

Towards Transferring Bulgarian Sentences with Elliptical Elements to Universal Dependencies *Issues and Strategies*

Petya Osenova and Kiril Simov

CLaDA-BG, IICT-BAS, Bulgaria

Syntax Fest, UD Workshop,

30 August 2019



Plan of the Talk

- Introductory words
- Related work
- Modeling Ellipsis in the original treebank
- Introducing the original model into UD
- Conclusions

Introductory Words

- BulTreeBank (BTB) — an HPSG-based treebank of Bulgarian (Simov et al., 2005) — encodes both constituent and head-dependant structure in each phrase
- The current conversion of the treebank into the Universal Dependencies (UD) annotation scheme does not include the sentences with elliptical elements.
- These sentences constitute about 7 % of the treebank.

Related Work

- ([Mikulova, 2014](#)) presents the typology of ellipsis in Czech in the dependency theory of Functional Generative Description – ellipsis is mainly modeled on deep (tectogrammatical) level
- ([Jelinek et al., 2015](#)) – a constituent-based analysis for handling ellipsis is proposed
- ([Osborne and Liang, 2015](#)) – dependency-based notion of catena is used

Related Work

- ([Schuster et al., 2017](#)) give arguments in favor of introducing distinct nodes for gapping constructions in the enhanced representation of UD guidelines version 2, instead of the previously used relations remnant and orphan
- ([Droganova and Zeman, 2017](#)) - varieties in the annotation of ellipsis within the UD treebanks
- ([Adam Przepiórkowski and Patejuk, 2019](#)) - challenges when transferring the linguistic information from LFG to UD

Modeling Ellipsis in the Original Treebank

- Ellipsis is viewed as an expression that lacks an overt element
- This element, however, is presupposed and thus recoverable or easily predicted by the context
- Ellipsis is in close relatedness to linguistic phenomena like coordination and substantivization
- The idea in BTB was to preserve full syntactic structures

Modeling Ellipsis in the Original Treebank

- Ellipsis was introduced through a mechanism of adding a special artificial node at the ‘place’ of ellipsis
- Connecting it with an index to the overt corresponding part (if there is such a part), or
- Connecting it at the sentence level only (if the ellipsis is recoverable in a broader context or from world knowledge)

Modeling Ellipsis in the Original Treebank

- Ellipsis was indicated on two levels:
 - Syntactic (V-Elip, N-Elip, A-Elip, PP-Elip, Prep-Elip) and
 - Discourse (VD-Elip, ND-Elip, PrepD-Elip).

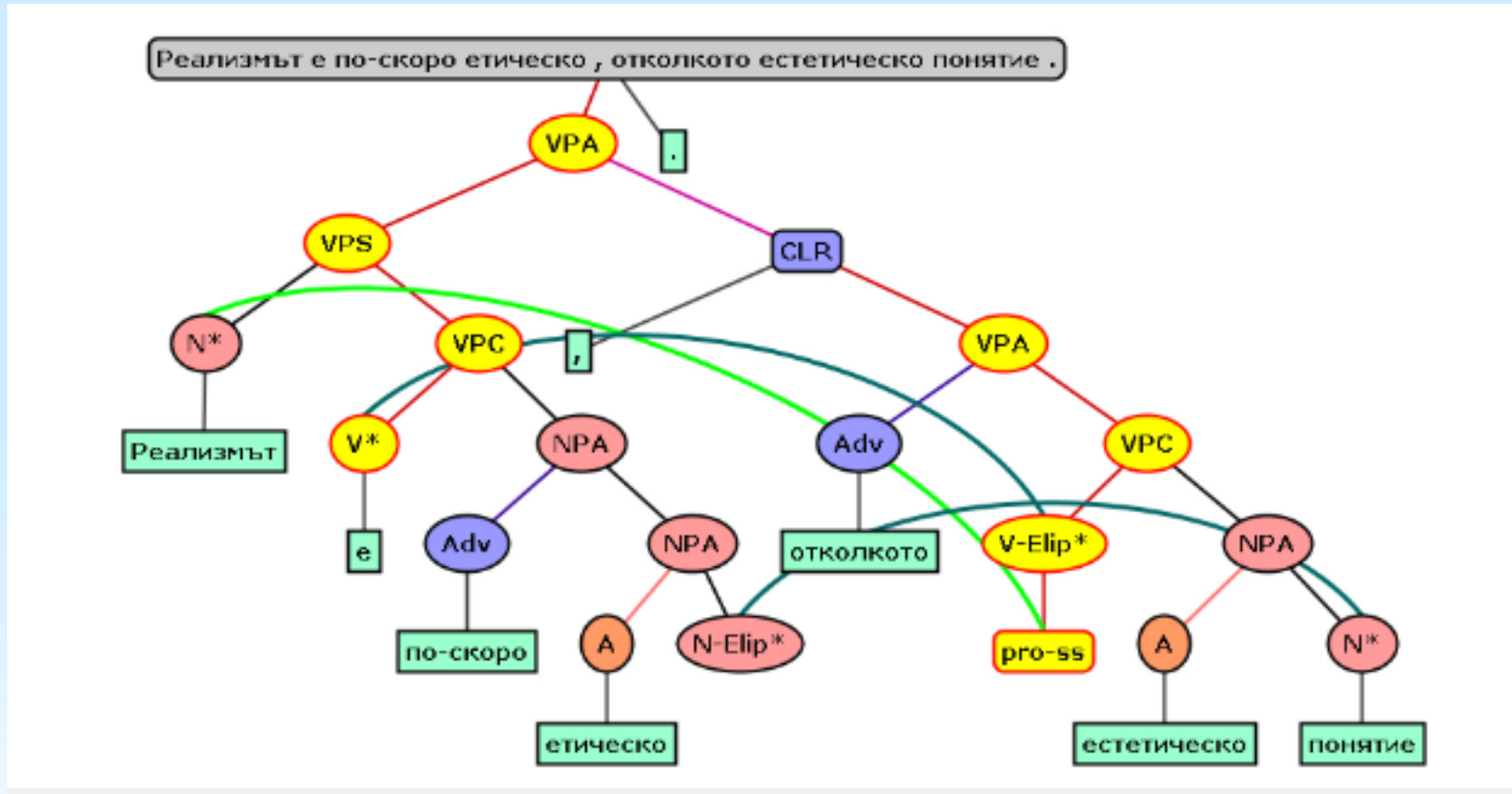
Verbal ellipsis was briefly discussed in (Osenova and Simov, 2018) in relation to handling enhanced dependencies

Modeling Ellipsis in the Original Treebank

In the original BTB the goal was to maximally restore the clausal structure

- **Coordination** – the cases were solved with predefined structures that can coordinate only if they have the same selectional restrictions (from both points of view - being heads or being dependants)
- **Substantivization**, it might be extended beyond the initially defined cases.

General Example



The realism is ethical **N-Elip** rather than esthetic **concept**

Types of Ellipses in BulTreeBank

Type of Ellipsis	Occurrences
N-Elip	327
V-Elip	262
VD-Elip	255
ND-Elip	70
PP-Elip	12
Prep-Elip	3
PrepD-Elip	2
A-Elip	1

Examples: structural ellipsis

- (1) Също така е развита химическата и техническата индустрия .
Also such is developed [chemical-the **N-Elip**] and [technical-the **industry**] .
'Also, the chemical and technical industries have been developed.'
- (2) Иван отиде в градината , но Петър не .
Ivan [went in garden-the] , but [Peter not **V-Elip**] .
'Ivan entered the garden but Peter did not.'

Examples: with specifics

- (5) Както Вулгарис _____ , така и княз Михайло
As Vulgaris **VD-Elip** [craved for Bulgarian lands] , such and prince Mihaylo
ламтеше за български земи .
craved for Bulgarian lands .
'Similarly to Vulgaris, prince Mihaylo was craving for Bulgarian land.'

- (6) Там нямаше заплаха, а _____ само радост .
There **was-not** threat, but **V-Elip** [was] only joy .
'There was not any threat, but only joy.'

Examples: with specifics

- (7) Наскоро замина където му бяха наредили
Recently **went-he** where him were ordered **V-Elip [to go]** .
'Recently he went to where he was ordered to go.'
- (8) Който се грижи за хората се грижат за неговите
Who se.REFL cares about people's **things**, people se.REFL care about his
N-Elip [things] .
'Who takes care of people's business, people take care about his [business].'

Introducing the Original Model into UD

- UD proposes the following strategies for handling ellipsis:
 - A **surface-based one** (in which a special *orphan relation* is used),
and
 - A **recovery-based one**
 - in which null elements for the elided material are used – as in the enhanced dependencies) or
 - promotion from the elided head to its dependants (when present) is introduced
 - In BTB the ellipsis has always been recovered, i.e. in this respect it followed somewhat a **non-surface-like** analysis

Introducing the Original Model into UD

- Null nodes for elided predicates: involves the addition of special null nodes in clauses with an elided predicate

I go to Varna, and you [V-Elip - go] to Sofia.

- In BTB such predicates are introduced as **V-Elip** nodes in an appropriate place in the structure. Thus, this label can be mapped directly into the so-called null nodes

Introducing the Original Model into UD

- There are two cases of usage of **V-Elip** - representation of elided single verbal form; and representation of elided phrase
- The first case is the more straightforward one
- In the second case in UD we need to introduce several null nodes in order to represent the whole VP
- In addition to the null nodes in BTB also some variation of the grammatical features is encoded. For the moment it is not clear how to represent these differing features in UD

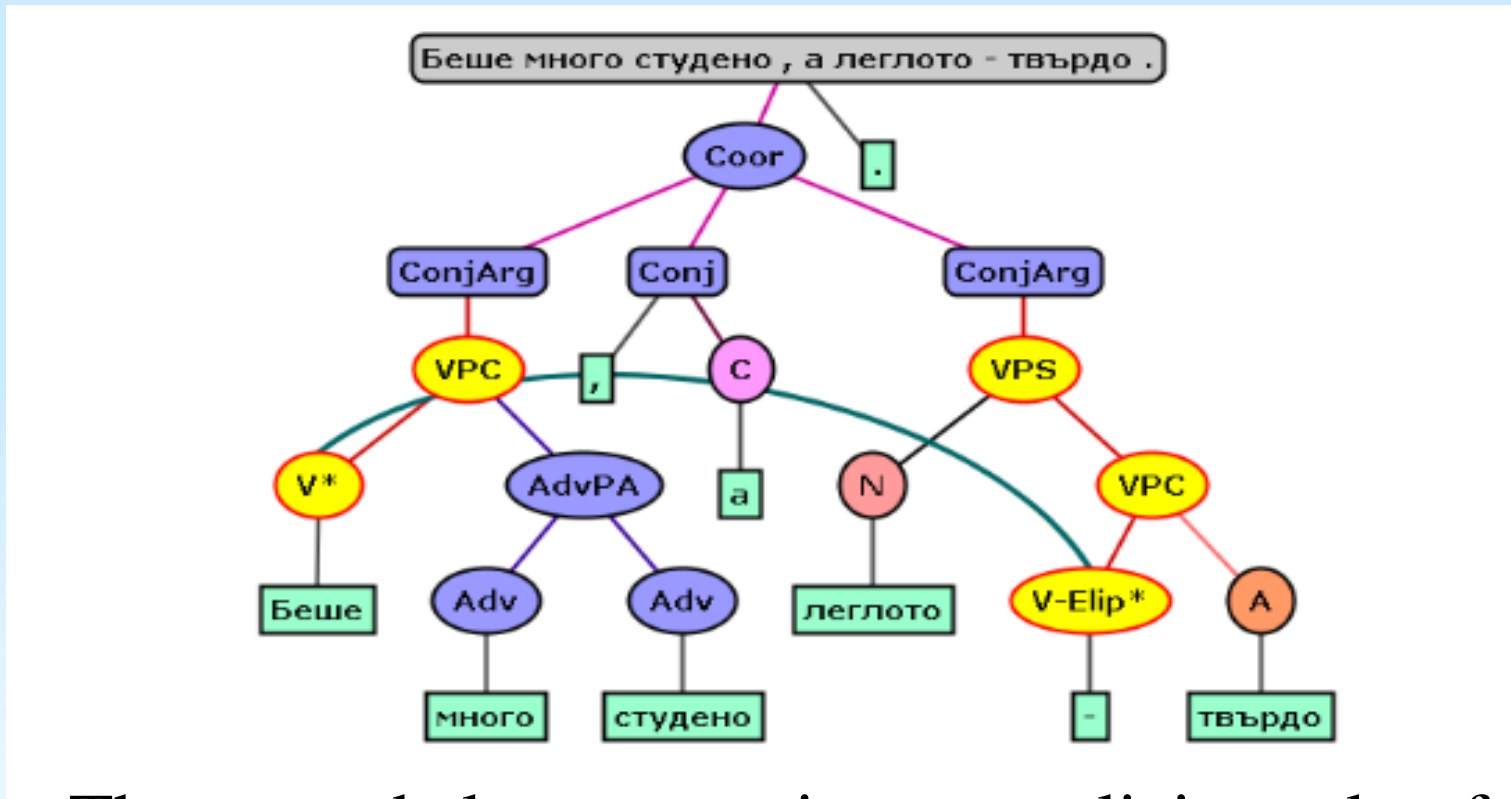
Introducing the Original Model into UD

- In contrast to V-Elip, the null nodes annotated with **VD-Elip** label in BTB provide discourse information that is difficult to identify by type (let alone the form) of the missing element(s)
- In this case within UD we could use *orphan* relation, but then the encoded information would be lost
- In order to preserve this information, we modify the *orphan* relation in order to specify the value of the discourse-restored value. For example, *orphan:cop* is used to represent the case of an elided copula licensed by discourse information

Observations

- The idea of using null elements instead of verbs or verbal groups does not cover all other cases with elided elements in UD.
 - In UD – mainly promotion of the dependent to head
 - In BTB – mainly ellipsis (promotion only in delimited cases as a) and b) below)
- In the case of BTB, the process of substantivization is restricted to: a) adjectives promoted to nouns; b) numerals in the structure *one of them; three of them*, etc.

Example: meaningful dash



The second clause contains an explicit marker for the place of the ellipsis (a dash)

Conclusions

- The current general principles behind UD for handling ellipsis are as follows:
 - elided element with no dependents is not processed at all
 - if it has dependants, then they are promoted as heads and
 - the promoted element uses the relation *orphan* when other functional elements are attached to it
- In BTB, besides the systematically applied **null-node-insertion-strategy**, ellipsis subtypes were added as a specification relation. Substantivation was kept mainly for the lexicalized dependants in the dictionary

Conclusions

- One possible direction of the UD development would be:
 - to extend the null node introduction
 - another one is to continue with the mixed strategy of treating ellipses in the basic and enhanced dependencies as it is now
- In both cases it would be useful to add more information on the ellipsis type and characteristics, and also to consider language specific features as it was done for other phenomena
- The proper treatment of ellipsis in an explicit way is important for the mono- and cross-lingual as well as for reasonable typological surveys across languages