The Universal Dictionary and Models of Meaning

Ivan KASSABOV, Sofia, Bulgaria

Abstract

The Universal Dictionary is a combination of an Encyclopedic and a Linguistic Dictionary intended to meet the demand for a simultaneous and adequate representation of both the objective-scientific and the linguistic "picture of the world".

In the field of lexicography we can prove the adequacy of the theory by creating a particular model to serve as a link between the ideal theory and the real object, natural language.

In this paper the methodology of acquiring and analyzing a linguistic corpus with reference to the Bulgarian language is presented. The analysis of the linguistic material demonstrates that the Universal Dictionary is representative of the cultural-historical and cultural-linguistic base of natural language.

1 Introduction

When we discuss the possibilities for a complete and, at the same time, concize representation of human knowledge of the world, we end up with the classical model of the encyclopedia as a systematic representation of human cultural history, usually treated from a contemporary point of view. On the other hand, every single natural language construes the physical and the cultural universe from the specific viewpoint of the community that speaks this language. Every language segments the continuum of the world in a specific way and organizes its content part into its own "picture of the world" within the context of the continuos cultural tradition of the respective civilization.

The linguistic "picture of the world" finds its systematic representation in the classical model of the Dictionary. I will adopt the standard view that the Encyclopaedia represents the world objectively and scientifically while the Dictionary represents the world on popular grounds and nands and naive concepts. Apart from this clear distinction between the Dictionary and the Encyclopedia, we should be particularly aware of the fact that a "pure" linguistic definition necessarily contains a partial "encyclopedism". Alternatively, "pure" Encyclopedias necessarily contain a partial "linguisticism", if I can say so. This problem has already been pointed out in view of its practical ramifications for the lexicographic theory, but I believe that it is the theory of semantics that has to provide a practical solution. Moreover, in my opinion, the solution lies in the nature and the way the linguistic sign functions, especially in the nature of the word as a fundamental linguistic sign.

From a more practical point of view, this approach motivates the need for the creation of a Universal Dictionary (i.e. an Encyclopedic one and a Linguistic one), which results from, and at the same time adequately represents, the proper combination of the objective-scientific and the linguistic "picture of the world".

2 Lexico-Semantic Modeling

The view of language as an organized system calls forth the need for constructing a model of the meaningful part of language, i.e. its lexico-semantic system. This model can work by itself, i.e. by computer. Such an approach is best revealed through he basic principle which underlies mathematic modeling: an object can be shown to exist truly if and only if there is an algorithm for its description. In other words, in the field of lexicography we can demonstrate the adequacy of the theory by creating a particular model, which may serve as the connecting pole between the ideal theory and the real object, i.e. natural language.

It is well known that in order for such a model to work, it has to have a finite number of units (no matter how great) and several simple ordering rules. It is also clear that these units have to be structured in such a way as to create a sufficiently rigid system. The almost unlimited number of units in natural language and their complexity requires that they be artificially limited down to the core component of the lexico-semantic system so that the model can meet the necessary and sufficient requirements for preserving the specific properties of the object. This limitation derives from the two objective qualities of the description, namely its completeness and its detailed elaboration. At the same time, the heuristic power of the model, especially when it functions, may compensate for the disadvantages of this limitation. The model shows new regularities and new standard characteristic features of the object, despite the fact that lexical semantics is the most difficult field for modeling.

3 Dictionaries as Lexico-Semantic Models

As a first step towards building up a model for the lexico-semantic field, I have constructed a *Semantics Dictionary-Minimum* [Kassabov 1990] which serves as a theoretical construct representative of the lexical system as a whole, where one can find an illustration of the structural organization of the lexico- semantic fields in a network system that reveals the specific "picture of the world" in Bulgarian.

0n the basis of this theoretical construct and following the same theoretical principles I have created a *Universal (linguistic-encyclopedic) Dictionary* [Kassabov, Simeonov 1999] which comprises 10 000 units or lexical entries. The latter are organized into 80 terminological fields (with 45 sub-fields) taking into account the peculiarities of the respective scientific domain. Each of these terminological fields contains only those terms which are necessary for the adequate representation of the respective scientific domain and are at the same time sufficient for the adequate representation of the encyclopedic knowledge of the world. The terms are arranged in alphabetical order and each of them is defined. Each lexical entry is divided into three parts – a terminological, a common linguistic and a metaphorical one. In the first part the entry is defined as a terminus technicus, in the second part it is defined with its everyday, naivist meaning and in the third part – with its metaphorical meaning (or the general, non- concrete meaning).

As a result of this approach, every word becomes a cross-section of three semantic fields, which are principally different from one another: the terminological-taxonomic, the popular-linguistic and the associative-metaphoric field. These fields are domains that belong to three principally different and co-existent "pictures of the world" in natural language. The scientific-encyclopedic

picture presents the scientific knowledge of the current state of the world perceived as objectively true. The popular-linguistic picture presents the common, naive knowledge of the world, which has been accumulated over the centuries and is socially significant for the present day. The associative-metaphoric picture is in fact a picture of the possible worlds in the domain of the imaginative, the poetic and the mythological. Such a state of the art requires that the semantization and the respective definition of each lexical entry be made to follow each of these three "keys" in accordance with the three respective registers, because the terms, the regular linguistic meanings and the metaphorical figurative meanings enter different types of semantic relations with the respective meanings of the rest of the words, the units of the lexico-semantic system.

4   Semantic or Cultural Primitives?

Examining the language as such a complex system from a global point of view, at least two questions arise:

1. How many sub-systems is this complex system comprised of, what are its levels and what is its hierarchy order as well as what are the units of each of the sub-systems and what is the basic unit for the whole system?

2. What are the basic functions of the system?

To find a satisfactory answer to these questions clearly a third thing has to be found out which would relate the subsystems and functions and thus make it possible to answer the fundamental question: What is the system's basic principle of functioning? This third thing could be either a unit, basic for the complex system, (as for example the word or utterance) or several (minimum number) units, common to the whole complex system, which are to provide for the system's isomorphous (and homomorphous) nature as well as to serve as a switch between the system as a schema and the system as a process and vice versa.

For our purposes we may disregard the well-known and already classical theories of language levels and their units in general linguistics and text linguistics as well as the status of "this frequently condemned but indispensable term" [Benveniste 1974] – the word. We may disregard as well the equally well-known and already classical theories of language functions.

We shall consider only the content plane (disregarding the expression plane), called *general* for reasons of convenience, in the context of semantics (regardless of its different applications).

The results from the analysis of the lexical-semantic (sub)system of language based on our work on two dictionaries assumed to be representative of the system as a whole, *Semantic Dictionary-Minimum* and *Universal Encyclopedic Dictionary,* allow for a new interpretation of the content plane. This interpretation is in the spirit of Greimas' [1966] work, with the difference that it is not in the direction of generation (from the famous "semiotic square") but in the opposite direction, i.e. reducing the whole to an elementary scheme.

The results are the following:

1. The lexical semantics of language as a whole may be reduced to a minimum of about 850 words-sememes distributed in 17 basic lexical semantic fields and organized around a core of about 300 words-sememes (meanings) with two nucleus units: MAN for nouns and BE for verbs. This clearly shows the anthropocentric nature of the system.

2. The combination of the two nucleus units produces the basic elementary predicative phrasal unit belonging to the sentence and the utterance: I AM A MAN. (The next step is: I SAY STH. ABOUT STH.)

3. The analysis of the metalanguage of the description (definitions) of all the items in *Semantic Dictionary-Minimum* shows that:

4. The nucleus verb BE is surrounded by the modal verbs WILL, MUST, MAY; by the verbs of perception PERCEIVE and FEEL, by the cognitive verbs THINK and KNOW, by the verbs of predication and narration SAY and SPEAK, and by the basic verbs expressing action and event in general DO and HAPPEN. All the remaining verbs can be interpreted through the definitions of those nucleus verbs.

5. The metalanguage of the definitions of the nouns is reduced to 7 nucleus units, words functioning as semes, i.e. HERO, ENEMY, ASSISTANT, FIGHT, MEANS, GOAL, SCENE. (A striking coincidence with the actor functions of Propp [1969] examined by Greimas [1966] and others and maybe with Fillmore's [1968] cases.)

6. The adjunctives (adjectives and adverbs) are reduced to the coordinates of space and time – deictics (or Jacobson's [1972] shifters).

The results of the semantic reduction of the lexical-semantic system lead to a principally new interpretation and to the formulation of a hypothesis about the principle inherent in the semantics of language as a system, a principle that is not only grammatical but narrative as well. This principle accounts for the transition from the lexical semantics to the semantics of the utterance in the different modalities and from them to narration and discourse through the "key" of the predication. It is important to explicitly state the fact that the semantics of the utterance as a unit of speech (parole) is inherent in the semantics of the word as a unit of language (langue), and vice versa, through common semantic units "semes" derived by means of an analysis of the metalanguage of the definitions. Thus it becomes clear that the above enumerated units are the real semantic primitives in the sense of Werzbicka's [1980] primitives.

5   Types and Scheme of Communication

The careful analysis of the verbal communication shows that the familiar scheme of Jacobson [1972]: Addresser – . . . – Code . . . – Addressee is a convenient way for the scholars to represent the cybernetic model of information transmission but is insufficiently explanatory regarding its fundamental nature. At least two issues deserve attention:

1. The transmission of information in a system cannot be independent of the organization and management of the information even with respect to cybernetics.

2. The sign character of language presupposes communication of the subject with the environment, then communication of the subject with himself and lastly, communication of the type "addresser – addressee", i.e. the following prerequisites for the language functions are present: representative, cognitive and expressive (or expressive-appellative according to Bühler [1934]). The twofold role of the addresser, namely being both the addresser and creator of discourse as well as the You-person of the addressed, perceiving and interpreting the discourse, finds special language expressions in each utterance (a problem investigated by Benveniste [1974] and Greimas [1966]). In other words, the producer of the utterance is also the author of the utterance and thus plays a double subject role, i.e. his role on the stage is both that of the actor and the stage director. In the long run, except for questions (metalinguistic function) and orders (appellative function) as well as phatic and magic function (according to Jacobson), all the rest that is deemed as language communication represent narrative discourses, i.e. stories about events (scenes) from the point of view of the narrator.

Thus, finally, we may formulate a hypothesis about the stage-likeness (with the acting treated in the widest sense) as a basic principle of the global semantics penetrating from the drama story into each narrative discourse, from the only predication-sentence form of every utterance into the organization of the lexical-semantic system of language which complexity is rooted in the complexity of their units, i.e. the words-signs, as well as in the double semiology (semiotic and semantic according to Benveniste) of their mechanisms in action, as the only possibilities to create all types of texts, from poetic to scientific and encyclopedic.

References

Benveniste, Emile (1974): *Problemes de linguistique generale, Problemes de linguistique generale.*

Bühler, Karl (1934) *Sprachteorie*, Jena,

Fillmore, Charles (1968): "The Case for Case". In: E. Bach, T. Harms, *Universals in Linguistic Theory*, New York, pp. 1-90.

Greimas, Algirdar (1966): *Semantique structurale*, Paris.

Jacobson, Roman (1972): "Shiftery, glagolnye kategoriy I russkiy glagol". In:*Prinzipy typologicheskogo analiza jazykov razlichnogo stroja,* Moskva..

Kassabov, Ivan (1990): *Semantic Dictionary-Minimum*, Sofia, Univ. Press.

Kassabov, Ivan and Simeonov, Krasimir (1999): *Universal Encyclopaedic Dictionary*, vol. I, Sofia, "Svidas" Publ. House,

Propp, Vladimir (1969): Morphologia skazky, Moskva.

Wierzbicka, Anna (1980): *Lingua Mentalis. The Semantics of Natural Language,* Sydney, New York, London.