A FORMAL PROCEDURE FOR BULGARIAN WORD FORM GENERATION

Elena Paskaleva

Sofia, Bulgaria

The generation procedure proposed aims at the modelling of the process of verbal and nominal inflexion in Bulgarian. As in most of the similar morphological models of inflexional languages the procedure uses a comparatively simple mechanism of description - a comparatively small number of initial objects among which only one relation (viz. concatenation) is assigned; the transitions generating a separate concrete word form (or class of word forms) are determined.

The procedure includes the following linguistic objects:

$S = \{s_1, s_2, \ldots, s_n\}$ - a set of stems in Bulgarian (these can be the stems of dictionary entries in a sufficiently full dictionary of Bulgarian or the stems of lexical items used by the native speaker in his language behaviour).

$K = \{k_1, k_2, \ldots, k_m\}$ - where k_i is a number of an inflexional type.

$G = \{g_1, g_2, \ldots, g_p\}$ - where g_i is grammatical meaning.

$F = \{f_1, f_2, \ldots, f_g\}$ - where f_i is an ending.

$T = \{t_1, t_2, \ldots, t_r\}$ - where t_i is the so called "theme", i.e. one of the following elements: a thematic vowel in the verbal conjugation, a form-building suffix in the nominal declension or an extension of the stem in putting an article to some masculine adjectives.

$A = \{a_1, a_2, \ldots, a_s\}$ - where a_i is a postpositional article.

$D = \{d_1, d_2, \ldots, d_w\}$ - where d_i is an inflexional suffix. (This deviation from the aim of the model to describe only the processes of inflexion is conditioned by the dual nature of the participle as a part of the verbal paradigm and at the same time following the adjectival declension. This is the reason why, as we shall see below, it is generated in two phases: "verbal stem → stem of a participle" and "stem of a participle → word form of a participle". In the first phase elements participate belonging to the verbal inflexion (the thematic vowel typical for the inflexional type of the verbal stem), and in the second, elements belonging to the adjectival declension.

The elements of S can be mapped onto K. This map is many-one. The element corresponding to s is denoted by k_s.

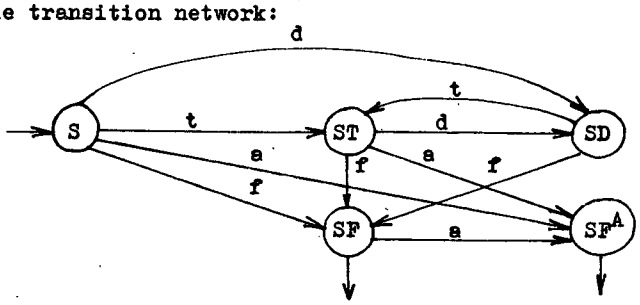
Between the elements of G and F, G and T, G and D, G and A correspondences exist which must be assigned in a table because of special language reasons - the ambiguity of the morphological elements and the different inflexion of one grammatical category.

For an initial symbol (S) of the generation procedure, some s_i is accepted accompanied by the corresponding k_{s_i} and the set $G_s \subset G$. By the concatenation of s with t, d, a and f (the sequence of operations can be followed on the transition network) the following linguistic objects are obtained:

a. as intermediate states of the generation procedure:
   ST - extended stem, i.e. "stem + theme" ($s_i \frown t_j$);
   SD - derivational stem, i.e. "stem (+eventually a theme) + word forming suffix" ($s_i \frown t_j \frown d_m$ or $s_i \frown d_m$).

b. as final states of the generation procedure:
   SF - word form;
   SF^A - word form with a postpositional article ($SF \frown a_r$).

In a most general form, the model for generation of Bulgarian word forms may be represented by the following very simple transition network:



The arcs of the network are marked with such an element of the sets T, D, F and A which is the second argument of the concatenation.

There being two final states is the result of the agglutinative character of the Bulgarian postpositive article, adjoined to an already generated word form. Deviation from the principles of agglutination we have in the direct transitions $S \rightarrow SF^A$ and $ST \rightarrow SF^A$, i.e. the article is adjoined to a stem (simple or extended). Such an adjunction we have in: a) putting an article to adjectives which in m.sg. have a zero ending and the article is added to the extended stem. b) adding an article to nouns and adjectives ending in masculine with iotized vowel (graphically represented by "vowel+я ") or with consonant. In the first case, the violation of the agglunitative character of the article is the result of the disregarding of morphonemic dependencies in the model generating only strings of letters (otherwise we would have a normal transition $SF \rightarrow SF^A$, i.e. repo - j → repo - j - a). In the second case, the direct transition $S(ST) \rightarrow SF^A$ is motivated by the linguistic unnaturalness of the resulting solution: the zero ending of the articleless masculine nouns and adjectives (final element of the word form) to be preserved also in the articled form as an intermediate element (брат-Ø →брат-Ø-а).

An alternative approach allowing to avoid this unnaturalness is to assume the identity "stem = word form" in the articleless form, i.e. the elimination of the zero ending as an element of the declension, which, however, would strongly affect the paradigmatic system.

The choice of concrete values of t, d, f and a for each s_i is determined by the value of k_i and the elements of G_s assigned in their initial state.

This network jointly represents the models of verbal and nominal inflexion. Treated separately, the three basic generative procedures are realized by the following transitions:

Verbs:

Finite forms
I. 1 → 2 → 3
II. 1 → 3
Participles
I.A. 1 → 4; B. 4 → 3 or 4 → 3 → 5 or 4 → 2 → 5
II.A. 1 → 2 → 4; B. 4 → 3 or 4 → 3 → 5 or 4 → 2 → 5

Nouns

I. 1 → 2 → 3 or 1 → 2 → 3 → 5
II. 1 → 3 or 1 → 3 → 5
III. 1 → 3 or 1 → 5

Adjectives

I. 1 → 3 or 1 → 3 → 5
II. 1 → 3 or 1 → 2 → 5
III. 1 → 3 or 1 → 5

The choice of procedures I, II or III within the framework of the basic generation procedures is determined by the elements of G_s and the value of k_i in the initial state. The existence of disjunction in procedures I, II and III for nominal generation is conditioned by the existence or absence of

an "article form" value among the elements of G_8. The transitions A and B of the generating procedure for participles are executed sequentially.