

The Formal Description Of Armenian Within The UNL System

Liana Hovsepyan
Yerevan State University
Yerevan, Armenia
lshovsepyan@rambler.ru

ABSTRACT

The purpose of this paper is to present the main principles of formal description of Armenian within the UNL system, some peculiarities of grammatical structure of Eastern and Western Armenian and their importance for typological researches.

Keywords

Computational linguistics, formalization, grammatical category, unified system

Significant work has been done in the recent years, starting from the 1960s towards formalization of the Armenian Grammar, but initially the description was fragmentary and not destined to be used in information technologies. Nowadays the problem is wider including not only the formalization of all levels of the Armenian grammatical system but also the more varied usage of the description results for retrieval, storing, modification, and processing information in Armenian in electronic form as well as for cooperating with other formalised description systems in machine readable form of any natural or artificial languages including an interlingua.

Taking into consideration the upper mentioned main problems the initial stage of developing formalized description of Armenian was the elaboration of basic description principles. It's useless to say that each natural language has a complex not homogeneous structure and for formalization it is not an easy object. Armenian also has many complexities first of all connected with variety of forms with one and the same meaning and polysemy of one and the same forms.

In other words, Armenian comes forth with a plenty of variants and variety of conditions of their choosing. Moreover, Armenian has a plenty of grammatical

categories and typological attributes including agglutination, inflectional forms as well as features common to analytical languages (it is remarkable that the approximate number of the main Armenian grammatical attributes includes over 250 units).

To solve these and other kinds of problems special principles were chosen. According to those principles first of all (1) the description object must be the normalized and codified Eastern and/or Western Armenian without dialectal, jargon and out-of-date forms, in other words the standard state of the language. Then (2) the description must be complete i.e. must present all levels of the language structure including exceptions, irregular forms and deviations (within the frames of the same system), (3) the description must be isomorphic and non-contradictory, i.e. different language levels must be presented with the same unitary principles and at last, (4) the description must be economic in the sense that by the help of limited number of rules it will be possible to present greater number of units. Using these principles the automatic synthesis and analysis programs for Armenian morphological system have been created that can work not only on their own but also within the computational languages environment. Those programs have successfully worked within the UNL system.

Nowadays, coming up to UNL system reconstruction problem, some questions on correction of some principles of language description arise. Mostly we'd like to stop on creating unified system and tagset of grammatical and conceptual categories for description of any separate natural languages, as well as for the UNL system. The task is important not only for solving specific practical and applied problems, but also from the point of view of general linguistics,

linguistic typology and universal linguistic model.

From this point of view G.Jahukyan's universal linguistic theory (substantial linguistics) and resp. the universal linguistic model is promising. The latter is "determined as a means of description of any object. The linguistic analysis and synthesis are done by the help of certain attributes selected by the inductive way and checked by the deductive way for all physical and psychical objects. If we explain an attribute in a narrow sense as a feature, then the following can be said: the conception of the feature in the substantial linguistics (according to Jakobson in wider sense the concept of characteristic) relates to the concept of relation (according to Elmslev). It is considered that they are related with reciprocal transference and substitution, i.e. the feature is based on certain relation and the relation on certain feature"¹.

The UNL that has been in use in recent decades² is also based on the relations of language units and is made to recognize automatically the texts content and to do semantic translation. The UNL is an independent system from natural languages; it can not and is not due to depict the structural peculiarities of all languages, but it is based on the general conceptual categories, on universal relations between linguistic units and on semantic characteristics (attributes) of grammatical categories, having itself no special grammatical structure. For a language UNL description the grammatical peculiarities of the natural language must be presented by corresponding conceptual categories, relations and attributes representing in the UNL the phenomena of the objective reality and relations between them, that in this or that language can be expressed by various lexical and grammatical means. However, no matter how independent is, UNL can not ignore natural language peculiarities and besides universalities (or frequentalia) common to all

languages, it must include unique peculiarities that are common to a single natural language (e.g. grammatical categories in Japanese as means of expressing politeness).

Thus, we find it important the combination of the upper mentioned deductive and inductive methods in the further enrichment and selecting of the corresponding conceptual categories, relations and attributes for the UNL.

As clarifying examples it's precise to mention some grammatical peculiarities especially common to Armenian that have not yet been studied completely in the UNL system of attributes.

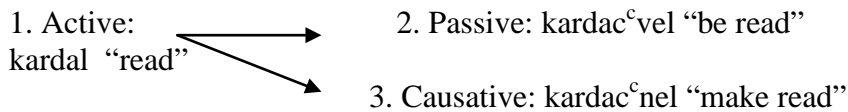
First of all we would like to mention the Causative subcategory within the verbal grammatical category of voice and within more complete limits of transitivity and intransitivity. In Armenian it has special manifestations both in content and expression levels (planes).

If, without details, we generalize and present all the formally expressed meanings of verbal voice in modern Armenian according to their word formation sequence we can confirm the following models of verbal voice multilateral oppositions.

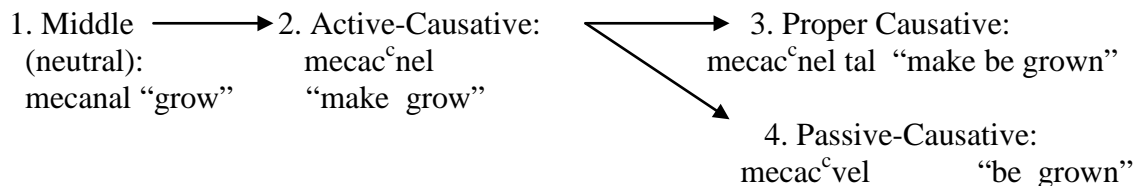
¹ *Г.Б.Джаукян, Универсальная теория языка, Прологомены к субстанциональной лингвистике, М., 1999, с. 32.*

² *Hiroshi Ucida, Meiyang Zhu, Tarcisio G. Della Senta, Universal Networking Language, Printed in Switzerland, 2005.*

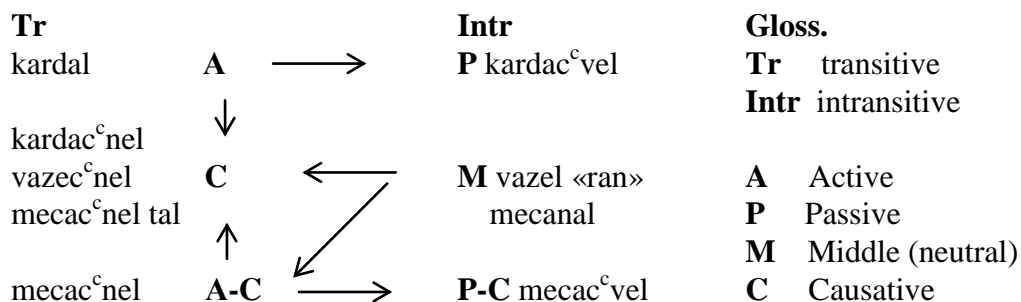
a) Three-member opposition:



b) Four-member opposition:



The general model of voice aspect interrelations, including opposition of transitive–intransitive, takes the following form:



From the typological point of view such behavior of the causative voice must somehow be reflected in the UNL conceptual system.

Of other grammatical realities of Modern Armenian can be mentioned the syncretic manifestations of the nominal category of actualization, where by the same system of articles there can be expressed the meanings of definitiveness, possessiveness, person and demonstrativeness. In deixis category the demonstrativeness has triple manifestation, i.e. it denotes either to the first or to the second persons an object is close or is far from both: ³Ûë «this (close to the first person)», ³Û¹ «that (close to the second person)», ³ÛÝ «that (far from both the first and the second persons)». Moreover, it is common not only to pronouns but also to pronominal adverbs: ³Ûëi»Õ «hier (close to the first person)», ³Û¹i»Õ «there (close to the second person)», ³ÛÝi»Õ «there (far from both the first and the second persons)», etc.

Intresting realities we find in the number category manifestation: pluralia tantum with -enk^c, -onk^c, -ank^c particles which show the

person, his relatives and kinsfolk as a totality. Thus, Vardanank^c denotes the military leader and hero St Vardan and his confederates, anerank^c (aner «father-in-law») – the father in law and his family, etc.

An original means of manifestation of possessiveness is a special category, where we see the quantity of the possessors and the possession (belongings). Moreover, that category has a more regular character in the Western Armenian than in the Eastern Armenian, where it is limited by monosyllabic words. In the Western Armenian there are present some models, where the plurality of both the possessor and possession is considered, independent of the quantity of syllables in the words, which designate the object of possession.

Possessive plurality is expressed by postpositive coalescent possessive articles -s (belonging to the 1st person), -d (belonging to the 2nd person), -n (belonging to the 3rd person), which follows the particle -ni, and formants of plurality -er/-ner according to the model: -er/-ner + ni + -s/-d/-n.

Thus, in Western Armenian the plurality of the possession is expressed: in the presence of the possessor in singular: girk^cers “my books”, namakners “my letters” and in the presence of the possessor in plural: girkcenis “our books”, namaknennis “our letters”.

Singularity of the possession in the presence of the possessors in plural is expressed by adding only the formant -ni and articles -s, -d, -n according to the following model: ni + -s/-d/-n: namaknis “our letter”.

The forms of possessive plurality can have the indirect case forms also, i.e., can be

declined: tunernus “of our houses”, namaknernes “from our letters”, etc.

All these unique manifestations of grammatical categories can take place in the list of UNL attributes or relations.

In addition, we’d like to mention the importance of elaborating of unified terminological and tag systems to evade creating additional programs for transference of natural languages descriptions into UNL terms of which programs we have been depending on up to this day.