Verbal fields in Hungarian simple sentences and infinitival clausal complements

Kata Balogh
DFG Collaborative Research Centre 991
Heinrich-Heine-Universität Düsseldorf
Katalin.Balogh@hhu.de

The paper presents an analysis of Hungarian sentence articulation driven by discourse-semantic functions such as *topic* and *focus*, hence the information structure of the utterance. In this paper extensions of the standard LTAG framework are proposed to represent the information structure driven positions in simple sentences. The elementary trees are generated by the meta-grammar, using the XMG tool for grammar writing. The verbal fields will also be investigated in complex sentences with infinitival clausal complements. I will discuss challenging phenomena, such as scrambling of the different verbal fields, and verbal modifier climbing.

1 Introduction

The core topic of this paper is the analysis of the verbal fields in Hungarian simple sentences and infinitival clausal complements. Hungarian is challenging for computational linguistic applications, given its flexible word order and discourse configurational type, where syntactic positions are not driven by grammatical functions, but rather by discourse-semantic functions such as *topic* and *focus*. In this paper I propose an analysis in Lexicalized Tree-Adjoining Grammar [LTAG; (Joshi and Schabes, 1997)] with an extension to the representation of information structure, the driving device for sentence articulation in Hungarian. The elementary trees are generated by the meta-grammar, using the eXtensible MetaGrammar tool [XMG; (Crabbé et al., 2013)].

Next to simple sentences, the paper also presents the proposal of an analysis of complex sentences with infinitival clausal components. The analysis of Hungarian infinitival clauses faces a list of interesting issues, regardless of the given framework. In the analysis we need to capture the fact that the embedded infinitival clause has its own verbal fields and that the arguments of the matrix verb and the embedded infinitival can be mixed. The analysis of infinitival clausal complements is challenging for standard LTAG, since it cannot handle discontinuity and scrambling phenomena (Becker and Rambow, 1990; Becker et al., 1991), caused by the mixing of the arguments in different verbal fields. This paper proposes some modifications in order to capture these phenomena: (i) using multi-component TAG [MCTAG] (Weir, 1988; Nesson et al., 2010), and (ii) extending the feature sets to represent stress-positions.

1.1 LTAG in a nutshell

Tree-Adjoining Grammar (TAG) is a tree-rewriting formalism, where the elementary structures are trees. A TAG is a set of *elementary trees* with two combinatorial operations: *substitution* and *adjunction*. The set of elementary trees is the union of a finite set of *initial trees* and *auxiliary trees*. A derivation in TAG starts with an initial tree (α) and proceeds by using either of the two operations. By *substitution* a non-terminal leaf node is replaced by an initial tree (β), while by *adjunction* an internal node is replaced by an auxiliary tree (γ).

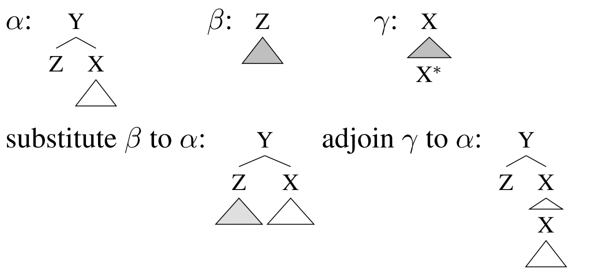


Figure 1: Substitution and adjunction

In a Lexicalized TAG [LTAG] each elementary tree contains at least one lexical element, its *lexical anchor* ($\diamond$). To increase the expressive power

Proceedings of the 12th International Workshop on Tree Adjoining Grammars and Related Formalisms (TAG+12), pages 58–66, Düsseldorf, Germany, June 29 - July 1, 2016.

of the formalism, adjunction constraints are additionally introduced to restrict whether adjunction is mandatory and/or which trees can be adjoined at a given node. In particular for natural language analyses another extension of TAG is proposed, using feature structures as non-terminal nodes. Among the reasons for a Feature-based TAG [F-TAG] two important ones are generalizing agreement and case marking via underspecification. A great advantage of F-TAG with respect to grammar writing is the result of smaller grammars that are easier to maintain, as well as the possibility of modeling adjunction constraints. The shape of the elementary trees is driven by linguistic principles (Abeillé and Rambow, 2000; Frank, 2002), reflecting the syntactic/semantic properties of linguistic objects. Syntactic design principles determine, for example, that subcategorization is expressed locally within the elementary tree of the predicate. In the grammar architecture *tree families* are defined, sets of tree templates representing a subcategorization frame and collecting all syntactic configurations the subcategorization frame can be realized in.

2 Verbal fields in simple sentences

With respect to information structure, Hungarian sentence structure distinguishes two fields: the *postverbal* and the *preverbal* field (see e.g. É. Kiss, 2005). The *postverbal field* by default hosts the 'argument positions' the order of which is free: the word order variations do not signal grammatical roles and are associated with the same semantic content. The *preverbal field* hosts the so-called 'functional projections', the order of which is fixed. Sentence initially we find the topic(s), followed by quantifier(s) and the narrow focus, which is placed in the immediate preverbal position:[1]

Topic* < Quantifier* < Focus < Verb [...]

A designated syntactic position immediately preceding the verb is important in Hungarian for several reasons. This position hosts narrow focus, sentential negation, the verbal modifier (verbal particles, bare nouns, infinitives etc.), partially in complementary distribution. In 'neutral sentences' – utterances without narrow focus or sentential negation –, the immediate preverbal position is occupied by the verbal modifier (VM) as illustrated in (1a). When the sentence contains sentential negation (1b) or narrow focus (1c) the VM stands postverbally.

(1) a. Pim meg-hívta Marit.
Pim VM-invited Mary.acc
'Pim invited Mary.'

b. Pim nem hívta meg Marit.
Pim neg invited VM Mary.acc
'Pim did not invite Mary.'

c. Pim MARIT hívta meg.
Pim Mary.acc invited VM
'It is Mary whom Pim invited.'

The verbal modifier is in complementary distribution with both the narrow focus and the sentential negation. However, narrow focus and negation can co-occur; see (2).

(2) Pim MARIT nem hívta meg.
Pim Mary.acc neg invited VM
'It is Mary whom Pim did not invite.'

The position of the VM depends on whether narrow focus or sentential negation is present in the sentence. The narrow focus and sentential negation must stand in the immediate preverbal position, forcing the VM to appear postverbally.

2.1 Functional positions

In Hungarian, sentence articulation is driven by discourse semantics rather than by grammatical function. Instead of defining structural positions for grammatical functions (e.g. subject) we need to define structural positions for topic and focus. Accordingly, the elementary trees in a given tree family need to encode the possible topic/focus structures. In the following, I will illustrate the system via the transitive tree family with two NP arguments. Both arguments can be either in postverbal position, in topic position and in focus position. Hence, the tree family of transitive verbs must contain the elementary trees for all structures: VP-NP-NP, NP_{top}-VP-NP, NP_{foc}-VP-NP, NP_{top}-NP_{top}-VP, NP_{top}-NP_{foc}-VP. Instead of merely listing the structures, the XMG tool is used to generate the set of elementary trees, so that generalizations on positions can also be expressed (see section 3).

2.2 Representing information structure

The representation of information structure comes on a par with the elementary tree, following the proposals by (Kallmeyer and Romero, 2008) and

[1] Next to topic, focus and quantifiers, the preverbal field also hosts sentential negation, optative operators (*bárcsak* 'if only') and interrogative operators.

(Kallmeyer and Osswald, 2012). Each elementary tree is linked to a frame-based semantic representation. The syntactic operations (substitution and adjunction) trigger the unification of the semantic representations, thereby deriving the meaning representation of the sentence. Next to the meaning representation, the representation of the information structure is proposed, mediated by the metavariables on the nodes.

The notion of *information structure* [InfS] covers a wide range of phenomena such as topic, comment, focus, background, given, new, contrast etc. The central aim in the current analysis is to represent InfS in a way that we can derive the basic focus structures: *sentence focus*, *predicate focus* and *narrow focus*. These structures are nicely reflected in the Hungarian sentence articulation by the structural distinction of the topic of the sentence and the comment, which we take as the *focus domain*. The focus domain is either as a whole the focus of the sentence (predicate focus) or it contains narrow focus.

(3) a. Q: What happened?
A: Meg-hívta Pim Marit.
VM-invited Pim Mary.acc
'[Pim invited Mary.]^F' → sentence focus

b. Q: What did Pim do?
A: Pim meg-hívta Marit.
Pim VM-invited Mary.acc
'Pim [invited Mary.]^F' → predicate focus

c. Q: Whom did Pim invite?
A: Pim MARIT hívta meg.
Pim Mary.acc invited VM
A′: MARIT hívta meg Pim.
Mary.acc invited VM Pim
'Pim invited [Mary]^F.' → narrow focus

In the elementary tree of the predicate, the feature INFS indicates the positions related to information structure. The values *topic* and *focus* indicate the positions for the sentence topic and the narrow focus respectively, while the value *pred* indicates the syntactic predicate.

Each elementary tree is linked to a representation of the information structure, uniformly given as an attribute-value matrix. The attributes TOP (topic) and FOC-DOM (focus domain) represent the *topic–comment* distinction. The focus domain can be further divided into FOC (indicating narrow focus) and NON-FOC (non-focus part of the focus domain).[2] The background part in the *focus–background* distinction can be derived as the unification of the NON-FOC and the TOP values.

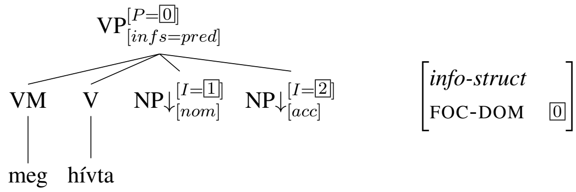


Figure 2: Elementary tree of *meg-hívta* for sentence focus

Figure 2 illustrates the elementary tree of the verb *meg-hívta* ('invited') for structures with sentence focus (3a). The resulting InfS representation signals the whole predication (P=0) being the focus-domain, and since it is not divided further, it represents sentence focus.

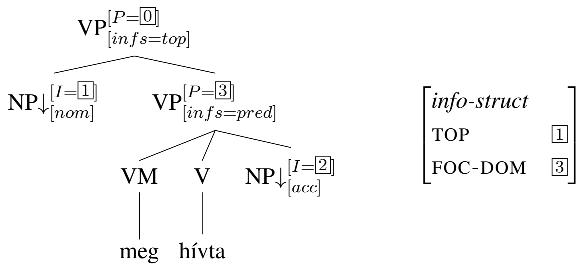


Figure 3: Elementary tree of *meg-hívta* for predicate focus

Figure 3 shows the elementary tree for predicate focus structure (3b). The InfS representation signals the syntactic predicate (P=3) as the focus-domain, while the element occupying the topic position (I=1) is represented as the sentence topic.

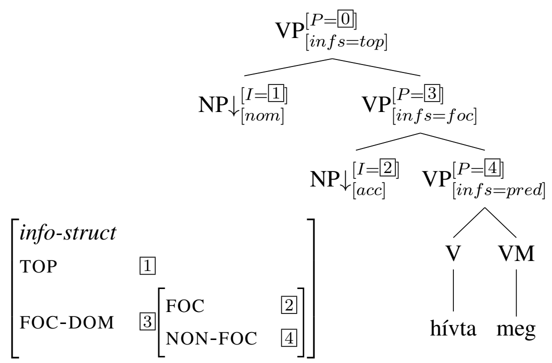


Figure 4: Elementary tree of *hívta meg* for narrow focus

Figure 4 illustrates the elementary tree of the verb for a construction containing both a topic and

[2]The notions of *focus domain* and *non-focus* are inspired by (Van Valin Jr., 2005).

a narrow focus (3c-A). The focus domain is divided into the narrow focus (FOC) and the non-focus, that contributes to the background.

The different information structures can be further mapped to the information status of the given elements, in terms of *givenness*. The representations of different topic/focus structures and their relation to the discourse is under investigation and will be presented in a later stage of the analysis.

3 Topic and focus positions

As already noted above, the tree families of the verbs must contain all possible structures reflecting the information structure of the utterance. The arguments of the verbs can occupy three structural positions, resulting in various structures. For example, transitive verbs must have two argument slots, each with three possibilities. Instead of merely listing these structures, the XMG tool is used to generate the set of elementary trees, so that generalizations can also be expressed.

3.1 XMG in a nutshell

An LTAG grammar is a set of elementary trees which contain most linguistic information. However, this set contains identical tree fragments, leading to multiple structure sharing. The XMG tool provides the elementary trees for a given grammar, such that it factors out redundant parts of a given tree set by identifying identical tree fragments in the set of elementary trees. An additional abstraction level is introduced, the *meta-grammar*, where generalizations can be expressed. The meta-grammar is a declarative system that combines re-usable tree fragments – *classes* – by conjunction and disjunction.

`Class ::= Name` $\rightarrow$ `Content`

`Content ::=`

  `Descr | Name |`

  `Content` $\wedge$ `Content | Content` $\vee$ `Content`

`Descr ::=`

  $n_i \rightarrow n_j \mid n_i \rightarrow^+ n_j \mid n_i \rightarrow^* n_j \mid n_i \prec n_j \mid$

  $n_i \prec^+ n_j \mid n_i \prec^* n_j \mid n_i[f_1 : v_1, ..., f_n : v_n] \mid$

  $n_i(c_1 : cv_1, ..., cf_n : cv_n)$ `| Descr` $\wedge$ `Descr`

The content of a class can be either a simple tree fragment or a conjunction / disjunction of two tree fragments. In the description of a tree fragment the dominance $\rightarrow$ and precedence $\prec$ relations of the nodes are given, where $\rightarrow^+$ and $\prec^+$ stand for their transitive closure and $\rightarrow^*$ and $\prec^*$ for their transitive, reflexive closure. At each node we refer to the features associated with it by $n_i[f_1 : v_1, ..., f_n : v_n]$ and each node can be marked for substitution, footnode, anchor, etc. by $n_i(c_1 : cv_1, ..., cf_n : cv_n)$.

```
class CanSubj
declare ?S ?VP ?NP
{<syn>{
node ?S (color=black) [cat=s] ;
node ?NP (color=black,mark=subst) [cat=np] ;
node ?VP (color=white) [cat=vp] ;
?S -> ?NP ; ?S -> ?VP ; ?NP » ?VP } }
```

Figure 5: Tree description of the class of canonical subject in English

Tree fragments can be combined by conjunction and disjunction resulting in tree templates, e.g.:

`Subject` $\rightarrow$ `CanSubject` $\vee$ `WhNpSubject`

`Object` $\rightarrow$ `CanObject` $\vee$ `WhNpObject`

In the combination of the tree fragments nodes get unified. Node equations are carried out by node polarization: annotating the nodes with colors (e.g. `color = black`), which declare implicitly how a given node can be unified with others. This method is based on a color matrix, according to which (i) a black node can unify with zero, one or more white nodes, producing a black node, (ii) a white node must be unified with a black node producing a black node, and (iii) a red node cannot be unified with any other node. The resulting tree fragment is a satisfying model only if it does not contain any white nodes.

3.2 Implementation using XMG

The class of transitive verbs must contain the verbal projection and two arguments, the subject (`Subj`) and the object (`Obj`).

`Trans` $\rightarrow$ `VProj` $\wedge$ `Subj` $\wedge$ `Obj`

Both arguments have three possible positions: postverbal (`ArgPos`), topic (`TopPos`) or focus (`FocPos`) position and get the appropriate case marking.

`Subj` $\rightarrow$ `(ArgPos` $\vee$ `TopPos` $\vee$ `FocPos)` $\wedge$ `Nom`

`Obj` $\rightarrow$ `(ArgPos` $\vee$ `TopPos` $\vee$ `FocPos)` $\wedge$ `Acc`

NP case marking is generalized via the classes `Nom, Acc`, ...as:

`Nom` $\rightarrow$ `n[cat:np,top:[case:nom]]`

`Acc` $\rightarrow$ `n[cat:np,top:[case:acc]]`

The core is the class of the verb projection, `VProj`, that must be either one of the three tree fragments:

`VProj` $\rightarrow$ `VProj1` $\vee$ `VProj2` $\vee$ `VProj3`

The above fragments are defined according to whether there is a verbal modifier present, and in which order it appears relative to the verb. Considering the VM-V order as default (non-inversion), the INV feature determines whether inversion is present (values *yes* / *no*). In case the sentence contains no VM, the INV feature is irrelevant / not applicable (value *na*).

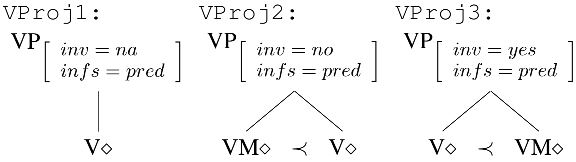


Figure 6: Tree fragments for verb projection

For the argument position the class `ArgPos` is defined. As illustrated in section 2, in the post-verbal field the order of the arguments is free. At the VP node the feature INFS represents the information structural function of the constituent, with value *pred* representing the syntactic predicate.

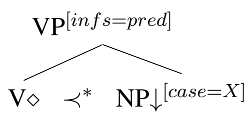
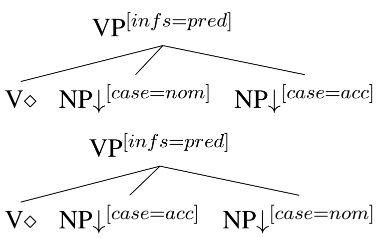


Figure 7: Tree fragment for argument position

The class of the verb projection combines with one or more argument positions, where the order of the arguments are free. This latter is made possible by the reflexive, transitive closure ($\prec^*$) of the precedence relation between the V node and the argument NP node. The case marking of the argument is underspecified.

By the above fragments we can generate the elementary trees for deriving all possible argument orders in the post-verbal field for a given predicate. For example, the tree family of a transitive verb will contain the structures (among others) for cases in which both arguments are postverbal:



In the pre-verbal field we have fixed structural positions for the *topic* and *focus*.[3] The classes for these positions are defined respectively as:

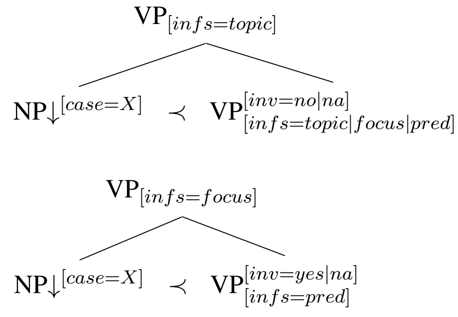


Figure 8: Tree fragments for topic and positions

According to the hierarchical order of the preverbal positions, a topic can be followed by another topic, the focus or the predicate, while focus can only be followed by the predicate. These ordering constraints are captured by the bottom features of the respective VP footnotes, while the top features of these nodes constrain the VM-V inversion. Focus induces inversion (*inv*=*yes*|*na*), while topics do not (*inv*=*no*|*na*). For both, a combination with a non-VM verb is of course possible.

The tree fragments above define the post-verbal position and the two functional positions (topic, focus) in a uniform way, capturing the generalizations behind them. Using the above tree fragments the meta-grammar generates the elementary trees, by which the grammar can derive all possible structures in simple sentences, possibly containing topic and focus positions. The non-grammatical structures – e.g. topic following focus, focus without VM-V inversion etc. – are ruled out correctly.

4 Infinitival embedded clauses

In this section I propose an analysis of infinitival complements and the structure and relation of the verbal fields of the matrix verb and the infinitival embedded verb. In Hungarian two types of control verbs can be distinguished: the ones that take main stress (e.g. *fél* 'is afraid') versus the ones that avoid main stress (e.g. *akar* 'want'), referred to as 'stress-bearing verbs' and 'stress-avoiding verbs' respectively. These two verb classes differ in syntactic behavior with respect to the placement of the verbal modifier of the embedded infinitive. In Hungarian the main stress falls on the *leftmost* element of the phonological phrase, hence, in neutral sentences the main stress falls on the VM.

[3]In Hungarian, (distributive) quantifiers have a designated position, too, between the topic(s) and the focus. Regarding the space limitations, the analysis of quantifiers is not discussed in this paper.

Example (4a) illustrates the stress-avoiding verb (*akar* 'want'), which induces the climbing of the VM of the infinitival verb.

(4) a. Pim el akarja (el-*)olvasni a levelet.
Pim VM wants (VM-)read.inf the book.acc
'Pim wants to read the letter.'

b. Pim (el*) fél el-olvasni a levelet.
Pim (VM) afraid VM-read.inf the book.acc
'Pim is afraid to read the letter.'

The VM *el-* is the modifier of the infinitival *olvasni*, but it must appear in the pre-verbal position of the matrix verb. In this way the VM *el-* receives the main stress instead of the matrix verb. In (4b) the VM of the infinitival stays in its own pre-verbal position, VM climbing is no grammatical. Koopman & Szabolcsi (2000) classifies different verb types as (i) the group of *auxiliaries*: that do not bear main accent and induce VM-climbing, (ii) *non-auxiliaries 1*: that bear main accent and do not induce VM-climbing and (iii) *non-auxiliaries 2* with mixed behavior.

In case of clausal complements we have multiple verbal fields, belonging to both the finite and the infinite verbs.

(5) [preV_1] V_1 [postV_1] [preV_2] V_2 [postV_2]

(6) András meg-tanította a diákokat a
Andrew VM-taught the students.acc the
mondatot csak LFG-ben elemezni.
sentence.acc only LFG.in analyze.inf
'Andrew taught the students to analyze the sentence only in LFG.'
(from (Szécsényi, 2009))

Syntactically interesting cases are the sentences in which arguments of the embedded verb are topicalized or focused. Szécseényi's (2009) example above provides evidence of the existence of the separate pre-verbal field of the embedded infinitive, hosting the focus expression *csak LFG-ben* 'only in LFG'. However, broader data suggest that the preferred position of a focused or topicalized argument of the infinitive is in the pre-verbal field of the matrix verb (see examples (7) and (8)).

In the following examples the embedded infinitive is *el-olvasni* 'VM-read.inf', containing a verbal modifier, while the matrix verb does not have a VM. Topics preferably stand in the preverbal field of the matrix verb (7b, 8b), however, it is also grammatical in the preverbal field of the embedded verb (7a, 8a). Focused constituents stand in the preverbal field of the matrix verb (7c, 8c). Next to these preferred positions, the sentence articulation is also sensitive to the type of the matrix verb (auxiliary vs. non-auxiliary), providing different structures.

(7) a. ?Pim fél [a levelet]^T el-olvasni.
Pim afraid the letter.acc VM-read.inf

b. Pim [a levelet]^T fél el-olvasni.
Pim the letter.acc afraid VM-read.inf
'Pim is afraid to read the letter.'

c. Pim [a LEVELET]^F fél el-olvasni.
Pim the letter.acc afraid VM-read.inf
'It is the letter, that Pim is afraid to read. '

(8) a. ?Pim el akarja [a levelet]^T olvasni.
Pim VM wants the letter.acc read.inf

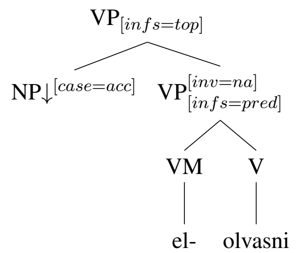
b. Pim [a levelet]^T el akarja olvasni.
Pim the letter.acc VM wants read.inf
'Pim wants to read the letter.'

c. Pim [a LEVELET]^F akarja el-olvasni.
Pim the letter.acc wants VM-read.inf
'It is the letter, that Pim wants to read. '

4.1 Proposal: from LTAG to TL-MCTAG

Closely related to scrambling, the position of the verbal modifier of the embedded verb in stress avoiding versus stress bearing control verbs poses interesting questions to the analysis.[4]

The structures reflecting the verbal fields of the control verbs and the infinitives can both be generated along the lines of the analysis of simple sentences. See, e.g., the elementary tree of *el-olvasni* 'to read' with a topicalized argument:



The core of the problem with an LTAG analysis here is caused by scrambling of the arguments and the verbal modifier in the different verbal fields of the matrix verb and the embedded verb. Using the standard LTAG formalisms it is not possible to capture VM-climbing (4a), since it proposes elementary trees for sentence embedding, such that sentential complements being represented as a footnode (S*), and the elementary

[4]Hungarian infinitival clauses pose several more interesting issue to deal with, for example the verb-object agreement between the matrix verb and the object of the infinitive. For more issues around infinitival clauses in Hungarian see (Koopman and Szabolcsi, 2000) and (Szécsényi, 2009).

tree of the matrix verb (*akar* 'wants') being adjoined into the elementary tree of the embedded verb. The standard LTAG analysis derives straightforwardly the structures, in which the matrix verbs and its arguments are not split. This involves sentences with a stress-bearing control verb with different possible topic/focus structures, e.g.

(9) a. Pim^T fél el-olvasni a levelet.
Pim like.mod VM-read.inf the letter.acc
'Pim is afraid to read the letter.'

b. A levelet^T Pim^F fél el-olvasni.
the letter.acc Pim like.mod VM-read.inf
'It is Pim, who is afraid to read the letter.'

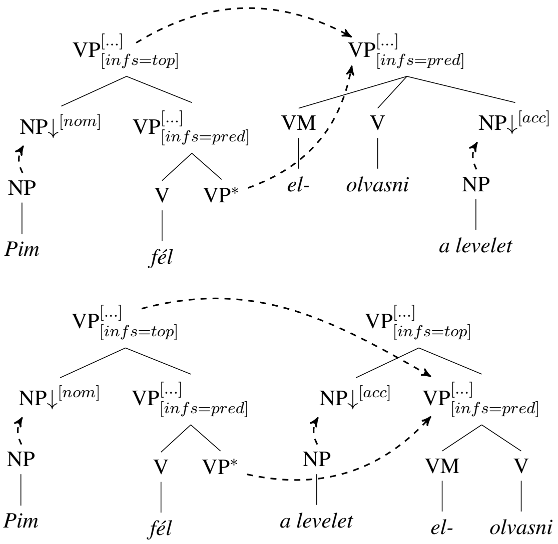


Figure 9: Standard LTAG analysis of (9a) and (9b)

The analysis in Figure 9 correctly derives the structures for stress-bearing verbs (4b), however this structure is not grammatical for stress-avoiding verbs as *akar* 'want'.

(10) *Pim akarja el-olvasni a levelet.
Pim wants VM-read.inf the letter.acc

This structure should be ruled out for stress-avoiding verbs, but must be derived for stress-bearing verbs. Furthermore, Hungarian has a number of control verbs compatible with both VM-climbing and VM in situ structures. For such verbs it should be possible to derive both structures .

(11) a. Pim szeretne el-olvasni egy levelet.
Pim like.mod VM-read.inf a letter.acc
'Pim would like to read a book.'

b. Pim el szeretne olvasni egy levelet.
Pim VM like.mod read.inf the letter.acc
'Pim would like to read a book.'

In cases of VM-climbing (4a,11b), the tree of the matrix verb (*akar*) should split into more pieces when adjoined to the tree of the embedded verb (*el-olvasni*). This core problem for the standard LTAG analysis also arises when the arguments of the two verbs are mixed in the pre-verbal field of the matrix verb. As shown before, the preferred position of the topicalized or focused argument of the infinitive is in the pre-verbal field of the finite verb, hence arguments of different verbs can mix. See, for example, the sentence in which the object of the infinitive is focused and the subject of the finite verb is topicalized.

(12) Pim a LEVELET^F fél el-olvasni.
Pim the letter.acc afraid VM-read.inf
'It is the letter, what Pim is afraid to read.'

The problems of the analysis discussed above are due to the fact that standard LTAG cannot capture discontinuity (scrambling, extraposition etc.) in general (Becker and Rambow, 1990; Becker et al., 1991). In order to overcome the problems of our analysis, the use of Multicomponent TAG [MCTAG] is proposed. MCTAG is a modified TAG formalism, that allows elementary structures as set of trees. For natural language grammars the tree-local MCTAG [TL-MCTAG] (Weir, 1988; Nesson et al., 2010) is considered, which comes with the restriction that all trees in the set have to attach to the same elementary tree. TL-MCTAG is strongly equivalent to standard LTAG, thus by using this formalism we can overcome the problems of scrambling and VM-climbing without losing any of the attractive formal and computational properties of LTAG.

The first necessary modification of our original analysis is the position of the VM. In order to make VM-climbing possible, we need more structure below the given VP node, hence the previously proposed flat structure must be revised.

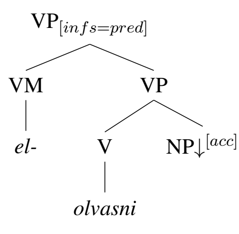


Figure 10: VM-V structure revised

Evidently, the structures with a topicalized/focused argument must be revised accordingly. Such a structure allows adjunction at the VP node right above the V, thereby allowing the VM-V verbal complex to split. This is one of the necessary conditions for capturing VM-climbing. However, this is merely the first step. The problem of the discontinuity of the finite verbs and its (topicalized) argument is still unresolved. Obviously, we need to deal with cases, where both the finite verbs and the embedded infinitive come with a split structure regarding the arguments and the verbal modifier. The solution is provided by TL-MCTAG, taking the elementary structures of the matrix verbs (*akar*, *fél*) as sets of trees.

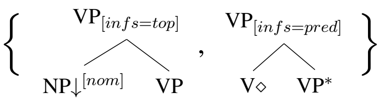


Figure 11: Elementary tree set for control verbs

Through this modification of the analysis we can derive the correct VM-climbing structure of stress avoiding verbs by allowing the first tree in the tree set to adjoin at the root VP node of the infinitival tree, and the second tree at the inner VP node right above the V. However, with these elementary tree sets we also derive the non-grammatical structures for both stress-bearing and stress avoiding verbs. The difference between these two verb classes relies on their relation with the prosodical structure of the sentence, that motivates the extension of the analysis by features representing stress positions. In Hungarian, the main stress falls on the left edge of the phonological phrase, hence in the default VM-V order the VM bears the main stress, while the V is unstressed. This is reflected in the elementary trees of verbs, e.g.:

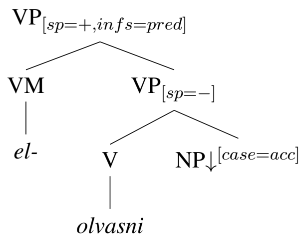
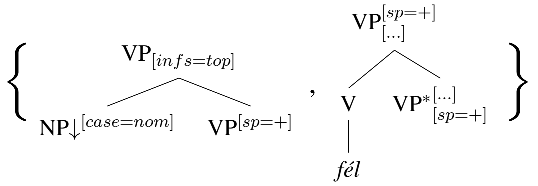
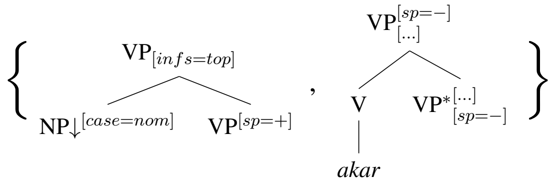


Figure 12: Features for stress positions

Stress-bearing verbs (*fél* 'is afraid') are marked for a stress position ([sp=+]) and thereby can only be adjoined at the root VP node of the tree of the infinitive. Adjoining at the inner VP node is ruled out by a feature clash between the footnode of *fél* and the target node.



Stress-avoiding verbs (*akar* 'want') are marked for a non-stress position ([sp=–]) and hereby the second tree in the set can only be adjoined at the inner VP node of the tree of the infinitive.



Verbs allowing for both structures come with a SP feature with an underspecified value, and thus can be equally derived in both structures.

5 Conclusion

This paper proposes an analysis of the verbal fields in Hungarian sentences articulation. I discussed several issues for an LTAG analysis of the information structure driven syntactic positions in simple sentences and in infinitival clausal complements. As information structure is the main device driving sentence articulation, an extension is proposed for representing different topic–focus structures. The elementary trees – reflecting the possible structures – are generated by the meta-grammar, using the eXtensible MetaGrammar tool. As shown in the paper, the flexible word order in Hungarian simple sentences is relatively easy to capture, generating all possible structures, while expressing the important generalizations on the functional positions.

Next to the analysis of the verbal fields in simple sentences, the paper proposed an analysis of complex sentences with infinitival clausal components. The analysis of Hungarian infinitival clauses faces a list of interesting issues. It is especially challenging for standard LTAG, since it cannot handle discontinuity and scrambling phenomena, caused here by the mixing of the arguments and the verbal modifier of different verbs. The proposed analysis shows that the challenges can be overcome by using TL-MCTAG extended with features to represent stress positions.

References

Anne Abeillé and Owen Rambow. 2000. Tree Adjoining Grammar: An overview. In Anne Abeillé and Owen Rambow, editors, *Tree Adjoining Grammars: Formalisms, Linguistic Analyses and Processing*, pages 1–68. CSLI Publications, Stanford.

Tilman Becker and Owen Rambow. 1990. Formal aspects of long distance scrambling. *Unpublished Paper*.

Tilman Becker, Aravind K. Joshi, and Owen Rambow. 1991. Long-distance scrambling and tree adjoining grammars. In *Proceedings of the 5th meeting of the European Chapter of the Association for Computational Linguistics (EACL)*, Berlin, Germany.

Benoit Crabbé, Denys Duchier, Claire Gardent, Joseph Le Roux, and Yannick Parmentier. 2013. XMG: eXtensible MetaGrammar. *Computational Linguistics*, 39(3).

Katalin É. Kiss. 2005. *The Syntax of Hungarian*. Cambridge University Press.

Robert Frank. 2002. *Phrase Structure Composition and Syntactic Dependencies*. MIT Press, Cambridge, MA.

Aravind K. Joshi and Yves Schabes. 1997. Tree-Adjoining Grammars. In G. Rozenberg and A. Salomaa, editors, *Handbook of Formal Languages*, pages 69–123. Springer, Berlin.

Laura Kallmeyer and Rainer Osswald. 2012. An analysis of directed motion expressions with lexicalized tree adjoining grammars and frame semantics. In L. Ong and R. de Queiroz, editors, *WoLLIC 2012, Lecture Notes in Computer Science LNCS 7456*, pages 34–55. Springer.

Laura Kallmeyer and Maribel Romero. 2008. Scope and situation binding in LTAG using semantic unification. *Research on Language and Computation*, 6(1):3–52.

Hilda Koopman and Anna Szabolcsi. 2000. *Verbal Complexes*. MIT Press.

Rebecca Nesson, Stuart M. Shieber, and Giorgio Satta. 2010. Complexity, parsing, and factorization of tree-local multicomponent tree-adjoining grammar. *Computational Linguistics*, 36(3):443–480.

Krisztina Szécsényi. 2009. *An LF-driven Theory of Scrambling in Hungarian Infinitival Constructions*. Ph.D. thesis, Szeged University, Hungary.

Robert Van Valin Jr. 2005. *Exploring the syntax-semantics interface*. Cambridge University Press.

David J. Weir. 1988. *Characterizing Mildly Context-Sensitive Grammar Formalisms*. Ph.D. thesis, University of Pennsylvania.