M A T E M A T I C K O - F Y Z I K Á L N Í F A K U L T A

P R A H A

ANNOTATION ON THE TECTOGRAMMATICAL LEVEL. ADDITIONS TO ANNOTATION MANUAL (WITH RESPECT TO PDTSC AND PCEDT)

MARIE MIKULOVÁ

ÚFAL Technical Report
TR-2013-52

ISSN 1214-5521

U N I V E R S I T A S C A R O L I N A P R A G E N S I S

Copies of ÚFAL Technical Reports can be ordered from:

Institute of Formal and Applied Linguistics (ÚFAL MFF UK)
Faculty of Mathematics and Physics, Charles University
Malostranské nám. 25, CZ-11800 Prague 1
Czech Republic

or can be obtained via the Web: http://ufal.mff.cuni.cz/techrep

Annotation on the Tectogrammatical Level
Additions to Annotation Manual
(with respect to PDTSC and PCEDT)

Marie Mikulová

Translation:
Blanka Cajthamlová
Majda Rysová

Content

Introduction

The full tectogrammatical annotation of the Czech texts was first presented *in The Prague Dependency Treebank 2.0* (PDT 2.0), published in 2006 (Hajič et al., 2006). The annotation principles and guidelines are described in the manual (available also in an English version), which is currently published as a technical report (Mikulová et al., 2005 (Czech version) and 2006a (English versions)) and in electronic form as an HTML document (on the Internet at http://ufal.ms.mff.cuni.cz/pdt2.0/doc/manuals/en/t-layer/html/index.html). In addition to the extensive annotation manual, a shortened reference book (Mikulová et al., 2006b and 2006c) is available.

A similarly based annotation has been used for other Prague treebanks. *The Prague Czech-English Dependency Treebank 2.0* (PCEDT 2.0; Hajič et al., 2011) contains parallel PDT-like annotations of English texts (Wall Street Journal part of Penn Treebank) and their professional translation into Czech. Further information can be found at the project web page (http://ufal.mff.cuni.cz/pcedt2.0/). *The Prague Dependency Treebank of Spoken Czech 2.0* (PDTSC 2.0; planned to be published in 2014) contains spontaneous dialogue speech, transcribed, reconstructed and further annotated in the PDT style.

PCEDT 2.0 and PDTSC 2.0 do not cover all features captured by the PDT 2.0. On the tectogrammatical level, we have annotated dependency structure, functors, verbal valency (substantial valency only in limited extent), references to lower layer and grammatical and textual coreference. In comparison with the PDT 2.0, we have not realized annotation of grammatemes, node types, topic-focus articulation and some other attributes (e. g. quot_type attribute; see Mikulová – Štěpánek, 2009).

Restrictions of annotation required some modifications and complements of existing rules (particularly the rules of tectogrammatical lemma and valency annotation). Our annotation of new texts (containing new syntactic phenomena, new examples) also brought new knowledge. Existing rules have been specified and modified and for some phenomena new rules have been established.

Note: There is a reference to the original annotation manual (Mikulová et al., 2006a) in each of the additions. The references start with the abbreviation MAN followed by the number and name of the chapter.

Acknowledgements

The research was supported by the Ministry of Education, Youth and Sport within the project LINDAT-Clarin LM2010013 and by Grant Agency of the Czech Republic within the project P406/2010/0875.

References to analytical level

Substitute for MAN 2.1 Relation between the tectogrammatical level and the lower levels

Legend for examples> <*a-aux-node*> *a-lex-node* [attribute = value]
The expression annotated as a-aux-node is in pointed brackets (< >). Underlined expression is annotated as a-lex-node. In every example, annotated a-nodes are, always for a single t-node, marked by pointed brackets and underlining. In square brackets ([ ]) behind the example, values of some other attributes of a given t-node (particularly its t-lemma) can be represented. Curly brackets ({ }) are used to indicate a t-node that was filled in.

1 Two types of references to a-level

The tectogrammatical level is interconnected with the analytical level. Relations between the levels are represented so that there is a reference leading from a t-node to every a-node that somehow influences the value of the t-node attribute. We distinguish two types of references to a-level depending on the type of attribute that influences the a-node. Annotated a-nodes, i.e. a-nodes that influence values of t-nodes attributes, are then marked as:

- **a-lex-node** = a-node that gives a lexical meaning, or its largest part, to the t-node.
  A-lex-node influences the value of *t-lemma* attribute.
  It is usually a a-node that represents an autosemantic word.
- **a-aux-node** = other referenced a-nodes.
  A-aux-nodes usually influence values of such attributes as *functor, subfunctor* and *gram*.
  These are usually a-nodes that represent functional words: prepositions, conjuntions, auxiliary verbs, anaphoric words.

There is always only one a-lex-node in a t-node (if it exist).
Reference can lead from more t-nodes to a single a-node.

2 Determination of a-lex-node and a-aux-nodes

2.1 Basic rules

Rule A:
Coreferred a-nodes in a t-node that represents **name**:

- **a-lex-node**: a-node that represents substantive, adjective, numeral, pronoun.
- **a-aux-nodes**: a-nodes that represent preposition
  (expression *se* – with reflexive deverbal word, see Rule D),
  (period behind the ordinal number – see Rule Q).

Examples:

(1) <*na*> <u>*zahradě*</u> [t_lemma = *zahrada*]
(=*in garden* [t_lemma = *garden*])

(2) NB! <*přes*> <u>*dvacet*</u> *korun* [t_lemma = *dvacet*]
(=*more_than twenty crowns* [t_lemma = *twenty*])

(3) <*na*> <u>*jabloni*</u> *a hrušni* [t_lemma = *jabloň*]
(=*on* <u>*apple*</u> *(tree) and pear (tree)* [t_lemma = *apple*])

(4) <*na*> *jabloni a* <u>*hrušni*</u> [t_lemma = *hrušeň*]
(=*on apple (tree) and* <u>*pear*</u> *(tree)* [t_lemma = *pear*])

(5) <*Ve*> <*srovnání*> <*s*> <u>*Pavlem*</u> *jsem lepší.* [t_lemma = *Pavel*]
(=*In comparison with* <u>*Pavel*</u>*, I am better.* [t_lemma = *Pavel*])

Rule B:

Coreferred a-nodes in a t-node that represents **<u>verb</u>** (predicate):

- **a-lex-node**: a-node that represents full verb.
- **a-aux-nodes**: a-nodes that represent: auxiliary verb,
  modal verb,
  supporting expression,
  subordinating conjunction,
  (expression *se* in passive – see Rule J),
  (expression *se* in dispositional modality – see Rule K),
  (expression *se* in reflexive verb – see Rule D).

Examples:

(6) *Zítra* <*nebudu*> <u>*dělat*</u> *nic.* [t_lemma = *dělat*]
(=lit. Tomorrow, I will_not do nothing. [t_lemma = *to_do*]*)*

(7) *Udělám to* <*proto*>*,* <*aby*> <*mohl*> <u>*přijít*</u>*.* [t_lemma = *přijít*]
(=I will do it in order for him to be able to come. [t_lemma = *to_come*]*)*

(8) <*Kdybys*> <*byl*> <*býval*> <u>*přišel*</u>*, nestalo by se to.* [t_lemma = *přijít*]
(=If you had come, it would not have happened. [t_lemma = *to_come*]*)*

(9) *Kdybys byl býval přišel,* <u>*nestalo*</u> <*by*> <*se*> *to.* [t_lemma = *stát_se*]
(=If you had come, it would not have happened. [t_lemma = *to_happen*]*)*

(10) <u>*Přišel*</u>*.* [t_lemma = *přijít*]
(=lit. (He) came. [t_lemma = *to_come*]*)*

Rule C:

Coreferred a-nodes with **<u>other t-nodes</u>** (that represent particles, interjections, coordinating conjunction):

- **a-lex-node**: a-node that represents particle, interjections, one-word coordinating conjunction.
- **a-aux-nodes**: none.

Examples:

(11) <u>*Pardon*</u>*.* [t_lemma = *pardon*]

(12) *na jabloni* <u>*a*</u> *hrušni* [t_lemma = *a*]
(=*on apple (tree) and pear (tree)* [t_lemma = *and*])

2.2 Special cases

2.2.1 Special cases I: T-nodes with multiword t-lemmas

Rule D:
Coreferred a-nodes in a t-node that represents **reflexive pronoun or verb**:

- **a-lex-node**: a-node that represents noun or full verb.
- **a-aux-nodes**: a-node that represents reflexive expression *se*.

Examples:

(13) *smát* <*se*> [t_lemma = *smát_se*] (*=to laugh)*
(14) *chovající* <*se*> [t_lemma = *chovající_se*] (*=behaving)*

Rule E:
Coreferred a-nodes in a t-node that represents **dependent part of an idiomatic expression** (*functor* = DPHR):

- **a-lex-node**: a-node that represents a governing word of the dependent part of an idiomatic expression (the word is part of the t-lemma). If the governing word of the dependent part of an idiomatic expression cannot be determined, the a-lex-node is the first full word (according to the surface structure) that belongs to the dependent part of an idiomatic expression.
- **a-aux-nodes**: a-nodes that represent other words belonging to the dependent part of an idiomatic expression.

Examples:

(15) *Běhá mi mráz <po> <zádech>.* [t_lemma = *mráz_po_zádech*]
(*=Shiver runs down my spine.* [t_lemma = *shiver_down_(the) spine*]*)*
(16) *Bere všechno <na> <lehkou> váhu.* [t_lemma = *na_lehkou_váhu*]
(*=He makes light of everything.)*

Rule F:
Coreferred a-nodes in a t-node that represents **multiword connective** (*nodetype = coap*):

- **a-lex-node**: a-node that represents the last (according to the surface structure) expression that belongs to a multiword connective.
- **a-aux-nodes**: a-nodes that represent other expressions that form a connective in the surface form of a sentence.

NB! Expressions that modify (even the multiword) connective are represented by individual t-nodes (*functor* = CM).

Examples:

(17) *<Buďto> půjdeme hned, nebo nepůjdeme vůbec.* [t_lemma = *buďto_nebo*]
(*=We will go either now or we will not go at all.* [t_lemma = *either_or*])
(18) *Udělám to, <a> to hned.* [t_lemma = *a_to*]
(*=I will do it, and now.* [t_lemma = *and*]*)*
(19) *Koupil chleba <a> rohlíky a housky.* [t_lemma = *a_a*]
(*=He bought bread and rolls and bagels.* [t_lemma = *and_and*]*)*

Rule G: (Applies only to PDT 2.0!)
Coreferred a-nodes in a t-node with the t-lemma of "**12 hours**" type:

- **a-lex-node**: a-node that represents adjective.
- **a-aux-nodes**: a-nodes that represent number.

Example:

(20) *<8> hodinová pracovní doba* [t_lemma = *8_hodinový*]
(=8 hours working time. [t_lemma = *8_hours*])

Rule H:
Coreferred nodes in a t-node with the t-lemma of **„number_number (_number)^{n}“** type:

- **a-lex-node**: a-node that represents first number.
- **a-aux-nodes**: a-nodes that represent other numbers.

Example:

(21) *telephone 345 <629> <456>* [t_lemma = *345_629_456*]
(22) *line 28 <40>* [t_lemma = *28_40*]

Rule I:
Coreferred a-nodes in a t-node with the t-lemma of **„van_Beethoven“** type:

- **a-lex-node**: a-node that represents name.
- **a-aux-nodes**: a-nodes that represent foreign preposition (or possibly apostrophe).

Existing list of prepositions (according to PDT 2.0): *von, van, O, d, de, da, di, zum*.
Example:

(23) *Ludwig <van> Beethoven* [t_lemma = *van_Beethoven*]

2.2.2 Special cases II: Passive, dispositional modality and reciprocity

Rule J:
A-node for the expression *se,* which is present in the sentence in order to express the **reflexive passive** is:

- **a-aux-node** in the effective root of the passive clause (in the t-node that represents governing verb in passive).

Example:

(24) *Košile <se> pere a kalhoty také.* [t_lemma = *prát*]
(=Shirt is_being washed and trousers too. [t_lemma = *to_wash*]*)*
(25) *Košile <se> pere a kalhoty {prát} také.* [t_lemma = *prát*] – in copied node
(=Shirt is_being washed and trousers (to wash) too. [t_lemma = *to_wash*]*)*

Rule K:
A-node for the expression *se,* which is present in the sentence in order to express **dispositional modality** is:

- **a-aux-node** in the effective root of the clause with dispositional modality (in the t-node that represents governing verb carrying dispositional modality, *gram/dispmod = disp1*).

Example:

(26) *Matematika <se> mi studuje dobře.* [t_lemma = *studovat (to study)*]
(=Mathematics is good to study for me.)

Rule L: NB! Difference from PDT 2.0!
A-nodes that represent formal indicators of **syntactic reciprocity** (expression *se* in various (prepositional) cases, expression *jeden druhý (=one, the other)*; see ADDreciprocity) are represented as:

- **a-aux-nodes** in newly established t-node with the t-lemma *#Rcp*.

Examples:

(27) *Pavel a Martin <na> <sebe> narazili {#Rcp*.PAT}.
(=Pavel and Martin ran into each other.)

(28) *Pavel a Martin narazili <jeden> <na> <druhého>* {*#Rcp*.PAT}.
(=lit. Pavel and Martin ran one into another.)

(29) *Martin a Radek se <mezi> <sebou> domluvili.* {*#Rcp*.ADDR}
(=Martin and Radek agreed mutually.)

2.2.3 Spicial cases III : Non-alphanumeric symbols

2.2.3.1 Special cases III : T-nodes representing non-alphanumeric symbols

These rules apply to t-nodes that represent non-alphanumeric symbols (*is_generated* = 0).

Rule M:
Coreferred a-nodes in a t-node that represents **one non-alphanumeric symbol**:

- **a-lex-node**: a-node that represents identical punctuation symbol.
- **a-aux-nodes**: none.

This rule applies to t-nodes with the t-lemma substitutes: *#Comma*, *#Dash*, *#Colon*, *#Slash*, *#Period*, *#Semicolon*, *#Ast*, *#Amp*, *#Percnt*.
Examples:

(30) *Mladost – radost.* [t_lemma = *#Dash*; *functor* = PRED *(=Youngness – happiness)*

(31) *20 %* [t_lemma = *#Percnt*]

Rule Na:
Coreferred a-nodes in the root of a paratactic structure (nodetype = coap), representing **punctuation symbol as a separating character of coordinated elements** (that are not connected by a conjunction):

- **a-lex-node**: a-node that represents the last punctuation symbol functioning as a separating character of coordinated elements.
- **a-aux-node**: none.

This rule applies to t-nodes with the t-lemma substitutes: *#Comma*, *#Dash*, *#Colon*, *#Slash*, or possibly: *#Period*, *#Semicolon*.
Example:

(32) *Máme psa, kočku, rybičky, křečka, andulku…*[t_lemma = *#Comma*]
(=lit. We have a dog, a cat, fishes, a hamster, a budgerigar…)

Rule Nb:
Coreferred a-node in a root of a paratactic structure (nodetype = coap), representing **punctuation symbol as a separating character of elements in apposition** (that are not connected by a conjunction):

- **a-lex-node**: a-node that represents the first punctuation symbol functioning as a separating character of elements in apposition.
- **a-aux-nodes**: none.

This rule applies to t-nodes with the t-lemma substitutes: *#Comma*, *#Dash*, *#Colon*, *#Slash*.
Examples:

(33) *Máme rádi Boženu Němcovou, autorku babičky.* [t_lemma = *#Comma*]
(=lit. We like Božena Němcová, the authoress of Babička.)

(34) *Boženu Němcovou, autorku babičky, máme rádi.* [t_lemma = *#Comma*]
(lit. Božena Němcová, the authoress of Babička, we like.)

Rule O:
Coreferred a-nodes in a t-node that represents **three dots**:

- **a-lex-node**: a-node that represents the first dot.
- **a-aux-nodes**: a-nodes that represent two remaining dots.

This rule applies to t-nodes the t-lemma substitute *#Period3*.
Example:

(35) *Máme psa, kočku, rybičky, křečka, andulku.* <.><.> [t_lemma = *#Period3*]
(=lit. We have a dog, a cat, fishes, a hamster, a budgerigar…)

Rule P:
Coreferred a-nodes in a t-node that represents **parenthesis**:

- **a-lex-node**: a-node that represents the left bracket.
- **a-aux-nodes**: none.

This rule applies to t-node with the t-lemma substitute *#Bracket*.
Example:

(36) *ÚFAL (Ústav formální a aplikované lingvistiky)* [t_lemma = *#Bracket*]

2.2.3.2 Special cases III : Non-alphanumeric symbols not represented by t-node

Rule Q:
A-node that represents the **dot after an ordinal number** is represented as:

- **a-aux-node** in a t-node that represents an ordinal number.

Example:

(37) *2* <.> *ledna* [t_lemma = *2*]
(=the 2nd of January)

Rule R:
A-node that represents **a hyphen in compounds** (in case of relation of determination between the hyphenated words) is represented as:

- **a-aux-node** in a t-node that represents a dependent hyphenated word.

Example:

(38) *stupnice c*<->*dur* [t_lemma = *c*]
(=scale of C-major)

There is no reference leading from a t-node to the other non-alphanumeric symbols that are not represented by a t-node. It is a 0 ⇒ a-node type.
It applies for following a-nodes:

- a-nodes that represent **punctuation symbols separating coordinated parts of a sentence** (except for the types described by rules: Na and Nb),
- a-node that represents **commas (punctuation symbols) that separate a dependent item** (clause, non-restrictive attribute),
- a-node that represents **a colon that introduces direct speech**,
- a-node that represents **quotation marks delimitating a segment of text**,
- a-node that represents **a bracket (other punctuation symbol) delimitating a parenthesis**,
- a-node that represents **a punctuation mark that indicates the end of a sentence**,
- a-node that represents **a dot after abbreviations**.

2.3 To references in mixed paratactic connections

Rules of annotation set that in mixed appositional connection, the syntactic component of parataxis has the same functor as the non-syntactic component of a paratactic connection.
Similarly, in coordinating/appositional connections with the relative pronoun *což*, the effective root of the sentence that is connected by the relative *což* gets the same functor as the second component.
Examples:
Váže těžké kovy.PAT_M, *jako*.APPS *je*.PAT_M *plutonium.*
(=lit. *(It) binds_to heavy metals such as plutonium.*)
O práci přijdou všichni.ACT_M, *kdo pracovali*, [*#Comma*.APPS] *což je*.ACT_M *asi 46 lidí.*
(=*All employees will lose their jobs, which is about 46 people.*)

A question emerges whether in cases where the modification is a determining functor of the second component of the paratactic connection, expressed by the prepositional form or by a dependent clause introduced by a subordinating conjunction (and this preposition/conjunction influences the functor) – whether this preposition/conjunction should be referenced in the *a/aux.rf* attribute also in the second syntactic component of the subordinating connection (similarly to the standard cases of paratactic connections: *jel do Prahy a <do> Brna* (=*(he) went to Prague and <to> Brno.*)). The question is connected with how the valency controls are determined in these cases. Actually, a question emerges whether the adopted rules of annotation (to solve these cases as mixed apposition and coordination) are designed properly.
Existing solutions:
Functor of the second, syntactically expressed, component of mixed coordination or apposition is identical to the functor of the first, non-syntactically expressed component.
In the *a/aux.rf* attribute, there is a reference to the functional word that determines the functor of components connected by parataxis also in case of the second component of the paratactic connection (mixed coordination, apposition).
Examples:
Koncern si tuto akvizici přál, <protože> společnost A se zabývá.CAUS_M *solárními články, což zapadá*.CAUS_M [aux: *protože*] *do odvětví společnosti B.*
(=*The concern longed for this acquisition because the A company deals with solar cells which is in tune with the sector of the B company.*)
<včetně> gigantů.ACMP_M, *<jako> jsou*.ACMP_M [aux: *včetně*] *společnosti A a B*
(=*including giants such as companies A and B*)
Zaměřili se <na> akcie.PAT_M, *<jako> jsou*.PAT_M [aux: *na*]....
(=*(they) focused on shares such as…*)
Akcie stouply <na> 55.PAT_M, *což je*.PAT_M [aux: *na*] *10% z celkové hodnoty.*
(=*Shares climbed to 55, which is 10 % of total value*)

Note: "Ellipsis of function words" that are referenced in the a/aux.rf attribute are not described in the manual nor in an appendix. It is for example:
jel <do> Prahy.DIR3_M *a Brna*.DIR3_M [aux: *do*]
(=*(he) went to Praha and Brno*)
<musel> utírat.PRED_M *nádobí a čistit*.PRED_M [aux: *musel*] *boty*
(=*(he) had to dry the dishes and clean shoes*)
<v> Praze.LOC_M, *Brně*.LOC_M [aux: *v*] *a Bratislavě*.LOC_M [aux: *v*]
(=*in Praha, Brno and Bratislava*)
These cases are explicit. The types, however, cannot be ignored when dealing with the "aidref" problems.

T-lemma

Supplement to MAN 3 Tectogrammatical lemma (t-lemma)

1 T-lemma for t0-nodes (is_generated = 0)

The correct form of the t-lemmas of t0-nodes (i.e. nodes with the value 0 in the attribute *is_generated*) is considered their basic word form (nominative, infinitive) in the first phase of annotation, so the t-lemma which is already in the tree. (Derivative) changes of t-lemmas of adverbs to adjectives, changes of t-lemmas in pronouns etc. (described in Manual mainly in Section 5.1, Syntactic and lexical derivations) will not be carried out (and changes made already automatically will be neither corrected or checked)! Similarly, we will not annotate the so called representative t-lemma (described in Manual 4.3.1).
The first phase of annotation captures the final form of the t-lemma in the following cases of t0-nodes.

1.1 Examples of multi-word t-lemmas

(Cases are described in Manual 4.3 Multi-word t-lemmas.)

- **reflexive words**: *se* (=REFL.) as part of the t-lemma of reflexive words (t-lemma: *slovo_se* (=lit. *word*_REFL); for example: *smát_se* (=*laugh*), *setkat_se* (=*meet*), *chovající_se* (=*behaving*))
- **multi-word conjunctions** like *buď_nebo* (=*either_or*), *a_a* (=*and_and*): due to the fact that more theoretically appropriate representations of coordinating conjunctions have not yet been chosen, the t-lemma will consist of the present forms of conjunctions in a sentence, we will not solve the so called representative t-lemma.
- **dependent part of idioms,** t-lemma of nodes with the functor DPHR like *na_lehkou váhu* (=*underestimate*). Note that t-lemma consists of forms of words!
- **type "van_Beethoven"**
- **type "number_number"**: telephone numbers, postcode.

If these cases miss some part of t-lemma, firstly, we annotate links to a-layer correctly, then we correct t-lemma – with the appropriate macros (macro: Alt+i, Alt+L, Alt+F).
If it was impossible to make a change in t-lemma with macros, but it was made manually, such node is marked by choosing a note *T-lemma with underscore*. It is necessary only to select the type of note.

1.2 Substitutional t-lemmas for non-alphanumeric symbols

It is necessary to solve the correct form of the t-lemmas for non-alphanumeric symbols (expressed in the surface representation of the sentence), for example: *#Colon* for colon etc. (MAN 4.4. T-lemma substitutes)

1.3 #PersPron

Expressed personal, possessive and reflexive pronouns will be assigned the t-lemma *#PersPron*.

1.4 Typo

Typo in the t-lemma will be corrected manually. A change in t-lemma will be marked by choosing a note *Typo*.

1.5 Incorrectly analyzed basic word forms

In case of incorrect morphological analysis of basic word forms (for example, the pronoun *je – them* is generated as the infinitive *být – to be*), the t-lemma is manually corrected. Change in t-lemma is marked by the selection of note *M-lemma*. It is necessary only to select the type of note.

1.6 T-lemmas of foreign-language expressions

T-lemma of foreign-language expressions is morphological form, even in case of non-alphanumeric symbols. The annotator changes the incorrect t-lemmas using a macro (Alt+F). If it is impossible to correct the wrong t-lemma using a macro, it is changed manually and the change of t-lemma is marked by selecting the notes *T-lemma*.

2 T-lemma for t1-nodes (is_generated = 1)

2.1 "Copied" t-nodes

It is supposed that all copied nodes will be properly completed in the first phase of annotation. Note the links in a-layer for copied nodes!

2.2 Completed t-nodes with substitutional t-lemmas

The final form of the following t-lemmas will be properly determined for t-layer in the first phase of annotation:

#AsMuch	*#Forn*	*#Rcp*
#Equal	*#Idph*	*#Separ*
#EmpNoun	*#Neg*	*#Some*
#EmpVerb	*#Oblfm*	*#Total*

It is valid for the annotated data, that all of these t-lemmas are correct.

The node complemented by the annotator for ellipses of obligatory actants will have a single t-lemma: ***#NewNode***. It means that the annotator does not solve whether the complemented node should have t-lemma: *#Cor*, *#QCor*, *#Unsp*, *#PersPron*, or *#Gen*. Assigning t-lemma *#NewNode* means that the t1-node may have any of these t-lemmas.

T-lemmas (*#Cor, #QCor, #Unsp, #PersPron, #Gen*) for automatically generated nodes will not be changed, but it is verified whether the node was generated correctly in terms of completeness of valency frame and whether it was assigned the correct functor. In case of extra generating, the annotator deletes the node. (We will check and correct t-lemma for all t1-nodes with t-lemmas *#Cor*, *#QCor*, *#Unsp*, *#PersPron* and *#Gen* in the next phase of annotation (within coreference annotation). Also all nodes with t-lemma *#NewNode* will get the final form of t-lemma.)

Annotation of verbal valency

Supplement to MAN 6.2 Valency

Valency is determined for verbs, nouns, adjectives and also adverbs if they have it. Annotation of valency means:

1. Determining and assigning a valency frame from a valency lexicon (i.e. the attribute *val_frame.rf* is filled)
2. Filling a valency frame in a tectogrammatical tree
   - nodes for expressed dependent modifications have filled in correctly the attribute *functor*
   - nodes for obligatory valency modifications unexpressed on the surface are added to the tectogrammatical tree

Valency frames for verbs were previously annotated manually and they were also assigned by automatic procedures and nodes for obligatory modifications unexpressed on the surface were also added.

1 Step 1: Control of the associated valency frame

The annotator does not create valency frames for verbs, he only checks the accuracy of automatically selected frame; if its incorrect, he selects the correct valency frame from the already existing valency frames that are in the lexicon. In any case, the annotator does not change anything in the PDT-vallex!

What to do if a correct valency frame is not in the vallex?
If the correct valency frame is not in the PDT-vallex, it may mean several cases that we describe below.

1.1 F (frame): the frame is missing in the vallex, no frame is assigned

Valency frame for the given meaning is not at all in the lexicon.
Solution: The annotator does not assign any frame to the verb. (Attribute *val_frame.rf* remains empty.) In the text of the note like *New valency frame*, he obligatory writes "F" and a proposal for a new frame as close as possible to the form that valency lexicon uses, including the example (or notes) by which he clearly explains the meaning. For DPHR functor, it is sufficient to write the "text" of the dependent part of idiom in parentheses.
Examples:
F ACT(.1) PAT(po+6) Nápoj chutná po třešních. (*=The drink tastes like cherries.*)
F ACT(.1) PAT(.4) ?ADDR(.3) DPHR(na míru) Šije mu úkoly na míru. (*=He prepares tasks for him tailored to his needs.*)

1.2 W (word): the word is missing in the vallex, no frame is assigned

A given word is not at all in the valency lexicon.
Solution: The annotator does not assign any frame to the verb. (Attribute *val_frame.rf* remains empty.) In the text of the note like *New valency frame*, he obligatory writes "W" and a proposal for a new frame as close as possible to the form that valency lexicon uses, including the example (or notes) by which he clearly explains the meaning.
Example:
W ACT(.1) PAT (.4) Pavel candal mléko na podlahu. (*=Pavel poured milk on the floor.*)

1.3 A (alternative): the form is missing in the vallex, the frame is assigned

In the valency frame for the given meaning of the verb, a description of the actant form is missing. The form expresses the given actant in the annotated sentence. This type also includes cases where the list of names for a compound predicate lacks some of the names in description of CPHR.
Solution: The annotator assigns an appropriate frame with a missing form to the verb. In the text of the note like *New valency frame*, he obligatory writes "A" and he states only the actant that lacks some form in the assigned frame. In the description of the form in this actant, he states only the missing form. He can also give an example.
Examples:
A PAT(zda[.v]) Váhá, zda začít. (*=He hesitates to start.*)
The description means that the assigned frame lacks a possibility to express the patient by the dependent clause beginning with the subordinating conjunction *zda* (lit. *whether*).
A CPHR({zátěž}) Kladl na něj přílišnou zátěž. (*=He made too big demands on him.*)
The description means that the assigned frame lacks the noun zátěž (demand) for the nominal part of the compound predicate.

1.4 N (new): non-obligatory actant is missing in the frame, the frame is assigned

Another non-obligatory (!) actant expressed in the annotated sentence is suitalble to the valency frame for the given meaning of the verb.
Solution: The annotator assigns an appropriate frame with missing non-obligatory (i.e. facultative) actant. In the text of the note like *New valency frame*, he obligatory writes "N" and he states only the description of the non-obligatory actant missing in the assigned frame. In the description of the form in this actant, he states the form that appeared in the annotated sentence (possibly another that is evident). He always gives an example.
Examples:
N ?EFF(na+4) Reorganizuje hypoteční sekci na samostatnou jednotku. (*=He reorganizes mortgage section to a separate unit.*)
The description means that the assigned frame (*ACT(.1) PAT(.4)*) lacks the non-obligatory effect with the form *na+4* (*to+4*).
N ?PAT(z+2;na+6) Trh prosperuje ze špatných zpráv. (*=The market thrives on bad news.*)
The description means that the assigned frame (*ACT(.1)*) lacks the non-obligatory PAT with the forms *z+6,na+6* (*from+6, on+6*).

1.5 X: a functor is missing in an alternative group, the frame is assigned

To the alternative group of functors (separated within the vertical bar) in the valency frame for the given meaning of the verb, another functor is suitable that is expressed in the annotated sentence. This type is also for cases with no alternative in the valency frame, but if we want to create this alternative (i.e. having two members).
Solution: The annotator assigns an appropriate frame with missing functor in the alternative group. In the text of the note like *New valency frame*, he obligatory writes "X" and the description of the last of the functors of the alternative group and he adds the description of the added functor missing in the assigned frame behind the vertical bar. In the description of form of this functor, the annotator states the form that appeared in the annotated sentence (possibly another that is evident). He always gives an example.
Examples:
X CPR(*)|MEANS(*) Vidět svět japonskýma očima. (*=See the world through Japanese eyes.*)
The description means that to the alternative group of the assigned frame (*ACT(.1) PAT(.4) MANN(*)|ACMP(*)|CPR(*)*), it is necessary to add the functor *MEANS(*)*.
X LOC(*)|ACMP(*) Akcie uzavřely s mírným poklesem.(*=Shares closed with a slight decline.*)
The description means that from the slot *LOC(*)* in the assigned frame (*ACT(.1) LOC(*)*), it is necessary to create an alternative group: *LOC(*)|ACMP(*)*.

General notes to the types 1.1–1.5:
To all types 1.1–1.5, the annotator may add another commentary in the note of the type *Vallex*. On the contrary, to the text of the note *New valency frame*, no other comments may be added. This type of note has an exactly prescribed form which must be observed!
Beware the form of the description, the right transcript of the functor and the form (especially the typos like CHRP). For each functor, there is always the form in parentheses, from this, it is obvious that the example always follows the closing parenthesis!

1.6 Other cases

- sense needs split in vallex
- word sense(s) hard to distinguish
- frame assigned, but doubts remain
- misspeling in examples
- wrong synonym/antonym/aspect counterpart
- others

Solution: The annotator assigns the valency frame that seems to be the most appropriate. He describes the problem in the text of the note *Vallex*.

2 Step 2: Control of the fulfillment of valency frames in a tree

The annotator checks the assignment of functors/actants of nodes for expressed modifications as well as nodes for modifications unexpressed on the surface. For automatically supplemented nodes for modifications unexpressed on the surface, he does not check the t-lemma (see Appendix *T-lemma*). He corrects incorrectly filled functors, adds the missing nodes, deletes the redundant.
In case that the node was not assigned the valency frame (the above-described type F and W), the annotator acts as if the node has been assigned the frame he suggested in the note *New valency frame*.

Valency annotation with "nonverbs"

Supplement to MAN 6.2 Valency

Note: Valency frames for nouns, adjectives and adverbs are not yet elaborated systematically in the PDT-wallex. Their elaboration is going to require much time and contemplation. Thus, valency of "non-verbs" was not annotated in the first phase of the annotation of PCEDT_en. This supplement comprehends rules of treating depending modifications of "non-verbs" in the first phase of the annotation of PCEDT_en. The instructions are based on the fact that the valency frames for nouns ending with *–ní* and *–tí* are widely elaborated in the valency lexicon and we have had much experience of them.

1 Deverbal nouns ending with *–ní* and *–tí*

In case of nouns ending with *–ní* and *–tí*

- it is marked in the *New Valency Frame* annotation note whether the annotator interprets the meaning of the noun as event (P) or non-event (nP).
- in its event meaning, the noun is assigned a valency frame provided that the annotator has found the corresponding (right) valency frame in the valency lexicon; otherwise, the noun is not assigned any frame.

Annotation: First of all, the decision is made in case of every noun whether its meaning is event or non-event in the given context. Nouns ending at *–ní* and *–tí* primarily express an event! It is assumed that the event meaning is more frequent. If the interpretation of the event meaning is questionable, the event meaning is preferred.

Meaning of noun is **event**:
There is a P (process) value in the *New Valency Frame* annotators' note.
If the corresponding valency frame - i.e. the frame derived from the valency frame of corresponding base verb (in case of frames for event meanings, the verbs are (usually) in parenthesis before examples) - was found in the PDT Valency Lexicon-vallex, the valency frame is assigned.
In all other cases (noun is not in the vallex, there is no appropriate valency frame, annotator did not accept the proposed frame for the event meaning) the valency frame is not assigned.
According to the corresponding valency frame (assigned or just assumed), the functors in the surface level of a sentence are assigned by means of an expressed modification, and nodes for obligatory valency modifications that are not expressed in the surface level of a sentence (only for those nouns there is a consistent annotation of reciprocity) are assigned as well.
Agreeing adjectival attributes are interpreted, they do not have the RSTR functor, they do have corresponding functors for free modifications (adjuncts). The RSTR functor goes with agreeing adjectives only when they are:

- demonstrative pronouns, emphasizing expressions (*ten, on* (=*he*))
- deverbal adjectives representing a dependent attributive clause (*soužití Romů trvající*.RSTR *století* (=lit. *coexistence of Romanies lasting for centuries*))

Meaning of noun is **non-event**
There is an nP (non-process) value in the *New Valency Frame* annotators' note.

2 Other nouns

In case of other nouns (including nouns ending with *–ní* and *–tí* that do not have event meaning), the functors are assigned only in the surface form of a sentence by means of an expressed modification. Valency modifications and all genitive modifications are assigned the RSTR functor.There are no newly established nodes in the tectogrammatical tree for modification that are not in the surface form of a sentence. Agreeing adjectival attributes are not interpreted, the have the RSTR functor. Only free adverbial modifications in prepositional cases are interpreted. On no account do annotators assign a valency frame. Reciprocity is not annotated for those nouns.

Examples:

dálkové ovládání televize.RSTR (=*remote control (for) television*)
balení másla.PAT *na lince v závodě* (event) (=*packing of butter on the line in a factory*)
jedno.RSTR *balení másla*.RSTR *na regále.*LOC (=*one package of butter on the shelf*)
myčka nádobí.RSTR (=*dish-washer*) / *myčka na nádobí*.AIM (=lit. *washer for dish*)
obvaz rány.RSTR *lékařem*.RSTR (=*bandaging of a wound by a doctor*)
buničitý.RSTR *obvaz na rány*.AIM (=lit. *cellulose bandage for wounds*)
struhadlo brambor.RSTR (=lit. *rasp of potatoes*)
struhadlo na brambory.AIM (=lit. *rasp for potatoes*)
můj.RSTR *dopis mamince*.RSTR (=lit. *my letter to mamma*)
výroba tužek.RSTR (=*production of pencils*)
výrobek ze dřeva.RSTR (=*product (made) of wood*)
čekárna nemocných.RSTR *na vyšetření*.AIM (=lit. *waiting room of the sick for examination*)
čekárna pro matky.BEN *s dětmi* (=lit. *waiting room for mothers with children*)
román o válce.RSTR (=*novel about war*)
zákon o rodině.RSTR (=*act on family*)
*návrh řešení.*RSTR (=*proposal for solution*)
žádost o změnu.RSTR (=*request for change*)
prosba o pomoc.RSTR (=*plea for help*)
záměna holčičky.RSTR *za chlapečka*.RSTR (=lit. *confusion of a girl for a boy*)
struktura firmy.RSTR (= *structure of company*)
rekord sportovce.RSTR (=*record of athlete*)
náš.RSTR *cíl zvítězit*.RSTR (=*our goal to win*)
spor poslanců.RSTR (=*dispute of deputies*)
dohoda mezi oběma koncerny.RSTR (=lit. *agreement between both concerns*)
23.RSTR *března 2008*.RSTR (=lit. *23. march 2008*) (but: *rok 1998*.ID (=*year 1998*.ID))

3 Adjectives and adverbs

Depending modifications of adjectives and adverbs are interpreted, i.e. they are assigned functors of arguments and free modifications (never RSTR). Valency of these word classes is annotated intuitively only (in accordance with instructions in the manual), without the valency lexicon, i.e. the words are not assigned valency frames.

Examples:

lidé čekající na vyšetření.PAT (=*people waiting for examination*)
muž pracující se dřevem.PAT (=*man working with wood*)
poslanec zvolený voliči.ACT (=*deputy elected by voters*)
dopis došlý poštou.MEANS (=*letter received by mail*)
je blíž umění.PAT (=*(he) is closer to art*)

Reciprocity

Substitute for MAN 6.2.4.2 Reciprocity

1 What the reciprocity is?

Reciprocity
= relation between two valency modifications (arguments or obligatory free modifications) that are put into the symmetric relation so that the relation between the two can be paraphrased on the basis of the model:
Petr a Pavel se potkali. = Petr potkal Pavla a (zároveň) Pavel potkal Petra.
(=lit. *Petr and Pavel* REFL *met. = Petr met Pavel and (simultaneously) Pavel met Petr.*)

1.1 Lexical reciprocity

= the relation of reciprocity between two modifications follows from the lexical meaning of a predicate. The relation of reciprocity between two valency modifications is required for many verbs, it is necessary for the performance of the designated activity.
Those verbs are analyzed as **inherently reciprocal verbs.**
Example: *Petr se setkal s Pavlem.*(=lit. *Petr* REFL *met Pavel*.)

1.2 Syntactic reciprocity

= the relation between two modifications expressed by syntactic (formal) means. A syntactic operation is performed above the valency frame of the predicate that results in the action of putting the modification of predicate in the relation of reciprocity.
This operation is called **reciprocalization.**
Modifications where the operation is applied are called **reciprocalized modifications** (or possibly **reciprocalized (valency) positions**).
Input conditions for reciprocalization:

- lexical homogeneity of modifications where the operation will be applied
- modifications where the operation is applied are of an identical topic-focus articulation

Output conditions for reciprocalization

- one reciprocalized expression (one reciprocalized valency modification) is not expressed
- plural or coordinated form in the second – expressed – reciprocalized position

As a result of reciprocalization, the position of one of the reciprocalized modifications disappears (is not expressed) from the surface form of the sentence. In the second reciprocalized position, both modifications in the relation of reciprocity are expressed together.
The operation of reciprocalization may be applied in case of both inherently reciprocal verbs and the verbs that lack this feature.
Examples:
Input: *Jan se líbá s Marií*. (=lit. *Jan* REFL *kisses Marie*.)
Output: *Jan a Marie se líbají*. (=lit. *Jan and Marie* REFL *kiss*.)
Input: *Jan líbá Mariinu ruku*. (=lit. *Jan kisses Marie's hand.*)
Output: reciprocalization cannot be accomplished, the condition of modifications being homogenous is not met.

2 Lexical reciprocity

2.1 Inherent reciprocal verbs

Inherent reciprocal verbs are the verbs where there is a presupposition of the relation of reciprocity between two valency modifications. However, there is only a potential of the reciprocal meaning, because it can be withdrawn by context.
Examples:
Petr se setkal s Pavlem. (=lit. *Petr met Pavel*.)
Petr se setkal s nezájmem. (=lit. *Petr met indiference*.)

Inherent reciprocal verbs are divided into inherent reciprocal verbs and inherent reciprocal variations of verbs (derived reciprocal verbs).

2.1.1 Inherent reciprocal verbs
= non-derived (primary) verbs with reciprocal meaning
They are:
- reciprocal reflexives (reflexiva tantum): *hádat_ se* (=*to quarrel*), *prát_se* (=*to fight*), *setkat_se* (=*to meet*), i.e. the verbs where the expression *se* (derivational formant with reciprocal meaning) is a part of lemma of the verb (indicated by the underline character) and the verb cannot be used without this expression.
- other verbs. The inherent meaning of reciprocity is present in other verb meanings as well: *zápasit* (=*to struggle*), *obchodovat* (=*to trade*), *diskutovat* (=*to discuss*), *polemizovat* (=*to argue*), *soupeřit* (=*to compete*), *soutěžit* (=*to compete*), *sousedit* (=*to neighbor*).

2.1.2 Inherent reciprocal variations of verbs (derived reciprocal verbs)
= verbs with reciprocal meaning, derived from the verbs with accusative or dative construction with the help of the derivational formant *se* or *si.*
The process of derivation of inherent reciprocal verb versions is analogical to the process derivation of reflexive verb versions (change of valency frame and attachment of the *se* formant to the lemma of the verb occur here as well).
Example: *roztrhnout* (=*to tear*) – *roztrhnout se* (=lit. *to tear* REFL)
Valency frame of the verb *roztrhnout*: ACT(.1) PAT(.4) *Někdo roztrhl pytlík*.
(=lit. *Someone tore the bag.*)
Valency frame of the derived reflexive *roztrhnout_se*: ACT(1) *Pytlík se roztrhl.*
(=lit. *The bag* REFL *tore*.)

2.1.2.1 Reciprocal verbs derived from the verbs with accusative construction

In case of verbs in nominative and accusative position, the relation of reciprocity between the modifications can also be expressed lexically – by derivation of inherent reciprocal version of the original verb, in the following way:

- the derivational formant *se* is attached to the base of the original verb,
- the accusative construction of the verb is replaced by the construction *s*+7.

Schematically:

nominative	verb	accusative
⇓	⇓	⇓
nominative	verb_*se*	*s*+instrumental

Example: *líbat* (=*to kiss*) ⇒ *líbat_se* (=*to kiss* REFL)

Jan	líbá (= kisses)	Marii (=Marie).
⇓	⇓	⇓
Jan	se_líbá (=REFL kisses)	s Marií (=with Marie).

Valency frame of the verb *líbat* (=*to kiss*): ACT(.1) PAT(.4)

Valency frame of the derived reciprocal verb *líbat se* (=*to kiss* REFL): ACT(1) PAT (s+7)

Examples of other derived reciprocal verbs: *objímat_se* (=*to hug* REFL), *potkat_se* (=*to meet* REFL), *navštěvovat_se* (=*to visit* REFL), and if it occurs (i.e. in data), then also: *fotografovat_se* (=*to photograph* REFL), *popisovat_se* (=*to describe* REFL), *vidět_se* (=*to see* REFL), *léčit_se* (=*to heal* REFL).

2.1.2.2 Reciprocal verbs derived from the verbs with dative construction

In case of verbs with modifications in nominative and dative position, the relation of reciprocity between those modifications can also be expressed lexically – by derivation of inherent reciprocal version of the original verb, in the following way:

- the derivational formant *si* is attached to the base of the original verb
- the dative construction of the verb is replaced by the construction *s*+7

Schematically:

nominative	verb	dative
⇓	⇓	⇓
nominative	verb_*si*	*s*+instrumental

Example: *telefonovat* (=*to phone*) ⇒ *telefonovat_si* (=*to phone* REFL)

Jan	telefonuje (=phones)	Marii (=to Marie).
⇓	⇓	⇓
Jan	si_telefonuje (=REFL phones)	s Marií (=with Marie).

Valency frame of the verb *telefonovat* (=*to phone*): ACT(.1) PAT(.4) ADDR(.3)

Valency frame of the derived reciprocal verb *telefonovat_si* (=*to phone* REFL): ACT(1) PAT(.4) ADDR(s+7)

Examples of other derived reciprocal verbs: *telefonovat_si* (=*to phone* REFL), *konkurovat_si* (=*to compete* REFL), *slíbit_si* (=*to promise* REFL), *věřit_si* (=*to believe* REFL), *blahopřát_si* (=*to congratulate* REFL), *volat_si* (=*to call* REFL), *sdělovat_si* (=*to tell* REFL), *pomáhat_si* (=*to help* REFL).

2.2 Lexical indicators of reciprocity

Lexical way of expressing reciprocity is a usage of expressions (adverbs) that express reciprocity: *spolu, společně, vzájemně, navzájem* (=*together*, *to each other*, *each other*, *one another*, *mutually*, *reciprocally*). Those expressions serve especially as an emphasis of reciprocal meaning in a syntactic reciprocity.

3 Syntactic reciprocity

3.1 Formal indicators of reciprocity

There are always two basic – obligatory – formal indicators of the accomplished operation of reciprocalization. In various cases, other formal indicators can be added to those formal consequences.

Basic formal indicators of syntactic reciprocity:

= they occur in a sentence each time the operation of reciprocalization is applied.

A. one reciprocalized valency position is left out
B. the other reciprocalized position is realized formally as:
   - coordinate (and hypotactic) structure (*Petr a Pavel* (=*Petr and Pavel*), *Petr s Pavlem* (=*Petr with Pavel*))
   - formal or only semantic plural of a noun (*kamarádi, dvojice* (=*friends, couple*)
   - *mezi*+instrumental form (*mezi Petrem a Pavlem* (=*between Petr and Pavel*))

Specific formal indicatirs of syntactic reciprocity:

= in a sentence, they are not present in each application of the operation of reciprocalization, yet they occur to various degree of regularity/necessity in various types of reciprocalization

C. the expression *se* in appropriate form.
   Se (strong form) is of the same case (even the prepositional case) as the reciprocalized modification. Provided that the unexpressed reciprocalized position is in the form of *s*+7, the form *s sebou* (=*with one*) is not used; the form *mezi sebou* (=*mutually*) is used instead. It cannot occur together with the formal indicator D.
D. the expression *jeden druhý* (=*one, the other*) in appropriate form.
   Jeden (=*one*) is of the same case (even the prepositional case) as the expressed reciprocalized modification, *druhý* (=*the other*) is of the same case (even the prepositional case) as the unexpressed reciprocalized modification. It cannot occur together with the formal indicator C.

NB! With reciprocal reflexives (reflexive tantum) and derived reciprocal verbs (reciprocal verbs with the forman *se, si* in the lemma), there is no need of any other formal indicators apart from the basic indicators A and B in reciprocalization.

3.2 Representing syntactic reciprocity in the tectogrammatical tree

Reciprocity is represented by a newly established t-node with the *#Rcp* t-lemma that is inserted to the tectogrammatical tree in the position of an unexpressed reciprocalized valency modification. The newly established t-node has the functor that corresponds to the unoccupied position according to the valency frame. The relation between the newly established t-node and the t-node in the expressed reciprocalized position is indicated as a relation of grammatical coreference. The C and D formal indicators of reciprocity (see 3.1) are not represented by an individual t-node in the tectogrammatical tree but the references to corresponding a-nodes are inserted in the *a/aux.fr* attribute with the newly established t-node with the *#Rcp* t-lemma. In contrast, lexical indicators of reciprocity (see 2.2) are represented by individual t-nodes and usually by the functor MANN.

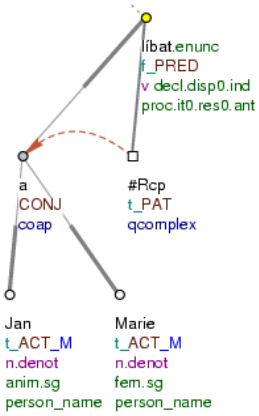


Fig.1: *Jan a Marie líbají <jeden druhého>.* (=lit. *Jan and Marie kiss <one another>*.)

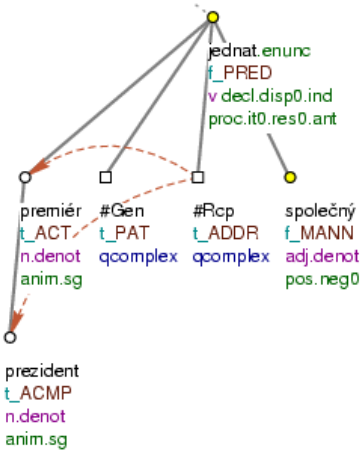


Fig.3: *Premiér s prezidentem spolu jednali.* (=lit. *Prime_ minister with president together negotiated.*)

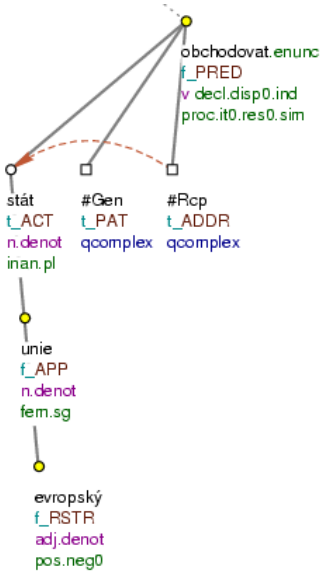


Fig. 2: *Státy Evropské unie <mezi sebou> obchodují.* (=lit. *States (of) European Union <between themselves> trade.*)

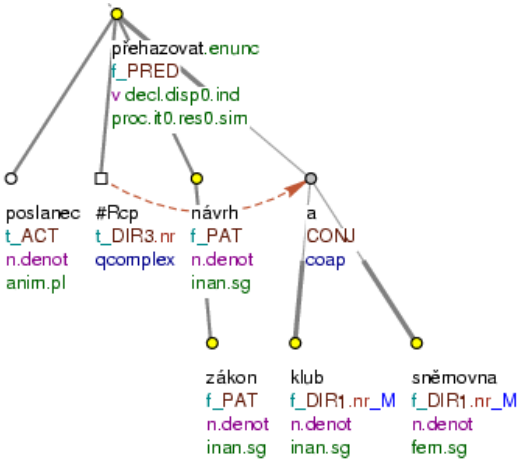


Fig. 4: *Poslanci přehazují návrh zákona mezi klubem a sněmovnou <z jednoho do druhého>.* (=lit. *Deputies throw_over bill between club and parliament <from one to another>*.)

4 Reciprocity with other classes than verbs

The process of reciprocalization occurs also with the (deverbative) nouns and adjectives, compare:
boj znepřátelených stran mezi sebou (=lit. *fight of enemy sides between each_other*)
lidé bojující mezi sebou navzájem (=lit. *people figting among themselves each other*)
For now, reciprocity with word classes other than verbs is analyzed in the tectogrammatical trees only:

- with deverbative substantives ending *–ní* and *–tí* that do not carry the meaning of time.
- with adjectives where the reciprocity is expressed through a specific formal indicator C and D *(mezi sebou, jeden druhý)*, i.e. where it is necessary to hide something!

Examples:
jednání Petra.ACT *a Pavla*.ACT *o prodeji domu* #Rcp.ADDR
(=lit. *negotiations (of) Petr and Pavel about sale (of) house*)
jednání států.ACT <*mezi sebou*> #Rcp.ADDR
(=lit. *negotiations of states* <*between themselves*>)
jejich.ACT *společné*.MANN *jednání se protáhlo* #Rcp.ADDR
(=lit. *their mutual negotiations* REFL *extended*)
setkání dvou prezidentů.ACT #Rcp.PAT (=lit. *meeting of two presidents*)
srovnání dvou nesourodých trhů.PAT #Rcp.EFF
(=lit. *comparison of two inhomogeneous markets*)
obchodování s ropou mezi státy.ACT #Rcp.ADDR (=lit. *trading with oil between states*)
lidé bojující <*mezi sebou*> *navzájem*.MANN #Rcp.ADDR
(=lit. *people fighting* <*among themselves*> *each other*)

In other cases, reciprocity is not analyzed. With other deverbative nouns where the reciprocity is expressed through a specific indicator C and D (*mezi sebou, jeden druhý*), i.e. where it is necessary to hide something, the reciprocity will not be analyzed. Formal indicators of reciprocity C and D will have the node in the tree, functor MANN, COMPL.
Examples:
boj znepřátelených stran.RSTR *mezi sebou*.MANN
(=lit. *fight of enemy sides between each_other*)
diskuze mezi čtenáři.RSTR *a spisovateli*.RSTR (=lit. *discussion among readers and authors*)
dohoda mezi zastupiteli.RSTR (=lit. *agreement among representatives*)
obchod s ropu.RSTR *mezi státy*.RSTR (=lit. *trade with oil between states*)
vztah Petra.RSTR *a Pavla*.RSTR (=lit. *relation of Petr and Pavel*)
vzájemný.RSTR *vztah Petra*.RSTR *a Pavla*.RSTR (=lit. *mutual relation (of) Petr and Pavel*)
spor poslanců.RSTR (=lit. *dispute of deputies*)

Ellipses

Supplement to MAN 6.12 Ellipses

1 Textual ellipsis of the governing noun in constituent coordination

Change! (MAN 6.12.1.2 Ellipsis of the governing noun)
Textual ellipsis of the governing noun in constituent coordination is no longer treated on the tectogrammatical level! The following rule applies: in the structure, coordination is located the lowest possible! We do not insert a new copied node!

1.1 Type "red and white wine"

Examples:

Fig. 1: *Pije červené a bílé víno.*
(=lit. *(He) drinks red and white wine.*)

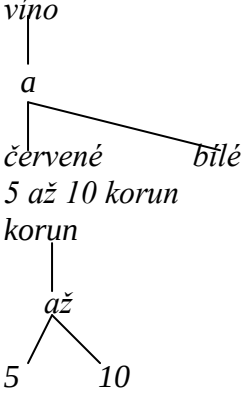


Fig. 2: *střední a východní Evropa*
(=lit. *Central and Eastern Europe*)

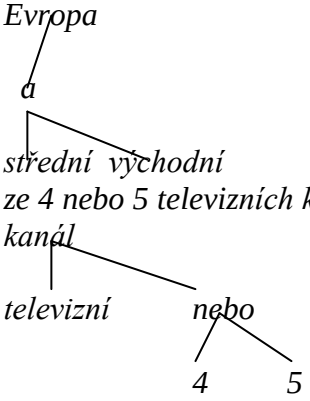


Fig. 3: *50 českých a moravských Romů*
(=lit. *fifty Czech and Moravian Romanies*)

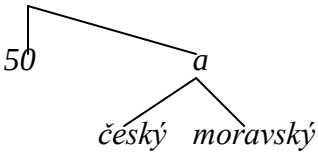


Fig. 4: *tisíc českých a moravských Romů*
(=lit. *a thousand Czech and Moravian Romanies*)

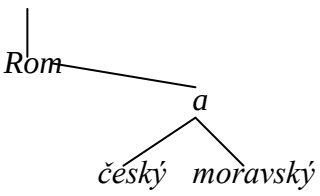


Fig. 5: *volali na první nebo druhé číslo*
(=lit. *they were calling the first or the second number*)

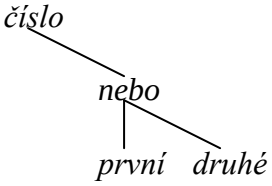
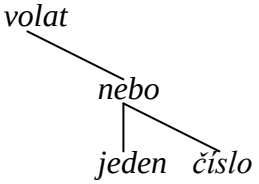


Fig. 6: *volali na jedno nebo druhé číslo*
(=lit. *they were calling one or another number*)



1.2 Type "Ministry of Labour and Social Affairs and Health"

Examples:

Fig. 7: *ministerstvo práce a sociálních věcí a zdravotnictví*
(=lit. *Ministry of Labour and Social Affairs and Health*)

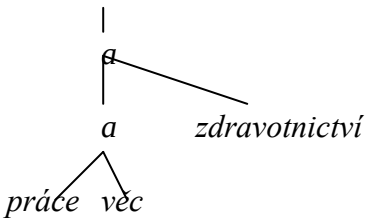


Fig. 8: *Dohodu podepsali ministři obrany ČR a Slovenska Antonín Baudyš a Pavol Kanis*
(=*The agreement was signed by ministers of defense of ČR and Slovakia, Antonín Baudyš and Pavol Kanis*)

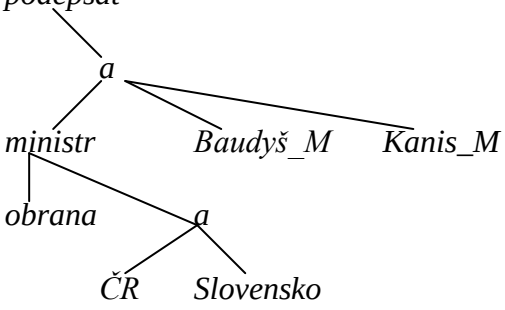
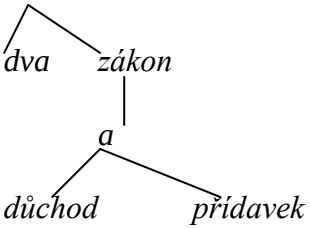


Fig. 9: *dva návrhy zákonů o důchodech a přídavcích na děti*
(= *two bills on pensions and child benefits*)



Exception**:** The only exception applies to **coordinated names** where the information about whether it is referred to one name (with coordination) or more (coordinated) names would be lost if annotated without ellipsis. It applies that two nodes with ID functor cannot be coordinated if it is referred to effective nodes of two different names.

Examples:

společnosti Toyota.ID *a [společnost] Fuji*.ID (=lit. *companies Toyota and (company) Fuji*)
pohádka Sněhurka.ID *a [pohádka] Zlatovláska*.ID
(=lit. *the Snow White fairytale and Goldilocks (fairytale)*)
litr benzínu Super.ID *a [benzínu] Speciál*.ID (=lit. *a liter of Super petrol and Special (petrol)*)
But: *pohádka Bolek*.ID *a [0] Lolek*.ID (=*Bolek and Lolek fairytale*)

2 Textual ellipsis of non-obligatory modification in paratactic structure

Change! (MAN 6.12.3 Ellipsis and the principle of shared modification)
The textual ellipsis of non-obligatory modification will continue to be analyzed with the help of "the principle of shared modifier"! The rule still applies that we analyze the non-obligatory modification as a shared modification only in cases that are totally unambiguous from the semantic viewpoint!
The cases of textual ellipsis of non-obligatory modification in paratactic structures can be analyzed as a shared modification in most cases. Now that we do not annotate the valency of nouns, there is no problem with "*sekretářka, ale zakladatelka firmy"* (=lit. secretary, though foundress of a company) type (see Fig. 11 and 12).

Examples:
Fig. 10: *zlevněné čepice a šály*
(=lit. *discounted caps and scarfs*)

Fig. 11: *sekretářka, ale zakladatelka firmy*
(=lit. *secretary, though foundress of a company*)

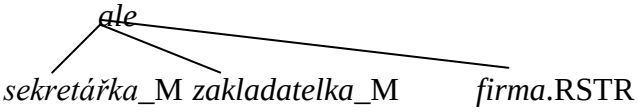
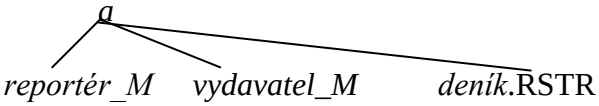


Fig. 12: *reportér a vydavatel deníku*
(=lit. *reporter and publisher of a journal*)



3 Ellipsis of the governing noun for the preposition "reasons for"

Following cases are construed as an ellipsis of the governing noun, thus the missing position is reconstructed by either t-node with t-lemma substitute *#EmpNoun* (grammatical ellipsis) or copied node (textual ellipsis).
Examples:
Všichni byli pro. (=All of them agreed.)
Důvody pro a proti. (=Pros and cons.)
Hlasoval proti. (=lit. He voted negatively.)

Special cases of coordination

Supplement to MAN 7.6.2.1 Special constructions represented as coordination or apposition

1 Appositional member in brackets

Also the content of the brackets is considered to be a part of apposition if it contains exactly the same information (said in another way) as it is mentioned before the brackets (it is possible to add "*it est*" between the members of apposition). See examples 1.1 1.2 a 1.3.
Annotation: The left bracket is captured as the root of coordination *(#Bracket*.APPS_P).
As root coordinate structure is captured left parenthesis (*#Bracket*.APPS_P). Expressions in brackets and the root of coordination will have *is_parenthesis* = 1.

1.1 Nominal phrase in the same case as the nominal phrase before brackets

Examples:
Poslanec Byron Dorgan (demokrat z Severní Dakoty) předložil v Kongresu zákon podporovaný také poslancem Lee Hamiltonem (demokratem z Indiany).
(*=The Deputy Byron Dorgan (Democrat of North Dakota) has submitted to Congress a law supported also by the deputy Lee Hamilton (Democrat of Indiana).*)
..už od dnů Alfreda Mahana (důstojníka amerického námořnictva a námořního historika)
(*=..since the days of Alfred Mahan (U.S. Navy officer and naval historian)*)
Námořnictvo má své vlastní letecké síly (flotilu letadlových lodí) a svou vlastní armádu (námořní pěchotu).
(*=the Navy has its own air force (fleet of aircraft carriers) and its own army (Marines).*)

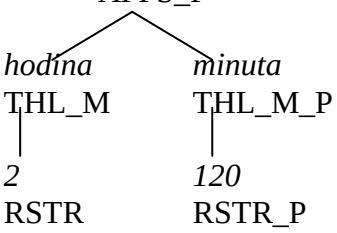
1.2 Abbreviation and its full version (and vice versa)

Examples:
ODS (Občanská demokratická strana) (*=CDP (Civic Democratic Party)*)
Občanská demokratická strana (ODS) (*=Civic Democratic Party (CDP)*)
Komerční banka (KB) (*=Commercial bank (CB)*)

1.3 Unit conversion

Examples:
2 hodiny (120 minut) (*=2 hours (120 minutes)*)
4 dolary (100 korun) (*=4 dollars (100 crowns)*)
Trvalo to 2 hodiny (120 minut). (*=It took 2 hours (120 minutes).*) (Fig. 1)

Fig. 1: *Trvalo to 2 hodiny (120 minut).* (*=It took 2 hours (120 minutes).*)



2 Appositional structure with the conjunction "as"

Type: "Váže těžké kovy, jako je plutonium."
(=lit. "It binds heavy metals as is plutonium.")
Node for the verb *být* (=*to be*) has the same functor as a node for nonverbally expressed terminal member of apposition (for the noun).
Exception: If a nonverbally expressed member of terminal apposition (noun) has a functor DENOM, the node representing the verb *být* (=*to be*) has a functor PRED.
Examples:
Váže těžké kovy.PAT_M, *jako je*.PAT_M *plutonium.* (Fig. 2)
(=*It binds heavy metals such as plutonium.*)
Počítačové společnosti, komodity.DENOM_M *podléhající ekonomickým výkyvům, jako jsou*.PRED_M *třeba auta, a akcie maloobchodních firem.* (Fig. 3)
(=*Computer companies, commodities subordinate to economic fluctuations such as, for example, cars, and shares of retail companies.*)

Fig. 2: *Váže těžké kovy, jako je plutonium.* (=*It binds heavy metals such as plutonium.*)

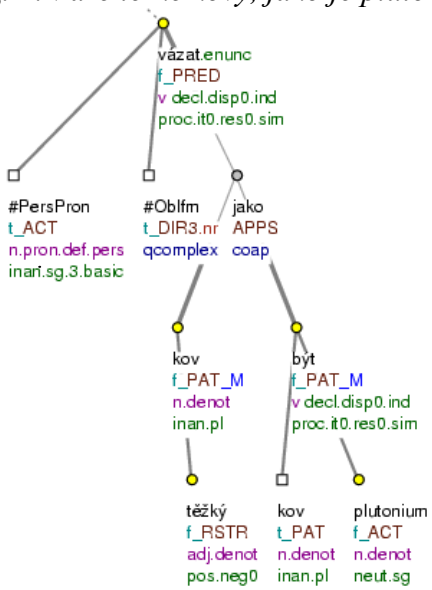
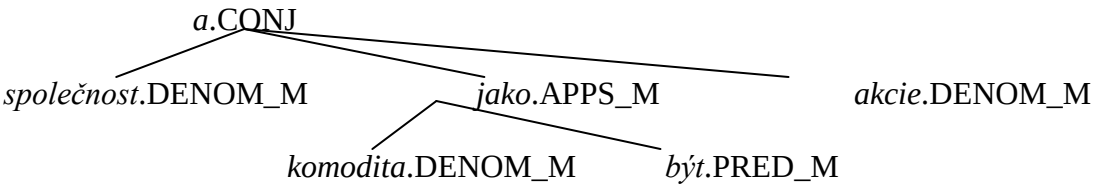


Fig. 3: *Počítačové společnosti, komodity podléhající ekonomickým výkyvům, jako jsou třeba auta, a akcie maloobchodních firem.*
(=*Computer companies, commodities subordinate to economic fluctuations such as, for example, cars, and shares of retail companies.*)



Syntactic idioms (phrasemes)

Supplement to MAN 6.8. Idioms (phrasemes)

The chapter 6.8. Idioms (phrasemes) in the Manual describes how to capture idioms (phrasemes) as lexical units. However, there are no rules how to capture syntactic idioms (phrasemes) (specific/unusual syntactic structures with a certain meaning that may be differently lexically expressed).
Capturing of the syntactic idioms (as well as lexical, in particular the "single-node") is therefore yet to be solved. In this supplement, we bring only the instructions for capturing of some syntactic idioms according to the available options of the tectogrammatical representation. However, the fact that it is an idiomatic (phraseological) construction (expressing the given meaning only as a whole) is not captured in the annotated data.

ROK CO ROK – YEAR AFTER YEAR
měsíc co měsíc, pátek co pátek – month after month, (lit.) Friday after Friday
"Rok co rok se budou setkávat s týž soudcem." – "Year after year, they will face the same judge."
year after year = every year
rok.TWHEN, na tom: *rok* <aux:*co*>.THO

DEN PO DNI – DAY AFTER DAY
měsíc po měsíci, týden po týdnu, minutu po minutě – month after month, week after week, minute after minute
"prodlužování kontraktu den po dni" – "extension of the contract day after day"
The meaning is probably "how often", "at what intervals"; it means "each day, every day." Thus:
den.THO *po dni*.TWHEN

KUS PO KUSU – PIECE BY PIECE
balík po balíku, úplatek po úplatku – package by package, bribe by bribe
"Prodat něco kus po kusu." – "Sell something piece by piece."
„Úplatek po úplatku nás vedou po cestě, kterou společnost vyšlapala.“ – "A bribe by bribe lead us down the path paved by the company."
We prefer MANN to TWHEN:
kus.MANN *po kusu*.TWHEN

DEN ODE DNE, PŘÍIPAD OD PŘÍPADU – DAY BY DAY, CASE BY CASE
"justice případ od případu rozšiřuje zákon o zneužívání" – "justice extends the Abuse Act case by case"
den.TWHEN, *ode dne*.THO

Note!
ČAS OD ČASU: *čas*.THO, *od_času*.DPHR – FROM TIME TO TIME
BOK PO BOKU: *bok*.MANN *po_boku*.DPHR – SIDE BY SIDE
RUKU V RUCE: *ruku*.MANN *v_ruce*.DPHR – HAND IN HAND
However, these are also not classic lexical idioms.

Meanings of extent

Supplement to MAN 7.13 Further specification of a functor

When expressing not accurate, but only approximate extent/number or if the extent/number is expressed by (non/closed) interval, the usual form of the given (both valency and non-valency) modification regularly changes. The theory assumes that these different shades of extent are captured by subfunctors. The problem is, among other things, that this phenomenon belongs to more functors (while subfunctors should be specific only for one functor). This is typical for the functor EXT (general non-valency meaning of "how much, how many"), for the functor THL (temporal functor with the meaning "how long"; spatial functor with the meaning "how far" does not exist), but we will meet this also in actants and functor RSTR (and possibly elsewhere). A draft summary is in the table at the end of this text.

List of prepositions expressing extent:

kolem+2 (*approximately*)	*do*+2 (*=up to*)	*na*+4 (=lit. *on*)
okolo+2 (*=approximately*)	*pod*+4 (*=less than*)	*po*+6 (=lit. *after*)
přes+4 (*=over*)	*nad*+4 (*=more*)	interval *od_do*.OPER
od+2 (*=from*)	*k*+3 (*=to*)	(*=from_to*)

Examples:

odliv byl 15 %.PAT (*=outflow was 15 %*. PAT)
odliv byl pod 15 %.PAT (*=outflow was below 15 %*. PAT)
odliv byl nad 15 %.PAT (*=outflow was above 15 %*. PAT)
odliv byl okolo 15 %.PAT (*=outflow was about 15 %*. PAT)
odliv byl přes 15 %.PAT (*=outflow was over 15 %*. PAT)

Temporary annotation (till the time of reevaluating the functors for expressing the meaning of extent and of determining the manner how we will capture the extent in the tectogrammatical layer):

Functor. Modification carrying an accessory meaning of extent (to temporal meaning, spatial meaning, meaning of patient etc.) gets a most suitable functor from the menu of funtctors (and under the current rules for assigning functors).

Links *a/aux.rf* to the prepositions of extent. The prepositions of extent should be referenced in the attribute *a/aux.rf* in the node to which the subfunctor of extent will belong. On the basis of this, we propose a few rules:

(i) In case of numerals in connection with counted objects, a reference to a preposition of extent is given to the numeral (preposition of extent is in *a/aux.rf* at the numeral, not at the counted object).

Examples:

vydělává 30 000.RSTR *korun*.PAT (*=he earns 30.000 crowns*)
vydělává <do> 30 000.RSTR *korun*.PAT (*=he earns up to 30,000 crowns*)
vydělává <od> 30 000.RSTR *korun*.PAT (*=he earns from 30,000 crowns*)
vydělává <přes> 30 000.RSTR *korun*.PAT (*=he earns over 30,000 crowns*)
vydělává <pod> 30 000.RSTR *korun*.PAT (*=he earns less than 30,000 crowns*)
vydělává od 30 000.RSTR_M *do 40 000*.RSTR_M *korun*.PAT (*od_do*.OPER) ¨
(*=he earns from 30 000 to 40 000 crowns*)
Byl u společnosti <přes> 20.RSTR *let*.THL (*=He was with the company for over 20 years.*)
Kupónová sazba ještě nebyla pevně určena, ale pravděpodobně bude stanovena <kolem> 8.RSTR *procent*.EFF
(*=The coupon rate has not yet been firmly determined, but is likely to be fixed at around 8 percent.*)

(ii) In other (more complex) cases, a reference to a preposition of extent is given to the control node of the modification.
Examples:
poklesnout <pod> hranici.PAT *2,90 marky* (*=fall below 2.90 marks*)
poklesnout <k> hranici.PAT *2,90 marky* (*=fall to 2.90 marks*)
Dostal <po> jablíčku.PAT (*=He got the apples.*)

Valency frame. The prepositions of extent belong to regular changes in surface-syntactic realizations that are not marked to the valency frame in case of valency modifications.
Examples:
Ani <kolem> deseti.RSTR *miliard*.ACT *nebude stačit.* (*=Not around ten billion will not be enough.*) [in PDT-vallex: ACT(.1)]
Dostal <po> jablíčku.PAT (*=He got the apples.*) [in PDT-vallex: PAT(.4)]
*Kupónová sazba ještě nebyla pevně určena, ale pravděpodobně bude stanovena <kolem> 8.*RSTR *procent.*EFF [in PDT-vallex: EFF(na+4)] (*=The coupon rate has not yet been firmly determined, but is likely to be fixed at around 8 percent.*)
poklesnout <pod> hranici.PAT *2,90 marky* [in PDT-vallex: PAT(na+4)]
(*=fall below 2.90 marks*)
*nabídnout <nad> 15.RSTR dolarů.*PAT (*=offer over 15 dollars*) [in PDT-vallex: PAT(.4)]

			"how much"	**"how long"**	**"how far"**
0	exactly		*Zaplatil pět korun.* (*=He paid five crowns.*)	*Trvalo to dvě hodiny.* (*=It took two hours.*)	*Bydlí tři kilometry odtud.* (*=He lives three kilometers away.*)
approx	approximately	kolem+2, okolo+2, asi, přibližně (=around, about)	*Zaplatil kolem pěti korun.* (*=He paid about five crowns.*)	*Trvalo to kolem dvou hodin.* (*=It took about two hours.*)	*Bydlí kolem tří kilometrů odtud.* (*=He lives about three miles away.*)
almost	almost	téměř, skoro, bezmála, málem (=almost)	*Zaplatil téměř pět korun.* (*=He paid almost five crowns.*	*Trvalo to téměř dvě hodiny.* (*=It took almost two hours.*)	*Bydlí téměř tři kilometry odtud.* (*=He lives almost 3 kilometers away.*)
less	less than	pod+4, méně než (=less than)	*Zaplatil pod pět korun.* (*=He paid less than the five crowns.*)	*Trvalo to pod dvě hodiny.* (*=It took less than two hours.*)	*Bydlí pod tři kilometry odtud.* (*=He lives less than 3 kilometers away.*)
more	over	přes+4, nad+4, více než (=over)	*Zaplatil přes pět korun.* (*=He paid over five crowns.*)	*Trvalo to přes dvě hodiny.* (*=It took over two hours.*)	*Bydlí přes tři kilometry odtud.* (*=He lives over three kilometers away.*)
lessincl	maximally	do+2, maximálně (=maximally)	*Zaplatil do pěti korun.* (*=He paid maximally five crowns.***)**	*Trvalo to do dvou hodin.* (*=It took maximally two hours.*)	*Bydlí do tří kilometrů odtud.* (*=He lives maximally three miles away.*)
moreincl	minimally	od+2, minimálně (=minimally)	*Zaplatil od pěti korun.* (*=He paid minimally five crowns.*)	*Trvalo to od dvou hodin.* (*=It took minimally two hours.*)	*Bydlí od tří kilometrů odtud.* (*=He lives minimally three miles away.*)

Tab. Expression of Meanings of Extent

The expressions *se* and *si*

Supplemnet to MAN 8.14. The expressions *se* and *si*

The expressions *se* (=*self / one another /* REFL) and *si* (=*self / one another/* REFL) may have various meaning in the sentence and their annotation varies accordingly.

1 Word-forming formant

The expression *se/si* is either motivated or unmotivated part of lemma of the verb.
Annotation: Multiword t-lemma of the verb: *verb_se* or *verb_si*; the expression *se/si* in the *a/aux/rf*.

1.1 Reflexive verbs (reflexiva/reciproka tantum)

The expression *se/si* is the unmotivated part of the verb. The verb cannot be used without those expressions. The expressions *se/si* cannot be replaced by the expression *sebe/sobě*, i.e. it is not the case of reflexivity.
Examples:
bát_se (=*to fear*), *usmát_se* (=*to laugh*), *snažit_se* (=*to try*), *stěžovat_ si* (=*to complain*), *odpočinout_si* (=*to rest*), *setkat_se* (=*to meet*), *prát_se* (=*to fight*), *hádat_se* (=*to argue*), *utkat_se* (=*to clash*), *přít_se* (=*to dispute*), *scházet_se* (=*to meet*), *loučit_se* (=*to say goodbye*), *domlouvat_se* (=*to arrange*), *podobat_se* (=*to resemble*)

1.2 Derived reflexive verbs

In the given meaning, we cannot imagine an external originator of the action; the verbs express involuntary actions performed towards oneself. The verbs can be used without the expressions *se/si*, but then they do not signify involuntary action.
Cf.: *Vlny se šíří prostorem.* (*šířit_se* (=*to disseminate*_ REFL))
(=*The waves are disseminated through space.*)
Rozhlas šíří vlny prostorem. (*šířit* (=*to disseminate*))
(=*The radio station disseminates waves through space.*)
Examples (processes in nature, processes performed on inanimate objects): *zlomit_se* (=*to brake*), *roztrhnout_se* (=*to split*), *rozsypat_se* (=*to get spilt*), *šířit_se* (=*to spread*), *rozlít_se* (=*to spill*), *rozpadnout_se* (=*to disintegrate*), *naplnit_se* (=*to fill*)
Větev se zlomila. (=lit. *The branch broke.*)
Pytel se roztrhl. (=lit. *The bag split.*)
Voda se vylila z břehů. (=lit. *Water overflowed the banks.*)
Sál se naplnil lidmi. (=lit. *The hall* REFL*filled (with) people.*)

Examples (processes performed on a person, but not caused by the person (it is not the case of processes performed by a person towards him/herself, i.e. reflexivity)): *trápit_se* (=*to suffer*), *zabít_se* (=*to kill*), *utopit se* (=*to drown*), *zranit_se* (=*to injure*)
Spadl do vody a utopil se. (=lit. *He fell into the water and drowned.*)
Celý rok se trápil. (=lit. *He suffered all year*.)

1.3 Derived reciprocal verbs

Verbs with reciprocal meaning, derived with the help of expression *se/si* with accusative *(se)* or dative (*si*) construction. The prepositional phrase *s+7* comes instead of the accusative/dative.

Derivation of the verbs with the **accusative** construction:
someone kisses someone → someone REFL kisses with someone
líbat ACT(.1) PAT(.4) → *líbat_se* ACT(.1) PAT (s+7) (=*to kiss* → *to kiss* REFL)
Examples: *líbat_se* (=*to kiss*), *objímat_se* (=*to hug*), *potkat_se* (=*to meet*), *pozdravit_se* (=*to greet*), *vítat_se* (=*to welcome*), *navštěvovat_se* (=*to visit*), *fotografovat_se* (=*to photograph*), *popisovat_se* (=*to describe*), *vidět_se* (=*co see*), *léčit_se* (=*to heal*), *natřít_se* (=*to put on*)

Derivation of the verbs with the **dative** construction:
someone telephones something to someone → someone REFL telephones something with someone
telefonovat ACT(.1) PAT(.4) ADDR(3) → *telefonovat_si* ACT(.1) PAT(.4) ADDR(s+7) (=*to telephone* → *to telephone* REFL)
Examples: *telefonovat_si* (=*to telephone*), *konkurovat_si* (=*to compete*), *slíbit_si* (=*to promise*), *věřit_si* (=*to believe*), *blahopřát_si* (=*to congratulate*), *volat_si* (=*to call*), *sdělovat_si* (=*to tell*), *pomáhat_si* (=*to help*)

1.4 Derived reflexive/reciprocal verbs

Verbs with reflexive/reciprocal meaning, derived with the help of expressions *se/si* with accusative construction (*se*). The accusative valency position is cancelled.
Such behavior was not detected in case of verbs with dative construction.

Derivation:
to acquaint someone with someone → "to acquaint oneself with someone" → to acquaint with someone
seznámit ACT(.1) ADDR(.4) PAT(s+7) → *seznámit_se* ACT(.1) PAT(s+7) (=*to acquaint* → *to acquaint* REFL)
Examples: *seznámit_se* (=*to acquaint*), *spojovat_se* (=*to unite*), *lišit_se* (=*to differ*), *zaměřit_se* (=*to focus*)

1.5 Actionsart

The expressions *se/si* are often used in connection with other means (prefixes) to form certain specific meanings of verbs.
Examples: *natahat_se* (=*to become exhausted*), *napracovat_se* (=*to work hard*), *zaposlouchat_se* (=*to listen intently*), *dočíst_se* (=*to (find out by) read(ing)*), *rozepsat_se* (=*to get writing*), *upít_se* (=*to drink oneself to death*), *vyběhat_se* (=*to become worn out*), *pospat_si* (=*to take a nap*), *zalyžovat_ si* (=*to have a ski*), *zatančit_si* (=*to have a dance*)

2 Syntactic formal means

The expression *se/si* as a means of various syntactic operations/transformations above the valency frame of a verb.
Annotation: The expression *se* in *a/aux.rf* with the t-node that represents the verb.

2.1 Passivization

Generalization of Actor (originator of the action).
Annotation: *#Gen*.ACT in position of Actor; the expression *se* in *a/aux.rf* with the node that represents the verb.
Examples:
Přirozený jazyk se popisuje formálními prostředky.
(=lit. *Natural language is described by formal means*.)
Tancovalo se až do rána. (=*Dancing went on until morning.*)
Diskutovalo se o novém objevu. (=*The new discovery was discussed.*)

2.2 Dispositional modality

It is a special type of modality that captures the relation (attitude) of the agent to the action. In Czech, this type of modality is carried by a special construction: the reflexive passive, the adverbial of manner and dative agent.
Annotation: The dative has the ACT functor.
Examples:
Přirozený jazyk se mi.ACT *nepopisuje nejlíp.*
(=*I do not take well to describing natural language*.)
Matematika se mu.ACT *studuje dobře.* (=*He takes well to studying mathematics.*)
*V téhle troubě se mi.*ACT *dobře peče.* (=*I can bake well in this oven.*)

3 Part of a sentence (sentence modification)

The expression *se/si* is a separate semantic unit that expresses the identity of a speaker of given position with another position, namely in reflexive or reciprocal constructions.
Annotation: t-node with the t-lemma substitute, *se/si* in *a/lex.rf*. !!!; grammatical correference.

3.1 Reflexivity

The (closest) subject of the action aims the action at itself.
Annotation: t-lemma substitute: *#PersPron*.
Examples:
Nakreslil se. (=lit. *(He) sketched himslef.*)
Viděl se v zrcadle. (=lit. *(He) saw hinself in the mirror.*)
Zaměřuje_se jen na sebe. (=lit. *(He/She) focuses only on himself/herself.*)
Oholil se. (=lit. *(He) shaved himself.*)
Koupil si auto. (=lit. *(He) bought himself a car.*)

3.2 Reciprocity

The identity of reciprocalized position with lexically-expressed reciprocalized position.
Annotation: t-lemma substitute: *#Rcp*.
Examples:
Petr a Pavel se_potkali mezi sebou. (=lit. *Petr and Pavel* REFL *met mutually.*)
Petr a Pavel potkali sebe navzájem. (=lit. *Petr and Pavel met each other.*)
Petr a Pavel se od sebe navzájem liší. (=lit. *Petr and Pavel* REFL *differ from each other.*)

Numbers

Substitute for MAN 8.10.2.3 Complex numerical expressions
and 8.10.2.4 Decimals and fractions

1 Type *two million five hundred thousand*

Complex numerical expressions of the type sto čtyřicet tisíc (=one hundred and forty thousand), dva miliony pět set tisíc (=two million five hundred thousand), třicet osm (=thirty eight) are analyzed as a paratactic structure. The root of the paratactic structure is a newly established node with #Separ t-lemma and CONJ functor. Elements of paratactic structure involve milliards, millions, thousands, hundreds, tens and ones. If a complex numerical expression contains tens or ones, it is not treated as a container, i.e. the numerical expression (analyzed as a paratactic structure) modifies the counted object as RSTR.
As a result of cancellation of valency of nouns annotation (see the relevant supplement), the counted object after the container numerals is assigned the RSTR functor, not MAT functor.

Examples:

Fig. 1: *sto čtyřicet tisíc židů*.RSTR
(=*one hundred and forty thousand of Jews*)

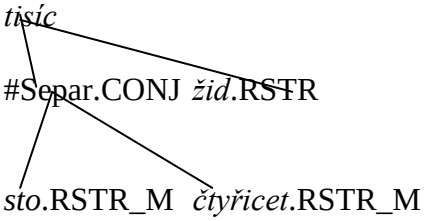


Fig. 2: *Vydělal 2 tisíce čtyři sta pět korun*.PAT
(=*He earned two thousand four hundred and five crowns*)

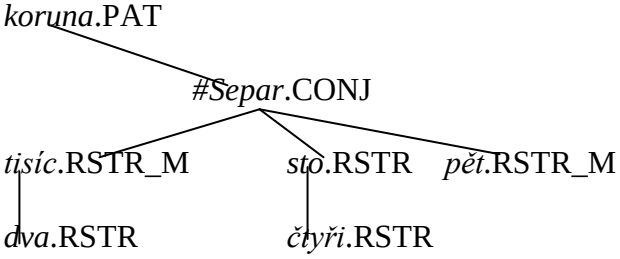


Fig. 3: *Žije zde jeden milion pět set tisíc lidí*.RSTR
(=*One million five hundred thousands people live here*)

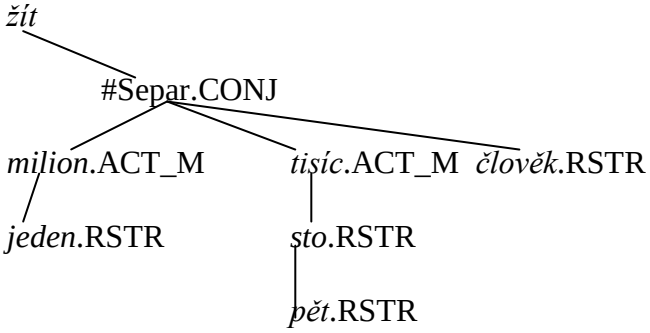


Fig. 4: *třicet osm žáků* (=*thirty-eight pupils*)

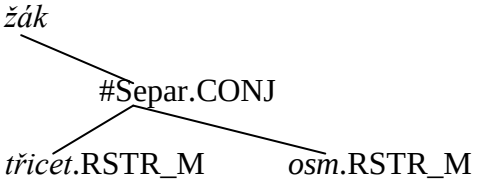


Fig. 5: *u sto dvaceti osmi miliónů dvě stě sedmdesáti čtyř tisíc lidí*
(=*with one hundred and twenty-eight million two hundred and seventy-four thousand people*)

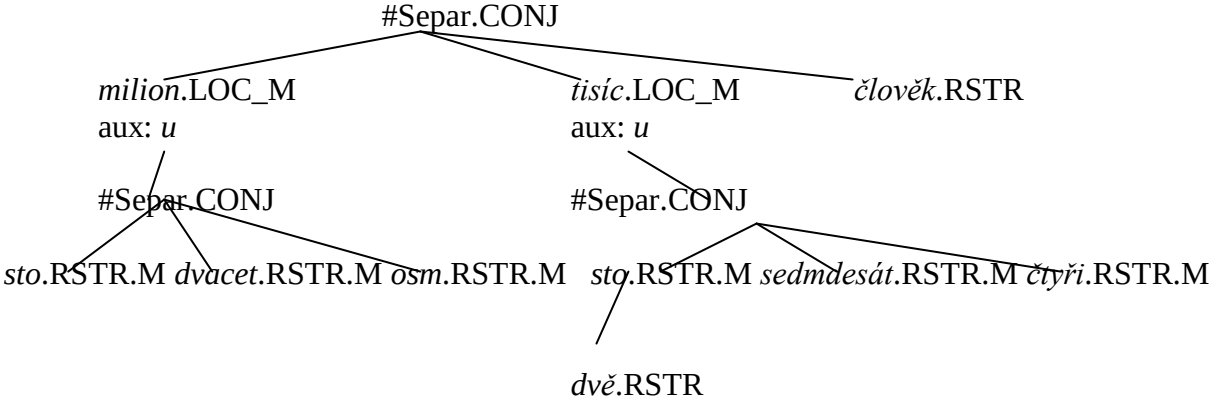
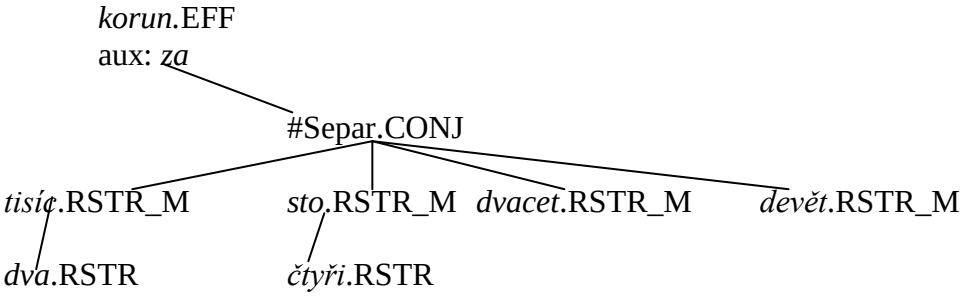


Fig. 6: *za dva tisíce čtyři sta dvacet devět korun*
(=*for two thousand four hundred and twenty-nine crowns*)



2 Numeral written in digits

A numeral written in digits is a single node, it depends on the counted object and it has the RSTR functor. If such numeral is not represented as a single node in data, its t-lemma is underscored!

Examples:

Mám 38 234.RSTR *korun*.PAT (=*I have 38 234 crowns.*)
Mám 38 234.PAT (=*I have 38 234.*)
Vydělal 2 405.RSTR *korun*.PAT (=*He earned 2 405 crowns.*)
Žije zde 1 500 000.RSTR *lidí*.ACT (=*1 500 000 people live here.*)
Vydělává 1 500 000.PAT *ročně.* (=*He earns 1 500 000 a year.*)

3 Decimal number

Decimal number is a single node.

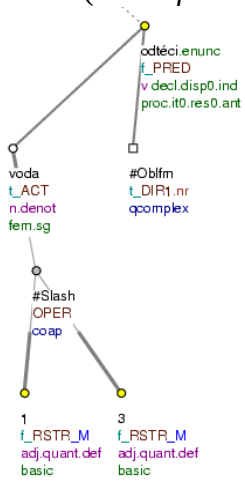
Examples:

Výroba se zvýšila o 2,3.RSTR *procenta.* (=*The production increased by 2.3 per cent*.)
Výstavu navštívilo 2,5.RSTR *tisíc lidí.* (=*2.5 thousand people visited the exposition.*)

4 Type *1/3 of the water*

Data in the form of mathematical operation are represented as a paratactic structure with the help of the OPER functor.

Fig. 7: *1 / 3 vody odtekla.* (*=1/3 of the water flowed away*)



5 Type *2m 10cm 4mm long pole*

Numeric data composed of the combination of numerals and physical units of the type *2m 10cm 4m, 1h 20 min* are represented as a paratactic structure, the root of the structure is a newly-established node with the #Separ t-lemma and CONJ functor.

Examples:

Tyč je dlouhá 2 m.EXT *10 cm*.EXT *4 mm*.EXT. (*=The pole is 2m 10cm 4mm long.*) (Fig. 8)

Film začíná ve 2 hod.TWHEN *35 min*.TWHEN (*=The film begins at 2 hrs 35 mins.*) (Fig. 9)

(Similarly: *Film začíná ve 2 hod*.TWHEN *a 35 min*.TWHEN
(*=The film begins at 2 hrs and 35 mins.*))

Vážil 53 tun.EXT *15 kg*.EXT (*=It weighted 53 t 15 kg.*)

(Similarly: *Vážil 53 tun*.EXT *a 15 kg*.EXT (*=It weighted 53 t and 15 kg.*))

Fig. 8: *Tyč je dlouhá 2 m 10 cm 4 mm.* (*=The pole is 2m 10cm 4mm long.*)

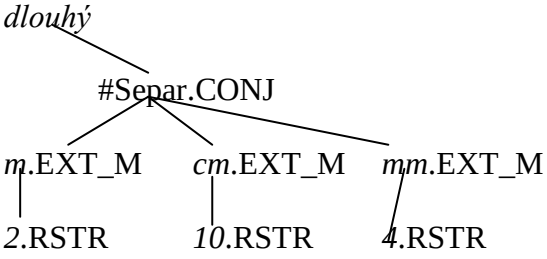
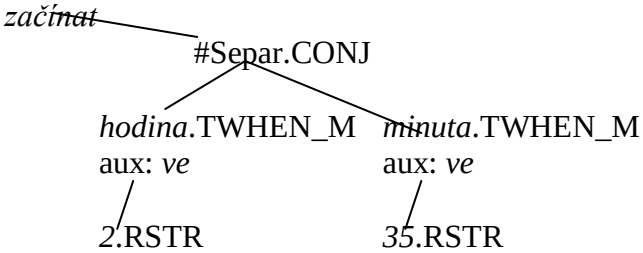


Fig. 9: *Film začíná ve 2 hod 35 min.* (*=The film begins at 2 hrs 35 mins.*)



6 Type *1 1/3 of water flowed out*

Numeric data of the type *1 1/3 vody odtekla* zachytíme jako souřadnou strukturu pomocí uzlu *#Separ*.CONJ.
Examples:
Fig. 10: *1 1/3 vody odtekla.* (*=1 1/3 of water flowed out*)

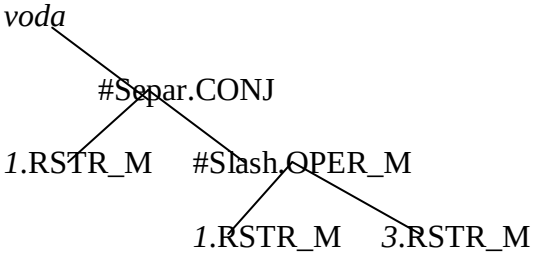
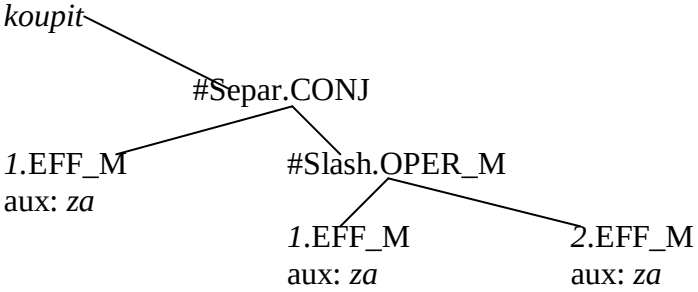
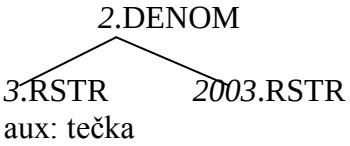


Fig. 11: *koupil za 1 ½* (*=he bought for 1 ½*)



7 Date *3. 2. 2003*

Data ve formátu *3. 2. 2003* zachycujeme podle následujícího příkladu-šablony.
Periods used with day and month are referred to in the a/aux.rf attribute.
Example:
Fig. 12: *3. 2. 2003*



8 Types *40 per cent* and *40 %*

Expressions of the type *40procentní* (*=40%* adj.) – *40 procent* (*=40 per cent*), *40%* (*=40%* adj.) - *40 %* are always represented as two nodes (one node represents the number, the other one represents the word/symbol), including the cases when it is one word in fact (*40procentní, 40%)*. On the contrary, one-word expressions, whole written in alphabetic symbols (*čtyřicetiprocentní* (*=fourtypercent* adj.) are represented as a single node. Thus, there is certain inconsistency in the annotation and we are aware of it.

Examples (nouns):
Fig. 13: *40 procent roztoku*
(*=40 per cent of the solution*)

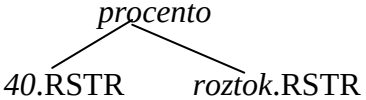


Fig. 14: *40 % roztoku*
(*=40% of the solution*)

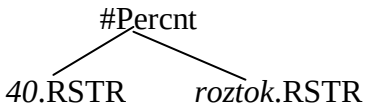
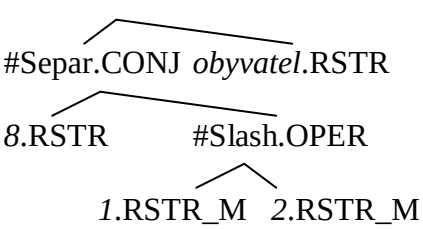


Fig. 15: *8 ½ % obyvatel* (*=8 ½ % of population*)



Examples (adjectives):

Fig. 16: *40procentní roztok* (*=40 per cent* adj. *solution*)

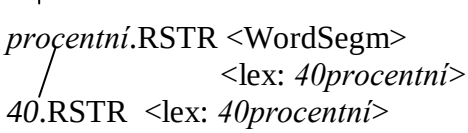


Fig. 17: *40% roztok* (*=40%* adj. *solution*)

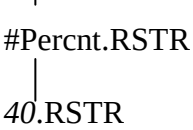
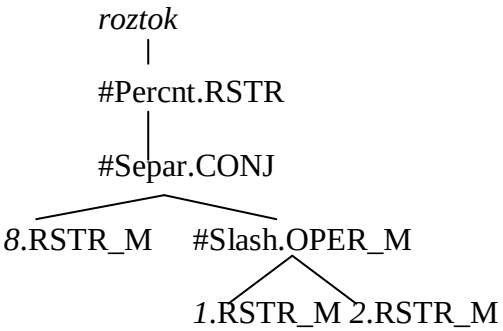


Fig. 18: *½% roztok* (*=8 ½%* adj. *solution*)



Foreign proper names

Supplement to MAN 8.9 Foreign-language expressions

1 List structure (*#Forn*)

As a list structure, with the help of a node with t-lemma *#Forn*, we analyze the types 1.1, 1.2 and 1.3 below. A Czech preposition is always referred with the node with the *#Forn* t-lemma as a a-aux-node. Unchanged forms of foreign names from the text i.e. the word form, even the declined one, serve as the t-lemma for nodes with the functor FPHR; punctuation marks that are part of the name do not have the tectogrammatical substitutional t-lemmas with # at the beginning! In controversial cases (for example when it is not possible to determine whether the word was used in its declined or non-declined form) we prefer to analyze the expression as a list structure.

1.1 Non-declinable foreign name (both multi-word and one-word names!)

Czech examples:
Z {#Forn.DIR1} *Uyuni*.FPFR *je to 22 km.* (=lit. *from Uyuni is it 22 km.*) (Fig. 1)
v deníku {#Forn.ID} *Financial*.FPFR *Times*.FPFR (=lit. *in journal the Financial Times*)
Přispíval do {#Forn.DIR3} *britských*.RSTR *Financial*.FPFR *Times*.FPFR (Fig. 2)
(=lit. *(He) was contributing to British Financial Times.*)
firma {#Forn.ID} *Eagle*.FPHR *Group*.FPHR *V*.FPHR ..FPHR *A*.FPHR ..FPHR (Fig. 3)
(=lit. *company Eagle Group V. A.*)

1.2 Multi-word foreign name where only its part can be declined

Czech examples:
do {#Forn.DIR3} *New*.FPHR *Yorku*.FPHR (=lit. *to New York*)
v{#Forn.LOC} *Rio*.FPHR *de*.FPHR *Janieru*.FPHR (=lit. *in Rio de Janeiro*)
v{#Forn.LOC} *San*.FPHR *Franciscu*.FPHR (=lit. *in San Francisco*)

1.3 Multi-word "non-European" proper names of people

(Applies to the names where a first name and a last name cannot be clearly distinguished and that are not declined or only their part is declined).
Czech examples:
čínský císař {#Forn} *Tung*.FPHR *chun*.FPHR *Chou*.FPHR (Fig. 4)
(=lit. *Chinese emperor Tung chun Chou*)
s jeho ženou {#Forn} *Li*.FPHR *Šu-sien*.FPHR (=lit. *with his wife Šu-sien*)
s předsedou vlády {#Forn} *Li*.FPHR *Pengem*.FPHR (=lit. *with Prime Minister Li Peng*)

2 What is not a list structure

As a list structure are not analyzed the following types.

2.1 Declined foreign names (i.e. with Czech morphological endings)

Czech examples:
v Oaklandu.LOC (=lit. *in Oakland*)
v Riu.LOC (=lit. *in Rio*)
Přispíval do britských.RSTR *Timesů*.DIR3(=lit. *(He) contributed to the British Times*)

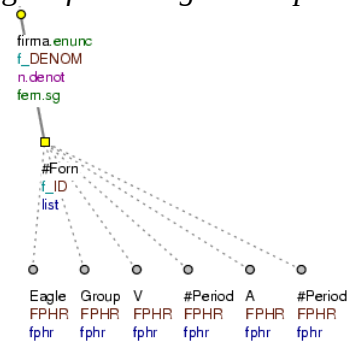
2.2 Foreign „European“ proper names of people

Czech examples:
George Bush (=lit. *Georgie Bush*)
van Gogh (=lit. *van Gogh*)
román Harukiho Murakamiho (=lit. *a roman of Haruki Murakami*)
prezidentu Jangu Šang-kchunovi (=lit. *to President Jang Šang-kchun*)

2.3 One-word foreign names in the position of appellative nominative

In the position of appellative nominative, the node representing one-word foreign name gets the ID functor and we do not add any newly established node with t-lemma *#Forn* to the tectogrammatical tree.
Czech example:
časopis Times.ID (= *the Times magazine*)

Fig. 1: *firma Eagle Group V. A.* (=*company Eagle Group V. A.*)



Error: instead of *#Period*, there is “.” in the t-lemma

Fig. 2: *Přispíval do britských Financial Times.*
(=*(He) was contributing to the British Financial Times.*)

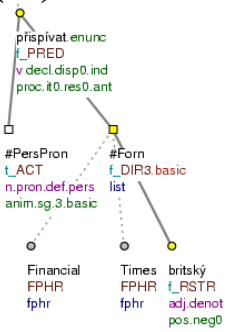


Fig. 3: *Z Uyuni je to 22 km.*
(=*From Uyuni is it 22 km.*)

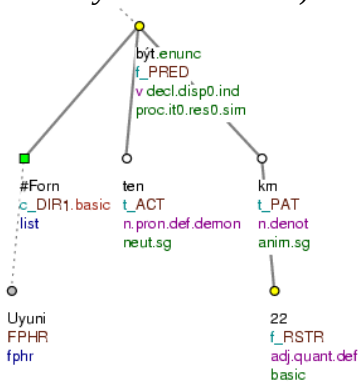
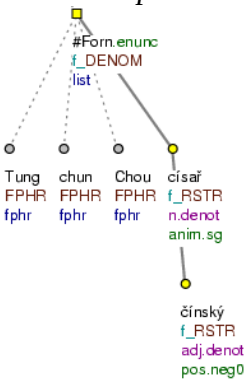


Fig. 4: *čínský císař Tung chun Chou*
(=*Chinese emperor Tung chun Chou*)



Abbreviations

Substitute for the MAN 8.15 Abbreviations

1 Abbreviations

= several initial letters of the abbreviated word usually with a dot at the end; some of them are fixed, i.e. they are always written in the same way.
Czech examples: *p.*, *čís.*, *v. r.*, *př. n. l.*, *a. s.*, *č. j.*, *čtk, atd., pozn. red.*. (similar examples in English: *Mr., No., m.p., B.C., sg., CTK (Czech Press Agency), etc.*). It also includes abbreviations of proper names: *M. Nováková, K. Vary* (in English: *M. Smith, L. Angeles)*

Annotation: No special rules for the annotation are established for this type of abbreviations; the only abbreviations subject to rules for the annotation of identifying expressions are abbreviations of proper names. The individual parts of certain abbreviations (*s. r. o. (ltd.); a. s.* (=*joint stock co.*); *hl. m.* (=*capital city*); *pozn. red.* (=*editor's note*)) are represented by separate nodes and are structurally analyzed; for example the abbreviation *s.r.o.* (=*ltd.)* is represented by three nodes. One node represents abbreviations such as *čtk* (=*CTK (Czech Press Agency)*), *atd.* (=*etc.*), *apod.* (=*etc.*). The t-lemma adopts the full or the abbreviated form; it maintains the form generated by the automatic parser (provided that it is not wrong). If the t-lemma adopts the wrong full form, it is corrected in order to be in the right full form.
Special rules in the manual:
mj. (=*among others*) – MAN 8.6.1 Meaning of a "restriction" (RESTR)
apod. (=*etc.*), *atd.* (=*etc.*), *aj.* (=*etc.*) – 6.6.2.1 Special constructions represented as coordination or apposition
tj. (=*i.e.*), *tzn.* (=*i.e.*) – coordinating connectives (APPS), see the list in the manual.

2 Initialisms

= indeclinable abbreviations created from initials written in capital letters.
Czech examples: *USA, ZOH* (=*Winter Olympic Games*), *UNICEF, MFF* (=*Faculty of Mathematics and Physics*), *ČR* (=*Czech Republic*) (similar examples in English*: USA, WTO*)

Annotation: An initialism is represented by a single t-node and is represented as an identification structure, i.e. the node representing an initialism has the functor ID and is dependent on the root node for the generic common noun that introduces the initialism (*a country, an organization, a faculty)*. If it is not introduced by the generic common noun, a new node with t-lemma *#Idph* is added to the tectogrammatical tree.
Examples:
V {#Idph.LOC} *ČR*.ID *včera nepršelo.* (=lit. *It did not rain yesterday in the CR.*)
{#Idph.ACT} *ČR*.ID *je malá země.* (=lit. *CR is a small country.*)
Pojedeme do {#Idph.DIR3} *USA*.ID (=lit. *We will travel to USA.*)
Organizace.ACT *OSN*.ID *bojuje proti bezpráví.*
(=lit. *The OSN organization fights against injustice.*)
Studuji na {#Idph.LOC} *MFF*.ID (=lit. *I study at MFF*)
Cyklista leží na oddělení.LOC *ARO*.ID (=lit. *A cyclist lies in ward ARO.*)
Cyklista leží na {#Idph.LOC} *ARO*.ID (=lit. *A cyclist lies in ARO*)

3 Acronyms

= abbreviations perceived as words, i.e. they are declinable or may become the base for further derivation.
Czech examples: *Čedok* (=*Czech travel agency*), *ARO* (=*Intensive care unit*) (similar examples in English: *AIDS, UFO)*

Annotation: If the abbreviated word is used in its declined form, it gets a functor according to its position in the sentence. If an indeclinable form is used, it is represented with the help of the same rules as in the case of initialism, i.e. as an identification structure.
Examples:
Na dovolenou jezdí s Čedokem.MEANS (=lit. *They go for holiday with Čedok.*)
„Jak na Nový rok, tak po celý rok," povzdechl si pacient na ARU.LOC
(=*"You spend whole year the way you spend New Year's Day," sighed a patient in ARO ward.*)

Hyphen

Supplement to MAN 8.18 Punctuation

The rule applied: something+dash/hyphen+dash (Czech examples: *BMP-1, srbsko-černohorský, c-dur, Rolls-Royce* (in English: *BMP-1, Serbo-Montenegrian, c-major, Rollce-Royce*)) is always represented as 3 nodes on the analytic level.

1 Compositional combination

Hyphen/dash combines words (symbols) that carry meaning on their own as well.

1.1 Words combined with a hyphen/dash are in a relation of parataxis

Czech examples: *anglicko-český slovník, zeleno-žluto-modrá láta, KDU-ČSL* (in English: *English-Czech dictionary, green-yellow-blue fabric, KDU-ČSL (Christian Democratic Union-Czech Party Liberal)*

Annotation: The group *something+dash/hyphen+something* is represented as **three nodes**, the node that represents the last dash/hyphen is the paratactic structure root node (*#Dash*.CONJ). Other possible dashes/hyphens (in a multiple combination) are not represented by single nodes neither are they annotated as a-aux-nodes.

Fig. 1: *zeleno-žluto-modrá látka* (=*green-yellow-blue fabric*)

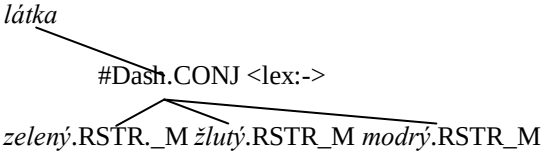
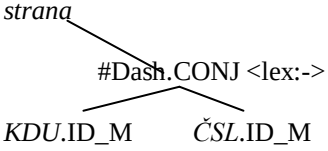


Fig. 2: *strana KDU-ČSL* (=*KDU-ČSL party*)



1.2 Combined words have a relation of determination

Czech examples: *c-dur, RM-systém, protein BMP-1, C-vitamin, alfa-záření, Rh-faktor, Praha-Smíchov, Brno-venkov, Havlíček-kritik.* (English examples: *c-major, BMP-1 protein, C-vitamin, alpha-radiation, Rh-factor, Portland-central (station), Manchester-west*)

Annotation: The group *something+dash/hyphen+something* is represented as **two nodes**: one node as governing and the other one as dependent. The node for a dash/hyphen is not represented as a single node on the t-level, however, it is annotated as an a-aux-node in the dependent node of the expression.

Fig. 3: *stupnice c-dur* (=*scale of c-major*)

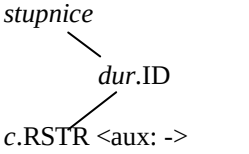


Fig. 4: *v RM-systému* (=*in the RM-system*)

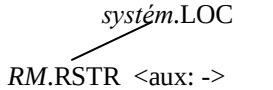
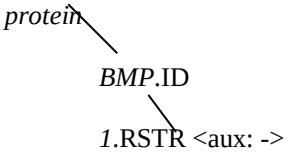


Fig. 5: *protein BMP-1* (=*BMP-1 protein*)



NB! There are special rules for multi-word proper names of people and cities: *Marie-Anna, Sklenářová-Malá, Frýdek-Místek* (English examples: *Mary-Ann, Bedford-Stuyvesant*)! Those expressions fall into the type 1.2 (see MAN 8.8.2.1 Specific rules for certain types of proper nouns).

2 One word

The expression cannot be decomposed into the meaning of individual parts (particularly in case of loanwords).
Annotation: The expression is analyzed by analogy according to the rules for idiomatic expressions (phrasemes) (as DPHR; MAN 6.8 Idioms (phrasemes)). The second, declined part is usually the governing node of the expression.
Examples:
Rolls-Royce (=lit. *Rolls-Royce*)
ping-pong (=lit. *ping-pong*)
moucha tse-tse (=lit. *tse-tse fly*)

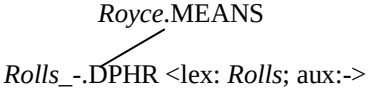


Fig. 6: *Na lov jedině Rolls-Roycem.*

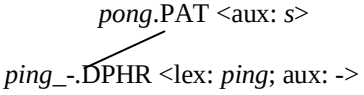


Fig. 7: *Končí s ping-pongem.*

New secondary prepositions

Supplement to MAN 8.17.1 Secondary prepositions

The list of approved secondary prepositions is appended to the large manual. The appendix includes the list of expressions that can serve as secondary prepositions in a specific text. However, the expressions from this list are not always secondary prepositions. The words mentioned as a part of secondary preposition can have their semantic and syntactic role in the clause and in such a case they are represented by an individual node in the tectogrammatical tree and do not make a part of the preposition.

Secondary prepositions that are not included in the list:
Even other expressions that are not included in the given list can act as secondary prepositions in the data. If the annotator "suspects" that a multi-word expression has a function of secondary preposition, s/he takes the following steps:

- multi-word expression shall not be captured as a secondary preposition; the comment of the *Secondary preposition* type shall be assigned to the node (one of nodes) that would not be an individual t-node if captured as a secondary preposition (and nothing else).
- potential secondary prepositions shall be subject to discussion and the list of secondary prepositions shall be extended gradually.

Various

1 ID belongs to numbers-labels

Change! ID (not RSTR) belongs to numbers-labels!
Examples:
linka 900.ID (=*line 900*)
telefon 342 434 234.ID (=*telephone 342 434 234*)
PSČ 561 64.ID (=*postcode 561 64*)
číslo 1.ID (=*number 1*)
Mazda 323.ID (=*Mazda 323*)
v roce 1993.ID (=*in year 1993*)
kurs 3.ID *ku*.OPER *5*.ID (=*exchange rate 3 to 5*)

2 Meaning: “how far”

Examples:
Odcestoval přes 30 km.farmore *na jih*. (=*He traveled over 30 km to the south.*)
Odcestoval okolo 30 km.farapprox *na jih*. (=*He traveled about 30 km to the south.*)
LOC
Umístil se na pátém místě.LOC.far (=*He placed fifth.*)
Umístil se kolem pátého místa.LOC.farapprox (=*He placed around fifth.*)
Umístili se od pátého místa.LOC.farmoreincl (=*He placed from fifth place.*)
Umístil do pátého místa.LOC.farlessincl (=*He placed to fifth place.*)
DIR3
Skočil dva metry.DIR3.far (=*He jumped two meters.*)
Skočil přes dva metry. DIR3.farmore (=*He jumped over two meters.*)
Voda sahá po kolena.DIR3.farlessincl (=*Water reaches the knees.*)
Subfunctors will be established later.

Fig. 1: *Vesnice leží 30 km na jih*.
(=*The village lies 30 km to the south.*)

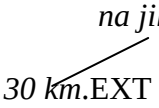
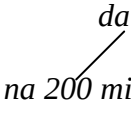


Fig. 2: *Posunuli se dva metry jižněji*.
(=*They have moved 2 meters to the south.*)

Fig. 3: *chvění bylo pociťováno až na 200 mil daleko*
(=*vibration was felt up to 200 miles away*)



3 Price for something, something for some price

Valency (the valency of nouns is not annotated):
price, bill, charge for something.RSTR
cena (=price) EXT(how much) RSTR(*for* +4)
cena 20 korun.EXT *za chleba*.RSTR (=*the price of 20 crowns for bread*)
RSTR(.2 (=*of*)
cena chleba.RSTR (=*the price of bread*)
vývozní cena naší oceli.RSTR (=*export price of our steel*)
ceny mouky.RSTR (=*prices of flour*)
RSTR(.2;.u;od+2) RSTR(.3)
(to award), medal) *cena sportovcům*.RSTR (=*the award to sportsman*)
poplatek (=charge)
EXT(how much) RSTR(for +4)
poplatek 20 korun.EXT *za vstupné*.RSTR
(=*the charge of 20 crowns for admision*)
poplatek za zavolání.RSTR (=*charge for the call*)
účet (=bill) EXT(how much) RSTR(for+4)
účet za tuto knihu.RSTR (=*a bill for this book*)
úrok (=interest) EXT(how much) RSTR(for+4)
úrok za půjčku.RSTR (=*interest rate for the loan*)

Non-valency:
value for something.REG **something for value.EXT**
6,5 mil. za podíl.REG (=*6.5 million for a share*)
podíl za 6,5 mil.EXT (=*share for 6.5 million*)
koruna za papír.REG (=*a crown for the paper*)
papír za korunu.EXT (=*paper for a crown*)
2,10 dolaru za akcii.REG (=*$ 2.10 per share*)
akcie za 2,10 dolaru.EXT (=*shares for $ 2.10*)
78 centů za libru.REG (=*78 cents per pound*)
libra za 78 centů.EXT (=*pound for 78 cents*)

4 EXT – ID – RSTR

pět.**RSTR** *čísel* (=*five numbers*)
číslo pět.**ID** (=*the number five*)
páté.**RSTR** *číslo* (=*the fifth number*)

The number of "things" is RSTR (it would be more suitable to assign the functor EXT):
30 000.RSTR *korun* (=*30.000 crowns*)
24.RSTR *dolarů* (=*24 dollars*)
dvě.RSTR *stanoviska* (=*two attitudes*)
přečetl jsem přes 100.RSTR *stránek*.PAT) (*přes* (=*over*) in *a/aux.rf* at RSTR)
(=*I have read over 100 pages.*)

In case of verbs, we follow the valency lexicon:
činí to 30 000 korun.EXT (=*it is 30,000 crowns*)
vydělává 30 000 korun.PAT (=*he earns 30,000 crowns*)
získal 30 000 korun.PAT (=*he received 30,000 crowns*)

Note! No PAT!
příjem zboží.RSTR (=*receipt of goods*)
příjem dvou.RSTR *korun*.RSTR (=*receive of two crowns*)
příjem dvě.RSTR *koruny*.EXT (=*receive of two crowns*)
příjem 30 000.RSTR *korun*.EXT (=*income of 30,000 crowns*) (also as payment)
zisk pěti.RSTR *dolarů*.RSTR (=*profit of five dollars*)
zisk pět.RSTR *dolarů*.EXT (=*profit of five dollars*)
výnos 102.RSTR *miliard*.EXT (=*yield of 102 billion*)
tržba 2.RSTR *miliardy*.EXT (=*revenue of 2,000,000,000*)
daň 3 000 korun.EXT (=*tax of 3,000 crowns*)
objem 10 kubíků.EXT (=*volume of 10 cubic meters*)
objem vody.RSTR *10 kubíků*.EXT (=*volume of 10 cubic meters of water*)
v hodnotě 30 000.RSTR *korun*.EXT (=*valued at 30,000 crowns*)
hodnota vody.RSTR (=*value of water*)
suma 3 000.EXT (=*the sum of 3,000*)
ve výši 30 000.RSTR *korun*.EXT (=*amounting to 30,000 crowns*)
délka 3.RSTR *metry*.EXT (=*length of 3 meters*)
celodenní minimum 2 141,1.EXT (=*all-day minimum 2,141.1*)
maximum 2 141,1.EXT (=*maximum of 2,141.1*)
Analogically, indefinite cardinal numerals:
více.RSTR *obchodních transakcí* (=*more transactions*)
více.RSTR peněz (=*more money*)
za kolik.RSTR *peněz* (=*how much money*)
Meanings of extent:
vydělává 30 000.RSTR *korun*.PAT (=*he earns 30,000 crowns*)
vydělává do 30 000.RSTR *korun*.PAT (=*he earn up to 30,000 crowns*)
vydělává od 30 000.RSTR *korun*.PAT (=*he earns minimally 30,000 crowns*)
vydělává pod 30 000.RSTR *korun*.PAT (=*he earns less than 30,000 crowns*)
Prepositions are in *a/aux.rf* at RSTR.

5 The end of a sentence is missing

If any sentence is missing its end (even in the original), then we annotate the sentence according to the normal rules of annotation as if the sentence was unfinished. In place of the missing part, there will be *#Emp**. The root of the tree gets the annotators‘ note *Other* with the text: *The end of a sentence is missing*.

6 “context(s)”

Example: *Typické Glassovo dílo je nejen s otevřeným koncem, je také často mnohoznačné, pokud jde o kontext(y).* (=*A typical work of Glass has not only an open end, it is often also ambiguous in terms of the context(s).*)
Interpretation: *in terms of the context or contexts*
In the sentence, we capture the coordination between the *context* and the copied *context*, the root of the coordination structure will be the node with t-lemma *#Separ* (functor according to the meaning, in this case DISJ). The copied node with t-lemma *context* will have the following values in the following attributes: *is_generated* = 1; *is_parenthesis* = 1; *a/lex.rf*: reference to *kontext*; *a/aux.rf*: references to: *o ( ) y*.

Segmentation of sentences to trees and segmentation of words to nodes

1 Segmentation of sentences to trees

1.1 Several sentences are in one tree

If there are several sentences in one tree (thus, there should be two, three trees etc.), the effective roots of both sentences will be hung as sisters on the technical root of the tree (*nodetype = root*). The comment such as *2 in 1* will be typed in the *anot_comment* of the *Sentence Segmentation* type by the technical root of the tree.
This is the way to solve just the cases when it is obvious that several sentences have been connected to a single tree (i.e. there are two or more sentences ended by a punctuation signifying the end of sentence in a single tree: *First sentence? Second sentence.*) Cases when several clauses are separated by e.g. a dash (*Praha (haš) – Sentence.*) are captured as a subordinate structure (coordination, apposition) and are not perceived incorrect sentence segmentation.

1.2 One sentence is divided into several trees

If one sentence is divided into several trees, all the nodes of the sentence are deleted in the trees where there are few nodes and the missing nodes are created again in the tree where there are the most nodes (i.e. they are "extracted" again from the analytical tree but their *is_generated* = 0). The sentence will be annotated in such a tree. The comment such as *1 in 3 < >* will be typed in the *anot_comment* of the *Sentence Segmentation* type by the technical root of the tree in which the sentence was annotated. The < or > mark serves to indicate whether the remaining parts appeared in the preceding or in the following tree.

2 Segmentation of words to nodes

2.1 One t-node represents several semantic units

If one t-node represents several semantic units (i.e. there should be two, three, etc. nodes), we create as many new nodes (they will be *is_generated* = 0) in the tree as we need for the correct annotation. T-lemmas of the nodes are to be corrected manually. If there is an error also at the analytical level, there are references from the newly created t-nodes to a single identical incorrect a-node. Such place is to be marked with a comment of the *Word Segmentation* type, reading *3 in 1.*
Example: *Serbian-Montenegrin* is one node but it should be three nodes.

2.2 One semantic unit is represented by several t-nodes

If one semantic unit is represented by several t-nodes, we select one t-node that is going to remain in the tectogrammatical tree and the rest of t-nodes is to be deleted (they are added as a-aux/lex-nodes). The t-lemma of the remaining t-node is to be corrected to the right value, and references to a-nodes are to be annotated correctly.
Example: there are two nodes in *van Beethoven*. It is not necessary to mark the standard cases of the *van Beethoven* type with the *Word Segmentation* comment.
In non-standard cases the comment has to be selected and we type in e.g. *1 in 2.*

Annotator's comment

The *annot_comment* attribute (the structure of *type* and *text* attributes) is a draft attribute that is used to mark trouble spots in data to which it will be necessary to get back during the final correction and solve the problem. The attribute will be deleted from the final form of the data. Certain types of problems can be anticipated and that is why the types of troubles were defined. In the *type* sub-attribute the annotator selects the type of the trouble. S/he can type in a comment in the *text* sub-attribute. If there is a phenomenon in the data that is marked as a trouble spot (the type of comment for this phenomenon is mentioned), it is the annotator's duty to mark the spot by selecting a proper type of comment. The text (commentary) about the comment is obligatory for certain types of comments.

1 Uncertainty in structure and function

1.1 Structure ?

When to use? This comment marks the spots where there is an uncertainty in structure of the tree, i.e. particularly in case of vague dependence, when the annotator is not sure about the rule that should be applied to annotate the sentence.
Text of comment: obligatory. The text should include the source of uncertainty, the rules between which the annotator hesitates.

1.2 Functor ?

When to use? This comment marks the nodes where the annotator hesitates concerning the selection of functor and wants to put a commentary to the uncertainty.
Text of comment: obligatory. The annotator puts a commentary in the text of the comment.

2 Incorrect t-lemma

2.1 T-lemma with underscore !

When to use? The comment marks the nodes where the original "automatic" t-lemma was changed to a multi-word t-lemma (to the t-lemma with underscore) and the established macros (Alt+L; Alt+F) could not be used during the change, i.e. the change had to be performed manually.
Text of comment: optional.

2.2 M-lemma M

When to use? This comment marks the nodes where the original "automatic" t-lemma was (manually) changed because its form was wrong for the reason of incorrect morphological analysis. For example: *is* has the *to be* t-lemma instead of *#PersPron*.
It does not apply to the nodes of the *fphr* type, corrected by the Alt+F macro.
Text of comment: optional.
When this type of comment is selected, it also means that there is an error in the morphological tag or another error at the a-layer. Therefore, the annotator does not (have to) type in anything in the comment of the *M-tag* type.

2.3 Typo !

When to use? This comment marks the nodes where the original "automatic" t-lemma was manually changed because of a typing error in the t-lemma.
Text of comment: optional.

2.4 T-lemma T

When to use? The comment marks the nodes where the original "automatic" t-lemma was manually changed for other reasons then these mentioned in 2.1 to 2.3.
Text of comment: obligatory. The reason for the change of the t-lemma is to be described.

3 Valency problems

3.1 New valency frame !

When to use? This comment marks the spots where it is necessary to change the vallex, and the change is clearly specified. It is also used to distinguish between event and non-event meanings of deverbal nouns ending with *-ní* and *-tí.*
Text of comment: obligatory and clearly specified (see ADDval_verb and ADDval_nonverb):

F	+suggestion of a new frame	(frame missing in vallex, no frame is assigned)
W	+ suggestion of a new frame	(word missing in vallex, no frame is assigned)
A	+new form	(an alternative missing in the record of form, frame is assigned)
N	+new functor	(nonobligatory actor missing in frame, frame is assigned)
X	+new functor	(functor missing in alternative group, frame is assigned)
P		(event meaning of deverbal noun ending with *-ní* and *–tí*)
nP		(non-event meaning of deverbal noun ending with *-ní* and *–tí*)

3.2 Vallex ?

When to use? This comment is used as a commentary to the vallex.
Text of comment: obligatory. Mention the commentary.

4 Problems with translation, sentence cannot be annotated

4.1 Translation !

When to use? This comment marks wrongly translated sentences (that have been, however, annotated):

- sentence has been annotated but it is necessary to delete (because they are superfluous) certain words/characters (represented by t-nodes) from the sentence.
- another problem with translation.

The comment is to be selected in the root of the tree (*nodetype = root*).
Text of comment: obligatory. It is obligatory to mention the words (characters) that are to be deleted from the sentence. Annotators can also add other comments about the translation.

4.2 Re-generated X

When to use? This comment marks a wrongly translated sentence that cannot be annotated for this reason. It has to be re-generated in a new version that is to be mentioned in the text of this comment.
The comment is to be selected in the root of the tree (*nodetype = root*).
Text of comment: obligatory and clearly specified. It includes new exact version of the sentence. The comment is not to include other information. The commentary can be typed in the *Not-annotated* comment.
If the comment of the *Re-generated* type is selected, it also means that the sentence has not been annotated at all, that is why the annotator does not have to select either the *Not-annotated* or the *Translation* comment.

4.3 Not-annotated X

When to use? This comment marks the sentences that have not been annotated for other reasons than translation. The comment is to be selected in the root of the tree (*root*).
Text of comment: obligatory. The reason why the sentence has not been annotated. The commentary about the version of the sentence is to be typed in the *Re-generated* comment.

5 Wrong segmentation

5.1 Word segmentation w

When to use? This comment marks the spots where a single t-node represented several semantic units (thus, there should be two, three, etc. nodes) and there is also an error at the analytical level. It is sufficient to select the comment for one of the "new" t-nodes. The comment is also to be used for non-standard spots where there are several nodes for a thing that should be captured by a single node.
Text of comment: obligatory and clearly specified. Comment of the *3 in 1, 1 in 2* type is to be typed in.

5.2 Sentence segmentation s

When to use? This comment marks the spots where there are several sentences in a single tree (thus, there should be two, three, etc. tree) and the spots where a single sentence is divided into several trees. The comment is to be selected in the root of the tree (*nodetype = root*).
For more information, see ADDsegmentation.
Text of comment: obligatory and clearly specified. Comment of the *2 in 1*; *1 in 2 >* is to be typed in.

6 Error in morphology

6.1 M-tag t

When to use? This comment marks the nodes where we found an error in morphological categories, in the tag (yet the t-lemma is correct).
Text of comment: optional.
If the error in the morphological tag influences the form of the t-lemma, the annotator selects the *M-lemma* comment and then it is not necessary to select the *M-tag* comment.

7 Secondary preposition

7.1 Secondary preposition

When to use? This comment marks the spots with a potential secondary preposition. It is assigned to one of the nodes that will not be an independent node of the tree if there is a secondary preposition.
Text of comment: obligatory and clearly specified. It includes exact version of the potential secondary preposition (and nothing else).

8 Other problem

8.1 Other ?

When to use? This comment marks the spots whit another problem (that the one defined above).
Text of comment: obligatory.

Types of comments and information whether the text of the comment is obligatory:

	Type of comment	When to use?	Text
M	**M-lemma**	incorrect t-lemma because of a morphological error	NO
L	**T-lemma**	other error in t-lemma	YES
t	**M-tag**	error in tag, t-lemma is OK	NO
!	**Typo**	typing error in t-lemma	NO
!	**T-lemma with underscore**	t-lemma manually corrected to multi-word	NO
!	**New valency frame**	missing valency frame (suggestion of a new frame)	YES
?	**Vallex**	problem in vallex (which)	YES
w	**Word segmentation**	incorrect segmentation of words to nodes	YES
s	**Sentence segmentation**	incorrect segmentation of sentences to trees	YES
!	**Translation**	error in translation, yet the sentence is annotated (describe it)	YES
X	**Re-generate**	incorrect translation (mention correct version of the sentence)	YES
X	**Not annotated**	sentence cannot be annotated (mention the reason)	YES
#	**Secondary preposition**	new secondary preposition	YES
?	**Structure**	uncertainty in structure (mention why)	YES
?	**Functor**	comment to the uncertainty in functor	YES
?	**Other**	other problem (mention which one)	YES

Templates

1 Templates for constructions from “Wall Street Journal”

1.1 ***Company X. Y., Los Angeles, California***

Fig. 1.1.1: *společnost X. Y., Los Angeles, California*
(=*company X. Y., Los Angeles, California*)

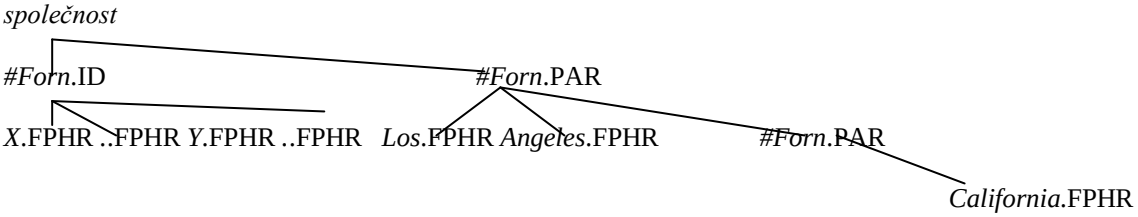


Fig. 1.1.2: *společnost X. Y., Los Angeles, Kalifornie*
(=*company X. Y., Los Angeles, California*)

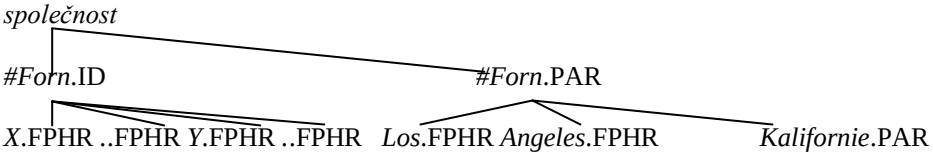


Fig. 1.1.3: *v nemocnici Columbia Hospital, Milwaukee,…*
(=*in the hospital Columbia Hospital, Milwaukee, …*)

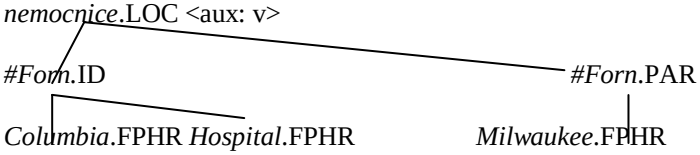


Fig. 1.1.4: *Tiny Tots Inc., sídlem v Campbellu v Kalifornii, ...*
(=*Tiny Tots Inc., seated in Campbell in California*)

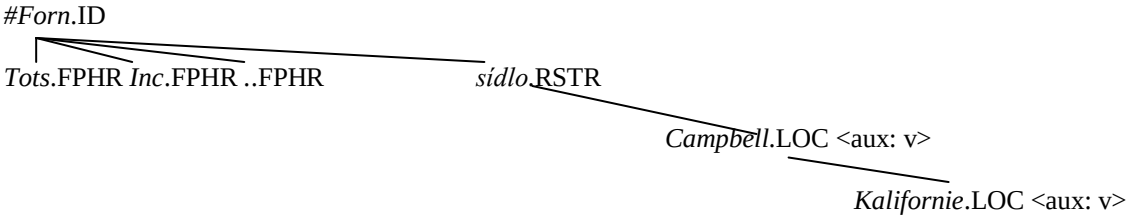


Fig. 1.1.5: *z firmy Health Care(,) se sídlem v Piscataway , N . J .*
(=*from the company Health Care(,) seated in Piscataway , N . J .*)

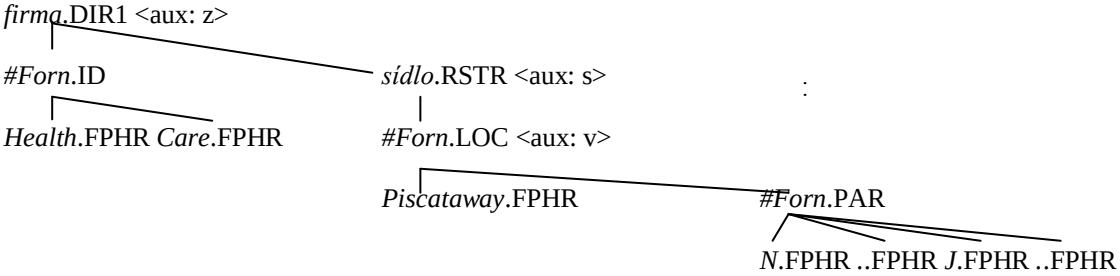
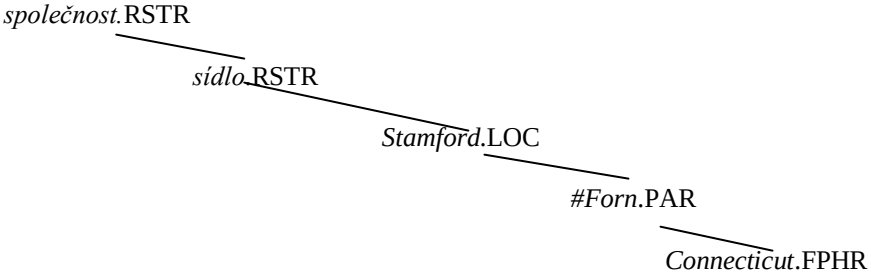


Fig. 1.1.6.: *Mluvčí společnosti(,) se sídlem ve Stamfordu, Connecticut, řekl...*
(=*A company spokesperson(,) seated in Stamford, Connecticut, said...*)



1.2 ***Earnings will increase by more than 20 % compared with 87.7 million dollars, or 1.25 dollars per share for the same period last year***

Fig.: 1.2.1: *Firma oznámila za třetí čtvrtletí čistý zisk 2,3 milionu dolarů , čili 68 centů na akcii, oproti čistému zisku 5,3 milionu dolarů , čili 1,61 dolaru na akcii, před rokem.*
(=*The company announced third-quarter net profit of $ 2.3 million, or 68 cents per share, compared with net profit $ 5.3 million, or 1.61 dollars per share, a year ago.*)

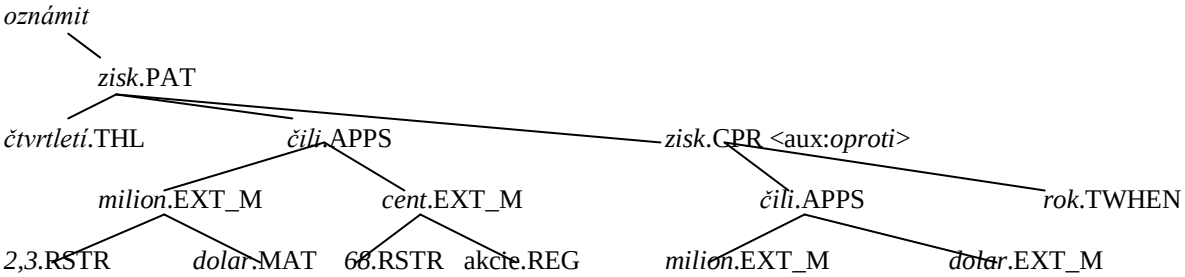


Fig. 1.2.2: *Zisk firmy XYZ se zvýší o více než 20 % oproti 87,7 miliónu dolarů, neboli 1,25 dolaru na akcii za stejné období loňského roku.*
(*=The profit of the company XYZ will increase by more than 20 % from 87.7 million, or $ 1.25 per share for the same period last year.*)

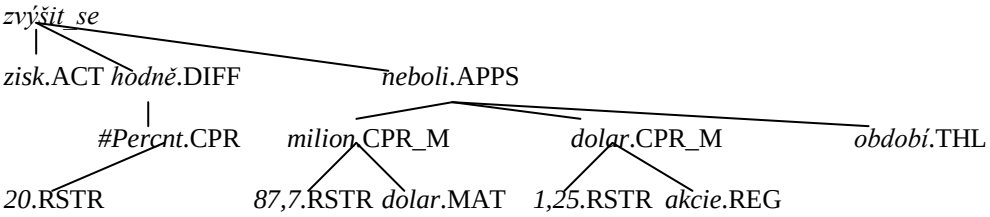


Fig. 9: *Ve třetím čtvrtletí roku 1988 Merck vydělal 311,8 miliónu dolarů, neboli 79 centů na akcii.* (*=In the third quarter of 1988, Merck earned 311.8 million, or 79 cents per share.*)

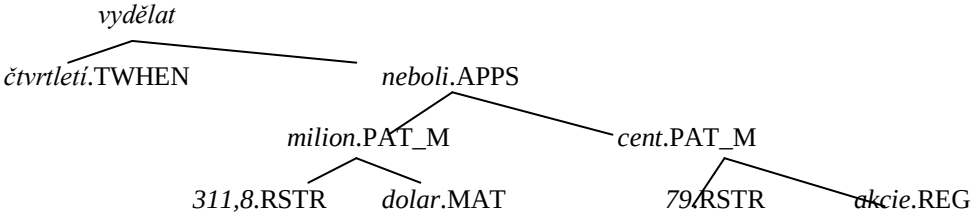


Fig. 1.2.3: *Společnost uvedla, že očekává výdělek od 14 do 17 centů na akcii, což je snížení z 25 centů na akcii v tomto období loňského roku.* (*=The company said it expects earnings from 14 to 17 cents per share, down from 25 cents per share in the same period last year.*)

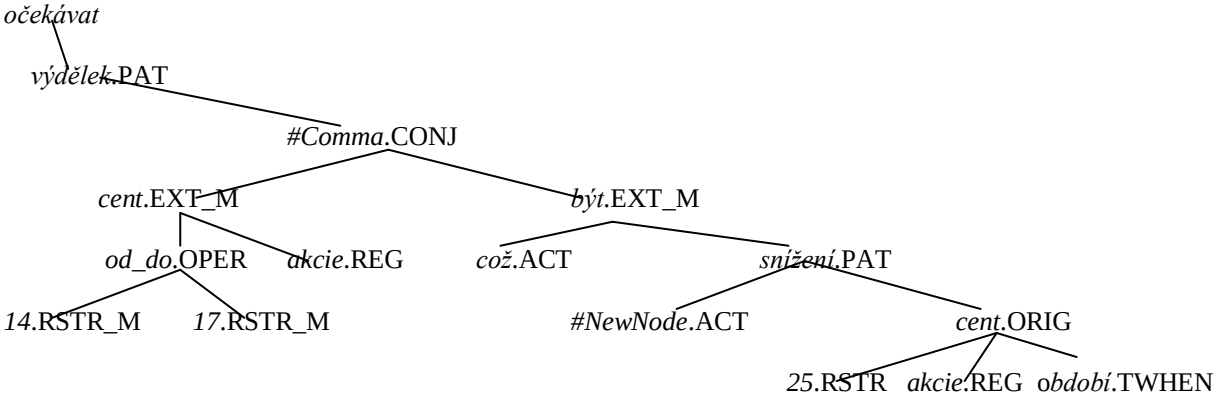
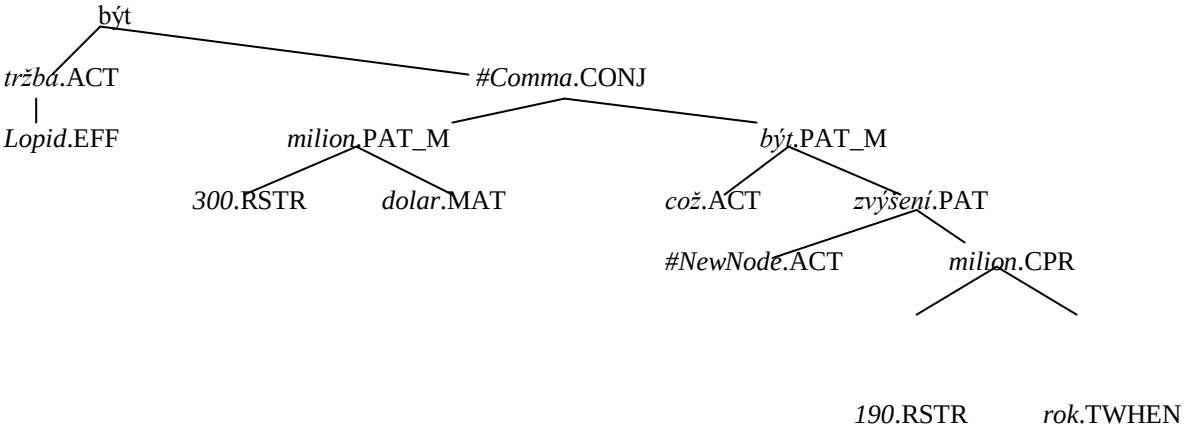


Fig. 1.2.4: *Očekává se, že tržby za Lopid budou letos okolo 300 miliónů dolarů, což je zvýšení oproti 190 miliónům v roce 1988.* (*=It is expected that sales of Lopid will be around $ 300 million this year, which is an increase from 190 million in 1988.*)



1.3 Exchange lists

1.3.1: *Columbia Savings & Loan (NYSE ; Symbol : CSV)*
1.3.2: *Třetí čtvrtletí, 30. září 1989: čistá ztráta 11,527 USD na akcii vs . čistý příjem: 37 centů na akcii*
(=*Third quarter, 30th September 1989: net loss 11,527 USD per share vs. net income: 37 cents per share*)
1.3.3: *Veřejně obchodované kmenové akcie: 19.6 milionu*
(=*Publicly traded ordinary shares: 19.6 million*)
1.4.4: *H&R Block (Newyorská burza cenných papírů; Symbol: HRB)*
(=*H&R Block (New York Stock Exchange, Symbol: HRB)*)
1.3.5: *Obor podnikání: Daňové poradenství*
(=*Scope of business: Tax advice*)
1.3.6: *Rok končící 30. dubna 1989:*
(=*Year ended April 30, 1989:*)
1.3.7: *Čistá ztráta: 100,2 milionů dolarů; 1,90 dolaru na akcii*
(=*Net loss: $ 100.2 million, $ 1.90 per share*)
1.3.8: *Příjmy na akcii: ztráta 8 centů versus ztráta 9 centů*
(=*Income per share: loss of 8 cents versus loss of 9 cents*)
1.3.9: *Pozn.: Do všech údajů "za akcii" byly započítány veškeré výsledky konverzních operací.*
(=*Note: To all "per share", all results of conversion operations were included.*)

Fig. 1.3.1: *Columbia Savings & Loan (NYSE; Symbol : CSV)*

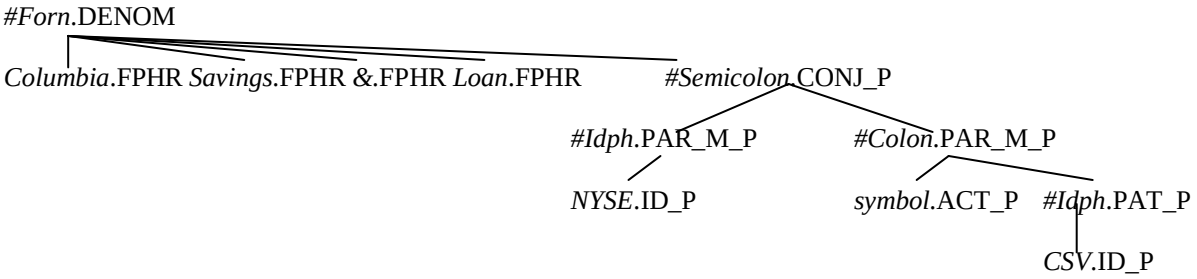


Fig. 1.3.2: *Třetí čtvrtletí, 30. září 1989: čistá ztráta 11,527 USD na akcii vs. čistý příjem: 37 centů na akcii*
(=*Third quarter, 30th September 1989: net loss 11,527 USD per share vs. net income: 37 cents per share* (a colon after *„příjem“* – "*income*" is neglected in this case))

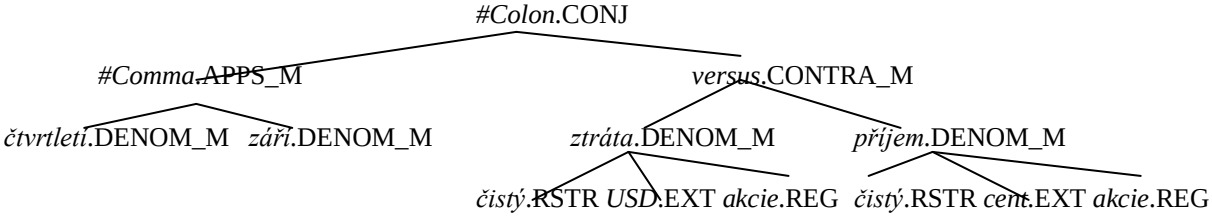


Fig. 1.3.3: *Veřejně obchodované kmenové akcie: 19,6 milionu*
(=*Publicly traded ordinary shares: 19.6 million*)

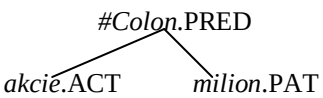


Fig. 1.3.4: *H&R Block (Newyorská burza cenných papírů; Symbol: HRB)*
(*=H&R Block (New York Stock Exchange, Symbol: HRB)*)

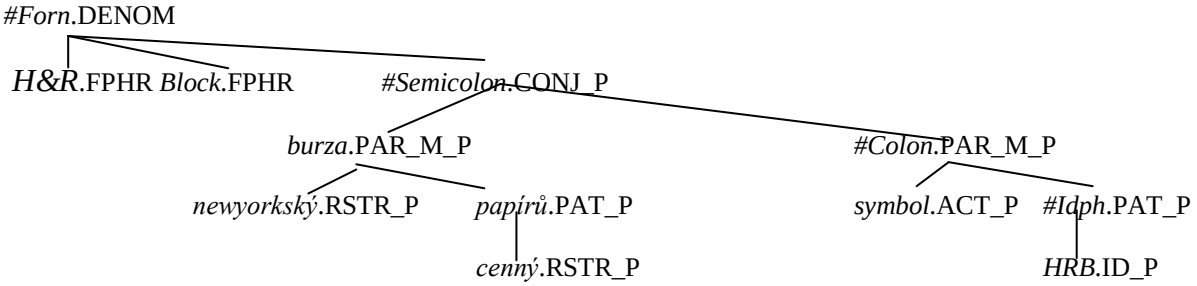


Fig. 1.3.5: *Obor podnikání: Daňové poradenství* (*=Scope of business: Tax advice*)

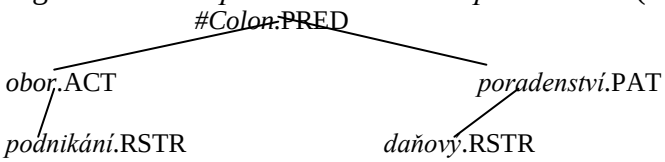


Fig. 1.3.6: *Rok končící 30. dubna 1989:* (*=Year ended April 30, 1989:*)

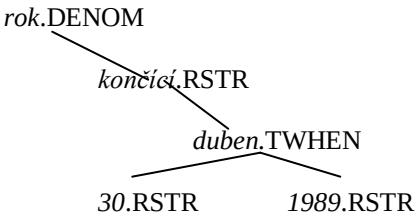


Fig. 1.3.7: *Čistá ztráta: 100,2 milionu dolarů; 1,90 dolaru na akcii*
(*=Net loss: $ 100.2 million, $ 1.90 per share*)

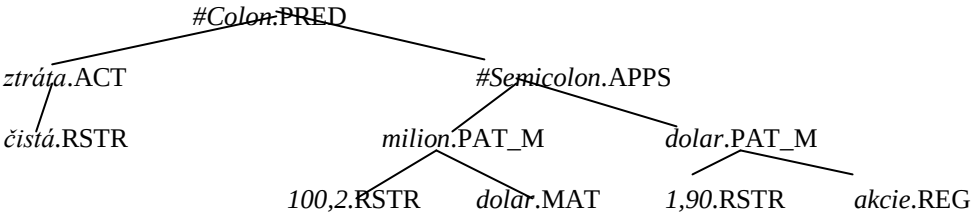


Fig. 1.3.8 *Příjmy na akcii: ztráta 8 centů versus ztráta 9 centů*
(*=Income per share: loss of 8 cents versus loss of 9 cents*)

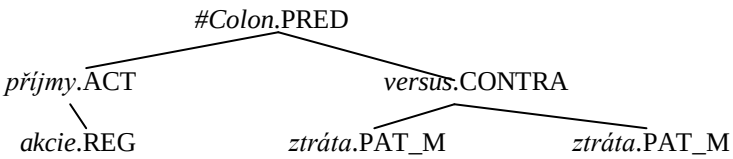
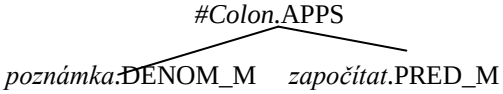


Fig. 1.3.9: *Pozn.: Do všech údajů "za akcii" byly započítány veškeré výsledky konverzních operací.* (*=Note: To all "per share", all results of conversion operations were included.*)



1.4 ***Federal Funds: The upper value of 8 %, the lower value of 7 %, the final offer of 7.5 %, the offer of 7.9 %***

Fig. 1.4.1: *FEDERÁLNÍ FONDY: horní hodnota 8 %, dolní hodnota 7 %, závěrečná nabídka 7,5 %, nabídka 7,9 %.* (*=FEDERAL FUNDS: the upper value of 8 %, the lower value of 7 %, the final offer of 7.5 %, the offer of 7.9 %.*)

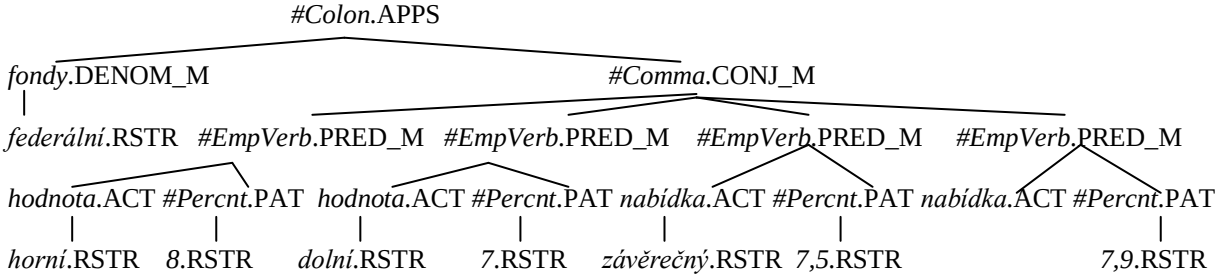


Fig. 1.4.2: *Federální fondy: 8 maximum, 8,4 % nabídka před uzavřením, 6 % nabízeno.* (*=Federal Funds: 8 maximum, 8.4 % offer before closing, 6 % offered.*)

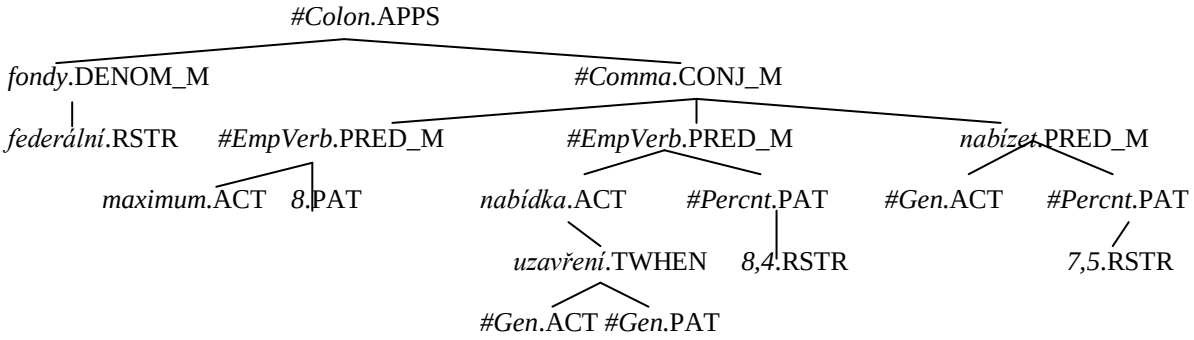
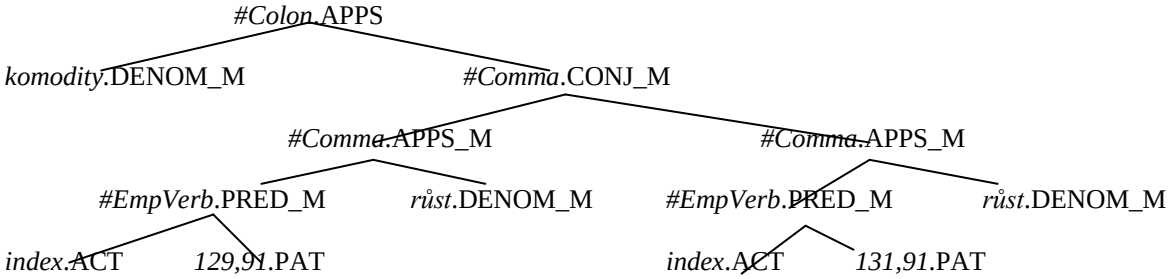


Fig. 1.4.3: *Komodity: Dow-Jonesův index termínových obchodů 129,91, růst o 0,28; index okamžitých cen 131,01, růst o 1,17.* (*=Commodities: Dow Jones futures index 129.91, an increase by 0.28; spot prices index 131.01, an increase by 1.17.*)



1.5 ***Lockheed - 9% bonds in the value of 300 million dollars payable on the 15th October 1999, valued at 99.90 so that the yield was 9.39 %***

Fig. 1.5.1: *Společnost XY, nabídka 108 000 veřejných akcií(,) prostřednictvím firem AB.* (*=Company XY, an offer of 108,000 of shares (,) through companies AB.*)

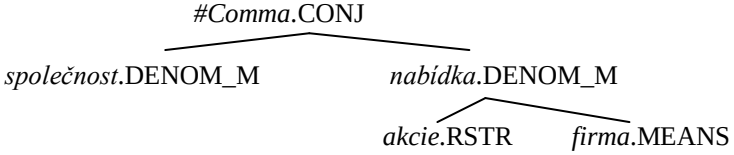


Fig. 1.5.2: *Lockheed - 9% dluhopisy v hodnotě 300 milionů dolarů, splatné 15. října 1999, oceněné na 99,90 tak, aby výnos byl 9,39 %.*
(*=Lockheed - 9% bonds in the value of 300 million dollars payable on the 15th October 1999, valued at 99.90 so that the yield was 9.39 %*)

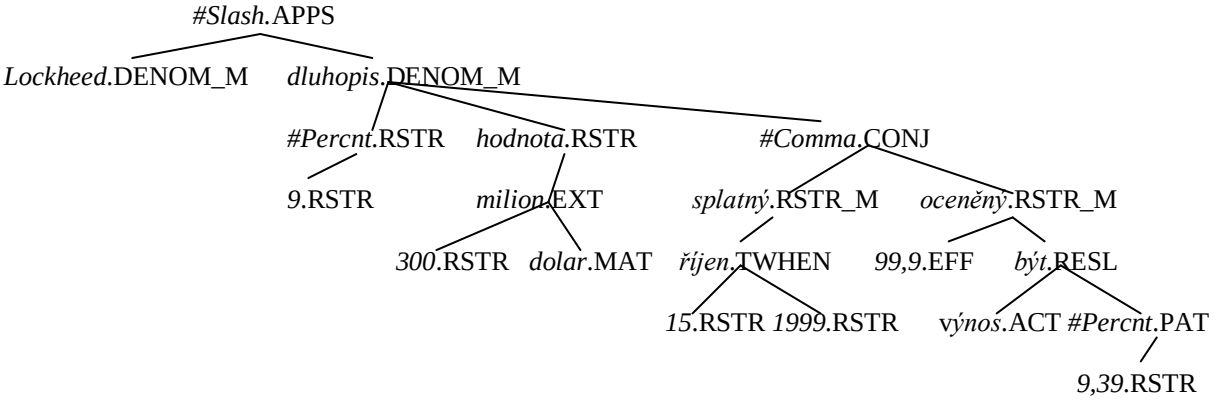


Fig. 1.5.3: *Toyobo Co. (Japan) – 9% dluhopisy splatné 2. listopadu 1993, nabízené prostřednictvím společnosti Nomura Ltd.* (*=Toyobo Co. (Japan) - 9% bonds due on the 2th November 1993, offered through the company Nomura Ltd.*)

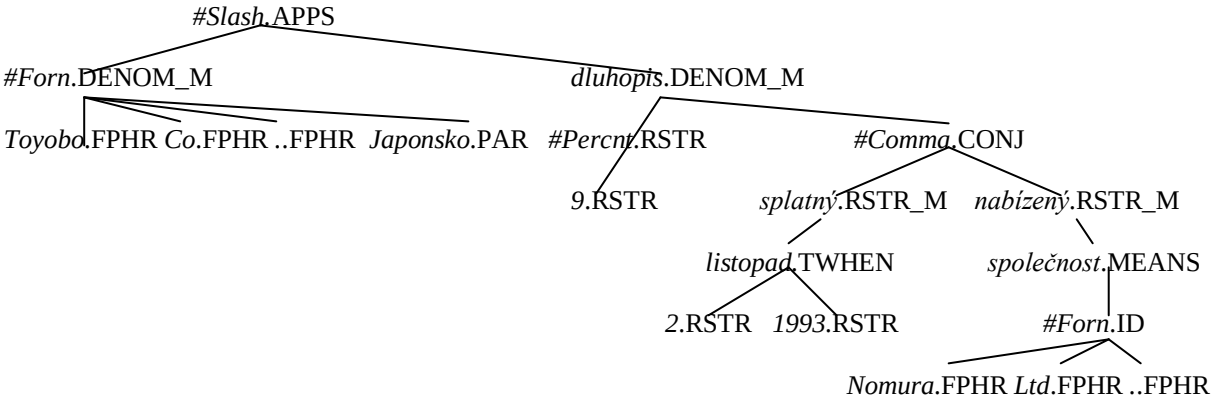


Fig. 1.5.4: *Beverly Hills - 75 milionů dolarů v obecných obligacích z roku 1989, série B (usnesení z 1987), prostřednictvím skupiny PaineWebber.*
(*=Beverly Hills - 75 million dollars in general bonds from 1989, series B (resolution from 1987), via PaineWebber Group.*)

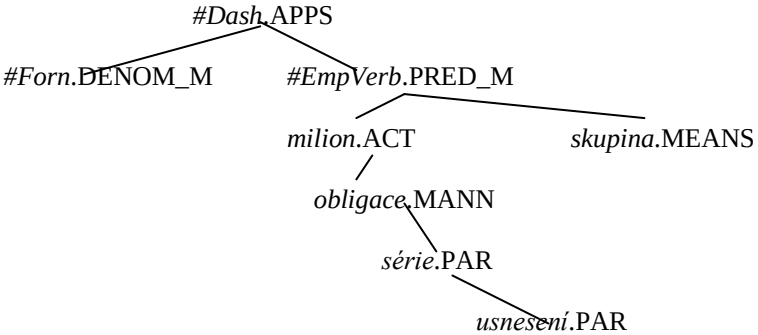
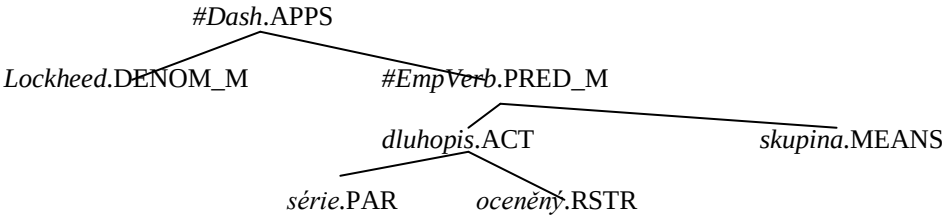


Fig. 1.5.5: *Lockheed - 9% dluhopisy v hodnotě 300 milionů dolarů, série A, oceněné na 99,90, prostřednictvím skupiny PaineWebber.* (*=Lockheed - 9% bonds worth 300 million dollars, Series A, valued at 99.90, via PaineWebber Group*)



1.6 ***Certificates of Deposit: 8.09 % one month***

Fig. 1.6.1: *Vkladové certifikáty: 8,09 % jeden měsíc*
(*=Certificates of Deposit: 8.09 % one month*)

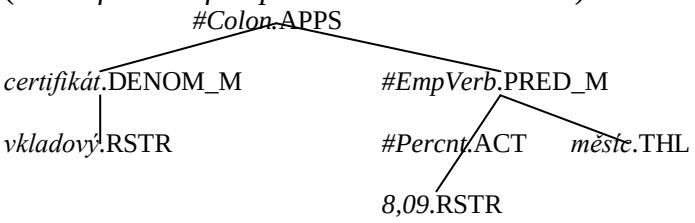
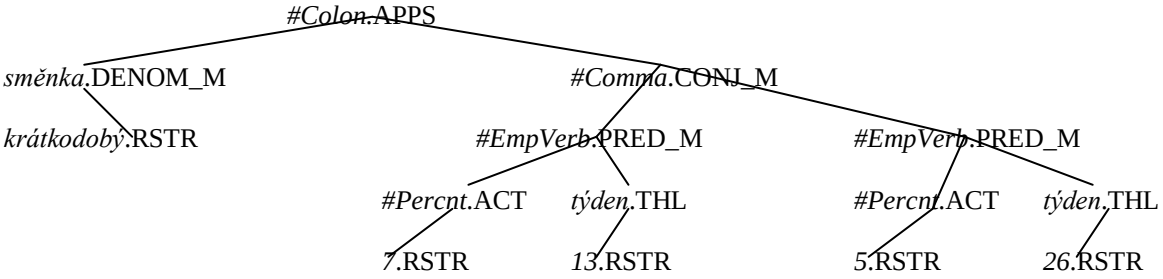


Fig. 1.6.2: *Krátkodobé směnky: 7 % 13 týdnů, 5 % 26 týdnů.*
(*=Short-term bills of exchange: 7 % for 13 weeks, 5 % for 26 weeks.*)



1.7 ***Leon Panetta (Democrat, California)***

Fig. 1.7.1: *s Leonem Panettou (demokrat, Kalifornie) už nespolupracuji*
(*=I cooperate with Leon Panetta (Democrat, California) no longer*)

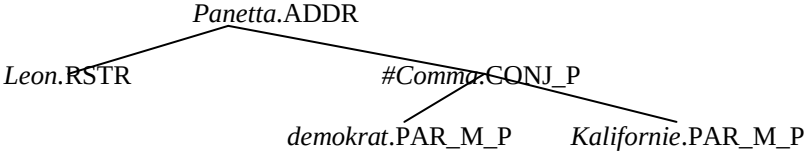


Fig. 1.7.2: *s Leonem Panettou (demokrat za Kalifornii) už nespolupracuji*
(*=I cooperate with Leon Panetta (Democrat from California) no longer*)

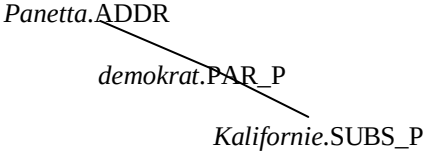


Fig. 1.7.3: *s Leonem Panettou (demokrat, California) už nespolupracuji*
(*=I cooperate with Leon Panetta (Democrat, California) no longer*)

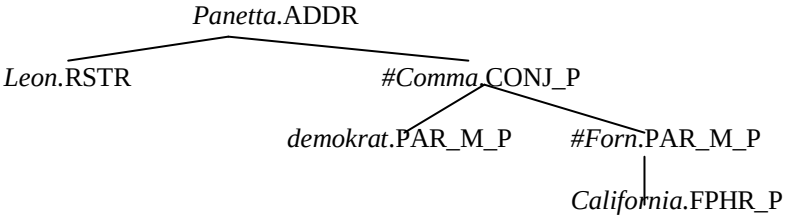
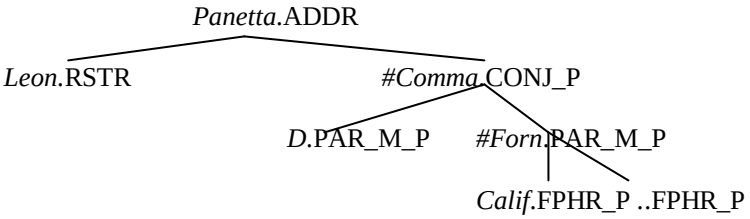
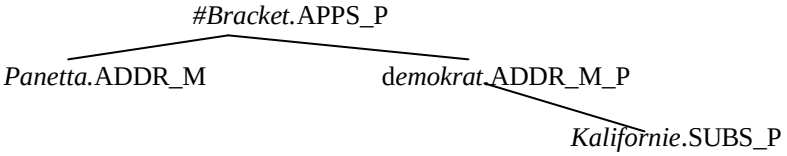


Fig. 1.7.4: *s Leonem Panettou (D., Calif.) už nespolupracuji*
(*= I cooperate with Leon Panetta (D., Calif.) no longer*)



Write to the note *Translation*: *untranslated abbreviations.*

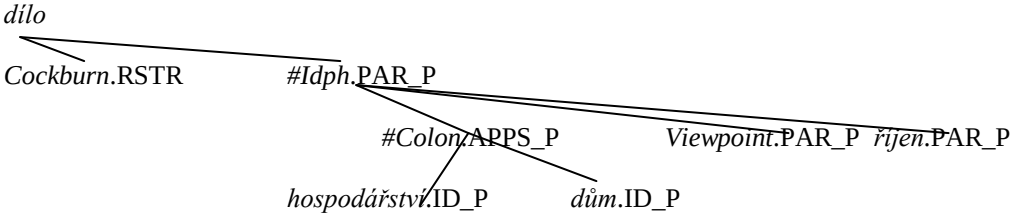
Fig. 1.7.5: *s Leonem Panettou (demokratem za Kalifornii) už nespolupracuji*
(*=I cooperate with Leon Panetta (Democrat from California) no longer*)



1.8 Bibliographic references

Fig. 1.8.1: *Názor Timothyho Gooda, autora "Do světa UFO" (Quill/William Morrow, 592 stran, 12,95 dolaru), je, že svět se směje příliš brzy.*

Fig. 1.8.2: *Děkuji vtipnému dílu Alexandra Cockburna(Hospodářství Spojených států: Dům postavený na písku rizikových obligací, Viewpoint, 19. října).*



2 Templates for the constructions of the Companions recordings

2.1 Opening the conversation

Fig. 2.1.1: *Jmenuji se Petra a budu si tu s vámi teď chvíli povídat.*
(=*My name is Petra and I'll talk here with you for a while.*)

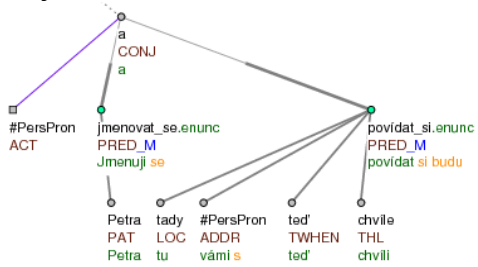


Fig. 2.1.2: *Ještě než začneme, je mě dobře slyšet?* (=*Before beginning, can you hear me well?*)

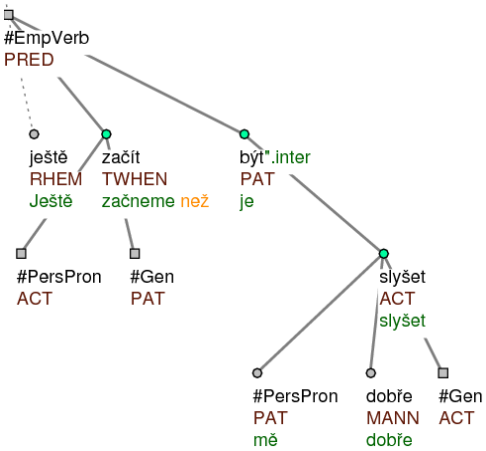


Fig. 2.1.3: *Ráda bych si s vámi popovídala o vašich fotkách.*
(=*I would like to talk with you about your pictures.*)

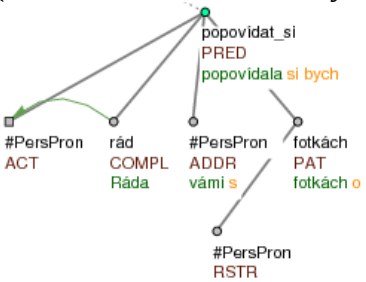


Fig. 2.1.4: *Začněme třeba s touhle. / Začneme třeba s touhle.*
(=*We will begin with this, for example / Begin with this, for example.*)

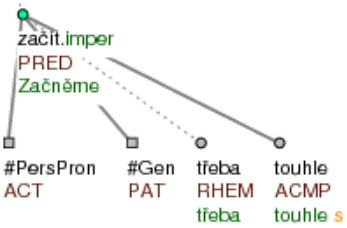
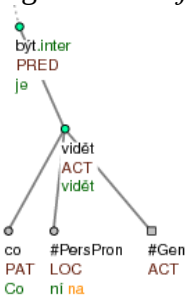


Fig. 2.1.5: *Co je na ní vidět?* (=*What is seen in it?*)



2.2 During the conversation

Fig. 2.2.1: *Povíte mi k tomu ještě něco?* (=*Will you tell me something else about it?*)

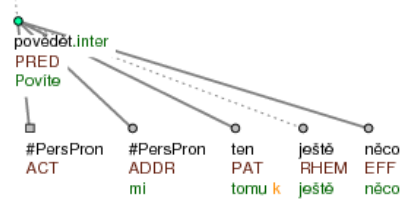


Fig. 2.2.2: *Chtěla byste k tomu ještě něco dodat?* (=*Would you like to add something?*)

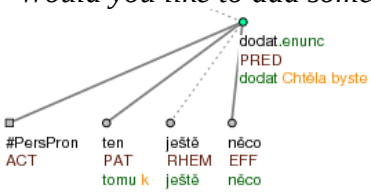


Fig. 2.2.3: *Podíváme se na další fotku.* (=*We will look at another picture.*)

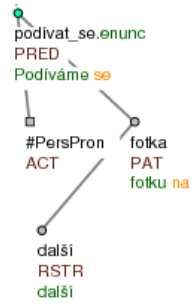


Fig. 2.2.4: *Přejdeme na další fotku.* (=*We go to the next picture.*)

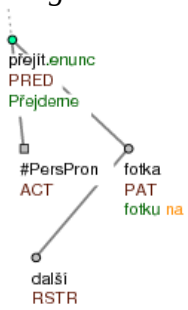


Fig. 2.2.5: *Jdeme na další.* (=*We go to the next one.*)

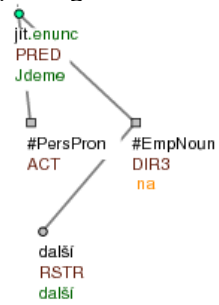


Fig. 2.2.6: *Jdeme dál.* (=*We go further.*)

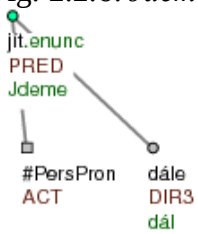


Fig. 2.2.7: *Copak to máme tady?* (=*What do we have here?*)

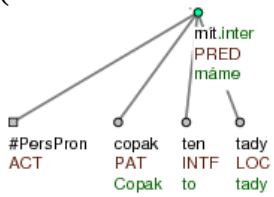


Fig. 2.2.8: *Co je tohle za fotku?* (=*What is this picture?*)

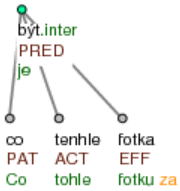


Fig. 2.2.9: *Co tady můžeme vidět?* (=*What can we see here?*)

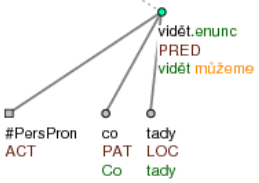
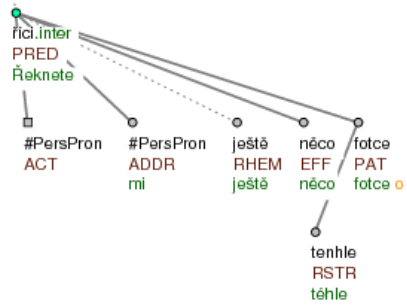


Fig. 2.2.10: *Dobře.* (=*Well.*)

Fig. 2.2.11:
Řeknete mi ještě něco o téhle fotce.PAT? (=*Will you tell me something more to this picture?*)
Chcete mi říct ještě něco k téhle fotce.REG?(=*Would you like to tell me something more to this picture?*)
*Co byste mi ještě řekl k této fotce.*REG?(=*What more would you tell me to this picture?*)



2.3 Closing the conversation

Fig. 2.3.1: *Bohužel nám vypršel čas.* (=*Unfortunately, our time is up.*)

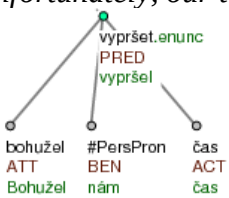


Fig. 2.3.2: *Tohle byla poslední fotka.* (=*This was the last picture.*)

Fig. 2.3.3: *Děkujeme vám za váš čas.* (=*Thank you for your time.*)

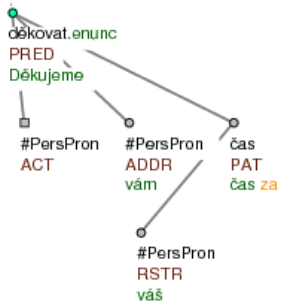


Fig. 2.3.4: *Moc hezky se mi s vámi povídalo.* (=*It was a pleasure for me to talk with you.*)

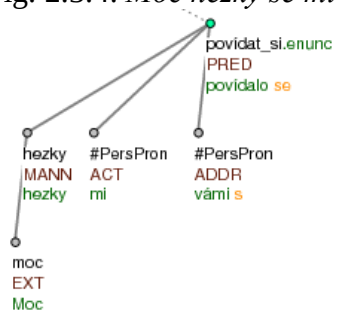


Fig. 2.3.5: *Počkejte chvilku, kluci vás přijdou vysvobodit.* (=*Wait a minute, the boys will come to free you.*)

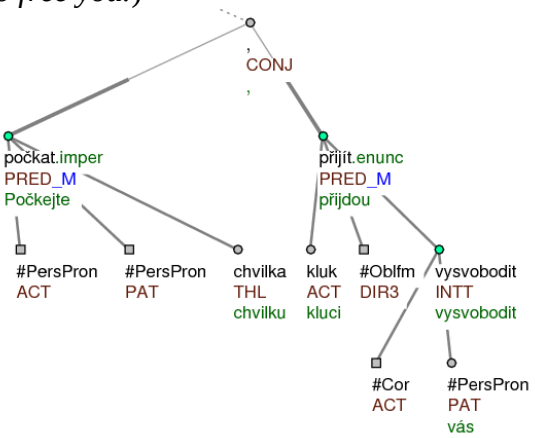


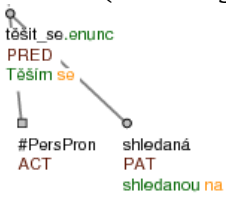
Fig. 2.3.6: *Konec kazety.* (=*The end of the tape.*)

2.4 Greetings

Fig. 2.4.1: *Dobrý den.* (=*Hello.*)

Fig. 2.4.2: *Na shledanou.* (=*Good bye.*)

Fig. 2.4.3: *Těším se na shledanou.* (=*Looking forward to see you again.*)



2.5 Courtesy phrases

Fig. 2.5.1: *Díky za váš čas.*
(=*Thank you for your time.*)

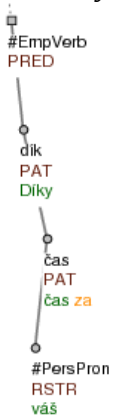


Fig. 2.5.2: *Děkuji.* (=*Thank you.*)

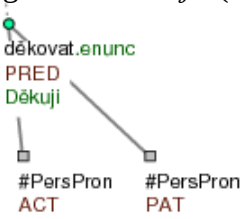


Fig. 2.5.3: *Rádo se stalo.*
(=*You are welcome.*)

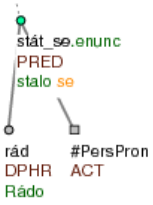


Fig. 2.5.4: *Velice mě těšilo.*
(=*Nice to have met you.*)

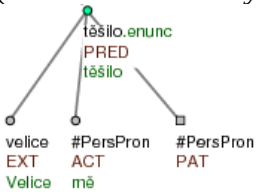
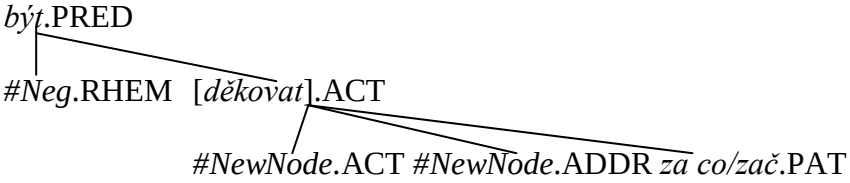


Fig. 2.5.5: *Není zač.* (=*You are welcome.*)



2.6 Please.

2.6.1 *(Děkuji.) Prosím.*
(=*(Thank you.) You are welcome.*)
Zdvořilost.

2.6.3: *(Mohu se tě něco zeptat?) Prosím.*
(=*(Can I ask you something?) Please.*)
Agreement, affirmation (= yes).

Fig. 2.6.5: *A to není prosím všechno.*
(=*And that is not all, please.*)
Emphasis.

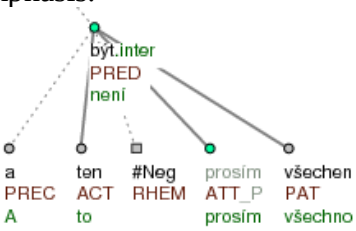
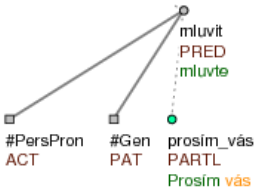
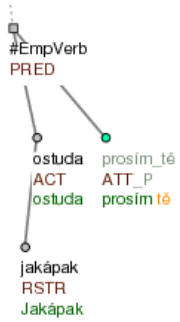


Fig. 2.6.2: *Prosím vás, mluvte.*
(=*Speak, please.*)



2.6.4: *Prosím? (Nerozuměl jsem.)*
(=*Please? (I could not understand.).*)
A question for misunderstanding.

Fig. 2.6.6: *Jakápak ostuda, prosím tě.* –
(=*What a shame, please.*)



References

Hajič et al. (2011):: *Prague Czech-English Dependency Treebank 2.0.* Data/software, Institute of Formal and Applied Linguistics, Prague, Czech republic, 2011.

Mikulová, M., Štěpánek, J. (2009): Annotation Procedure in Building the Prague Czech-English Dependency Treebank. In: *Slovko 2009, NLP, Corpus Linguistics, Corpus Based Grammar Research*. Slovenská akadémia vied, Bratislava, Slovakia, ISBN 978-80-7399-875-2, pp. 241-248.

Hajič et al. (2006):: *Prague Dependency Treebank 2.0.* Software prototype, Linguistic Data Consortium, Philadelphia, PA, USA, ISBN 1-58563-370-4, www.ldc.upenn.edu, 2006.

Mikulová et al. (2006a):: *Annotation on the tectogrammatical level in the Prague Dependency Treebank. Annotation manual.* Technical report no. 2006/30, ÚFAL MFF UK, Prague, Czech Rep., 1287 pp., 2006.

Mikulová et al. (2006b):: *Annotation on the tectogrammatical level in the Prague Dependency Treebank. Reference book.* Technical report no. 2006/32, ÚFAL MFF UK, Prague, Czech Rep., 193 pp., 2006.

Mikulová et al. (2006c): *Anotace na tektogramatické rovině Pražského závislostního korpusu. Referenční příručka.* Technical report no. 2006/31, ÚFAL MFF UK, Prague, Czech Rep., 183 pp., 2006.

Mikulová et al. (2005):: *Anotace na tektogramatické rovině Pražského závislostního korpusu. Anotátorská příručka.* Technical report no. 2005/TR-2005-28, ÚFAL MFF UK, Prague, ISSN 1214-5521, 1185 pp., 2005.

The ÚFAL/CKL Technical Report Series

ÚFAL

ÚFAL (Ústav formální a aplikované lingvistiky; http://ufal.mff.cuni.cz) is the Institute of Formal and Applied linguistics, at the Faculty of Mathematics and Physics of Charles University, Prague, Czech Republic. The Institute was established in 1990 after the political changes as a continuation of the research work and teaching carried out by the former Laboratory of Algebraic Linguistics since the early 60s at the Faculty of Philosophy and later the Faculty of Mathematics and Physics. Together with the "sister" Institute of Theoretical and Computational Linguistics (Faculty of Arts) we aim at the development of teaching programs and research in the domain of theoretical and computational linguistics at the respective Faculties, collaborating closely with other departments such as the Institute of the Czech National Corpus at the Faculty of Philosophy and the Department of Computer Science at the Faculty of Mathematics and Physics.

CKL

As of 1 June 2000 the Center for Computational Linguistics (Centrum komputační lingvistiky; http://ckl.mff.cuni.cz) was established as one of the centers of excellence within the governmental program for support of research in the Czech Republic. The center is attached to the Faculty of Mathematics and Physics of Charles University in Prague.

Technical Reports

The ÚFAL/CKL technical report series has been established with the aim of disseminate topical results of research currently pursued by members, cooperators, or visitors of the Institute. The technical reports published in this Series are results of the research carried out in the research projects supported by the Grant Agency of the Czech Republic, GAČR 405/96/K214 ("Komplexní program"), GAČR 405/96/0198 (Treebank project), grant of the Ministry of Education of the Czech Republic VS 96151, and project of the Ministry of Education of the Czech Republic LN00A063 (Center for Computational Linguistics). Since November 1996, the following reports have been published.

ÚFAL TR-1996-01 Eva Hajičová, *The Past and Present of Computational Linguistics at Charles University*
Jan Hajič and Barbora Hladká, *Probabilistic and Rule-Based Tagging of an Inflective Language – A Comparison*

ÚFAL TR-1997-02 Vladislav Kuboň, Tomáš Holan and Martin Plátek, *A Grammar-Checker for Czech*

ÚFAL TR-1997-03 Alla Bémová at al., Anotace na analytické rovině, *Návod pro anotátory (in Czech)*

ÚFAL TR-1997-04 Jan Hajič and Barbora Hladká, *Tagging Inflective Languages: Prediction of Morphological Categories for a Rich, Structural Tagset*

ÚFAL TR-1998-05 Geert-Jan M. Kruijff, *Basic Dependency-Based Logical Grammar*

ÚFAL TR-1999-06 Vladislav Kuboň, *A Robust Parser for Czech*

ÚFAL TR-1999-07 Eva Hajičová, Jarmila Panevová and Petr Sgall, *Manuál pro tektogramatické značkování (in Czech)*

ÚFAL TR-2000-08 Tomáš Holan, Vladislav Kuboň, Karel Oliva, Martin Plátek, *On Complexity of Word Order*

ÚFAL/CKL TR-2000-09 Eva Hajičová, Jarmila Panevová and Petr Sgall, *A Manual for Tectogrammatical Tagging of the Prague Dependency Treebank*

ÚFAL/CKL TR-2001-10 Zdeněk Žabokrtský, *Automatic Functor Assignment in the Prague Dependency Treebank*

ÚFAL/CKL TR-2001-11 Markéta Straňáková, *Homonymie předložkových skupin v češtině a možnost jejich automatického zpracování*

ÚFAL/CKL TR-2001-12 Eva Hajičová, Jarmila Panevová and Petr Sgall, *Manuál pro tektogramatické značkování (III. verze)*

ÚFAL/CKL TR-2002-13 Pavel Pecina and Martin Holub, *Sémanticky signifikantní kolokace*

ÚFAL/CKL TR-2002-14 Jiří Hana, Hana Hanová, *Manual for Morphological Annotation*

ÚFAL/CKL TR-2002-15 Markéta Lopatková, Zdeněk Žabokrtský, Karolína Skwarská and Vendula Benešová, *Tektogramaticky anotovaný valenční slovník českých sloves*

ÚFAL/CKL TR-2002-16 Radu Gramatovici and Martin Plátek, *D-trivial Dependency Grammars with Global Word-Order Restrictions*

ÚFAL/CKL TR-2003-17 Pavel Květoň, *Language for Grammatical Rules*

ÚFAL/CKL TR-2003-18 Markéta Lopatková, Zdeněk Žabokrtský, Karolina Skwarska, Václava Benešová, *Valency Lexicon of Czech Verbs VALLEX 1.0*

ÚFAL/CKL TR-2003-19 Lucie Kučová, Veronika Kolářová, Zdeněk Žabokrtský, Petr Pajas, Oliver Čulo, *Anotování koreference v Pražském závislostním korpusu*

ÚFAL/CKL TR-2003-20 Kateřina Veselá, Jiří Havelka, *Anotování aktuálního členění věty v Pražském závislostním korpusu*

ÚFAL/CKL TR-2004-21 Silvie Cinková, *Manuál pro tektogramatickou anotaci angličtiny*

ÚFAL/CKL TR-2004-22 Daniel Zeman, *Neprojektivity v Pražském závislostním korpusu (PDT)*

ÚFAL/CKL TR-2004-23 Jan Hajič a kol., *Anotace na analytické rovině, návod pro anotátory*

ÚFAL/CKL TR-2004-24 Jan Hajič, Zdeňka Urešová, Alevtina Bémová, Marie Kaplanová, *Anotace na tektogramatické rovině (úroveň 3)*

ÚFAL/CKL TR-2004-25 Jan Hajič, Zdeňka Urešová, Alevtina Bémová, Marie Kaplanová, *The Prague Dependency Treebank, Annotation on tectogrammatical level*

ÚFAL/CKL TR-2004-26 Martin Holub, Jiří Diviš, Jan Pávek, Pavel Pecina, Jiří Semecký, *Topics of Texts. Annotation, Automatic Searching and Indexing*

ÚFAL/CKL TR-2005-27 Jiří Hana, Daniel Zeman, *Manual for Morphological Annotation (Revision for PDT 2.0)*

ÚFAL/CKL TR-2005-28 Marie Mikulová a kol., *Pražský závislostní korpus (The Prague Dependency Treebank) Anotace na tektogramatické rovině (úroveň 3)*

ÚFAL/CKL TR-2005-29 Petr Pajas, Jan Štěpánek, *A Generic XML-Based Format for Structured Linguistic Annotation and Its application to the Prague Dependency Treebank 2.0*

ÚFAL/CKL TR-2006-30 Marie Mikulová, Alevtina Bémová, Jan Hajič, Eva Hajičová, Jiří Havelka, Veronika Kolařová, Lucie Kučová, Markéta Lopatková, Petr Pajas, Jarmila Panevová, Magda Razímová, Petr Sgall, Jan Štěpánek, Zdeňka Urešová, Kateřina Veselá, Zdeněk Žabokrtský, *Annotation on the tectogrammatical level in the Prague Dependency Treebank (Annotation manual)*

ÚFAL/CKL TR-2006-31 Marie Mikulová, Alevtina Bémová, Jan Hajič, Eva Hajičová, Jiří Havelka, Veronika Kolařová, Lucie Kučová, Markéta Lopatková, Petr Pajas, Jarmila Panevová, Petr Sgall, Magda Ševčíková, Jan Štěpánek, Zdeňka Urešová, Kateřina Veselá, Zdeněk Žabokrtský, *Anotace na tektogramatické rovině Pražského závislostního korpusu (Referenční příručka)*

ÚFAL/CKL TR-2006-32 Marie Mikulová, Alevtina Bémová, Jan Hajič, Eva Hajičová, Jiří Havelka, Veronika Kolařová, Lucie Kučová, Markéta Lopatková, Petr Pajas, Jarmila Panevová, Petr Sgall,Magda Ševčíková, Jan Štěpánek, Zdeňka Urešová, Kateřina Veselá, Zdeněk Žabokrtský, *Annotation on the tectogrammatical level in the Prague Dependency Treebank (Reference book)*

ÚFAL/CKL TR-2006-33 Jan Hajič, Marie Mikulová, Martina Otradovcová, Petr Pajas, Petr Podveský, Zdeňka Urešová, *Pražský závislostní korpus mluvené češtiny. Rekonstrukce standardizovaného textu z mluvené řeči*

ÚFAL/CKL TR-2006-34 Markéta Lopatková, Zdeněk Žabokrtský, Václava Benešová (in cooperation with Karolína Skwarska, Klára Hrstková, Michaela Nová, Eduard Bejček, Miroslav Tichý) *Valency Lexicon of Czech Verbs. VALLEX 2.0*

ÚFAL/CKL TR-2006-35 Silvie Cinková, Jan Hajič, Marie Mikulová, Lucie Mladová, Anja Nedolužko, Petr Pajas, Jarmila Panevová, Jiří Semecký, Jana Šindlerová, Josef Toman, Zdeňka Urešová, Zdeněk Žabokrtský, *Annotation of English on the tectogrammatical level*

ÚFAL/CKL TR-2007-36 Magda Ševčíková, Zdeněk Žabokrtský, Oldřich Krůza, *Zpracování pojmenovaných entit v českých textech*

ÚFAL/CKL TR-2008-37 Silvie Cinková, Marie Mikulová, *Spontaneous speech reconstruction for the syntactic and semantic analysis of the NAP corpus*

ÚFAL/CKL TR-2008-38 Marie Mikulová, *Rekonstrukce standardizovaného textu z mluvené řeči v Pražském závislostním korpusu mluvené češtiny. Manuál pro anotátory*

ÚFAL/CKL TR-2008-39 Zdeněk Žabokrtský, Ondřej Bojar, *TectoMT, Developer's Guide*

ÚFAL/CKL TR-2008-40 Lucie Mladová, *Diskurzní vztahy v češtině a jejich zachycení v Pražském závislostním korpusu 2.0*

ÚFAL/CKL TR-2009-41 Marie Mikulová, *Pokyny k překladu určené překladatelům, revizorům a korektorům textů z Wall Street Journal pro projekt PCEDT*

ÚFAL/CKL TR-2011-42 Loganathan Ramasamy, Zdeněk Žabokrtský, *Tamil Dependency Treebank (TamilTB) – 0.1 Annotation Manual*

ÚFAL/CKL TR-2011-43 Nguy Giang Linh, Michal Novák, Anna Nedoluzhko, *Coreference Resolution in the Prague Dependency Treebank*

ÚFAL/CKL TR-2011-44 Anna Nedoluzhko, Jiří Mírovský, *Annotating Extended Textual Coreference and Bridging Relations in the Prague Dependency Treebank*

ÚFAL/CKL TR-2011-45 David Mareček, Zdeněk Žabokrtský, *Unsupervised Dependency Parsing*

ÚFAL/CKL TR-2011-46 Martin Majliš, Zdeněk Žabokrtský, *W2C – Large Multilingual Corpus*

ÚFAL TR-2012-47 Lucie Poláková, Pavlína Jínová, Šárka Zikánová, Zuzanna Bedřichová, Jiří Mírovský, Magdaléna Rysová, Jana Zdeňková, Veronika Pavlíková, Eva Hajičová, *Manual for annotation of discourse relations in the Prague Dependency Treebank*

ÚFAL TR-2012-48 Nathan Green, Zdeněk Žabokrtský, *Ensemble Parsing and its Effect on Machine Translation*

ÚFAL TR-2013-49 David Mareček, Martin Popel, Loganathan Ramasamy, Jan Štěpánek, Daniel Zemana, Zdeněk Žabokrtský, Jan Hajič *Cross-language Study on Influence of Coordination Style on Dependency Parsing Performance*

ÚFAL TR-2013-50 Jan Berka, Ondřej Bojar, Mark Fishel, Maja Popović, Daniel Zeman, *Tools for Machine Translation Quality Inspection*

ÚFAL TR-2013-51 Marie Mikulová, *Anotace na tektogramatické rovině. Dodatky k anotátorské příručce (s ohledem na anotování PDTSC a PCEDT)*

ÚFAL TR-2013-52 Marie Mikulová, *Annotation on the tectogrammatical level. Additions to annotation manual (with respect to PDTSC and PCEDT)*