The Catena Approach: From Syntax to Compound Morphology

Petya Osenova and Kiril Simov
Linguistic Modelling Deparment, IICT-BAS
Acad. G. Bonchev 25A, 1113 Sofia, Bulgaria
petya@bultreebank.org and kivs@bultreebank.org

During the previous meeting we have presented a catena approach as a mechanism for encoding MWEs in the Bulgarian treebank. We extend the scope of application of catena. The catena approach is also very appropriate for modeling the connection among compounds and their syntactic counterparts in Bulgarian. In (Gross, 2011) the notion of 'morph catena' has been explicitly introduced. By granting a node to each morpheme[1], the author makes the problematic morpheme a dominant element over the other depending morphemes. Thus, all these morphemes are under its scope. The catena set includes also the intended meaning.

Here we have in mind examples like the following: a) compound deverbal noun whose counterpart can be expressed only through a free syntactic phrase (билколечение ('herbcuring', curing by herbs), *билколекувам (*'herbcure.1PERS.SG', to cure with herbs) and лекувам с билки ('cure.1PERS.SG with herbs', to cure with herbs); and b) compound deverbal noun whose verbal counterpart can be either a compound too, but verbal, or a free syntactic phrase (ръкомахане ('handwaving', gesticulating), ръкомахам ('handwave.1PERS.SG', gesticulate) and махам с ръка ('wave with hand', gesticulate).

A previously done survey in (Osenova, 2012) performed over an extracted data from a morphological dictionary (Popov et al., 2003) shows that in Bulgarian head-dependant compounds are more typical for the nominal domain (with a head-final structure), while the free syntactic phras-ing is predominant in the verbal domain. Also, regarding the occurrence of dependants in the compounds, subject is rarely present in the verbal domain, while complements and adjuncts are frequent - гласоподавам 'votegive.1PERS.SG', vote - where 'vote' is a complement of 'give'. On the contrary, in the nominal domain also subjects are frequently present, since they are transformed into oblique arguments - снеговалеж 'snowrain', snowing.

Irrespectively of the blocking on some compound verbs, there is a need to establish a mapping between the nominal compound and its free syntactic phrase counterpart. Both expressions are governed by the selection-based rules. Thus, the realization of the dependants in the syntactic phrases relies on the valency information of the head verb only, while the realization of the dependants in the nominal or verb compounds respects also the compound-building constraints.

A mechanism is needed which relates the external syntactic representations with the internal syntax of the counterpart morphological compounds. Moreover, some external arguments which are missing in the compound structures may well appear in the free syntactic phrases, such as: ръкомахам с лявата ръка 'handwave.1PERS.SG with left.DEF hand', I am gesticulating with my left hand, where the complement ръка (hand) is further specified and for that reason is explicitly present. Thus, we can imagine that in the lexicon we have the deverbal noun compounds as well as verb compounds, presented via morphological catena. These words are then connected to the heads of the corresponding syntactic phrases (again in the lexicon), but this time the relations are presented via a syntactic catena tree. We can

---

[1]Such as, histor-ic-al novel-ist where the morpheme 'ist' dominates the rest of the morphemes, thus resolving the bracketing paradoxes of the type [historical [novel-ist]] and [[historical novel]-ist]

think of the morphological catena as a rather fixed one, while of the syntactic catena as a rather flexible one, since it would allow also additional arguments or modifiers in specific contexts.

Let us see in more detail how this mapping will be established. The first case is the one where the deverbal nominal compound connects directly to a syntactic phrase (with no grammatical verb compound counterpart). The morph catena will straightforwardly present the tree of: билк-о-леч-ен-ие. However, in the syntactic catena a preposition is inserted according to the valence frame of the verb лекувам (cure): лекувам с билки ('cure.1PERS.SG with herbs', to cure with herbs). Using catena, we can safely connect the non-constituent phrase лекувам с (cure with) with the root morpheme of the head in the compound - леч. Also, all the possible modifiers of билки (herbs) in the syntactic phrase would be connected to the head morpheme билк.

The second case is the one where the nominal compound has mappings to both - verb compound and syntactic phrase. The connection among the nominal and verb compounds is rather trivial, since only the inflections differ. (ръкома-хане ('handwaving', gesticulating), ръкомахам ('handwave.1PERS.SG', gesticulate) and махам с ръка ('wave with hand', gesticulate): рък-о-мах-а-не vs. рък-о-мах-а-м. The connection with the syntactic phrase follows the same rules as in the previous case.

Here is the representation of the lexical unit for compound nouns: (билколечение ('herbcuring', curing by herbs):

[**form:** < билколечение >
catena:
(MorphVerbObj-C
(MorphVerb-C [1]билк-) (MorphoObj-C [2]леч-)
)
derivational catena:
(VPC-C
(V-C [2]лекувам (PP-C (P-C с) (N-C [1]билки)))
)
semantics:
лекувам_rel(e,x,y,[4]билки) & номинал_rel(e)
valency:
< mod; (PP (P с) [4](NP ModP* (N билки) ModS*)) :
ModP* or ModS* is not empty >
]

In this example we present two relations. First, the morph catena is presented with its roots (the role of affixes omitted for simplicity). Then, the catena reflecting the derivational syntactic phrase is shown. The correspondences are marked with tags [1] and [2]. The second relation is at the se-mantic level, where the semantics of the syntactic phrase (лекувам_rel(e,x,y,[3]билки)) is represented fully, and additionally the event is nominalized by the second predicate номинал_rel(e). In the valency list we might have a PP modifier (corresponding to the indirect object in the verb phrase) of the compound only if the preposition is с (by), the head noun of the preposition complement is the same as the noun in the verbal phrase билки (herbs) and there is at least one modifier of the noun. Thus phrases like: билколечение с български билки ('herbcuring with Bulgarian herbs', curing with Bulgarian herbs) and бил-колечение с билки, които са събрани през нощта ('herbcuring with Bulgarian herbs that are collected during the night', curing with Bulgarian herbs that were collected at night) are allowed. But phrases with dublicate internal and external arguments like билколечение с билки ('herbcuring with herbs', curing with herbs) are not allowed. Many of the other details are left out here in order to put the focus on the important relations. Among the omitted phenomena are the representation of the subject and patient information as well as the inflection of the compounds.

As a result, we propose a richer valence lexicon, extended with information on mappings between compounds and their counterpart syntactic phrases. The morph catena remains steady, while the syntactic one is flexible in the sense that it encodes the predictive power of adding new material. When connectors (such as prepositions) are added, the prediction is easy due to the information in the valence lexicon. However, when some modifiers come into play, the prediction might become non-trivial and difficult for realization.

References

Thomas Gross. 2011. Transformational grammarians and other paradoxes. In Igor Boguslavsky and Leo Wanner, editors, *5th International Conference on Meaning-Text Theory*, pages 88–97.

Petya Osenova. 2012. The syntax of words (in Bulgarian). In Diana Blagoeva and Sia Kolkovska, editors, *"The Magic of Words", Linguistic Surveys in honour of prof. Lilia Krumova-Tsvetkova*, Sofia, Bulgaria.

Dimitar Popov, Kiril Simov, Svetlomira Vidinska, and Petya Osenova. 2003. *Spelling Dictionary of Bulgarian (in Bulgarian)*. Nauka i izkustvo, Sofia, Bulgaria.