ANALYSIS BY REDUCTION OF ANALYTICAL PDT-TREES

M. Plátek, K. Oliva, D. Pardubská

NCMA 2017, Prague

Introduction I

long-term goal

- systematic effort to support the development of Functional (Generative) Description of Czech language (FGD) by a formally-theoretical apparatus

today's goals

- **proposal of formal tools** for an exact direct introduction of analysis by reduction of lexicalized trees, and for a study of the complexity of this analysis by reduction

  !! analysis by reduction **was considered to be an informal method**, based purely on the linguistic intuition - introspection

- preparation of a verified material for an incremental proposal and testing of a formal description of a grammar of Czech language, and/or for a grammar-checker for Czech language

  !! formal description by special types of restarting automata using certain types of meta-instructions (similar to ITAT 2013, FG 2014)

Introduction II

we focus on analytical trees (A-trees)

- A-trees are constructed by the method of Prague Dependency Tree-bank (PDT) with dependencies and coordinations (only)

- the analysis by reduction of A-trees, in particular the observations of how coordinations are treated, create the main novelty of the contribution

- we present observations about A-trees trying to show that the complexity of reductions of all (Czech) A-trees is limited in several senses

This presentation is informal, the technicalities are left out.

TFAR and SFAR

we introduce

- full analysis by reduction of A-trees (TFAR - Trees Full Analysis by Reduction)
- Sentence Full Analysis by Reduction (SFAR) that works with tagged sentences (i.e. with strings of word forms and punctuation marks enriched with morphological and syntactic information) instead of A-trees

for TFAR and SFAR we introduce

- several types of complexity measures
- complexity constraints formulated by properties of reductions

  we mainly work with constraints typical for TFAR here

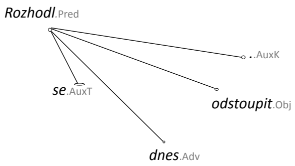
- two types of (non)stability for introduced constraints which enables us to formulate new exact observations and propositions about syntax of Czech A-trees

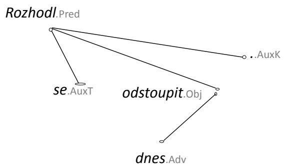
(without formal grammars or automata)

Sentence (1) and two possible translations and A-trees

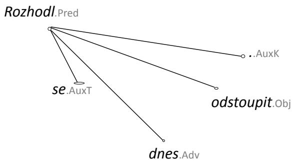
(1) *Rozhodl* .Pred *se* .AuxT *dnes* .Adv *odstoupit* .Obj *.* .AuxK
'(He) decided – REFL – today – (to) resign – .'

' He decided today to resign. ' He decided to resign today.





A-trees



A-tree
a tree structure with oriented edges and total (horizontal) node ordering

Orientation: The paths in an A-tree are oriented from leaves to the root (bottom up).

Nodes represent syntactically labeled lexical and punctuational items.

We are limited to tree structures with dependencies and coordinations only (without ellipses).

The total (horizontal) ordering of nodes expresses the word-order.

TFAR of an A-tree and SFAR of a tagged sentence

TFAR (of an A-tree) and SFAR (of a tagged sentence) are characterized by the following principles:

(i) TFAR (SFAR) consists of (maximal) branches (sequences) of continual reductions ; by a reduction we mean an ordered pair of A-trees (sentences) $T_1 \vdash T_2$, where T_2 arises from T_1 by performing a sequence of operations of two types: delete and shift.
Every reduction performs at least one delete.

!! since only deletes and shifts are considered, the forms of individual words (and punctuations), their morphological characteristics and their syntactic categories stay unchanged

(ii) Reductions preserve the correctness of structure; any reduction applied to a correct A-tree (sentence) results (needs to result) in a correct A-tree (sentence).

The remaining principles for TFAR and SFAR.

(iii) Reductions of TFAR (SFAR) do not belong to an a priori given set of so called forbidden reductions (exceptions).

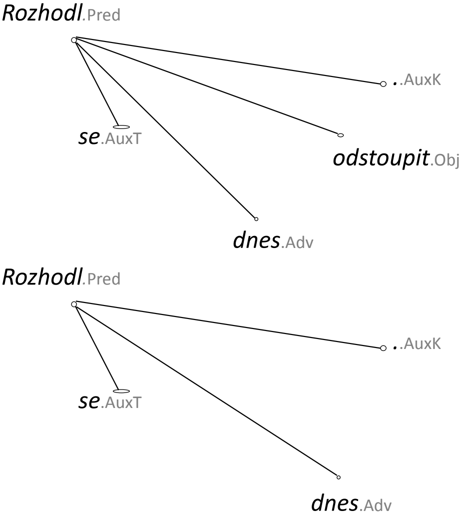
!! an example of such a forbidden reduction (exception) is a deletion of a sole preposition

(iv) Each reduction is minimal in the sense that omission of at least one operation from the reduction would violate the correctness of resulting A-tree (ii) or change the reduction to a forbidden one (iii) (or both).

(v) TFAR (SFAR) consists of all possible branches of reductions fulfilling the principles (i) to (iv).

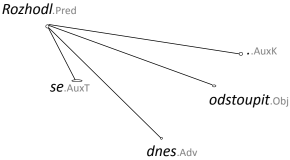
The reduction $T1_1 \vdash T1_3$.

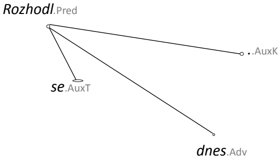
- He decided today to resign.
- He decided today.



The reduction $T1_1 \vdash T1_3$.

‘ He decided today to resign. $T1_1$ ‘ He decided today. $T1_3$



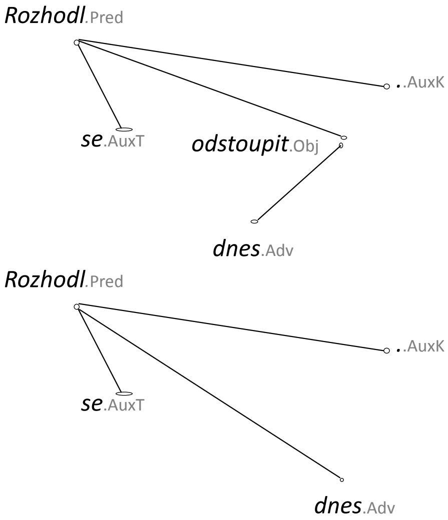


- The reduction deletes the node with the word *odstoupit/resign*, which is a leaf of $T1_1$.
- $T1_3$ does not contain a new edge (since a leaf was deleted).
- This formal reduction of a leaf node (and its appropriate edge) in the framework of TFAR corresponds to an informally understood reduction of meaning.
- The edge complexity of a reduction is the number of new edges created by the reduction.

  The edge complexity of the reduction $T1_1 \vdash T1_3$ is 0.

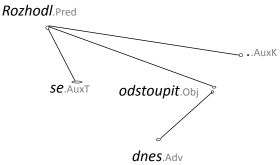
The reduction $T1_2 \vdash T1_3$

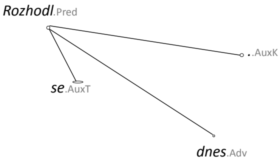
- He decided to resign today.
- He decided today.



The reduction $T1_2 \vdash T1_3$.

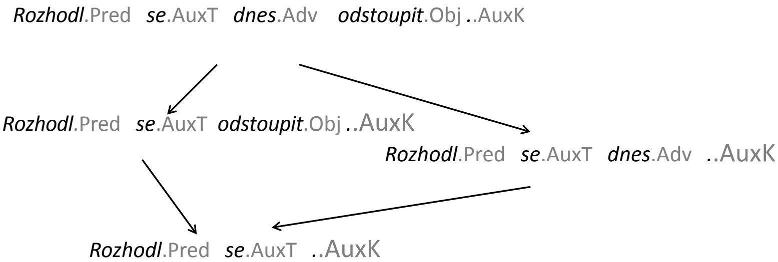
‘ He decided to resign today. $T1_2$ ‘ He decided today. $T1_3$





- The reduction deletes the node with the word *odstoupit/resign*, which is an internal node of $T1_2$.
- $T1_3$ contains an edge which is not in $T1_2$.
- Such reduction induces a change of meaning which cannot be understood as pure "reduction of meaning" (as it was in the previous case)
- The edge complexity of the reduction $T1_2 \vdash T1_3$ is 1.

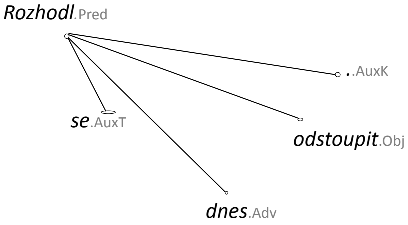
The SFAR of the sentence (1)

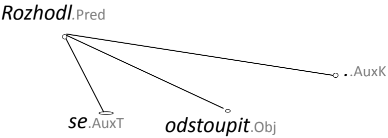


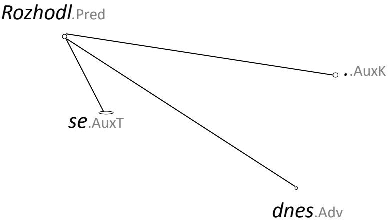
The SFAR of the sentence (1) creates a lattice.

TFAR of the A-tree $T1_1$, the first part

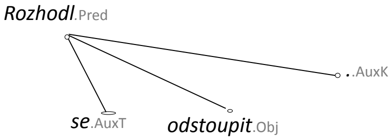
(1) *Rozhodl*.Pred *se*.AuxT *dnes*.Adv *odstoupit*.Obj ..AuxK

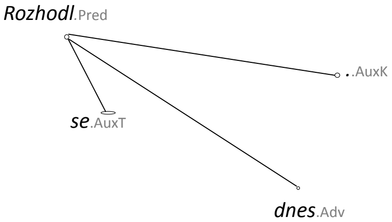


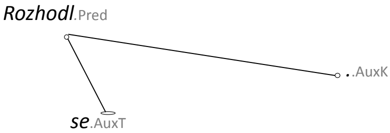




TFAR of the A-tree $T1_1$, the second part







no new edge was created within the reductions

ed-complexity of (the TFAR of) A-tree $T1_1$

TFAR(t, T_P, Z_P) denotes the TFAR of t with respect to T_P and Z_P

often only TFAR(t)

T_P - the set of correct Czech A-trees, $t \in T_P$
Z_P - the set of Czech prohibited reductions (exceptions)

- ed-complexity of the reduction is the number of new edges arising by a reduction

  All reductions of TFAR($T1_1, T_P, Z_P$) have the ed-complexity equal to 0

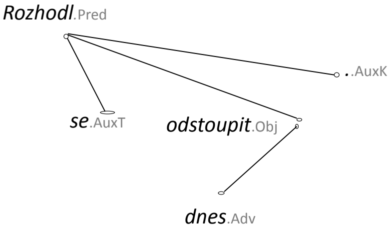
- ed-complexity of TFAR(t, T_P, Z_P) is the maximum of the ed-complexity of all its reductions

- ed-complexity of t is the ed-complexity of TFAR(t, T_P, Z_P)

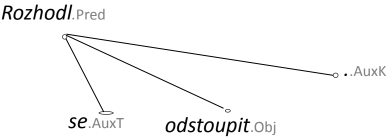
  $\hookrightarrow$ the ed-complexity of $T1_1$ is equal to 0

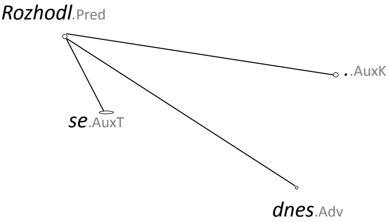
TFAR of $T1_2$, the first part.

(1) *Rozhodl*.Pred *se*.AuxT *dnes*.Adv *odstoupit*.Obj ..AuxK

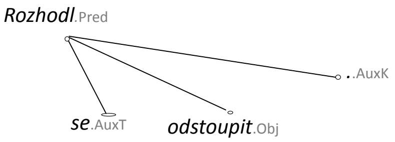


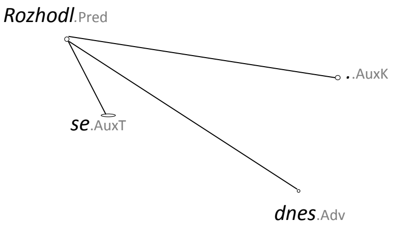
new edge has arisen

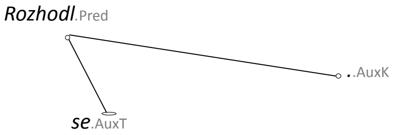




TFAR of $T1_2$, the second part.







TFAR of $T1_2$ and its ed-complexity

- One reduction of TFAR($T1_2$) has the edge complexity equal to 1, the other reductions of TFAR($T1_2$) have the ed-complexity equal to 0
  $\rightarrow$ the ed-complexity of $T1_2$ is equal to 1.

- $T1_2$ is a pure dependency tree, i.e. it has only dependency edges. TFAR of $T1_2$ contains a reduction with a new dependency edge, i. e. the resulting tree bears a meaning which is not a pure reduction of the meaning of the original tree.

- The natural linguistic requirement on the intuitive analysis by reduction is to exclude the reductions with such a shift in the meaning.

  !! For grammar-checking we need not work with this requirement.

Let i is a natural number, TFAR(s) = TFAR(s, T_P, Z_P), and ed denotes the edge complexity of reductions. Then

- TFAR(s, $ed \leq i$) denotes the subset of TFAR(s) such that ed of its reductions is not greater than i
- we say the A-tree s (resp. TFAR(s)) is T-stable for $ed = i$, if TFAR(s) = TFAR(s, $ed \leq i$)

Generally, T-stability expresses for the given constraint the total consistence with TFAR(s).

We say that the A-tree s (resp. TFAR(s)) is Mn-stable for $ed = i$ if the set of irreducible A-trees from TFAR($s, ed \leq i$) is equal to the set of irreducible A-trees from TFAR(s).

Mn-stability is introduced in order to express a weaker type of consistence to TFAR(s) than T-stability. It should serve as an (upper) bound for a suitable approximation by individual or composed constraints for the linguistically intuitive analysis by reduction.

First type of results.

We present two types of results:
1) propositions (of a mathematical type)
and
2) linguistic observations (L-observations).

The next propositions follows directly from definitions.

Proposition A. $T1_1$ is T-stable (therefore also Mn-stable) for $ed \geq 0$.

Proposition B. $T1_2$ is Mn-stable for $ed \geq 0$
and
$T1_2$ is not T-stable for $ed = 0$.

Proposition C. $T1_2$ is T-stable for $ed \geq 1$.

Second type of results: L-observation

L-observation 1. Let $s \in T_p$ be an A-tree without coordinations. Then s is Mn-stable for $ed = 0$, i.e. we have not found any A-tree in T_P which is not Mn-stable for $ed = 0$.

Our interpretation of the previous observation is that the intuitive analysis by reduction can be on pure dependency A-trees performed (simulated) with the ed-constraint equal to 0.

L-observation 2. Let $s \in T_p$ be an A-tree without coordinations. Then TFAR(s) creates a lattice (contains exactly one irreducible sentence).

Propositions and L-observations for coordinations

are different from

propositions and L-observations for dependecies.

We will show that by an example of an embedded coordination.

An example of an embedded coordination.

(5) *Pracujeme*.Pred. Co *a*.Cr.Co *myslíme*.Pred.Co *i*..Cr *jednáme*.Pred. Co..AuxK

'(We) work – and – think – and (also) – act –.'

'We work, and think, and (also) act.'

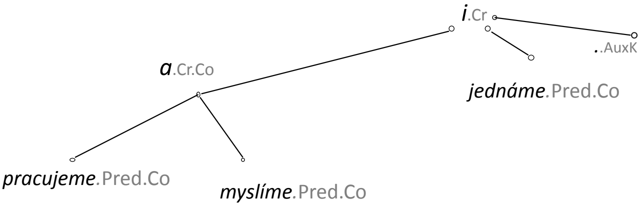


Figure: A-tree $T5_1$ with an embedded coordination without any dependency edge.

SFAR of the sentence (5) with two irreducible sentences.

Pracujeme. Pred.Co *a*.Cr.Co *myslíme*. Pred.Co *i* .Cr *jednáme*. Pred.Co **.** .AuxK

Pracujeme. Pred.Co *i*. Cr *jednáme*. Pred.Co **.** .AuxK

myslíme. Pred.Co *i* .Cr *jednáme*. Pred.Co **.** .AuxK

SFAR of the sentence (5) with an embedded coordination, and with two irreducible sentences.

The reduction $T5_1 \vdash T5_2$.

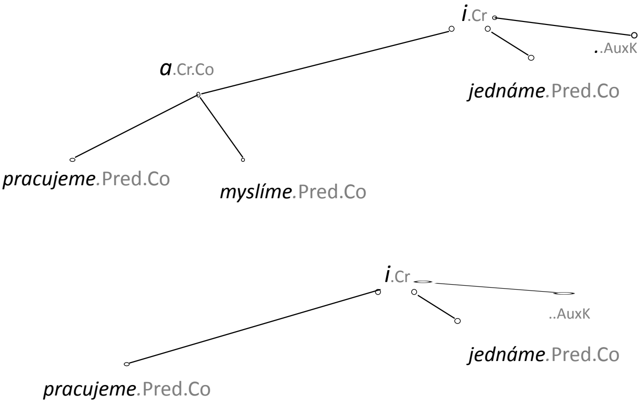


Figure: First reduction with the edge-complexity 1.

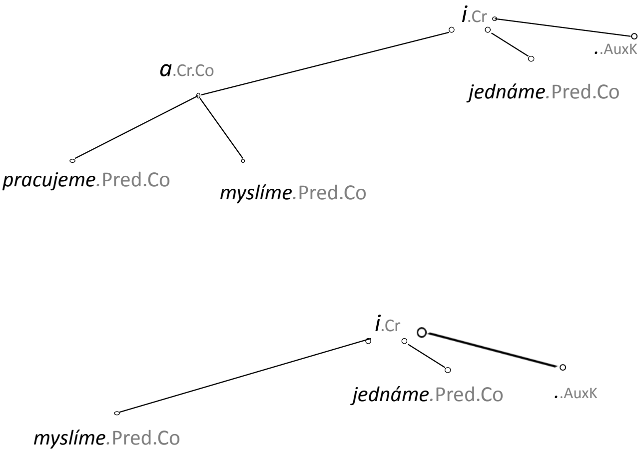


Figure: Second reduction with the edge complexity 1.

On Mn-stability.

Proposition D. $T5_1$ is T-stable for $ed \geq 1$.

Proposition E. $T5_1$ is not Mn-stable for $ed = 0$.

L-observation 3. Any A-tree from T_p with embedded coordinations, is not Mn-stable for $ed = 0$.

Remark. Both presented A-tree reductions with ed-complexity 1 correspond to an acceptable reduction of meaning. That makes a difference to the reductions of dependencies with the ed-complexity greater than 0.

Summary

We have presented in this talk results based on the edge constraints. In the proceedings contribution we have studied four other types of constraints with similar types of results. It follows from our previous work that all A-trees from T_P should be T-stable for reductions with at most 7 deletions and 2 shifts. This observation means that we should be able to write finite many meta-rules for correctness-preserving restarting automata which will be able to simulate (constrained) TFAR and SFAR for A-trees from T_P and their sentences.

We have focused on the constraints and properties which stress the difference between dependencies and coordinations.

Conclusion

Main aim of our contribution was to present new techniques for the exact study of dependency based syntax.
Let us note that the concepts of SFAR and TFAR can have a direct impact for a further development of grammar-checking.

We have used methods which are not far from analytical models methods from fifties and sixties of the last century (Marcus, Novotny, Nebesky, Kunze).

Thank you for your attention

Multiple coordinations

The graph-(dis)continuity of a reduction

An example of multiple coordinations

(3) *Je*.Pred *dědou*.Obj.Co ,.AuxX *otcem*.Obj.Co *a*..Cr *strýcem*.Obj.Co..AuxK

‘(He) is – (a) grandfather – , – (a) father – and – (an) uncle.’

‘He is a grandfather, a father, and an uncle.’

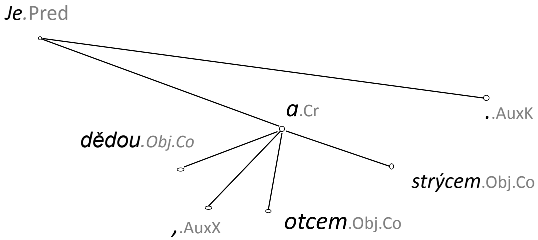
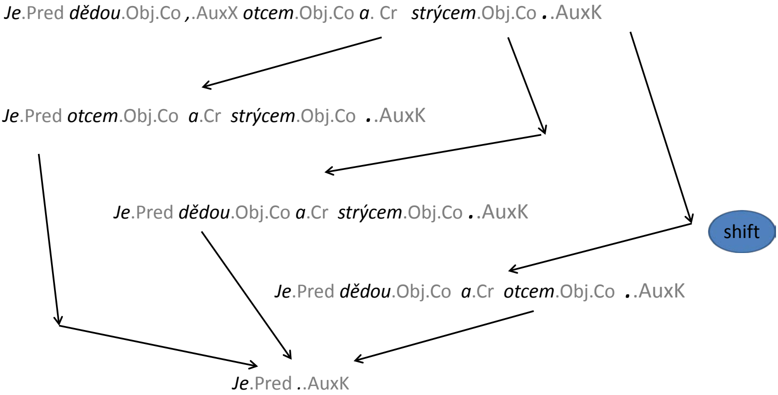


Figure: A-tree $T3_1$.

SFAR of the sentence (3)



A reduction of a multiple coordination.

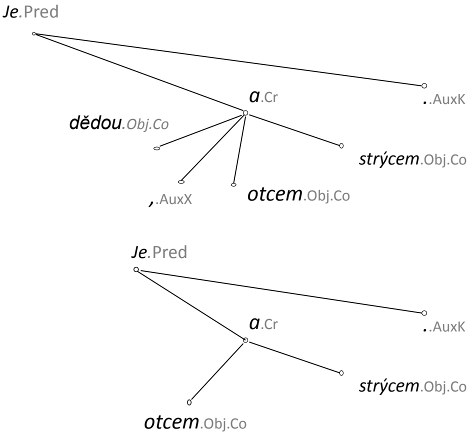


Figure: $T3_1$ reduced to $T3_2$.

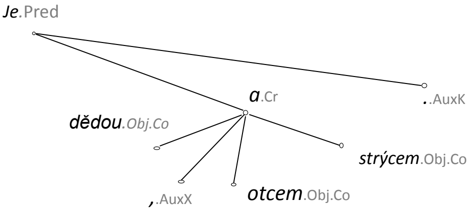
Two types of measures of discontinuity of a reduction

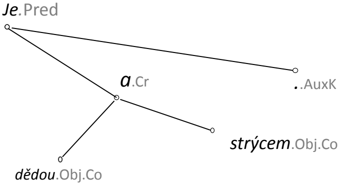
We denote as *nc* the number of components of the graph which was removed during a reduction.

We denote as Δ the minimal number of nodes which complete the graph removed by a reduction to a continuous sub-tree of the reduced A-tree.

The reduction $T3_1 \vdash T3_2$ fulfills $ed = 0$, $nc = 2$, and $\Delta = 1$.

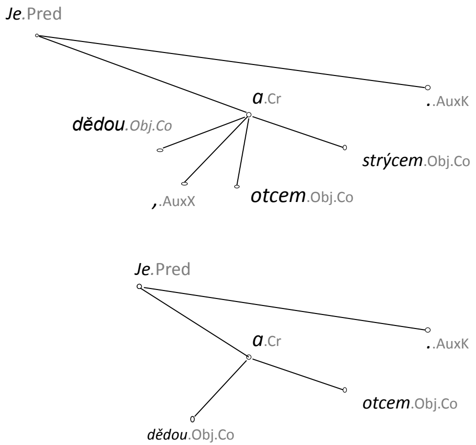
The reducion $T3_1 \vdash T3_3$.





This reduction fulfills $ed = 0$, $nc = 2$, and $\Delta = 1$.

The reducion $T3_1 \vdash T3_4$.



This reduction fulfills $ed = 0$, $nc = 2$, and $\Delta = 1$, and it uses one shift.