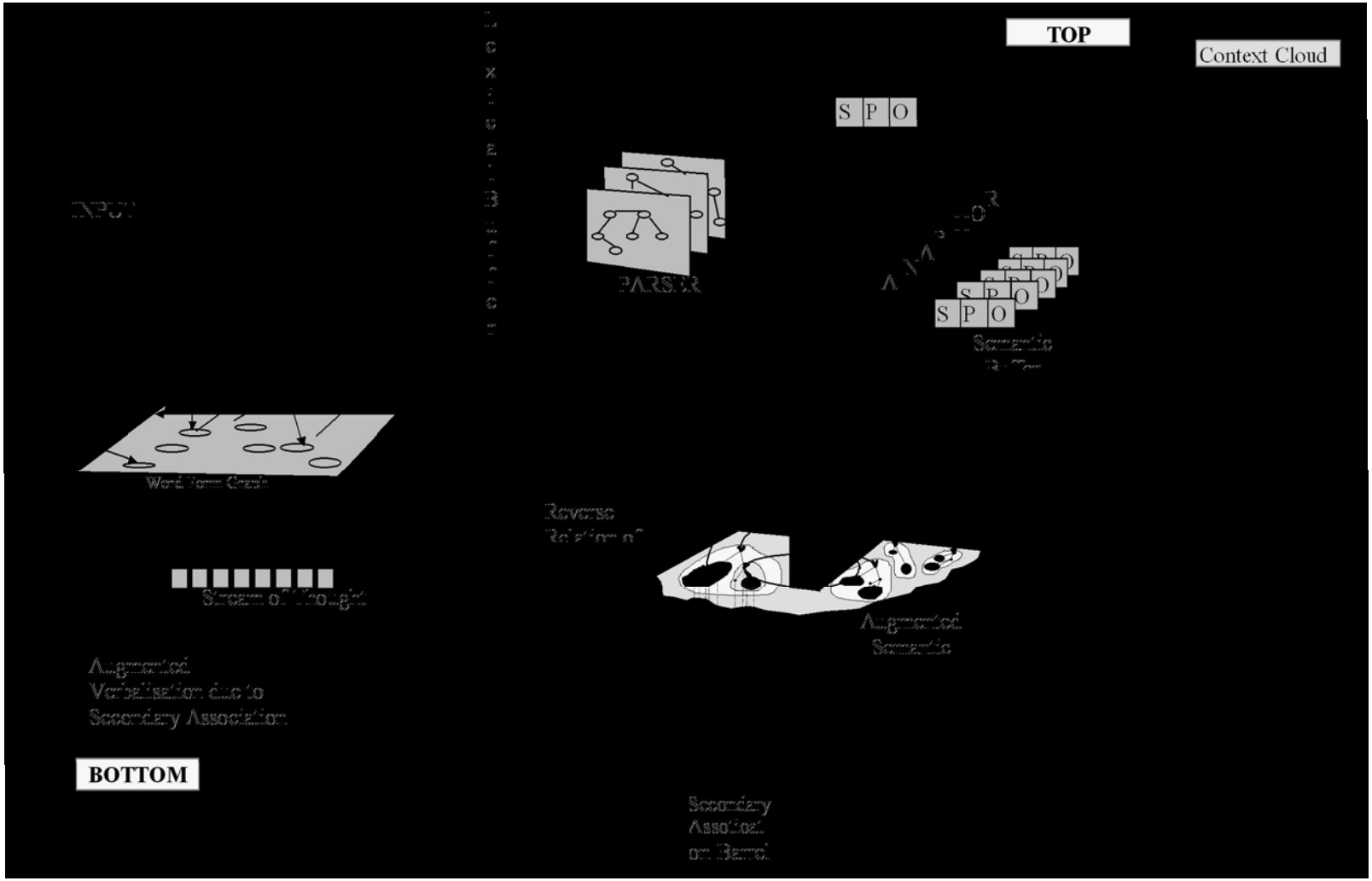

Fig. 1

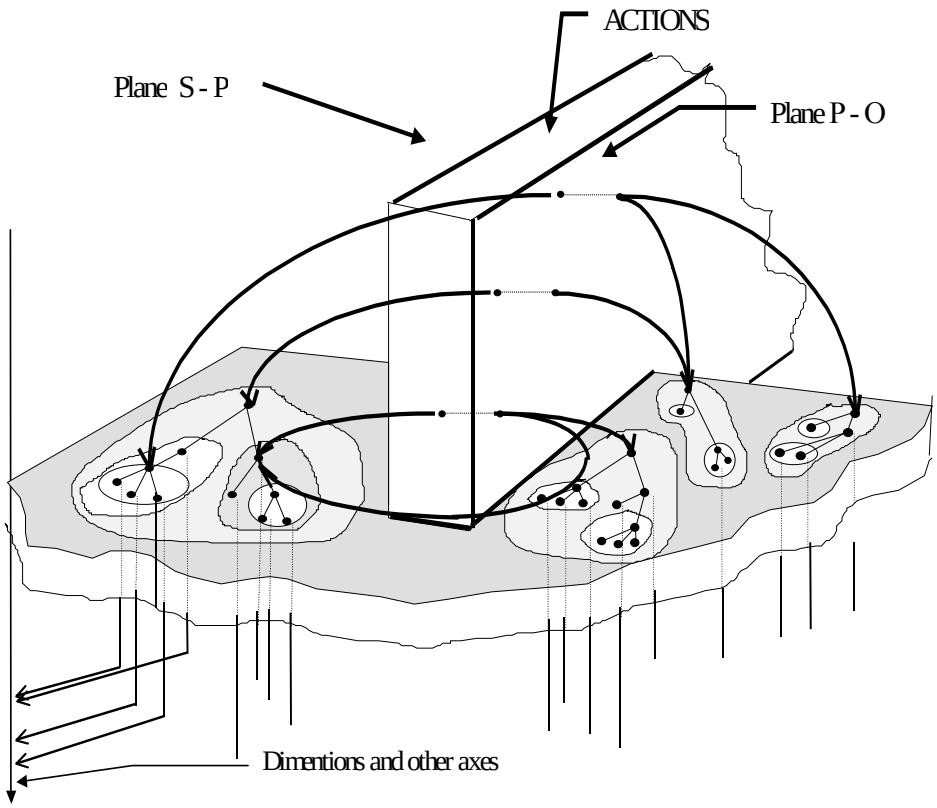

Fig. 2

Parser. The parser serves as a syntactic analyser for the actually received message by the listener. After the parsing process several syntactic trees are created, which are to be used in the consecutive steps by the semantic analyser in order to determine the semantic roles of the constituents.

Semantic analyser. The semantic analyser has to reveal the semantic structure of the sentence (message). By using the information provided by the semantic network - the thematic foci and the different syntactic trees it has to decide about the proper semantic interpretation of the sentence in the particular context.

Generally speaking this framework is a hybrid model unifying the symbolic and the connectionist approach.

4. Main processes and functions

4.1. Global survey of the model's processes and functions

There are several processes that take place in this model. In order to elucidate its functioning we will trace the processing of a message within the framework.

First. After a sentence element from the input (a word form) has reached the Graph of word forms it activates several different word forms via association links between them. For instance a word form like *girl* will increase the expectation of such word forms as *beautiful, smart, nice, sings, rides*, and so on. The expectation of the next word form to arrive at the input raises. A certain level of expectation has been achieved in the association cups (Fig. 3).

Second. The word forms of the message activate the corresponding node in the semantic network with its attributes. For our example *girl* the activated nodes will be *female, human being* and possible attributes would be *soft voiced, sensitive*, and so on. This is the so called *function of concept (conception) identification*.

Third. The feed-back between the semantic network and the activation cups of the word-form graph activate whole lexemes in the graph (in our example all the word forms that are to be derived from *girl* will be activated: *girls, girl's* ...). The *feed back function* is presented as a separate channel pouring its content (calculated weighted numbers) in the activation cup.

Fourth. There is a process of secondary association forming during the sentence understanding. When elements with common features are frequently mentioned, e.g. *Lion, Bear, Wolf*, attributes like *strong, fierce* etc. come to mind. This is modelled by the *distillation process of verbalisation* of features and nodes from the semantic network which takes place in the association barrel presented in Fig. 1. The result of this *verbalisation (lexicalisation) function,* which as all the activation components is also a number, is poured through the secondary activation channel into the activation cup (Fig. 3).

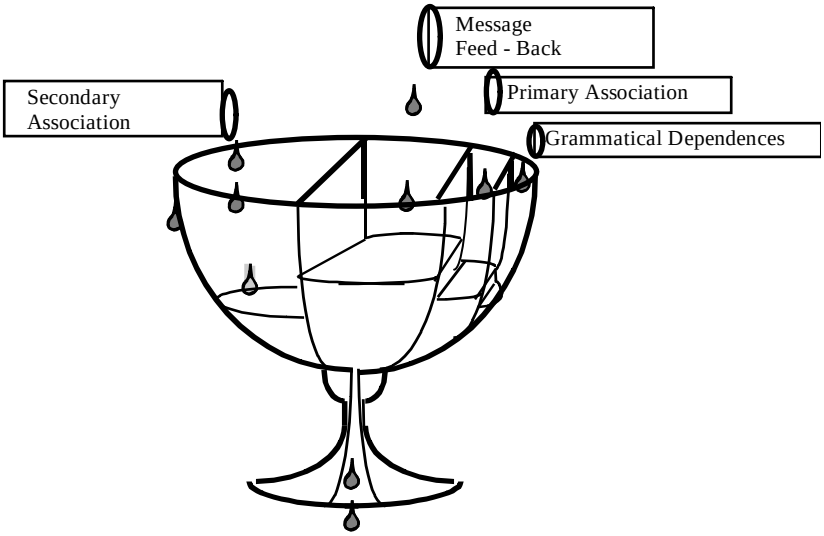

Fig. 3

Fifth. The actually received message (sentence) is being *parsed* by the parser. The set of several syntactic trees is provided for further processing to the semantic analyser. One has to take into account that the syntactic analysis of a sentence or even a phrase is ambiguous in the general case (e.g. *the English history teacher*). In order to resolve the possible ambiguity and to involve the context for the purposes of the inter-

pretation the framework has to perform the next processing functions.

Sixth. The semantic analyser has to use the semantic tree, the stream of thoughts and the identified thematic foci in order to decide upon the semantic roles of the constituents in the syntactic trees, to resolve anaphora relations and to extract the "sense" of the message in a general descriptive way. This *process of disambiguation and* interpretation will eventually lead to a context driven comprehension.

4.2. Set-theoretic representation of the modules, processes and functions in the model

Module – *word-form-graph WG.* This module is represented in the form of a multigraph (fig. 3b), called the graph of word-forms relationships. This graph establishes connections (edges of the graph) in the set of all word-forms WG (Fig. 3a) between the different ***word-forms W*** (vertices of the graph), which represent the lexic of the language under consideration.

Lexeme L (L ⊂ WG)

A lexeme is defined as a subset **L** of the set **WG**, consisting of all word-forms W_L, belonging to the lexeme from the linguistic point of view. Every word-form W belongs to at least one lexeme L.

Set LG of all lexemes

The set **LG** of all lexemes is a set of all subsets L_i (Fig. 4). This set corresponds to the set of wordforms **WG** without being a union of distinct subset, i.e. one and the same element (word-form) W can be a member of different lexemes, the intersection of two arbitrary chosen different lexemes is not necessary the empty set [0].

Representative of a lexeme Wr

Every lexeme has precisely one "representative" word-form **Wr**, the basic grammatical word-form which connects the lexeme with the notion(s) corresponding to it. The representative of the lexeme **Wr** has at least one corresponding element in the set of notions NSet. Each lexeme has exactly one representative of the lexeme **Wr**.

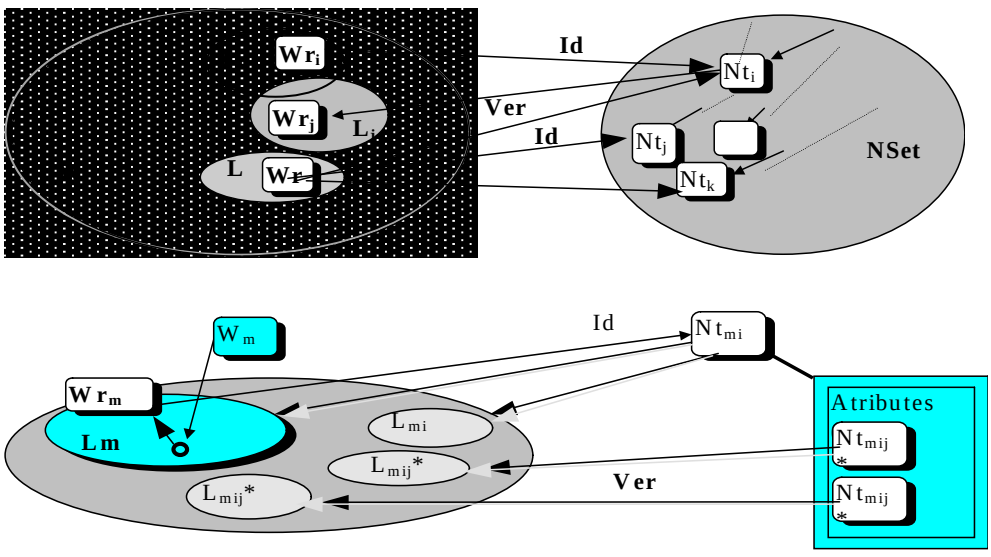

Fig.4

The set of all lexemes **LG** is fully represented by the **set of all the representatives of lexemes WrL.**

The processes and functions of notion identification, lexicalisation of notions and attributes through feed back and secondary association are presented schematically in Fig. 4.

5. The Association-Cup

The idea of the association cup is a crucial concept for this model. In the association cup (Fig. 3) the processes of expectation calculation and interaction between the word form graph, the augmented semantic network and the stream of thoughts take place. The activation numbers, which are calculated on the basis of weighted association (primary and secondary) dependencies between word forms, concepts and attributes, are processed in the activation cup in order to keep the graph of word forms in a certain state of expectancy.

Another main role of the activation cup is to create the stream of thought which is indispensable for analysing the context dependencies and for deciding upon Theme - Rheme relationships within the sentences.

The third function of the activation cup is to model the processes of oblivion by means of slowly decreasing the activation level - a metaphor that has to give credit to the phenomenon of forgetting and focusing the attention of the mind to different topics.

6. Expected results and enhancements

From the presented framework the following results are eventually to be expected:

- A consistent model for comprehension and context dependencies in conversations which reflects natural phenomena in a non-contradicting manner
- A syncretic description of language phenomena, that unites different major concepts in describing language and thought.
- A working tool for imitating association mechanisms and lexical dependencies on a subconscious basis
- An instrument for exploring the semantic, syntactic and lexical relations within different language descriptions. This model should offer the opportunity to explore the adequacy of different lexical, syntactic and semantic formalisms of different natural languages.

Possible enhancement of this model concern a context dependent adaptation of the association description and a knowledge-based system for reasoning within the context of the association as well as modelling of the influence of the subconscious on the comprehension.

References

1. Chomsky, N. Knowledge of Language: Its Nature, Origin and Use. New York, Praeger Publishers, 1986.
2. Chomsky, N. Syntactic Structures. The Hague, Mouton, 1956.
3. Crystal, D. A dictionary of linguistics and phonetics. Oxford, Basil Blackwell Inc., 1991.
4. Dick, S. Functional Grammar. Amsterdam, North-Holland Publishing Company, 1978.

5. D i r v e n, R.. G. R a d d e n. Filmore's Case Grammar: A Reader. Heidelberg, Julius Groos Verlag, 1987.
6. F r e u d, S. Vorlesungen zur Einfuehrung in die Psychoanalyse. Frankfurt am Main, Fischer, 1992.
7. G a z d a r, G., E. K l e i n, G. P u l l u m. Generalized Phrase Structure Grammar. Oxford, Basil Blackwell Publishers, 1985.
8. J a m e s, A. Natural Language Understanding. Menlo Park, California, Benjamins, 1987.
9. K o k i n o v, B. A hybrid model of reasoning by analogy. – In: Analogical Connections. J. Barden, K. Holyoak (Eds.) Norwood, New Jersey, Ablex Publishing, 1994.
10. R a d f o r d, A. Transformational Grammar. Cambridge, Cambridge University Press, 1988.
11. R a d d e n, G. The cognitive approach to natural language. – In: M. Putz (Ed.) Thirty Years of Linguistic Evolution. Amsterdam, Benjamins, 1992.

Модель асссоциативной бокали – обзор

Велина Славова, Теодор Кюмджиев

Новый болгарский университет, 1635 София

(Р е з ю м е)

Задача статьи является формированием гибридной символьной модели иммитационных процессов при обработке речевых сообщений. Модель позволяет создания общей схемы, применимой при построении компьютерных програм для понимания языковых текстов. Схема включает: мультиграф словоформ, семмантическая сеть, грамматический анализатор, модуль смыслового и контекстного анализа. Авторы предлагают и дальнешее развитие предложенной модели.