

Modeling of derivation in the multilingual expert system
 N. Amirezashvili*, L. Lortkipanidze†, L. Samsonadze‡

Our aim is to create a system for automatic language generation. We present the work aimed at building the **Multilingual Expert System of Language Modeling (MESLM)**. This task requires a considerable amount of knowledge. The core of each kind of human knowledge is a word: “Something central into the mechanism of a language... ” [5]. By means of words we can acquire and transmit the knowledge. But all words have a „linguistic sign as a bilateral entities consisting of two components, one known as the signified, the other as the signifier “[4]. Or in other words, each of word form has its own substance and expression. We can say that morphological knowledge about the word is its own morphological essence.

In the process of structuring the knowledge acquisition system it is necessary to have substances of each elementary component – word. So, the first step for **MESLM** system is accumulation of knowledge about the morphological level of the language. On the other hand, for acquisition of knowledge about morphology of a language we use **Artificial Intelligence (AI)** techniques (such are, for example knowledge-based systems). We come to such a decision because only within the bounds of Expert System it is possible to create the universal structure of multilingual morphology. Moreover, such system solves problems of knowledge acquisition with the help of experts, what, in one’s way, can permit individual approach of the linguist to the grammar of language being under consideration.

MESLM system can be considered as a tool imitating human thoughts and creating morphological rules for any language processor. Although some **AI** techniques have been applied for processing natural languages, it is very difficult to use them, especially, with inflectional languages because special skills are required for application of expert's knowledge. It can’t be said about our system. **MESLM** is used to derive knowledge from source language by the expert without teaching special high-level languages or some formal operators. In contrast to other typical expert systems, here it is possible to acquire knowledge on the basis of sample paradigms, which rely on a personal interview with a domain expert carried out by a knowledge engineer. After the pattern recognition the text will be segmented into words and non-lexical material and the lemma will be determined. The form of each word will be converted as a morphological rule and listed into two lexicons as an entry head (Initial words) and as a rule.

Morphological representation in the **MESLM** system can be compared with well-known Finite-State Morphology [2], or Net representation of morphologic processor [1], or other models which use the net representation for generation. In contrasts to them, on the left-hand (or on the first level, in the paper [2]) by each arc we have sheaf of sets of grammatical categories and on the right-hand (or on the second level, in the paper [2]), we have so-called “Morphological Functions” (**MF**). Briefly, **MF** operators reflect morphological process just needed for to build a desirable word-form with corresponding grammatical characteristics. Important aspects of language generation on inflection level in the **MESLM** system are presented in [3].

Linguists who are not good either at programming or language modelling simply can get a solution through the interactive knowledge based **MESLM** system. System has four main architectural components: the Language Knowledge Base, Knowledge Engineer, the Knowledge Acquisition Module, and the User Interface. In the Language Knowledge Base is represented and stored the necessary information to analyze symbols in the language as a combination of lexical meaning and values of grammatical parameters. Knowledge Engineer contains the resident specific knowledge that is used in the domain to solve linguistically motivated inventory that is taken into account in the phenomena observed in a large number of languages. Knowledge Acquisition Module enables experts to store their knowledge in the knowledge base or expert

* Dpt. of Language and speech systems, Institute of Control Systems, ninomaskh@yahoo.com

† Dpt. of Language and speech systems, Institute of Control Systems, liana@gw.acnet.ge

‡ Dpt. of Language and speech systems, Institute of Control Systems, liasams@yahoo.com

system to deduce new knowledge from existing knowledge through a machine learning process. The User Interface within the bounds of expert system provides the dialogue and the input/output of the knowledge base.

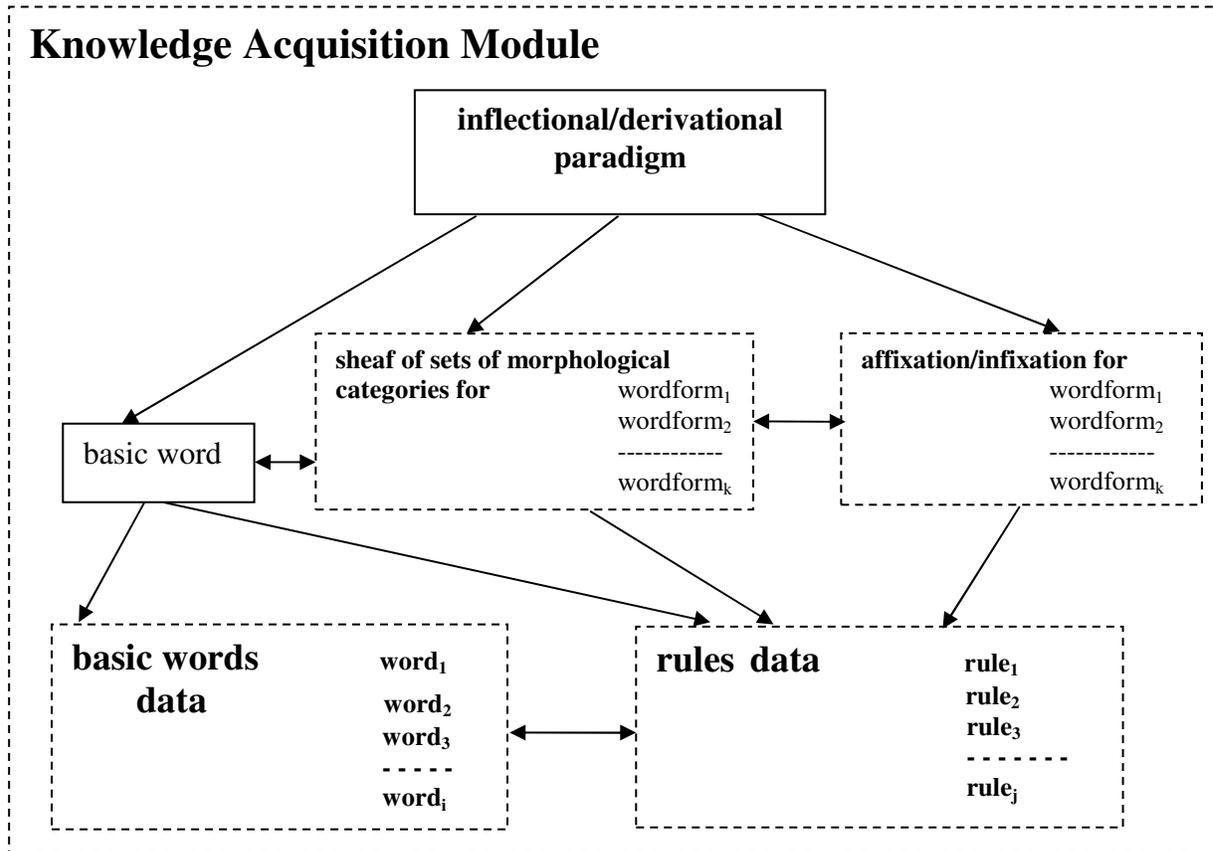


Fig 1. Knowledge Acquisition Module

Knowledge in the MESLM system is represented as the rules. For this in our system we have two kinds of lexicons: rules data and basic words data, (See Fig. 1). In the Knowledge Acquisition Module from inflectional/derivational paradigm for each word-form information about morphological categories and corresponding cases of affixation/inflexion will be captured. After that the rule data and the basic word data will be refreshed.

Acquisition of knowledge by the expert system is provided in accordance with one and the same scheme (See Fig. 2)

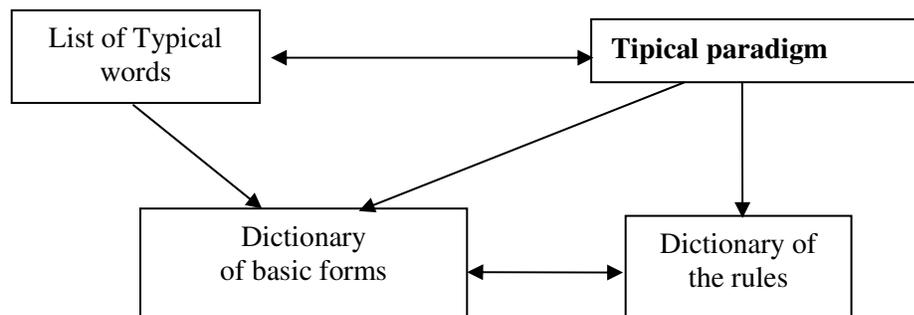


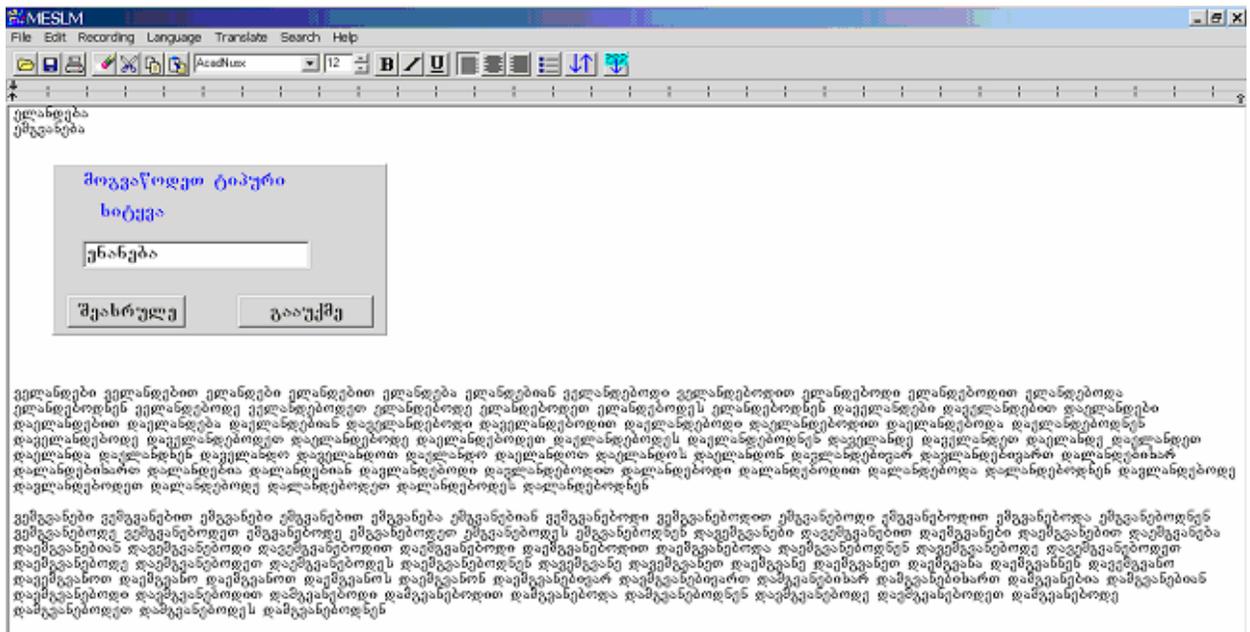
Fig. 2. Acquisition of knowledge by the MESLM expert system

We shall consider the said process for the verb. With the purpose of acquisition of knowledge, as we have noted earlier, to the system there should be supplied typical paradigm of the verb. Assume that this is the verb *enaneba* 'to regret'. Based on the typical paradigm, the system will acquire the knowledge about relevant morphological rule; save it in the Dictionary of basic forms and in the dictionary of the rules. Simultaneously, to the basic form there will be

given the number of relevant rule; the system, using the rule once saved, allows for examination of the paradigms of other basic words of the same type. When the expert makes sure that the List of typical word forms is selected correctly, to the dictionary of basic words, upon order of **MESLM** system software, there will be added the List of Typical words, in which, to the each one there will be recorded the number of relevant rule.

- elandeba* 1 'it seems to him'
- emgvaneba* 1 'he likens to smb.'
- enaneba* 1 'he is sorry for smth.'

This procedure will be repeated for the typical paradigms of all parts of speech. It should be noted that the linguist will be able, at any time, use the rule recorded once and accordingly increase the List of typical word forms.



When the expert makes sure that the List of typical word forms is selected correctly, to the dictionary of basic words, upon order of **MESLM** system software, there will be added the List of Typical words, in which, to the each one there will be recorded the number of relevant rule.

- ელანდება 1
- ემგვანება 1
- ენანება 1

This procedure will be repeated for the typical paradigms of all parts of speech. It should be noted that the linguist will be able, at any time, use the rule recorded once and accordingly increase the List of typical word forms.

A conceptual framework of representation in the **MESLM** system for derivation is not much distinct from inflexion. In both variants the interface remains the same and the user plays the role of a linguist from which only natural analytical ability is required. The inventory takes into account a paradigmatic representation of word morphology observed in a large number of languages. The morphological parameter values used by a source language, and means of their realization, differentiate one language from another but can play a role in knowledge/elicitation process for words substances and their expressions.

At present inflection part of English, Russian and Georgian morphological models have been designed allowing the **MESLM** system to support morphological generator, synthesis, analysis and lexical translation within these languages. As regards the derivation, some issues will be discussed.

As we know, derivation is a process where a new morphological word with its own inflectional paradigm is produced. So, the paradigm, describing derivation morphology must in parallel consider the introduction of a new model of word declension or conjugation. But derived

word might have the same inflection rule, which already has been acquired in the rule database. Consequently, the **MESLM** system must take into account also this fact as well yet supporting both of variants. The example 1 gives us derivation part of paradigm for Russian verb *разгибать* – 'unbend'. (On the right is given its English translation).

Исходная форма: разгибать	'Basic form: unbend'
Исходная часть речи: глагол	'Basic part of speech: verb'
Деривационная форма: разгибатель	'Derivational form: extensor'
Деривационная часть речи: существительное	'Derivational part of speech: noun'
Модель словоизменения соответствует: табель	'Model of inflection corresponding to: table'

Example 1.

Whether the derived word has its own paradigm or not **MESLM** system determines from 5th string of this entry:

Модель словоизменения соответствует: табель 'Model of inflection corresponding to: table'

The word after colon (*табель* – 'table') determines this solution. If the word is new the system gives for derivation word (*разгибатель*– 'extensor') its number of the rule. In other case, which is shown in Example 2, the 5th string will contain the same derivation word after the colon. At the same time, it is necessary to write flexible paradigm of derivational word form after the heading information.

Исходная форма: средний	'Basic form: middle'
Исходная часть речи: прилагательное	'Basic part of speech: adjective'
Деривационная форма: среднее	'Derivational form: average'
Деривационная часть речи: существительное	'Derivational part of speech: noun'
Модель словоизменения соответствует: среднее	'Model of inflection corresponding to: average'

	Един.	Множ.		'Singular	'Plural'
Именительный	среднее	средние	'Nominative	average	averages'
Родительный	среднего	средних	'Genitive	average's	averages''
Дательный	среднему	средним	'Dative	-	-'
Винительный	среднее	средние	'Accusative	-	-'
Творительный	средним	средними	'Ablative	-	-'
Предложный	среднем	средних	'Prepositional	-	-'

Example 2.

Regular derivational forms and their corresponding grammatical meanings in one language don't have the analogue in another one and vice versa. A multilingual dictionary of abbreviation is needed to discover possible divergences of grammatical categories in the different languages and accumulation of grammatical terms for interlingual lexicon of abbreviations. In **MESLM** system for abbreviations of grammatical categories English language is used. The system provides easy composition of appropriate terms and abbreviations. So, on the one hand there are grammatical categories in the source language and their English correspondents and on the other – their abbreviations. (See the Table 1. for some grammatical categories of Georgian language.) In the course of knowledge acquisition these correspondences can be supplemented at any time, yet only before using the data for grammatical analysis or synthesis.

Table 1.

Grammatical categories on source language	Grammatical categories on interlingual language – English	Abbreviations
<i>motxrobiti brunva</i>	Narrative Case	Narr_C
<i>vitarebiti brunva</i>	Circumstantial, Particular Case	Part_C
<i>cindebuliani brunva</i>	Prepositional Case	Prep_C
<i>tandebuliani brunva</i>	Enclitic Case	Enc_C

Finally it should be noted that **MESLM** system has been implemented in **C++ Builder**. It consists of four parts: Knowledge Acquisition, Multilingual Coordination, Language Generation and Lexical Translation modules. The first two must be listed before other ones. They can be implemented for any of languages. By this time morphological knowledge acquisition and multilingual coordination for English, Russian and Georgian languages has been done. It provides opportunity of application the system for analysis, synthesis and lexical translation.

References

- [1] G.Chikoidze. 1998. Bi-functional English Morphologic Processor. Proceedings of Institute of Control Systems. Tbilisi.
- [2] Kenneth R. Beesley, L. Karttunen. 2003. Finite-State Morphology: Xerox Tools and Techniques. Cambridge Press.
- [3] L. Lortkipanidze. 2003. “Record and reproduction” of morphological functions. Proceedings of the Fifth Tbilisi Symposium on Language, Logic and Computation. Tbilisi.
- [4] I. A. Mel’cuk. 1995. The Russian language in the meaning – text perspective. Moscow.
- [5] F. de Saussure. 1977. Cours de Linguistic Generale. In Russian. Moscow.