Hungarian verbal complexes and the pre-verbal field: towards an MCTAG analysis

Kata Balogh
HHU Düsseldorf

16th Szklarska Poreba Workshop
20-23 February 2015

Motivation for (L)TAG

• TAGs are mildly context-sensitive
  - parsing in polynomial time
  - generation of crossing dependencies
  - constant growth property (semilinearity)

• large coverage TAG grammars
  - English and Korean (XTAG; Joshi et al.)
  - French TAG (Crabbé’s PhD-thesis;)
  - German (GerTT; Kallmeyer & Lichte)

• grammar implementation with TAG
  - XTAG tools (UPenn) → parser, editor, viewer, . . .
  - XMG + TuLiPA (Tübingen)
    - TuLiPA: Tübingen Linguistic Parsing Architecture (Parmentier et al, 2008)

(L)TAG: Basics

• Tree Adjoining Grammar (TAG) is a set of elementary trees
  - a finite set of initial trees
  - a finite set of auxiliary trees

• two combinatorial operations
  - substitution: replacing a non-terminal leaf with an initial tree
  - adjunction: replacing an internal node with an auxiliary tree

• LTAG: Lexicalized TAG
  - each elementary tree contains at least one lexical item

Outlook

• goal: Hungarian grammar using TAG + XMG
  (Tree-Adjoining Grammar + eXtensible MetaGrammar)

• today: some problems and analysis around free word order
  - a very quick introduction to (L)TAG
    - (Lexicalized TAG)
  - grammar writing with XMG
  - Hungarian data: verbal complexes and the pre-verbal field
  - some results and proposed analysis using XMG & MCTAG
    - (Multi-component TAG)
(L)TAG: Basics

- to increase the expressive power: **adjunction constraints**
  - whether adjunction is mandatory and which trees can be adjoined:
    Null Adjunction (NA), Obligatory Adjunction (OA), Selective Adjunction (SA)
- **feature structures** as non-terminal nodes; reasons wrt TAG:
  - generalizing agreement and case marking (via underspecification)
  - modeling adjunction constraints ⇒ smaller grammars that are easier to maintain

```
    cat  np
   [case  nom]
    [agr  [pers  3]
     [num  sg]]
      [cat  s]
        [cat  vp
         [agr  [pers  3]
          [num  sg]]]

       she
      sings
```

Linguistic analyses with LTAG

- the ideal grammar formalism ⇒ linguistically adequate:
  - **phenomena**: linearization, agreement, discontinuity, ellipsis, ...
  - **generalizations**: valency, active/passive diathesis, alternations, ...
  - intuitive implementation
- LTAG = set of elementary trees
  What is an elementary tree, and what is its shape?

  ⇒ Syntactic design principles (Frank, 2002):
    - Lexicalization
    - Condition on Elementary Tree Minimality (CETM)
    - Fundamental TAG Hypothesis (FTH)
    - θ-Criterion for TAG
  ⇒ Semantic design principles (Abeillé & Rambow, 2000)
  ⇒ Design principle of economy

Syntactic design principles

**Fundamental TAG Hypothesis (FTH)**
Every syntactic dependency (subcategorization, binding, ...) is expressed locally within an elementary tree.

**θ-Criterion for TAG**
(a) If H is the lexical head of an elementary tree T, H assigns all of its θ-roles in T.
(b) If A is a frontier non-terminal in T, A must be assigned a θ-role in T.

⇒ Valency/subcategorization is expressed within the elementary tree of the predicate: either a substitution node or a footnode

```
S
NP↓ VP

likes
```

```
S
NP↓ VP

wants
```

Grammar architecture

```
Morph Database

Syntactic Database

Tree Database
```

inflected form
root form, POS, inflectional information
list of tree templates and tree families
Tree templates and tree families

- **Tree templates**: e.g. for the declarative transitive verb and the transitive verb with object extraction
  - marks the lexical insertion site

```
S
NP↓ VP
V↓ NP↓
|   |
|   |
```

- **a tree family**
  - is a set of tree templates,
  - represents a subcategorization frame, and
  - unifies all syntactic configurations the subcategorization frame can be realized in

Hungarian and LTAG

- **LTAG** → fixed positions for grammatical functions
- **flexible word order?**
- **larger tree families, larger set of elementary trees; e.g.**

```
S
VP
NPnom↓ NPacc↓

V
sings
```

... etc.

XMG

- **eXtensible MetaGrammar** ⇒ specifying an F-LTAG
  - **LTAG set of elementary trees** ⇒ most information contained in the elementary trees
  - **XMG** ⇒ generate the elementary trees for a given grammar
- **meta-grammar** ⇒ expressing generalizations
  - **additional abstraction level**
  - **factoring out reusable tree-fragments: classes; e.g.**
    - **Subject** position in English or **Topic/Focus** positions in Hungarian
      ⇒ appearing in elementary trees of verbs with different subcategorization frames
  - **classes (tree-fragments) can be combined by conjunction and disjunction**

---

Hungarian

- **flexible word order & discourse configurationality**
  - **w.r.t. information structure**: post-verbal and pre-verbal field
    - **post-verbal field**: “argument positions” ⇒ order is free
      1. Adott Pim egy könyvet Marinak.
         gave Pim.nom a book.acc Mary.dat
      2. Adott Marinak Pim egy könyvet.
         gave Mary.dat Pim.nom a book.acc
      - all 6 permutations: ‘Pim gave a book to Mary.’
    - **pre-verbal field**: “functional projections” ⇒ fixed order
      1. Topic* < Quantifier* < Focus < Verb < ...
      3. Marinak mindenki egy könyvet adott.
         Mary.dat everyone.nom a book.acc gave ‘It was a book, that everyone gave to Mary.’
by combining tree fragments → tree templates; e.g.
Subject → CanSubject ∨ WhNpSubject
Object → CanObject ∨ WhNpObject
ActiveTransVerb → Subject ∧ ActiveVerb ∧ CanObject

● description language for tree fragments
   class CanSubj
declare ?S ?VP ?NP
   { <syn> {
     node ?S (color=black) [cat=s] ;
     node ?NP (mark=nbst,color=black) [cat=np] ;
     node ?VP (color=white) [cat=vp] ;
   } }

● dominance (→) and precedence (>>) also with transitive closure
● color codes for specifying node equations

XMG & Hungarian

● pre-verbal field → fixed positions for topic & focus
● arguments can be in topic position
  SubjTop    ObjTop    ... etc.

  S
  NP↓ VP

  • NP [case nom] < S
  • NP [case acc] < S

● arguments can be in focus position
  SubjFoc    ObjFoc    ... etc.

  S
  NP↓ VP

  • NP [case nom] < VP [vminv +/nil]
  • NP [case acc] < VP [vminv +/nil]

● also implemented: verbal modifiers, sentential negation

XMG & Hungarian

● post-verbal field → free argument order
● example (without verbal prefixes); the tree fragments:
  SProj, NoVMVerb

  S
  VP

  V

  • subject, object and oblique argument in post-verbal (argument) position
  SubjArg    ObjArg    OblArg

● DiTrVerbPV → SProj ∧ NoVMVerb ∧ SubjArg ∧ ObjArg
  providing all 6 elementary trees

Verbal complexes

● verb + verbal modifier (VM) → verbal complexes
● verbal modifiers: broad group
  ● verbal prefixes
  (4) Pim meg-látogatta Marit.
      Pim Pref-visited Mary.acc
      ‘Pim visited Mary.’

  ● infinitives without VM
  (5) Pim úszni akar.
      Pim swim.inf wants
      ‘Pim wants to swim.’

  ● DPs, adjectives, bare nouns
  ● syntax: complementary distribution with negation and focus
  ● semantics: secondary predication, relation to aspect
Syntactic position

- in neutral sentences (without Foc and Neg)
  pre-verbal position
  (6) Pim meg-látogatta Marit.
      Pim Pref-visited Mary.acc
      ‘Pim visited Mary.’

- in non-neutral sentences (with Foc and/or Neg)
  post-verbal position
  (7) Pim nem látogatta meg Marit.
      Pim not visited pref Mary.acc
      ‘Pim did not visit Mary.’
  (8) Pim MARIT (nem) látogatta meg.
      Pim Mary.acc (not) visited pref
      ‘It is Mary, whom Pim (did not) visited.’

Verbal complexes and clausal complements

- two types of control verbs
  - e.g. fel ‘is afraid’ → take main stress
  - e.g. akar ‘want’ → avoids main stress

- different behavior wrt the VM of the embedded infinitive
  (9) Pim (el*) fel el-olvasni a levelet.
      Pim (Pref) is-afraid Pref-read.inf the letter.acc
      ‘Pim is afraid to read the letter.’
  (10) Pim el akarja (el*) olvasni a levelet.
      Pim Pref wants (Pref-)read.inf the letter.acc
      ‘Pim wants to read the letter.’

- Koopman-Szabolcsi (2004) classification:
  - Auxiliaries: no main accent, VM-climbing (e.g. akar ‘want’)  
  - Nonauxiliaries 1: main accent, no VM-climbing (e.g. fel ‘is afraid’) 

Infinitival complements

- two verbs ⇒ both with pre-verbal and post-verbal fields
  [... pre-V ...] matrix-V [... post-V ...] [... pre-V ...] inf-V [... post-V ...]

- preferred position of a focused/topicalized argument of the embedded
  verb is in the pre-verbal field of the matrix verb

- need to deal with scrambling ⇒ Multi-component TAG (MCTAG)
Nonauxiliaries

Example: fél ‘is afraid’

(11) Pim fél el-olvasni a levelet.

Pim is-afraid Pref-read.inf the letter.acc

‘Pim is afraid to read the letter.’

- neutral sentences (no Foc) → standard LTAG analysis

```
NP
  S
  VP
  VM V NP
  V S
      ʃ
      fél
```

(12) ?Pim fél [a levelet]\textsuperscript{T} el-olvasni.

Pim is-afraid the letter.acc Pref-read.inf

(13) Pim [a levelet]\textsuperscript{T} fél el-olvasni.

Pim the letter.acc is-afraid Pref-read.inf

‘Pim is afraid to read the letter.’

(14) *Pim fél [a LEVELET]\textsuperscript{F} olvasni el.

Pim is-afraid the letter.read.inf Pref

(15) Pim [a LEVELET]\textsuperscript{F} fél el-olvasni.

Pim the letter.acc is-afraid Pref-read.inf

‘It is the letter, that Pim is afraid to read.’

Auxiliaries

Example: akar ‘want’

(16) Pim el akarja el-olvasni a levelet.

Pim Pref wants read.inf the letter.acc

‘Pim wants to read the letter.’

(17) *Pim akarja el-olvasni a levelet.

Pim wants Pref.read.inf the letter.acc

- standard LTAG analysis cannot derive the ‘VM-climbing’

```
NP
  S
  VP
  VM V NP
  V S
      ʃ
    akar
```

(18) ?Pim el akarja [a levelet]\textsuperscript{T} olvasni.

Pim Pref wants the letter.read.inf

(19) Pim [a levelet]\textsuperscript{T} el akarja olvasni.

Pim the letter.acc Pref wants read.inf

‘Pim wants to read the letter.’

(20) *Pim el akarja [a LEVELET]\textsuperscript{F} olvasni.

PimPref wants the letter.acc read.inf

(21) Pim [a LEVELET]\textsuperscript{F} akarja el-olvasni.

Pim the letter.acc wants Pref-read.inf

‘It is the letter, what Pim wants to read.’
Multi-component TAG

- standard (L)TAG cannot analyze
  - discontinuity (extraposition, extraction, scrambling)
  - ellipsis (gapping, subject deletion, right node raising)
- Scrambling: challenge ⇒ variability in word order; German example:
  - daß ihm Peter den Kühlschrank heute zu reparieren versprach
däß ihm den Kühlschrank Peter heute zu reparieren versprach...
  - (‘that Peter promised him to repair the fridge today’)
- Problem: if ihm is considered to be an argument/complement of versprach, the tree for versprach has to split into three pieces when conjoined with the tree of zu reparieren
- possibilities in an (L)TAG-Analysis:
  - zu reparieren adjoins to versprach ⇒ contradicts θ-criterion
  - ihm adjoins to zu reparieren ⇒ contradicts θ-criterion

MCTAG - Basics

- multi-component TAG → elementary structures are sets of trees
- tree-local MCTAG (TT-MCTAG)
  - all trees in the set have to attach to the same elementary tree
  - strongly equivalent to TAG
- set-local MCTAG
  - all trees in the set have to attach to the same elementary tree set
  - weakly equivalent to LCFRS and simple RCG
- non-local MCTAG
  - the fixed recognition problem is NP-complete (even with lexicalization and dominance links)
  - mainly TT-MCTAGs are considered for natural language grammars due to complexity issues

MCTAG - German scrambling

TT-MCTAG can handle scrambling up to two levels of embedding, i.e. three verbs with one complement each forming a coherent construction. [Joshi et. al 2000]

 daß ihm den Kühlschrank Peter zu reparieren versprach
  (‘that Peter promised him to repair the fridge’)

VP
NP_{acc}↓

VP
NP_{dat}↓

VP*
NP_{nom}↓

VP*
VP*

V

zu reparieren

versprach

TT-MCTAG for Hungarian

- Auxiliaries vs. Nonauxiliaries (fél ‘is-afraid’ vs. akar ‘want’)
- proposal using TT-MCTAG
- elementary trees of the infinitival verb el-olvasni ‘Pref-read.inf’

obtained by XMG as before
• tree sets for fél 'is-afraid' and want 'want' and

\[
\begin{align*}
\{ & \quad S \quad | \quad V P_{\text{av-str} -} \to fél \\
NP_{\text{nom}} \quad S^* \quad V \quad VP^* \to \text{fél} \\
\} \\
\{ & \quad S \quad | \quad V P_{\text{av-str} +} \to \text{olvasni} \\
NP_{\text{nom}} \quad S^* \quad V \quad VP^* \to \text{akar} \\
\}
\end{align*}
\]