A MANUAL FOR TECTOGRAMMATICAL TAGGING OF THE PRAGUE DEPENDENCY TREEBANK

Eva Hajičová, Jarmila Panevová and Petr Sgall

In cooperation with A. Böhmová, M. Ceplová and V. Řezníčková

Translated by Z. Kirschner, E. Hajičová and P. Sgall

ÚFAL/CKL Technical Report TR-2000-09

December 2000

Introduction: Three layers of tagging of the Prague Dependency Treebank

This manual is supposed to introduce into the practice of syntactic tagging in the framework of the *Prague Dependency Treebank* (henceforth PDT). After a brief Introduction, a list of used symbols is given (Sect. 1) followed by a description of the automatic procedure dealing with grammatemes (Sect. 2.1), and by instructions covering further transducing (non-automatic, for the time being) of morphemic and analytic data to the tectogrammatical level. Section 2.2.1 concerns morphological grammatemes, and the subsequent sections (2.2.2 - 2.6) represent what is supposed to be of maximal importance for the majority of annotators: the parts dealing with functors and syntactic grammatemes. In the concluding Section 3 the topic-focus articulation is treated.

The tagging of PDT, a corpus compiled on the basis of the Czech National Corpus (in preparation at the Institute of Czech National Corpus at the Faculty of Philosophy, Charles University, under the guidance of F. Čermák and in cooperation with other research institutions) is conceived as a three-layer system of tags (Hajič, Hajičová, Rosen 1996): the individual layers can be characterized as follows:

(i) morphemic tagging capturing relatively disambiguated values of morphemic categories; let us note that also a result of a full morphemic analysis is available, i.e., complete sets of values of individual forms without disambiguation: e.g., the form *dobrým* gets "I.SG or D.PL", yet for the tag just one of the two possibilities is chosen according to the given context;

(ii) syntactic tags at the so-called analytic level, capturing the functions of individual word forms as they are expressed in the surface shape of the sentence; in the analytic tree structures (ATSs), every word token and punctuation mark has a corresponding node and is analyzed as for its POS and morphemic value, as well as for the main syntactic functions ('analytic functors', 'Afuns'); among the values of Afun, Subj, Obj, Adv are not classified in a more subtle way;

(iii) syntactic tags at the tectogrammatical level (TGTSs) rendering the deep (underlying, tectogrammatical) structure of the sentence, i.e., its syntactic structure proper (with a detailed classification of functors, see below).

The level (i) is described in detail particularly in the writings by Hajič and Hladká (1997, 1998). The analytic syntactic level has been dealt with, e.g., in the writings by Hajič (1998), Hajič, Hajičová, Panevová and Sgall (1998), Hajič and Hajičová (1997). In addition, a manual has been prepared for manual analytic annotation (Bémová et al. 1997).

Contents

1. SPECIFICATION OF TECTOGRAMMATICAL TAGS 1

1.1 THE CHARACTERISTICS OF TECTOGRAMMATICAL TREE STRUCTURES 1

1.2 LIST OF TECTOGRAMMATICAL TAGS 1

(a) Morphological grammatemes 2

Specific grammatemes and further symbols 3

(b) Functors 4

(i) With the uppermost node of the tree structure 4

(ii) With the main verb of the sentence 4

(iii) With dependents of verbs (sometimes also of nouns) 4

Participants (arguments, inner modifications) 4

Adjuncts (free modifications) 4

(iv) With nouns only 6

(c) Functors for coordination, apposition and parenthesis 6

(d) Syntactic grammatemes 7

(i) With participants (arguments) 7

(ii) With free modifications (adjuncts) 7

With locative and directional 7

With temporal adjuncts 8

With other adjuncts 9

(iii) With coordination and parenthesis (as auxiliary markers/tags) 9

(e) Lexical parts of the tags 9

2. CONVERSION OF ANALYTIC TREE STRUCTURES TO TECTOGRAMMATICAL STRUCTURES 11

2.1 AUTOMATIC PROCEDURE OF ADJUSTING THE TREES 11

2.1.1 The first phase of the automatic procedure 11

(i) Main verb 12

(ii) Modality 12

(iii) Aspect 13

(iv) Gender, number 13

(v)–(xxi) Other issues 14

2.1.2 The second phase of the automatic procedure 16

2.2 MANUAL CONVERSION OF ATSS TO TECTOGRAMMATICAL SYNTACTIC STRUCTURES (TGTSS) 17

2.2.1 How to convert morphological grammatemes 18

2.2.2 How to assign functors and syntactic grammatemes 20

Introductory remarks 20

Further remarks 20

The governing verb 20

Sentences without governing verbs 21

Apposition 21

Direct speech 21

Numerals 21

Reference words 22

Relative clauses 22

Temporal functors 22

Functors with nouns 22

Syntactic grammatemes 23

Phrasemes 23

Coordination 24

2.2.2.1 Endings of cases, infinitive, adverbials 24

2.2.2.2 Prepositions (primary and secondary) ... 27
2.2.2.3 Subordinating conjunctions and connecting expressions ... 36
2.2.2.4 Agreement with the noun ... 39
2.3 LEXICAL PARTS OF THE TAGS ... 40
2.4 COORDINATION, APPOSITION AND PARENTHESIS ... 42
2.5 DELETIONS ... 43
A. When and whereto restore elements ... 44
A.1 When to restore a node ... 44
A.2 Where to place a restored node ... 46
B. What lemma is assigned to the restored node ... 46
B.1 Filling in lexical units ... 46
B.2 The general participant Gen expressed by zero ... 47
B.3 Verbs of 'control' ... 49
B.4 Restoration of a pronominal, anaphoric element ... 50
B.4.1 Zero subject (pro-drop) ... 50
B.4.2 Deleted omissible obligatory participant ... 51
B.4.3 Inomissible obligatory participant ... 51
B.4.4 Restoration of a pronominal node ... 52
B.4.5 Deleted pronouns of laziness ... 52
C. Notes concerning more complex examples ... 52
2.6 COREFERENCE ... 53
A. Textual coreference ... 53
B. Grammatical coreference ... 54
3. TOPIC-FOCUS ARTICULATION ... 57
Concluding remark ... 63
Appendix: A list of adverbs (incomplete) ... 65
References ... 71

1. SPECIFICATION OF TECTOGRAMMATICAL TAGS

1.1 The characteristics of tectogrammatical tree structures

Tectogrammatical tree structures (TGTSs) are based on dependency syntax; the tagging at this level is guided by the following principles:

(a) a node of a tectogrammatical tree represents an autosemantic (lexical) word; the correlates of synsemantic (functional, auxilliary) words are attached to the autosemantic words to which they belong (that is to say, auxiliary verbs and subordinating conjunctions to the verbs, prepositions to nouns, etc.); an exception concerns by coordinating conjunctions, which, in TGTSs, are treated in the same way as in the analytic trees; therefore, no further dimensions for coordination and apposition are considered, a two-dimensional tree-structure is adhered to;

(b) in the cases of deletion in the surface shape of the sentence, nodes are introduced into the tectogrammatical tree to 'recover' a deleted word;

(c) no non-projective structures are admitted at the tectogrammatical level (non-projectivity is supposed to be solved by movement rules between the tectogrammatical tree and the morphemic string);

(d) not only the direction of the dependence on the governing node (dependence to the left, dependence to the right) is taken into account, but also sister nodes are ordered (from left to right).

Thus, the tagging results in a dependency tree. This tree differs from a theoretically pure tectogrammatical representation in the way the coordination is treated (see the paragraph (a) above), and in some other points, too, see below.

We expect that a few dozens of model TGTSs with complete tagging will be prepared (model collection, MC). Another, large set of TGTS will be represented by the so-called basic or large collection (LC), based on an automatic procedure and 'tuned' manually. As regards the latter collection, it has been provisionally refrained from handling some not yet fully theoretically and/or empirically mastered phenomena, as will be seen from the following explanations (e.g., the coverage of topic-focus articulation will not be complete, in some cases the possibility of more than one analysis of functors will be preserved, etc.).

In the LC, morphemic grammatemes will be dealt with in a rough manner only, within the limits of the first version of automatic transduction from the analytic level; they are being derived straightforwardly from the morphemic values.

The transposed use of forms (historical present tense and the present *pro futuro*, epistemic validity of Deontmod, singular validity of *pluralia tantum*, etc.) will not be captured in the LC, while in the MC all this is supposed to get treated.

The automatic procedure is expected to prepare syntactic grammatemes for the subsequent manual treatment by storing synsemantic words and the case value of the noun within a special attribute; this is to remain so in the LC for the time being, while in the model collection a more profound treatment is anticipated.

As far as the DICTIONARY is concerned, we expect it to be compiled step by step by supplying the data obtained in the course of tagging the corpus. We assume that the lexical entries will contain, in addition to lemmas and morphemic data, also the valency frames of words in which, among others, at least elementary information on subcategorization will be present (on whether a participant with a given functor can be a N, V, A, etc.).

1.2 List of tectogrammatical tags

We present here an outline of the list of tectogrammatical tags with brief notes on important details of the conversion from ATSs to TGTSs. A tectogrammatical tag consists of lemma, i.e. a symbol referring to the lexical value of the word proper (just its orthographic form, for the time being), and of indices, i.e., the values of attributes falling into two groups - grammatemes and functors. The grammatemes correspond, above all, to morphological categories, whereas the functors represent syntactic functions (as regards this difference cf. the

writings on Functional Generative Description, e.g., Sgall 1967, Panevová 1980, Sgall et al. 1986). For the labels of individual categories, English terms and abbreviations are used to make them compatible (at this phase of research) with the English terms used in our writings on Functional Generative Description as well as in those on tectogrammatical tagging.

The values of the grammatemes and functors are written in capitals here (e.g., DEB), while with the lexical values of specific symbols only the first letters come from the upper case set (e.g., Neg for negation).

Formal (technical, empty) symbols supplied automatically:

??? - the value has not been treated yet

NA - non-applicable, the attribute cannot be applied in the given context, it is not to be filled in (e.g., tense with nouns)

NIL - primary value of a given attribute (e.g.: the case in question is not direct speech, quoted word, negative syntactic grammateme ACMP, REG).

(a) MORPHOLOGICAL GRAMMATEMES

and further symbols in a similar position - in LC only those treated in the automatic procedure will appear, while in MC all will be contained.

category (attribute)	values	explanation
Sentmod	ENUNC	enunciation
	EXCL	exclamatory (*Tam jich bylo!* 'What numbers of them were there!')
	DESID	optative (*kéž* 'if only', *ať* 'let him', *nechť* etc.)
	IMPER	imperative
	INTER	interrogative
Verbmod	IND	indicative
	IMP	imperative
	CDN	conditional
Deontmod	DEB	debitive (*muset* 'must')
	HRT	hortative (*mít (povinnost)* 'be obliged')
	VOL	volitive (*chtít* 'want')
	POSS	possibilitive (*moct* 'can, be able')
	PERM	permissive (*smět* 'may, be allowed')
	FAC	facultative, ability (*dovést* 'can', *umět* 'know')
	DECL	declarative, without modal verb
Tense	SIM	simultaneous (present)
	ANT	anterior (past)
	POST	posterior (future)
Aspect	PROC	processual (progressive, imperfective)
	CPL	complex (complete, perfective)
	RES	resultative (perfect: *mám/je uklizeno* 'I have done with cleaning/it (the place) has been left tidy')
Iterativeness - for MC only:		
	IT1	iterative
	IT0	non-iterative, single act
Number	SG	singular
	PL	plural (also with *rukama* 'with hands', etc., residues of dual in Czech)

Gender (only with nouns and substantival or substantively used pronouns)

	ANIM	masculine animate
	INAN	masculine inanimate
	FEM	feminine
	NEUT	neuter
Degrees of comparison	POS	positive
	COMP	comparative
	SUP	superlative

Specific grammatemes and further symbols

attribute	values	explanation
tfa	T	contextually bound node (prototypically less dynamic than its governing node; i.e., "given", non-contrastive)
	F	contextually non-bound ("new") node
	C	contrastive T; applied regardless of projectivity
fw	PREP	with the verb of a subordinate clause
	CNJ	the conjunction will automatically be stored; with a noun the same holds for prepositions
phraseme	PHRi	i = surface serial number of its first part
	NIL	
quoted	QUOT	quoted word
	NIL	
ord	ordinal "sequential" number using decimal point: if a deleted node is inserted between 1. and 2., it obtains ORD 1.1; between 1. and 1.1 it is assigned 1.01, etc. this procedure takes place automatically in the course of building up the tree;	
del deletion:	ELID	elided, deleted: it is deleted in the outer shape of the sentence, unmodified;
	ELEX	expounded deletion: it indicates that the antecedent is modified, that some of the members that depend on it can be added to the deleted element in the full interpretation;
	EXPN	the node has not been deleted, yet something deleted depends on it that needn´t be reconstructed (especially if coordination is the case, see Sect. 2.4);
	NIL	the node has not been deleted
antec antecedent:	functor of antecedent with grammatical coreference	
	NIL	
coref coreference:	lemma of the antecedent of coreference	
	NIL	
cornum	number of antecedent (see ORD above)	
corsnt coreference in sentence		
	PREVi	antecedent in a preceding sentence; three values are distinguished, viz. PREV1, PREV2 and PREV3, if the antecedent is in the immediately preceding sentence, in the last but one sentence, or in the third sentence before the given sentence, respectively
	NIL	antecedent in the sentence just being analyzed
dsp direct speech	DSP	top node of a direct speech clause with inverted commas from both

		sides
	DSPP	direct speech, partial: the top node of the first or last sentence in a longer direct speech (with a left or right quotation mark only)
	DSPI	interrupted direct speech
	NIL	direct speech is not the case

(b) FUNCTORS

Participants ('arguments') are listed here first, then adjuncts (free modifications); the abbreviations of the latter are ordered alphabetically.

Note: There is a blank space in the attribute functor to place a second possible value there: if the choice of the first functor is uncertain, a question mark can be placed here, e.g.

PAT and DIR1 - uncertainty: PAT or DIR1? (from where)

PAT and ? - probably PAT, the annotator cannot know for sure;

as a matter of fact, using the second place should be avoided if possible;

(i) With the uppermost node of the tree structure

root	SENT	the uppermost node of the tree standing above the governing verb of the whole sentence; its lemma contains the identification number of the sentence

(ii) With the main verb of the sentence

predicate	PRED	main verb
denomination	DENOM	title (a noun in Nominative case) as the governing node of a verbless sentence
sentence particle	PARTL	*Ano* 'Yes', *N*e 'No', adverbs and interjections
vocative sentence	VOC	*Jirko!* 'George!'
vocative in apposition	VOCAT	*Pojď*.PRED *sem, Jirko*.VOCAT*!* 'Come here, George!' (VOCAT and PRED both depend on APPS)
empty verb	EV	the governing word of a verbless sentence in the remaining cases

Note: Every main clause in a compound (coordinated) sentence is handled as including a main verb etc.; thus nodes with PRED can be coordinated, as well as nodes with DENOM or those with another of the above mentioned functors.

(iii) With dependents of verbs (sometimes also of nouns)

Participants (arguments, inner modifications)

actor/bearer	ACT	agentive, deep subject
patient	PAT	patient, deep object - *prošli celý les* 'they traversed the whole wood', but *prošli lesem* 'they passed across/through the wood' --> DIR2
addressee	ADDR	*komu* 'to whom'
effect	EFF	result (*zvolí ho předsedou, za předsedu* '(they) elect him as, for chairman')
origin	ORIG	origin *z čeho* 'of, from s.t.' (not *odkud* 'from where')

Adjuncts (free modifications)

accompaniment	ACMP	*s, bez* 'with, without'
aim	AIM	purpose (*aby, pro něco* 'so as to, in order to, with the aim of')
attitude	ATT	*s radostí,* 'with pleasure', *vhodně* 'aptly', *právem* 'rightly')
benefactive, -tory	BEN	*pro koho, proti komu* 'for, against s.o.'

cause	CAUS	
comparison	CPR	*než* 'than', *jako* 'as'
complement	COMPL	depends on the verb; see Sect. 2.2.2.
concession	CNCS	*ačkoli* 'although'
condition	COND	real condition: *jestli, -li, jestliže*, *když* 'if'
confrontation	CONFR	*kdežto* 'whereas', *zatímco* 'while', or, as the case may be *jestliže* 'if'
counterfactual	CTERF	counterfactual condition: *kdyby* 'if+Past'
criterion	CRIT	standard: *podle něj* in the sense of ´according to what he said´
difference	DIFF	difference: *oč* 'in, by'
dir(ectional)		
1-from	DIR1	from where (but not *udělat co z čeho* 'make st. from st.': this is ORIG)
2-which way	DIR2	*prošli lesem* 'they walked through the wood'; but see PAT
3-where to	DIR3	*do* 'into', *k* 'to', etc.; but not: *změnit nač* 'change into st.' (EFF)
part of phraseme	DPHR	dependent part of phraseme without a clear syntactic function
ethical dative	ETHD	free dative, subjectivizing: *děti nám nechodí včas* ´we don´t have the children coming in time´, *Já ti mám knih!* 'I do have lots of books, I tell you'
extent	EXT	degree: *hodně* 'very', *velmi mnoho* 'very much', *trochu* 'a bit'
heritage	HER	inheritance: *po otci* 'after father'
intensification	INTF	a 'connecting' element, 'false subject': *To Karel ještě nepřišel?* 'Is it so that Charles hasn´t arrived yet?' *To prší.* 'What a rain!' *Ono táhne.* 'It is draughty here'
intent	INTT	intention: *šel se koupat* 'He went for a bath'; *poslali ho nakoupit* 'they sent him out shopping'
locative	LOC	place where: *jednání uvnitř koalice* 'negotiations within the coalition'
manner	MANN	way, mood, manner: *ústně* 'orally', *psát česky* 'to write in Czech'
means	MEANS	instrument, tool: *psát rukou* 'to write by hand', *na počítači* 'to type on computer', *tužkou* 'to write with a pencil', *pohnout rukou* 'to move the hand-Instr.'
adverbial of modality	MOD	*asi* 'perhaps', *možná* 'maybe', also *To je myslím zlé* (without commas) 'which I deem bad' (lit.: 'that is I-think bad')
norm	NORM	*ve shodě s* 'in agreement with', *podle* 'according to'
reference to preceding text	PREC	*tedy, tudíž* 'thus', *protože* 'since', *naopak* 'on the contrary', *také* 'as well as', similarly: *když, jenže, taky, neboť*, *vždyť* (typically at the beginning of a sentence, if they do not join clauses into a complex sentence)
regard	REG	*se zřetelem* 'with respect to', *bez ohledu na* 'irrespective of'
rhematizer	RHEM	focalizer: *i* 'even', *také* 'also', *jenom* 'only', *nejen* 'not only', *vůbec* 'altogether'
restriction	RESTR	*kromě, mimo* 'but for, except'; mind the difference from RSTR that concerns restrictive adjuncts only
result	RESL	outcome: *opálen do hněda* 'tanned brown', *prsty ztuhlé, že je nenarovná* 'fingers stiff never to get straight'
substitution	SUBS	*místo koho/čeho* 'instead (in place) of'
temporal:		
when	TWHEN	*loni* 'last year', *napřesrok* 'next year', *vstupuje v platnost dnem podpisu* 'it comes into effect on the day of signature'

since when	TSIN	*od té doby, co* 'since the time that', *platí ode dne podpisu* 'becomes effective since the day of signature'
till when	TTILL	*až_do* 'till', *dokud_ne, než* 'until'
how long	THL	*četl půl hodiny* ´he was reading for half an hour', *celou zimu* 'the whole winter', *po_tu*(_celou)*_dobu*/*čas* 'for the (whole) time', *dokud/pokud* 'as long as', *za_dobu, kdy* 'for the time when'
for how long	TFHL	*na dva dny* 'for two days', *na dobu/čas_kdy* 'for the time when', *na věky* 'for ages'
how often	THO	*často* 'often', *mnohokrát* 'many times'
parallel, contemporary	TPAR	*během* 'during', *zatímco*/*mezitím co* 'while', *za celý večer (zápas)* 'during the whole evening (match)'
from when	TFRWH	*Zbylo od Vánoc cukroví* 'There are some sweets left from X-mas', *Z dětství si nepamatuji nic* 'From my childhood I do not remember anything', *Vstupenka z pátku* 'A ticket from Friday'
to when	TOWH	*Odlož výuku na pátek* 'Put off the classes till Friday'; *Demonstrace je svolána na šestou hodinu* 'The demonstration has been called up for six o'clock'

(iv) With nouns only

appurtenance	APP	whose, of whom/what: *Jirkova sestra* 'George´s sister', *dům mých rodičů* 'the house of my parents'
descriptive	DES	a non-restrictive adjunct: *zlatá Praha* 'Golden Prague', *kočky, patřící k savcům* 'cats, belonging to mammals'
identity	ID	*pojem čas(u)* 'concept (of) time', *parník Hradčany* 'the steamboat Hradčany'; it may be a whole sentence or an infinitive (as titles)
material	MAT	partitive: *hrnek čaje* 'a cup of tea'
restrictive	RSTR	restrictive adjunct: *včerejší noviny* 'yesterday newspapers'
vocative sentence	VOC	*Jirko!* 'George!'
vocative in apposition	VOCAT	*Pojď sem, Jirko!* 'Come here, George!'

(c) FUNCTORS FOR COORDINATION, APPOSITION AND PARENTHESIS

conjunction	CONJ	*a* 'and', Comma, *přičemž* 'while', *jak - tak*, *jednak - jednak* 'both - and'
disjunction	DISJ	*nebo, anebo* 'or', *ani* 'neither, nor', specific use of *od - přes - (až) do/k/po* 'from - through - to', *ani X - ani Y* (with a negative verb) 'either - or'
gradation	GRAD	*i* 'even', *a také,* 'and also', *ani* 'even'
adversative	ADVS	*ale* 'but', *však* 'however', *sice - ale* 'it is true - though'
consequence	CSQ	*a proto* 'and therefore', *a tak, a tedy* 'and so', *takže* 'so that', *pročež* 'which is why…'
reason	REAS	*neboť*, *totiž*, *vždyť* 'since'
apposition	APPS	*Jirka, můj přítel* 'George, a friend of mine'; with AuxY in the ATS: *tj.*'ie.', *totiž* 'thus', *a to* 'namely', *jako* 'as', Comma
parenthesis	PAR	an inserted segment without a syntactic relation to other elements of the sentence (but enclosed in commas, thus differing from MOD, see Sect. 4.1.2 above): *myslím* 'I think', *věřím* 'I believe'

(d) SYNTACTIC GRAMMATEMES

(i) With participants (arguments)

functor	grammateme	commentary
ACT	NIL	unmarked actor
	GNEG	*Není peněz* '(There) is no money'
	DISTR	*Na každé větvi viselo po jablíčku* 'Apples were hanging one by one on each branch' (lit.: On each branch hang by an-apple)
	APPX	*Na sta mušek rozžehlo si světla v trávě* 'Fireflies in the hundreds turned on their lights in the grass' (lit.: About hundreds of-flies turned-on their lights in grass), *Přišlo tam na desítky odpůrců zákona* 'Opponents of law turned up in the tens' (lit.: Came there about tens of-opponents of-law.)
	GPART	*Vody ubývá* 'Water (Genitive) is running low'
	GMULT	*Tam bylo lidí!* 'What numbers of people were there!' (lit.: There were people-Genitive)
	VCT	*"Vlasto," ozývalo se ze všech stran.* '"Vlasta!" could be heard from all sides'
PAT	NIL	unmarked Patient
	GNEG	Genitive of negation: *Neřekl mu ani slova.* (*ani* has the functor value RHEM) 'He didn´t tell him one word-Genitive', *Ta vesnice nemá vody* 'That village doesn´t have water-Genitive'
	DISTR	*Dal každému dítěti po jablíčku* 'He gave each child (lit.: by) one apple'
	APPX	approximative: *Roznesl na sto letáků* 'He delivered as many as about one hundred leaflets'
	PNREL	relational predicate noun, with copula only, see Sect. (xii) in 2.1.: *Byl tajemníkem* 'He was a secretary'
	GMULT	*Ten má knih!* 'What a number of books he has!'
	VCT	*Volali: "Vlasto!"* 'They were calling: "Vlasta!"'

(ii) With free modifications (adjuncts)

With locative and directional

see Fig.1. The case value (A(ccusative), D(ative), G(enitive), I(strumental), L(ocative)) is given here as a help to determine the functor; it does not constitute a part of the symbol.

LOC	DIR2	DIR3	DIR1
(where)	(which way)	(where to)	(from where)
na+L	*přes*+A	*na*+A	*z/s*+G
(on, at)	(over, across)	(on, to)	(from, at)
visí na zdi 'hang on the wall' *leží na stole* 'lie on the table'			
v+L	I, *skrz*+A	*do*+G (na+A)	*z*+G
(in)	(by, through)	(to, into)	(from)
v Praze, *na Smíchově* 'in P., in S.'		*do lesa*, *na Smíchov* 'to the wood, to S.'	

LOC (where)	DIR2 (which way)	DIR3 (where to)	DIR1 (from where)
u+G (at, by)	*podél*/*kolem*+G (along, (a)round)	*k*+D (to)	*od*+G (from)
nad+I (over, above)	*nad*+I, *přes*+A (over, across)	*nad*+A (over, above)	*znad* (from over)
pod+I (under, below)	*pod*+I (under, below)	*pod*+A (under)	*zpod* (from below)
	před+I (in front of, before)	*před*+A (in front of, before)	*zpřed* (from before)
	za+I (behind)	*za*+A (behind)	*zpoza* (from behind)
	mezi.1+I (among)	*mezi*.1+A (among)	
	mezi.2+I (between)	*mezi*.2+A (between)	
	naproti+D (opposite)		
	mimo+A (out)		
	vedle+G (beside)		
	kolem+G (round)		
	blízko+G (near)		

Figure 1

Instead of syntactic grammatemes with LOC and DIR, (Czech) prepositions (in lower case letters) are written, or, as the case may be, with numerical indices (if they stand as primary expressions for more grammatemes). In the MC primary prepositions are chosen even in situations where some other preposition is used on the surface in a secondary function; under 'primary preposition' the preposition from the leftmost column is understood. Thus, *na Spořilově* is tagged as LOC.*v*, *do Prahy* and *na Spořilov* as DIR3.*v*, *podél lesa* 'along the wood' DIR2.*u*, etc.

With temporal adjuncts

functor	grammateme	commentary
TWHEN	NIL	'whenever'*, v době/okamžiku/chvíli, kdy(ž)* 'at the time (moment) when', lexicalizations of the type *za svítání* 'at dawn'*, za Přemyslovců* 'under Přemyslides'*, s příchodem* 'with the arrival', *na odchodu* 'at the departure'*, v chůzi* 'when walking', *o sobotách* 'on

		Saturdays'
	AFT	*(dříve) než* 'before when', *(předtím) než* 'before when', *před* 'before'
	BEF	*až, poté, co, po* 'after'
	JBEF	*jakmile* 'just after', *(hned) jak* 'as soon as' (meaning: just before)
	APPX	*kolem/okolo poledne* 'about noon'
	INTV	*mezi šestou a sedmou* 'between six and seven', *mezi pondělkem a středou* 'between Monday and Wednesday'
THO	NIL	*(vždycky) při (každém) příchodu* '(always) with (every) arrival'
	AFT	*(vždycky) po (každém) příchodu* '(always) after (every) arrival'
	BEF	*(vždycky) před (každým) příchodem* '(always) before (every) arrival'

With other adjuncts

EXT	NIL	extent: *zaplatit na halíř* 'pay to the (last) penny' (lit.: to pay to the heller)
	APPX	lesser degree of precision: *je jich na sto* 'there is about a hundred of them', *váží to kolem*… 'it weighs about…'
	MORE	*nad padesát* 'over fifty'
	LESS	*pod padesát* 'under fifty'

With certain further adjuncts a 'positive' and a 'negative' grammateme is distinguished:

ACMP	NIL	accompaniment: *s* 'with'
	WOUT	*bez* 'without'
BEN	NIL	benefactive: *pro* 'for'
	AGST	*proti* 'against' (*bojovat* 'to fight', *akce* 'action')
CPR	NIL	comparison: *v_porovnání_s* 'in comparison with', *jako* 'as'
	AGST	*v_protikladu_k* 'in contrast to'
	DFR	with comparatives: *větší než Jirka* 'taller than George'
REG	NIL	regard: *se zřetelem k* 'with regard to'
	WOUT	*bez zřetele k* 'regardless of'

(iii) With coordination and parenthesis (as auxiliary markers/tags)

attribute Reltype ('type of syntactic relation'):

values:	CO	with all members of a coordinated structure,
	PA	with a parenthesis.

(e) LEXICAL PARTS OF THE TAGS

As has been already stated (in Sect. 1.1), we assume that the dictionary will be coming into existence gradually in the course of tagging the corpus, and that the lexical entries will, in addition to lemmas and morphemic information, contain valency frames of words including, among others, data on subcategorization (at least elementary: whether the modification with the given functor can be a N, V or A, etc.).

For the time being, some open questions still remain as far as derivation is concerned. Its most productive types should be covered by deriving from the basic lemma not only the forms of the given word, but also such derivatives as, e.g., with the verb *psát* 'write', *píšící,* 'writing (A)', *psaný,* 'written', *psaní* 'writing' (N), or feminines as *ředitelka* 'female director', deminutives as *stolek, stoleček* 'small table, very small table', adverbs as *dobře* 'well', perhaps also *přímo* 'directly'; negative derivatives as *nevelký* 'not large', *nedávno* 'not long ago'; however, sometimes it is not clear where to draw a dividing line: e.g., *nepřítel* 'non-friend, enemy' is not exactly a productive type. For the present we confine ourselves to taking as a purely "syntactic" derivation e.g. *můj* 'my': *já*.APP, 'I.APP' (as the case may be, with some other functor) and *otcův* (*otec*.APP,… 'father's').

Adverbs derived in a productive way from adjectives with corresponding meanings, such as *hezky* 'nicely'*, česky* 'in Czech'*, čistě* 'purely' have the lemmas of the adjectives.

In this manual lemmas are provisionally written as basic dictionary forms, but the spaces are underlined, e.g., *a_to, smát_se, t_j* (for "tj.").

Specific lexical symbols:

Neg	for negation (also for the prefix with verbs, but not with N, A): *nepíše* 's.o. doesn't write' is analysed as Neg.*psát,* but, e.g., *neotesanost* 'boorishness', 'ill-mannered behaviour' or *nemalý* 'not small' are lemmas
Gen	for general participant
Emp	for "empty verb" (in a verbless sentence)
se_Recp	for reciprocal *se, sebe, sobě, sebou* (in more detail see Sect. 2.6)
Cor	for the tectogrammatical counterpart of the subject of an infinitive with the verbs of control (with *zamýšlet* 'plan'*, radit* 'advise', etc.)
Comma	for comma with asyndetic coordination or apposition
Dash	dash
Colon	colon (as an apposition conjunction only, i.e. not with direct speech)
Slash	forward slash
Brackp	pair of brackets
Brackl	left bracket (for special cases where Brackp does not suffice)
Brackr	right bracket (for special cases where Brackp does not suffice)

2. CONVERSION OF ANALYTIC TREE STRUCTURES (ATSs) TO TECTOGRAMMATICAL STRUCTURES (TGTSs)

The procedure of the translation of analytic structures (ATSs) to the tectogrammatical ones (TGTSs) is conceived as a process having two steps:

(i) the first step consists in automatic preprocessing of analytic structures in the course of which they get rid of redundant nodes (in so far as this can be done automatically; a part of the automatic procedure, its second phase, takes place only after the trees of the large collection have been constructed manually);

(ii) the second step is represented by manual adjustments to the ultimate tectogrammatical structures; thus, the output of the automatic procedure ("pruning") serves as the input for manual preparation of training data; basic instructions for this preparation can be found in Sects. 2.2 - 2.6.

As a rule, it is the morphological grammatemes that are processed automatically and the tree is automatically deprived of the nodes that are redundant for the underlying structure.

In the large collection (LC), it is mainly the functors that are treated manually; the deleted nodes with lemmas are supplied and the topic-focus articulation is recorded. In the model collection (MC) also textual coreference and the marked, exceptional values of the grammatemes of tense, modality, number, as well as the values COREF, CORNUM and ANTEC with co-reference are dealt with.

Among the exceptions to the above basic scheme there are especially the following ones:

(a) The automatic treatment concerns also:

- the functors ACT, ADDR and PAT in basic configurations (also Instr after a copula --> PAT.PNREL),

- the functors INTF and ETHD,

- with *se* having 'Afun' Obj or Subj in a simple active clause the lemma Gen at the node having the functor ACT is introduced automatically,

- numerals (*pět lidí* 'five people': the numeral will depend on the noun),

- figures,

- quotation marks (inverted commas);

- such technical lemmas as Neg etc. (see Sect. 2.3) are also supplied.

(b) In the LC the following data are added manually:

(i) gender and number to the (potentially deleted and restored) pronoun *on* ´he´(see 2.5.A.1(b)), gender to the pronouns *já* ´I´, *ty* ´you´(=´thou´), *my* ´we´, *vy* ´you´ in agreement with the verb or, as the case may be, with an adjectival complement); gender and number is also assigned to *kdo* 'who' if it differs from the prototypical values ANIM and SG, resp. (the latter are added automatically in the second phase of the automatic analysis); this assignment is done separately, by a single annotator in the second pass;

(ii) the lemma of the antecedent will be stored as the value of the attribute COREF with grammatical coreference (see 2.5.B.3), i.e., with the lemmas Cor, *se* '-self', *svůj* 'his-Refl', *který, jenž* 'which', *kde* 'where' (e.g., *V Pelhřimově, kde jsme…´*In P., where we…´), *kam* 'where to', *odkud* 'where from', etc., and also with the predicative complement; should the antecedent be coordinated, the lemma of the conjunction is placed in the COREF.

2.1 Automatic procedure of adjusting the trees

2.1.1 The first phase of the automatic procedure

This part consists of several steps:

(i) Main verb

(a) the auxiliary symbol AuxS gets cancelled; into the lemma the number of the sentence is placed and the attribute functor is assigned the value SENT,

(b) the main verb of the sentence obtains the functor PRED.

(ii) Modality

The main verb is found (that is, the finite verb forms at the top level of the tree, i.e., in the main clauses).

From morphological data the following information is automatically taken over:

Verbmod	IND	indicative
	IMP	imperative
	CDN	conditional

CDN is also assigned to such constructions as *Nechtěl, aby přišli* ´He did not want them to come´.

The nodes for auxiliary verbs (AuxV) are cancelled in this step. Individual elements of tectogrammatical morphology (the values of grammatemes, etc.) by which auxiliary verbs connected with the main verb are to be replaced, are enumerated in a list.

Copula is regarded as a transitive verb with optional PAT: *Jirka.*ACT *byl malý.*PAT 'George was little'; *Jirka.*ACT *byl na zahradě.*LOC 'G. was in the garden'; *Lidí.*ACT *bylo pět.*PAT 'There were five people'; - but: *Lidí.*ACT *přišlo pět.*RSTR 'Five people came'.

> Note: Even in sentences such as *Byl na zahradě* '(He) was in the garden' the same verb as copula is to be seen; there is no ´verbum existentiae´ as a special lexeme in our approach.

Sentmod - (a) Where coordination (compound sentence) is not the case, the analytic value AuxK assigned to sentence final punctuation marks is rewritten as the following grammateme values at the main verb:

exclamation mark, fullstop, semicolon, colon:

if the leftmost node in the sentence is *ať, nechť, kéž* 'let' --> DESID

fullstop, semicolon, colon without *ať, nechť* --> ENUNC

exclamation mark: if Pred contains Verbmod IMP --> IMPER

else --> EXCL

question mark --> INTER

(b) If coordination is involved, the final symbol gets changed to Sentmod with the last verb (as if there were no coordinative relation) and with the other verbs it is distinguished according to Verbmod: Verbmod IMP gives Sentmod IMPER; if Verbmod is either IND or CDN without optative particles *kéž, ať, nechť,* then Sentmod ENUNC results; with CDN introduced by these particles DESID will be assigned.

Examples:

*Já půjdu.*ENUNC *a ty zamkni.*IMPER *dům!* 'I'll be leaving (now) and you lock up the house!'

*Ty zamkni.*IMPER *dům a já půjdu.*ENUNC 'You lock up the house and I'll be leaving.'

*On je v pořádku.*ENUNC*, ale ty máš.*EXCL *ránu!* 'He is OK, but you, what a shocking sight you are!'

Deontmod:

muset 'must'	--> DEB
mít 'be obliged' with an infinitive as analytic Obj	--> HRT
chtít 'want' with an infinitive as analytic Obj	--> VOL
moci 'be able'*, dát se* 'be possible'	--> POSS
smět 'may'	--> PERM

dokázat, *dovést*, *umět* 'can, know' --> FAC

The analytic grammatemes of tense, number, Verbmod and Deontmod are assigned in accordance with the values of the modal verb.

Note 1: Deontmod FAC with *už čte, píše* 'He/she is already reading, writing' cannot be assigned automatically (similarly as the use of modal verbs for probabilistic, epistemic modality), nor is manual adjustment envisaged in this case; it will appear in the MC only.

Note 2: The Czech *lze/nelze* 'is (im)possible' is treated as the following illustrations show: *Lze*.PRED *to*.PAT *splnit*.ACT 'It is possible to fulfil that'; *Něco*.ACT *takového*.RSTR *nelze*.PRED 'Such a thing is impossible'; also *je možné* 'it is possible', *je nutné*/*o* 'it is necessary', *je záhodno* 'it is advisable', *je třeba* 'it is needed' and so on is treated in this way.

(iii) Aspect

Impf --> PROC (processual)

Perf --> CPL (complex)

(a) if ATS contains AuxV + V pass. part. IMPF --> V.PROC

(b) if ATS contains AuxV + V pass. part. PERF --> V.CPL

(c) if ATS contains *být* 'be' + pass. part. PNOM --> V.RES (PNOM becomes mother, the node *být* disappears)

Note: (c) concerns the type *dveře jsou otevřeny* 'the door has been opened', *oběd je uvařen* 'the lunch (is) has been cooked', not the type *dveře jsou otevřené* 'the door is open', where the morphemic tag is Adj; here the copula *být* remains, with *otevřený*.PAT.

(d) Infinitives with the verb *mít* (*má uvařeno, má oběd uvařen* 'he/she has done with cooking, done with cooking the lunch' will be adjusted only manually, see 2.2.1, both in the MC and in the LC; they are reduced to one node (*uvařit* 'cook') and the automatically assigned Aspect value PROC gets changed to RES; the value of tense corresponds to that of the auxiliary verb.

(iv) Gender, number

with nouns: they are retained from the analytic level; the same holds for substantivized adjectives:

(a) *bytná* 'landlady', *hajný* 'gamekeeper', *hostinský* 'innkeeper', as well as e.g. *(nad)lesní, pokladní, pokojská, ponocný, vrchní, výčepní* (based on nouns), účetní, *Novákovi(c)* 'the Nováks';

(b) *krušovické, plzeňské* (kinds of beer named after the breweries Krušovice, Plzeň), *hovězí* 'beef', *telecí* 'veal', *vepřové* 'pork', *žitná* 'rye brandy', etc.

(c) *cestovné* 'travelling expenses', *nemocenská* 'health insurance fees', *odlučné* 'living-out maintenance', as well as e.g. *odstupné, výkupné, výpalné*, *kapesné, taneční, (mimo) jiné*, *stravné*;

(d) *cestující* 'traveller', *podezřelý* 'suspect', *nemocný* 'sick', as well as *neslyšící, obžalovaný, odsouzený, postižený, pracující, přednášející, příbuzný, raněný, studující, vedoucí, věřící, závislý (na drogách), žalovaný, kolemjdoucí* etc.

Note: the class is open; especially as concerns the type (d), the list is being constantly completed.

adjectives - the longer forms are taken as lemmas if there are any: *spokojen* 'satisfied', *zdráv* 'healthy' gets the lemma *spokojený*, *zdravý*;

the values for gender and number remain unchanged with superlatives, too, provided they do not depend on the noun, e.g.:

Budou tam jen ti nejlepší (ANIM.PL)

Bude tam jen ta nejlepší (FEM.SG)

'Only the best ones/one will be there'

Nejchytřejší z dívek bude/budou přijata/přijaty

'The smartest of the girls will be accepted'

here the gender corresponds to the dependent noun, the number is determined on the basis of context (here according to agreement);

even if adjectives are used in phrasemes, gender and number correspond to the morphemic form:

byl v úzkých (INAN.PL) 'he was in a tight corner',

platil hotovými (INAN PL) 'he paid in cash',

přišel s veselou (FEM SG) 'he arrived in a cheerful mind',

ťal do živého (NEUT SG) 'he cut to the quick)'

s dobrou se potázal (FEM SG) 'he had a good passage';

pronouns *ten(to)* 'that, this', *některý* 'some', *všichni* 'all', *naši* 'ours', *vaši* 'yours', etc. also occur in the positions of nouns;

gender is asigned to "genderless" pronouns (*já, ty, my, vy)* according to what univocally follows from agreement:

(a) agreement with the adjective dependent on a pronoun having any functor (*já ubohý, mně ubohému* 'I, the miserable').

(b) with subject (Afun Subj) gender is assigned according to the agreement with an adjective dependent on the copula (*my jsme nezávislí* 'we are independent') as well as with a verbal participle (*vy jste přišli* 'you have come'); *Přišli jsme* 'we-came' --> my.ANIM.PL, *Přišly* 'they-came' --> on.FEM.PL; *každý z nás* 'every one of us' has *každý* without gender (gender follows from agreement only), but *my* 'we' gets my.ANIM.PL;

numerals cardinal *dva - čtyři* 'two - four':

the values for gender and number remain unchanged, provided they do not depend on the noun;

gender and number with adjectives (including adjectival pronouns and numerals) and verbs are not cancelled for the time being, to make it possible to assign the respective values to a zero subject on their basis in the second phase of the automatic procedure.

Other issues

(v) The AuxP´s (the nodes for prepositions) get cancelled and the preposition is added to the attribute FW (Prep) of the noun on which the preposition depends in the ATS.

(vi) The node for subordinating conjunction AuxC gets cancelled and the conjunction is stored with the attribute FW (Conjunction) of the verb; with coordination of clauses the conjunction is

supplied to all members of the coordinated structure.

(vii) Degrees of comparison are rewritten automatically from the morphological data (POS, COMP, SUP).

(viii) With AuxT, *se/si* is attached to the lexical value of the governing word, e.g., *bát_se* 'be afraid of'.

(ix) All nodes labelled AuxX are cancelled if they do not immediately follow a noun (in this position the commas are preserved for manual treatment of the dividing line between a restrictive and a descriptive attribute; then they are deleted).

(x) Constructions with numerals:

Counted object = mother node, numeral = daughter:

substantival numerals: the numerals *pět* 'five' and higher, up to *devadesát devět* 'ninety nine'; further *čtvrt, (ne)mnoho, (ne)málo, několik, kolik, tolik* 'a quarter, (not) many, little, some, how many, as much', respectively, etc.: the counted noun is the governor, while the dependent numerical expression obtains the functor value RSTR;

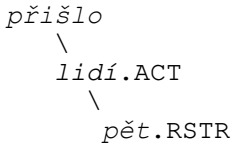
such numerals as *čtvero, patero,..., několikero, tolikero, hodně (lidí), dost (lidí)* 'four sorts (of), five sorts,…, several sorts, so many sorts, numbers (of people), enough (of people)', respectively, behave in the same way;

on the other hand, the following words behave as nouns: *milión* (and others ending in *-ión*), further *miliarda, polovina/polovice/půl(e), třetina, tisícina* 'billion, a half, a third, a thousandth', respectively, and also *tucet, veletucet, kopa, řada, spousta, hromada, zástup, dav, dvojice, trojice,* etc.;

the same holds for *sto, tisíc*, *trocha/u* (*s celým stem lidí* 'with the whole hundred (of) people', *byly tam tisíce* (pl.) *lidí* 'there were thousands of people'); i.e., *sto* 'hundred' etc. is the governor, and the counted object is dependent, having the functor value MAT; it is only in the MC that such configurations as *se sto lidmi, s trochu/trochou lidmi* 'with a hundred (of) people, with some/a few (of) people', respectively (should they occur) are analyzed in the same way as *s pěti lidmi* 'with five people'.

The same holds for analytic non-projectivities:

Lidí přišlo pět 'As for the people, (only) five arrived' is changed into a projective restrictive attribute:

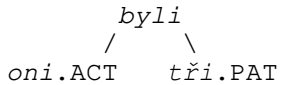


In a similar way:

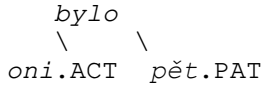
Piv.ACT *mi stačí deset*.RSTR 'as regards glasses of beer, ten is enough for me'; *bundu*.PAT *chci mít jednu*.RSTR 'as to jackets, I want to have (just) one'.

The situation is more simple with:

Byli tři 'They were three'



Bylo jich pět 'There were five of them'



(xi) An ordinal numeral together with the following full stop is represented by a single node with the relevant functor (RSTR); the same functor appears with *rok 1999*.RSTR 'the year 1999'.

(xii) Inverted commas (both normal and simple): they get cancelled if they occur only once in the given sentence and if

(a) there is a V.Obj between them; the verb is assigned PAT and either the grammateme DSP (direct speech) (if the inverted commas are placed on both sides - left and right) or the grammateme DSPP (a part of direct speech - if the quotation mark occurs on one side only) is assigned;

(b) there is just one word or a group of words between them involving one governing item (yet not a finite verb form and not being introduced by a colon); this item is assigned the value QUOT.

> Note: (1) If the direct speech sequence consists of more than two sentences, the intermediate sentences (without inverted commas) are not marked in a special way.
>
> (2) Such instances as *"Přijdu zítra," řekl Jirka, "protože..."* ("I'll come tomorrow," said George, "because…") are analyzed in the MC as: *Jirka řekl: "Přijdu..."* (George said: "I´ll come…").

(xiii) Afun PNOM at a noun in Instrumental --> the functor PAT carries the syntactic grammateme PNREL (Predic. Noun Relational); in other cases, PNOM --> functor PAT.

(xiv) AuxO with the pronouns *já* 'I', *ty* 'you', *my* 'we', *vy* 'you' in Dative --> functor ETHD (Ethical Dative): *On nám nedělá dobrotu.* 'We don´t have him behaving well'.

AuxO with lexemes *ten* 'that'*, on* 'he' --> functor INTF (intensifier): *On tam Jirka nebyl* 'He wasn't there, Jirka'*; Ono prší* 'It's raining, it is'*; To prší* 'What a rain!'

(xv) Subtrees constituted by complex numerals (e.g., 2350 specified in words) are replaced by a single node.

(xvi) Afun Subj with a verb in active voice --> ACT; in addition:

- if the form of Subj is Genitive and the verb is negated, the syntactic grammateme GNEG is assigned to the ACT: *Není peněz* 'there is no money';

- without negation:

(a) if the exclamation mark is present (EXCL with the main verb), ACT.GMULT results,

(b) else: ACT.GPART;

- if Subj in Locative follows the preposition *po,* it obtains the syntactic grammateme DISTR: *Na každé větvi viselo po jablíčku* 'Apples were hanging one by one on each branch'; lit.: 'On each branch hang by an apple';

- if Subj is in Accusative with the preposition *na*, it is assigned APPX: *Na sta mušek rozžehlo si světla v trávě* 'Fireflies in the hundreds turned on their lights in the grass'.

(xvii) If the verb is in active voice and Obj in Accusative and/or Dat are present --> PAT, ADDR respectively; passive is rendered in the same way as active (in ATS, the passive participle depends on AuxV as PredN); tense and modality of AuxV are retained, the aspect is taken over from the participle; at this stage the difference between active and passive can be recognized from the relation between ATS and TGTS only.

If, with the verb ín passive voice, the Obj is in Instr --> ACT.

(xviii) With *se*:

(a) if Afun is AuxT, the node is cancelled, *_se* is added to the lemma;

(b) if Afun is AuxR --> Gen.ACT

(c) if Afun with *si* is Adv --> *se*.BEN

(d) if Afun with *si* is AuxO --> *se*.ETHD

(e) else *se/si* is left with '???' for manual treatment;

(xix) With AuxY_PA: the word is not cancelled (it is not an auxiliary); it is assigned the functor PAR and the syntactic grammateme PA

With XX_PA (where XX is not AuxY): the syntactic grammateme PA is assigned; this grammateme is assigned to all parts of an inserted structure (ie. all nodes in parentheses, between dashes etc.)

(xx) NIL is added:

(a) to syntactic grammatemes except for those with LOC (i.e., with the functors LOC and DIRx the '???' remains, elsewhere NIL appears),

(b) in place of a lemma with the functor APPS, unless there is an element like *tj., tedy* 'i.e., hence';

(c) to DEL, ANTEC, COREF, CORNUM, CORSTN, Direct Speech, Phraseme, Quoted;

(d) to Reltype, unless coordination, apposition, parenthesis is the case;

(e) to Iterativeness with verbs (with other parts of speech NA is added).

(xxi) The words *a podobně, ap., apod.* 'similarly', *aj.* 'etc.'*, atd.* 'and so on' are divided into two nodes: *a* 'and' becomes the lemma of the node with the functor CONJ and *podobně* 'similarly', *jiné* 'other', *tak_dále* 'so on' gets the position of the rightmost element of the coordinated construction.

2.1.2 The second phase of the automatic procedure

The second phase of the automatic procedure is supposed to take place in the LC after the manual treatment:

(i) After the gender and number values are transferred in the LC according to the agreement (see above), the values are cancelled (i.e., NIL is supplied) with adjectives and adjectival pronouns that depend on a noun, or are in the predicate (PAT after copula), or carry the functor COMPL;

adjectival pronouns and adjectives, therefore, keep the gender and number information only when used as nouns (in a substantival function): *Ty modré dej do krabičky* 'Put the blue (ones) into a box';

gender and number do not get cancelled with substantival adjectives, superlatives etc., see Sect. (iv) in 2.1.1 above;

the pronoun *kdo* 'who' gets the values ANIM and SG if the values still were '???'; *co* 'what' gets NEUT and SG;

with possessive pronouns *jeho, její, jejich* 'his, her, their', if they depend on a noun, the lemma *on* 'he' is assigned together with the gender and number of the base pronoun:

jeho --> *on*.ANIM.SG

její --> *on*.FEM.SG

jejich --> *on*.XY.PL

similarly, the lemma, gender and number are assigned as follows:

můj, má, mého… 'my' --> *já*.XY.SG 'I'

tvůj 'your' --> *ty*.XY.SG 'you'

náš 'our' --> *my*.XY.PL 'we'

váš 'your' --> *vy*.XY.PL 'you'

matčin 'mother's' --> *matka*.FEM.SG 'mother' (also with *matčini* (PL), *etc.)*

otcův 'father's' --> *otec*.ANIM.SG 'father'

(ii) Sentmod with dependent content clauses: with the aid of a list of main verbs in the frameworks of which dependent question, command and announcement (or, more broadly, a content clause) can occur as objects and with the aid of a list of connecting expressions for ENUNC (*že* 'that'), IMPER (*ať, nechť, aby* 'let', 'so that'), INTER (*zda* 'whether', interrogative pronouns and adverbs), etc.; cf. 2.2.1(d).

(iii) Within coordination, modalities as well as tense are adjusted if they differ with individual coordinated verbs.

(iv) Secondary values of syntactic grammatemes are filled in (in place of NIL) wherever this is possible according to the prepositions: *bez, proti* 'without, against'; this also concerns at least some of the locative or directional grammatemes (according to the preposition and case: *do, mezi,…* 'into, between',…).

(v) The remaining nodes for commas, hyphens, inverted commas, brackets, colons and dashes get cancelled.

(vi) The preposition or conjunction from the attribute FW is transferred to the attribute of the syntactic grammateme, if it fits there according to the chart and list of syntactic grammatemes in Sect. 1.2(d).

(vii) With lemma *se* (Refl) the lemma of the ACT is assigned to COREF.

(viii) Wherever lemma '???' remains with a verb and a noun in coordination, the lemma '???' is to be replaced by the lemma of the left-hand or right-hand sibling.

A future version of automatic analysis is being prepared, based on the experience from the present stage of the tagging, which will take over some of the tasks of the hitherto manual procedure.

2.2 Manual conversion of ATSs to tectogrammatical syntactic structures (TGTSs)

Note: If an error is found in the ATS, we leave the tree unchanged, but the correction must be registered.

2.2.1 How to convert morphological grammatemes

In the LC (with the exceptions pointed out below), only those values of morphological grammatemes are assigned which are obtained automatically. Some values, therefore, may, for the meantime, not agree with theoretical requirements.

The following manual assignments will be carried out by a single annotator in a separate pass through the LC:

(a) constructions of the type *má uvařeno* 'has done cooking', where the two nodes are reduced to one, to which the value of aspect RES (instead of PROC) is assigned. Two cases must be distinguished: in such constructions as *má oběd uvařen* 'he has his lunch cooked' it must be decided from the context whether it was himself who had done the cooking or whether someone else had cooked the lunch for him; in both the cases the nodes unite into one *uvařit.*RES; the subject obtains the functor ACT if the cook was himself, else it is assigned ADDR; the tense conforms to the auxiliary verb, e.g.:

(i) *Honza*.ACT/ADDR *už uvařit*.RES *oběd.*PAT 'Johny ACT/ADDR already cook.RES lunch.PAT'.

(ii) Gen.PAT *Honza*.ACT/ADDR *už uvařit*.RES

If ADDR is chosen, Gen.ACT is added:

(i) Gen.ACT *Honza*.ADDR *už uvařit*.RES *oběd.*PAT

(ii) Gen.ACT Gen.PAT *Honza*.ADDR *už uvařit*.RES

(b) the assignment of gender to "genderless" pronouns (*já, ty, my, vy, kdo, co* '*I, you (thou), we, you, who, what*', respectively*)* as well as gender and number to the pronoun *on* 'he' and to zero subject as determined by agreement, will be carried out by a single annotator in the second pass of manual annotation, see above in Sect. 2.1.(iv):

on 'he': ANIM or INAN, *ona* 'she': SG.FEM or PL.NEUT

který 'which': ANIM or INAN, unless it is used as an adjective (e.g., *který hoch* 'which boy' will obtain gender with *hoch* only, but: *všichni, kteří*.ANIM 'all who');

ti, kdo.PL.ANIM 'those, who', *všichni, co*.PL.ANIM 'all that';

já, ty, my 'I, you (thou), we' will have ANIM or FEM (INAN or NEUT seem likely to appear in fairy tales etc. only);

Přišli jsme 'we have come' --> *my*.ANIM.PL,

Přišly 'they have come' --> *on*.FEM.PL;

každý z nás 'everyone of us' has *každý* 'everyone' without gender (this follows from agreement only), but *my*.ANIM.

Adjectival pronouns and adjectives obtain gender in substantival use only: *Nemocní leží v přízemí* 'The sick are lying on the ground-floor.' *Naši nejsou doma* 'Our people are not at home' lit.: 'Our (ones) are not…'.

For verbal forms such as *dělal* 'he did/was doing' as well as *dělá* 'does/is doing', etc. the following instructions are formulated:

- the gender (more precisely: ANIM/INAN, FEM/INAN) is transferred from the verb to the subject pronoun,

- also to a subject that is ambiguous, as *on* 'he, it'*, ony* 'they' or that has the form of zero or of a genderless pronoun;

(c) possessive pronouns and possessive adjectives have a special position: *můj* 'my'*, tvůj* 'your'*, svůj* 'his' (reflexive)*, náš* 'our'*, váš* 'your', do not reflect differences of gender (of agreement: *můj/má/mé* 'my'; only in the MC gender is taken over from *já* 'I': *Řekla jsem, že můj stůl je starý* 'I (FEM) said that my table was old' - *můj* gets *já*.SG.FEM.APP or .ACT etc.; the gender is assigned by a single annotator in the second pass of manual annotation everywhere, the number is assigned only with the pronoun *váš* 'your-SG or PL'.

Similarly:

jeho: SG.ANIM/INAN/NEUT 'his'; slash means "or" here - the decision will be based on context

její: SG.FEM 'her'

jejich: PL.ANIM/INAN/FEM/NEUT 'their'

In all these cases the lemma is *on*.

jehož: SG.ANIM/INAN/NEUT 'whose'

jejíž: SG.FEM 'whose'

jejichž: PL.ANIM/INAN/FEM/NEUT 'whose'

In all these cases the lemma *který* 'who' is assigned.

Similarly with possessive adjectives:

otcův 'father's', *Janův* 'John's', *matčin* 'mother's', etc. - lemma, gender and number (always SG) will correspond to the base noun.

(d) gender is assigned to numerals and pronouns in such constructions as *jeden z*...'one of…', *každý z*...'every of…'.

In the MC, in addition to the above items also the following issues are treated:

(a) secondary and transposed use of

(i) tense and aspect: 'praesens historicum' obtains ANT, 'praesens pro futuro' is assigned POST; *pokvete* 'will flower', *půjde* 'will go', *poletí* 'will fly' is assigned the values POST (for tense) and PROC (for aspect);

(ii) number with 'pluralia tantum' (*nůžky* 'scissors', *kalhoty* 'trousers') obtain SG or PL according to the given context: *(jedny) nůžky* '(one pair of) scissors' obtain SG;

(b) the category ITER: in the MC IT1 will be assigned to both the sentence *Když vejde do dveří, rozsvítí světlo* 'After entering the door he switches the light on', and *Vždycky rozsvítí světlo v 9 hodin* 'He always switches the light on at 9 o'clock'

(c) Deontmod:

In the MC, modal verbs (and negation) are treated (in conformity with Buráňová's treatment in Panevová, Benešová, Sgall 1971, p.127) in the following way:

DEB: *muset* 'must', *být nucen* 'be forced to', *nemoci ne-* 'cannot not to', *neumět ne-* ´not to know how not to´, *nedovést ne-* ´not to be able not to´

HRT: *mít* 'ought to'

POSS: *moci* (meaning 'it is possible'), *dá se* 'it can be done'

PERM: *smět, moci* 'may' (meaning *mít dovoleno* 'be allowed to')

FAC: *dovést, umět, moci* 'can' (in the sense *mít schopnost* 'be able')

The automatic procedure assigns FAC to structures with *umět, dovést* 'can, be able' only; in the MC, FAC is also assigned in contexts in which it is not expressed by means of modal verbs, but where the so-called habitual capacity is the case (*Dítě už chodí* 'The child walks already'; *Ta trouba dobře peče* 'That oven bakes well'.

VOL: *chtít* 'want', *hodlat* 'intend'

> Note 1: Such phrases as *je možné* 'it is possible', *je nutné* 'it is necessary' are understood as connections of lexical units (with copula).
>
> Note 2: For the time being, a special treatment of combinations of modal verbs (*může chtít* '(he) can want', etc.) is not envisaged; the infinitive will be represented as PAT. In a similar way the negation of an infinitive will be treated: *může nechtít*.PAT '(he) is able not to want'.
>
> Note 3: Not even in the MC the use of modal verbs to express epistemic (probabilistic) modality is taken into account (*Musí tady někde být* 'He must be somewhere here'; *To by mohl být Jenda* 'That could be Johny').

Essentially the treatment corresponds to the following chart (i stands for initiator, a for actor):

	i = obj.circumstances	i differs from a	i = a
DEB	<----------	*musí* 'must'	---------->
HRT	<------ *má* 'shall'	---------->	*chce* 'wants to'
POSS	*může* 'can'	*může/smí* 'may'	
FAC			*může/umí* 'can'

(d) Sentmod in subordinate content clauses:

In the MC, Sentmod will appear also with the dependent verb of a subordinate content clause following a verb of saying (*verbum dicendi* - in a narrower sense, i.e., where dependent question, command or announcement can stand in competition), e.g., *Řekl* (ENUNC), *zda/kdy/proč přijde* (INTER) 'He said whether/when/why was he coming' (the same holds with *koho viděl* 'whom did he see', *kam přišel* 'where did he come' - with all interrogative pronouns and adverbials), *Řekl* (ENUNC)*, ať přijde/aby přišel* (IMPER) 'He said that he come/He told him to come', *Řekl* (ENUNC), *že přijde* (ENUNC) 'he said he would come', *Řekl* (ENUNC), *kéž by přišli* (DESID) 'He said if only they came/might they come'.

(e) Quoted - QUOT: in the MC the head of a quoted group of words is assigned this functor value if the group is enclosed in quotation marks (even in simple ones) and if it does not constitute a complete clause; if it is just one word, it obtains QUOT in the automatic procedure.

(f) Dual: *nohama* and *nohami* represent two different lemmas:

noha1 'leg' (a limb with Instr PL in *-ama*) and *noha2* (a leg of a piece of furniture with Instr PL in *-ami*).

2.2.2 How to assign functors and syntactic grammatemes

Introductory remarks

In what follows we proceed according to morphemic means by which the functors are expressed; therefore, we first deal with the endings of cases, then with prepositions, subordinating conjunctions and, lastly, with the functions of agreement - in the attribute, the complement and the relation between predicate and subject.

The means of expression having syntactic functions, i.e. the endings of cases and of the infinitive (later, as the case may be, transgressives and syntactic derivatives) will be in part converted to functors already by the automatic procedure (the typical instances of the main participants, i.e., ACT, ADDR, PAT, and also PAT.PNREL); the rest are converted manually, even in the LC; for exceptions see below the paragraph on the assignment of functors to nouns. In the attribute functor we have a space for another value possible: that is where also a question mark can be placed (in case the choice of the first functor is uncertain).

In case of a structural uncertainty the annotator decides according to his/her understanding of the text.

Further remarks

The governing verb

The governing verb of the sentence obtains the functor PRED (if the governing verbs participate in a coordination construction, each of them will be assigned this functor).

The copula *být* (as well as the 'verbum existentiae') has an obligatory ACT and an obligatory PAT; the latter is absent e.g. in *je pozdě*.TWHEN 'it is late', *něco je daleko*.LOC 'st. is far', *vzdálenost je pět minut*.EXT *chůze*.APP 'the distance is five minutes of walk'.

A further group of verbs has an optional ACT: *jde* (*mi*.ACT) *o něco*.PAT 'I am concerned about st.'; similarly *jedná se* (*mi*.ACT) *o něco*.PAT 'st. is at stake', also *jedná se o, běží o*; ACT is obligatory e.g. in *došlo k nehodě*.ACT 'an accident occurred'; see also Panevová (1998b) on *bolí mě* 'I feel pain', *táhne* 'there is a draught', etc.

Sentences without governing verbs

The functors VOC, EV and DENOM:

If in the ATS the top of the tree (under #) some other element than a finite verb is found (VF, i.e., a personal form), the treatment proceeds (in both MC and LC) as follows:

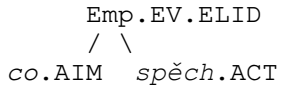
(a) if the top node is N in Nominative, it gets the functor DENOM;

(b) if the top node is N in Vocative, then it gets the functor VOC;

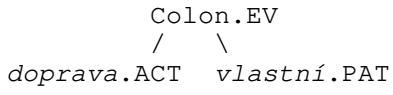
(c) if the top node is INTJ, the particle *ano* 'yes', 'no', then it gets the functor PARTL;

(d) else a node is added (probably always as governing a node with ExD in ATS) with the functor EV and with the lemma Emp (or Dash or Colon, as the case may be); should in the analytic tree more edges occur under #, then a node with the lemma Emp (=empty verb) and the functor EV will be placed above them, the words below getting "regular" tags.

Example: *Nač spěch?* 'Why the hurry?'



Doprava: vlastní 'Transportation: individual'



If under # coordination or apposition occurs, the procedure described under (a), (b), (c) above concerns individual components of the coordination or apposition construction. ExD is always replaced by the functor corresponding to the type of dependency of the given node on its newly established governor: Nom, as a rule, corresponds to ACT, Acc to PAT, etc. If the node with ExD is a N in Voc, it obtains the functor VOCAT and is recorded as standing in apposition with the rest of the sentence.

Apposition

Apposition has APPS as the value of the functor and Comma (Brackp, Dash, as the case may be) as the lemma of the conjunction if there is only a comma (etc.) instead of a conjunction; if the connectors *tj*. 'i.e.'*, tedy* 'then, consequently'*, či(li)* 'or' are present in the sentence, then their lexical values are assigned as the lemma.

Direct speech

Direct speech (i.e. its head verb) is assigned the functor PAT or ACT, according to its relation to the verb of saying:

Řekli: Nechoď.PAT *tam!* 'They said: Don't go there' - but *Ozvalo se: Nechoď*.ACT *tam!* 'It-was-heard: Don't go there'.

As verbs of saying also such expressions as *usmál se* 'he smiled', *ušklíbl se* 'he grinned', etc. are classed.

If the direct speech has no verb of saying, Emp with the functor EV is supplied.

Such instances as "*Přijdu zítra," řekl Jirka, "protože…"* '"I'll come tomorrow," said George, "because…"' will be analyzed as: *Jirka řekl: "Přijdu, protože…"* 'George said: "I'll come because…"'.

Numerals

The automatic procedure converts *pět* '*five*' etc. into attributes, see Sect. 2.1(x); however, some other numerals behave as nouns on which quantitative data depend: *milión* 'million' as well as *sto* 'hundred' and *tisíc* 'thousand' (*s celým stem lidí* 'with a whole hundred of people', *byly tam tisíce lidí* 'there were thousands of people', *trocha/u* 'a bit, a few'; i.e., *sto* 'a hundred' etc. represent the governing elements, the counted object is dependent with the functor MAT. Only in the MC such configurations as *se sto lidmi* 'with a hundred of

people'*, s trochu/trochou lidmi* 'with a few people' are analyzed in the same way as *s pěti lidmi* 'with five people'.

Reference words

The node of the pronoun *to* 'it', if it refers to a subordinate (adverbial or content) clause (i.e., not to a relative clause), is cancelled both in the MC and LC; the verb of the dependent clause depends on the verb of the governing clause, which obtains a corresponding functor (the elements in brackets [] are left out in the TGTSs):

*Přemýšlel [o tom, že] přijde.*PAT 'He pondered whether s/he would come', lit. '…[about it that] s/he would come'

*Zajímavé bylo [to, že] přišla.*ACT 'It was interesting that she came'

*Bavil se [tím, že] přišla.*MEANS 'He was amused by her coming'

*Udělal něco [proto, aby] přišla.*AIM 'He did something to make her come'.

Relative clauses

With relative clauses, however, the reference word is a genuine pronoun on which the verb of the clause depends as RSTR, DES etc.: *Vypravoval o těch, které tam potkal* 'He was telling about those whom he met there'.

Sometimes, if ambiguity is involved, it is necessary to decide: *Přemýšlel o tom, kdo přijde* 'He was thinking about who was to come' - this either means "about the man who was to come" (with the relative clause depending on the pronoun), or "about which of the men in question were to come". In the latter case, the content clause depends on the verb *přemýšlel* 'was thinking' as PAT.

The words *tam* 'there', *odtud* 'from there', *tudy* 'this way' have always a node of their own and a functor (LOC or DIRi); a dependent local clause can have different relative adverbs (*kde* 'where'*, kam* '(to) where'*, odkud* 'from where'*, kudy* 'which way'); the verb of the relative clause on which these adverbs depend has the functor RSTR (unless there is a clear reason for using some other functor; e.g., *dívala se tam.*DIR3*, odkud.*DIR1 *ses ozval.*RSTR 'She was looking in the direction from where you had made yourself heard').

Temporal functors

Note that *za jak dlouho* 'in how long a time' is neither a special functor nor a grammateme, it is a contextual variant of *jak dlouho* 'how long' with complex (perfective) aspect: *psal ten úkol dvě hodiny* 'he was writing that paper (for) two hours', *napsal ten úkol za dvě hodiny* 'he wrote that paper in two hours'.

TSIN 'since when?' is also expressed by such compound conjunctions as: *od_té_ doby,_co* 'since the time that', as well as *od_těch_dob* 'since then', *od_toho_času,_kdy* 'since the time when'.

Functors with nouns

Wherever it is semantically suitable, "verbal" functors are assigned to the dependents of a noun: *kabát do deště* 'raincoat', lit.: a coat to rain (DIR3), *chata v lese* 'a cabin in the wood' (LOC), *kapky na trávení* 'drops for digestion' (AIM), *názor/návod na co* 'opinion about/instruction for st.' (PAT), *dopis matce* 'letter to mother' (ADDR), *dopis pro matku* 'letter for mother' (BEN), *román (od) Tolstého* 'a novel by Tolstoy' (ACT), *otcův* (APP) *román od Tolstého* (ACT) 'Father's (APP) novel by Tolstoy' (ACT), *Rodinova (ACT) socha Balzaca* (PAT) 'Rodin's statue of Balzac', *královnina* (APP) *socha Shakespeara* (PAT) *od Brancusiho* (ACT) 'the Queen´s statue of Shakespeare by Brancusi', *Masarykův* (PAT) *portrét* 'portrait of Masaryk', *Centrum Viléma Mathesia* (APP) 'Vilem Mathesius Centre'.

Certain nouns have PAT in their valency frames, although no ACT is present there; example: *příklad čeho* 'example of'*, případ čeho* 'case of'*, clo na zboží* 'duty fee for commodity'*, potvrzení na* 'certificate for'*, poplatek za* 'fee for'.

The nouns of action are approached in a "lexicalist" way, as lexical derivations; they are not transformed into a dependent predication, but they in part have obligatory participants in the same way as verbs, so that in the LC

deletions are restored with them (even in the position of 'control' Cor is supplied to them, see Sect. B.3 in Chapter 2.5).

Remarks on the assignment of particular functors to nouns:

(A) Action nouns, in principle, share the frames of the basic verb; their ACT (if it is not a General Actor) has the form of Instr (*ošetření lékařem* 'treatment by a doctor'), Genitive (or Possessive Adjective, as its variant) or *od*+Genitive (*ošetření lékaře*/*lékařovo ošetření* 'treatment of the doctor'/'doctor's treatment', *ošetření*/*pomoc od lékaře* 'treatment/help of the doctor').

(B) Agentive nouns do not have an ACT (the Actor of the action is 'absorbed' by the suffix *-tel*/*-ač* etc.); according to the structure of their frames, this group of nouns can be divided into three subgroups:

(i) pure agentive nouns, with an obligatory PAT (*učitel tance* 'teacher of dance', *předkladatel zákona* 'a proposer of a law', *vyznavači metalu* 'lovers of metal (music)', *zástupce ministerstva* 'a representative of the ministery', *pisatel* 'the writer';

(ii) PAT is possible, but not necessary: *překladatel* 'translator', *autor* 'author', *režisér* 'director', *prodavač* 'shop assistant';

(iii) PAT is not possible: e.g. a name of a profession (*nakladatel* 'publisher', *kovář* 'blacksmith', *kupec* 'dealer').

(C) The names of artifacts are themselves names of results, and therefore they do not have EFF in their frames. They have optional ACT, PAT and ORIG. ACT can be expressed by a noun in Genitive, a Possessive Adjective, by a preposition *od*+Genitive; ORIG is rendered by *od*+Genitive; PAT is rendered by *o*+Locative (*román o čem* 'novel about', *báseň* 'poem', *dopis* 'letter') or by a noun in Genitive or Possessive Adjective (*portrét* 'portrait', *obraz* 'picture'). The latter forms may also render APP.

ACT may be present also with such names of artifacts as *budova* 'building', *kusy nábytku* 'pieces of furniture', *kusy oděvu* 'pieces of dress'; in *Gočárova stavba* 'Gočár's building' *Gočár* is ACT of *stavba*, in contrast to *Gočárovo*/ *Priessnitzovo sanatorium* 'Gočár's/Priessnitz' sanatorium', where the dependent is APP (the sanatorium was named on their honour). With the names of artifacts, ACT is only optional: in *Vezu si s sebou tři romány* '(I) take with me three novels', *Mám v pracovně čtyři židle* '(I) have four chairs in my office' there is no ACT with the nouns *román* and *pracovna*, respectively.

(D) It is a rare case that a noun has no PAT but has an ADDR in its frame (they, of course, may have also an ACT): *dárek mamince* 'gift to mother', *dopis rodičům*, 'letter to parents', *věnování příteli* 'dedication to friend'; PAT is here 'absorbed' by the type of derivation: what is donated, is a gift etc.

ID, MAT and APP are participants specific to nouns; ID is optional; MAT is optional e.g. with *sklenice* 'glass', *talíř* 'plate', *košík* 'basket' but obligatory e.g. with *skupina* 'group', *zbytek* 'rest', *část* 'part'; APP is an ajunct (free modification), obligatory e.g. with *bratr* 'brother', *povrch* 'surface', *začátek* 'beginning'. The functor ID can have the form of an infinitive or of a clause (as a heading of a text, or similarly).

> Note: For a number of newly grammaticalized prepositions new functors will probably be necessary; in the MC it might be useful to introduce them step by step, while in the LC under newly grammaticalized prepositions we understand those complex expressions that are listed in Sect. 2.2.2.2. below.

Syntactic grammatemes

Syntactic grammatemes are consistently introduced in the MC only. In some cases they are inserted automatically with the functors ACT and PAT (cf. above, Sect. 2.1(xiii) and (xvi)). The rest are assigned manually in the MC. - NIL supplied by the automatic procedure is changed manually into a secondary value (e.g., WOUT, BEF) in places where necessary; instead of syntactic grammatemes with LOC and DIR in the MC a (Czech) preposition in lower case letters is written with an index according to the Table from Sect. 1.2(d)(ii); at the same time, secondary uses of the prepositions etc. will be checked.

Phrasemes

A phraseme is denoted as PHRi, where i equals 1 with all the parts of the first phraseme in the sentence, or, as the case may be, 2 with the second phraseme in the same sentence etc. Inside a phraseme PHRi also a part

without functor (assigned NIL) can occur, which depends on the main part of the phraseme: *křížem*.MANN *krážem*.NIL 'criss-cross'.

Coordination

With coordination the **conjunction** is to be represented as **lemma**, the kind of coordination as **functor**; it is necessary to distinguish between CONJ and GRAD with the conjunctions *i*, *ani* 'as well as, nor', e.g., *muži i.CONJ ženy* 'men and women', *přestupky, ani*.GRAD *zločiny* 'trespasses nor crimes'. In case *i, ani* do not join two parts (or clauses), RHEM is assigned: *Byly tam i ženy* 'there were also women as well'*; Netrestají ani zločiny* 'They do not punish even crimes'.

In the LC, manual insertion of syntactic grammatemes is not carried out; NIL or '???' are replaced by respective values only in MC.

2.2.2.1 Endings of cases, infinitive, adverbials

As primary functions (here: p) such **unmarked** instances are regarded with which no confinement to a specific class of contexts is assumed; the rest of the functions are illustrated by typical examples with context criteria (specified tentatively); with most of these secondary functions further research will probably confirm that they are restricted lexically (phraseologically) and therefore that it will be possible to make an inventory of contexts which allow them.

Prepositionless cases:

Nom	p: **ACT**	also *(Nač) ten spěch* '(What's) all the hurry for';
	ID	*parník Hradčany* 'the steamboat H.', *pojem čas* 'the concept time'; the criterion: indeclinable Nom, may alternate with Gen;
	PAT	following a copula: *je učitel* 'he is a teacher', *je mladý* 'he is young';
	COMPL	*vrátil se nemocný* 'he came back ill', *šli jeden.*COMPL *za druhým* 'they were following one another', see 2.2.2.4;
	(Note: As for the lexical meanings, with *být* 'be' the auxiliary function is differentiated from the lexical one, without making a difference between the copula and the verb of existence.)	
Genit	p: **APP**	*klobouk mého bratra* 'my brother's hat'*, střecha domu* 'the roof of the house'*, předseda (tajemník, sekretariát) Aliance* 'the chairman (secretary, office) of the Alliance' *nerozhodnost mnoha lidí* 'irresoluteness of many people', *Centrum V. Mathesia* 'Centre of V. Mathesius'; the criterion: meaning is similar to 'belonging to'
	RSTR	*krasavice vysoké postavy* 'a beauty of tall stature'; the criterion: *jaká?* 'what?';
	ID	*pojem čísla* 'the concept of number', *otázka laickosti* 'the question of laity', *trest smrti* 'penalty of death'; the criterion: the concept (question, penalty) referred to is number (laity, death), or: number is a kind or an instance of concept; note: The construction *otázka jeho ubytování u nás* 'the question of his accommodation with us' is ambiguuos between ID (the question connected with his accommodation) and PAT (…whether we will accommodate him with us).
	MAT	*sud okurek* 'a barrel of cucumbers', *série omylů* 'a series of mistakes';
	ACT	*vznik světa* 'origin of the world'*, román Jáchyma Topola* 'a novel by Jáchym Topol'; (what, who originated, wrote, did); *knih (těch, kteří…) je pět* 'books (those that…) are five', similarly as *knihy jsou čtyři* 'the books are four';
	ACT.GPART	*vody ubývá* 'water is running low';
	ACT.GMULT	*Tam bylo knih!* 'What numbers of books were there!' (lit.: There were books-Gen);
	PAT	*stvoření světa* 'creation of the world'*, cena benzinu* 'price of petrol', *pomník T. G. Masaryka* 'monument of T.G.M'*, bál se toho, že zaprší* 'he was afraid (of it) that

there would fall a rain' (*zaprší* is PAT depending on *bát_se* 'be afraid'); *toho* 'of it' is left out (see above in Sect. 2.2.2 – Introductory remarks on the so-called reference words); with *bát se, aby ne* 'be afraid lest' PAT is without negation;

PAT.GMULT *Ten měl knih!* 'What (heaps of) books he had!'

TWHEN.NIL *toho dne* 'on that day', *roku 1994* 'in the year 1994', *každého pát*ku '(on) every Friday';

CPR.DFR *Jitro je rozumnější večera* 'the morning is wiser than the evening'

Dat p: **ADDR** *dát, koupit, psát komu co* 'give, buy, write s.o. s.t.', *dárek sestře* 'a gift for the sister', *dopis kolegovi* 'a letter to a colleague'; *divit se tomu, že svítá* 'marvel that dawn is coming' (lit. marvel at it (Dat) that…), see above on the so-called reference words; similarly with the other cases;

BEN *spravit někomu kolo* 'to repair a bicycle for s.o.', *pomník padlým* 'a war memorial' (lit. a memorial to the war dead);

PAT *vyučování angličtině* 'teaching English';

ETHD *Děti nám nechodí včas* 'we don't have the children coming in time' (lit.: the children don't come in time to us (Dat)); *Já ti mám knih!* 'I do have lots of books, I tell you!';

Acc p: **PAT**

THL *spal celou noc* 'he slept all night'*; pobyl tři týdny* 'he stayed three weeks'*; počkej okamžik* 'wait a moment'*; běžel tu trať tři hodiny* 'he ran that distance three hours';

TWHEN *Stalo se to ten den* 'It happened that day'*; každou hodinu* 'every hour' (this is not THO, because "how long" is expressed here by the lexeme *každý* 'every'*;*

EXT *vysoký šest stop* 'six feet tall'*, stojí dvacet korun* 'it costs twenty crowns'*, platí málo* 's/he/it/they pay(s) little (money)';

Voc if there is a N (A, Pron) in Vocative as a dependent of the node # in the ATS, the functor **VOC** is assigned to it: *Ty*.VOC *nešťastná*.RSTR*!* 'you unhappy!';

if a N (A) in vocative represents the governing word of a direct speech construction, it obtains a functor according to its relation to the verb of saying (usually **PAT**) and the syntactic grammateme VCT is added – *Volali: Jirko*.PAT.VCT*!* 'They called: George!';

if a node marked ExD is in Vocative case (in some other position), it obtains the functor **VOCAT** and it is recorded as an apposition with the rest of the sentence: *Jirko, podej mi to!* 'George, give me that!'

Loc has always a preposition; it is dealt with in Sect. 2.2.2.2.

Instr p: **MEANS** *jet tramvají* 'go by a tram'*, krájet nožem* 'carve with a knife';

ACT *byl učitelem pokárán* 'he was reprimanded by the teacher';

PAT.PNREL *je učitelem* 'he is a teacher';

PAT.NIL *stal se učitelem* 'he became a teacher'*, jevil se roztržitým* 'he appeared absent-minded'*, hrozit válkou* 'threaten with war';

DIR2 *jít lesem* 'pass through a wood';

TWHEN.NIL *časem se to hodí* 'time and again it comes in handy'*, chvílemi* 'at times'*, každým okamžikem* 'at any moment';

THL *Dnem i nocí dlí u lože dítěte* 'day and night s/he is staying (up) at the child's

		bedside';
	MANN	*podívat se úkosem* 'look askance at s.o.'*, opakovat šeptem* 'repeat in a whisper';
	CPR.NIL	*Krev se řinula proudem* 'blood ran in streams';
	CAUS	*rozplývat se štěstím* 'melt for happiness'*, rozstonat se chřipkou* 'fall ill of flu';
	REG	*rodem Jihočech* 'a South Bohemian by birth'*, rovnat se komu čím* 'equal s.o. in stg.';
	CRIT	*řídit se zásadami* 'follow the principles';
	EFF	*nazvali ho utrhačem* 'they have called him slanderer'*, zvolili ho předsedou* 'they elected him (as) chairman';
Infin	p: **PAT**	*rozhodl se* 'he decided to'*; snaha* 'effort to'*;* also with phase verbs, such as *začít* 'begin', *končit* 'finish' as well as with all constructions involving negative infinitives: *může tam nejít* '(s/he) is allowed not to go there'*, nemůže tam nejít* '(s/he) is not able not to go there' (only with a positive Infinitive the value of modality results automatically from purely modal verbs, so that only the node of the full verb remains);
	ACT	*jít tam znamená…* 'to go there means…'; *nelze chodit pozdě* 'it is impossible to come late'
	PRED	infinitive for Imperative: *Nemluvit!* 'Not to speak'; in MC also the grammatemes are assigned acording to the context;
	COND	*nebýt vás, nebyl bych tady* 'if it were not for you, I wouldn't be here' (lit.: not to be of you,…);
	INTT	*šel se koupat* 'he went bathing', *poslali ho pást krávy* 'they sent him to graze the cows' (this differs from AIM: *aby* 'so that' is more closely connected with the governing verb, INTT and AIM can stand side by side and they cannot be coordinated: *šel nakoupit, aby měli co jíst* 'he went shopping so that they may have something to eat'; in a similar way INTT differs from DIR3 *do* '(in)to' – *šel na houby do lesa* 'he went picking mushrooms into the wood');
	EV	in sentences such as *Požadovat*.PAT *vyšší úroveň*! 'A higher level is to be required!', where the infinitive depends as PAT on EV; in MC also the grammatemes are added according to the context
	RESTR	see the conjunction *než* below (Sect. 2.2.2.3)
adverbs	see the list in Appendix A;	
		the criterion: try questions, not only *jak* 'how', but: *jakým způsobem* 'in what manner': **MANN** *s jakým cílem* 'with what aim': **AIM** *s jakým účinkem* 'with what result'**: RESL** *v jaké míře* 'to what extent': **EXT** etc.;
	examples:	
	p: **MANN**	*dobře* 'well', *česky* 'in Czech', *právnicky* 'juridically', *lexikálně* 'lexically'…
	LOC, DIRn	*tady* 'here', *tam* 'there'…
	THO	*často* 'often', *zřídka* 'rarely', *dvakrát* 'twice'
	TWH and other temporal adverbs	*teď* 'now'*, odedávna* 'long since',…
	EXT	*velmi* ´very´
	PREC	*spíš* 'rather' (although having AuxY!), *spíš než*'rather than' CPR.DFR

REG	*příjmově* 'as for the income', *národnostně* 'nationally'
CAUS	*proto* 'therefore', *schválně* 'on purpose'
RESL	*slavně* 'triumphantly', *skandálně* 'scandalously', *zřetelně* 'distinctly', *viditelně* 'visibly' (criterion: "so that it was…")
ATT	*bohužel* 'regrettably', *radostně* 'cheerfully'
RHEM	*jen*(*om*) 'only/just', *i* 'as well as' (provided it does not join clauses), *také/y* 'also', *právě* 'just/precisely', *zejména* 'particularly', *především* 'above all', *trochu* 'a bit/a little', *takřka* 'practically/almost'

2.2.2.2 Prepositions (primary and secondary)

Those secondary prepositions that, for the time being, failed to be represented at the analytic level will have to be treated later, together with the phrasemes, and it may be necessary to supply further functors and syntactic grammatemes. Those which have been already covered are converted to a form similar to primary prepositions.

In the position of a syntactic grammateme the preposition is written in lower case letters.

během 'during'	TPAR;
bez 'without'	ACMP.WOUT: *bez brýlí* 'without spectacles', *bez toho, že by (aby)* 'without -ing';
	ATT: *bez obalu* 'without cover';
bez ohledu na 'regardless'	REG.WOUT;
bez zřetele k 'irrespective of'	REG.WOUT;
blízko	p: LOC.*blízko*: *Jsou blízko nás* 'they are near us';
	DIR3.*blízko*: *Položil to blízko vázy* 'he laid it near the vase';
	DIR2.*blízko*: *Prošli blízko Chrudimi* 'they passed near Chrudim', *cestou* 'by way of';
	MEANS: *cestou experimentů* 'by way of experiments';
co do 'as to/as for'	REG.NIL: *co do výsledku* 'as for the result'
díky 'thanks to'	CAUS: *díky pomoci* 'thanks to the help'
dle	see *podle* 'according to'
do	p: DIR3.*v*
	TTILL (*až*) *do pádu Říma* 'till the fall of Rome';
	EXT *do posledního místa* 'to the last place', *do krve* 'till one bleeds', *sytosti* 'to one's heart's content', *do krajnosti* 'to the extreme', *účtovat výlohy do výše*.EXT *čeho*.APP 'to bill expenses up to the level of';
	EFF *roztáhnout se do elipsy* 'to dilate into an ellipse', *vyrůst do velkých rozměrů* 'to grow to large dimensions'
	REG.NIL *Sbor je do počtu slabý* 'the corps is weak in number';
	RESL *Opálená do hněda* 'sunburnt brown';
	PAT *Zamilovat_se* 'fall in love', *začíst_se* 'get buried in a book', *dát_se* 'set about';
doprostřed čeho 'in(to) the middle of'	DIR3.*uprostřed*;
dovnitř čeho 'into'	DIR3.*v*;
k 'to'	p: DIR3.*u*;
	LOC.*u ke konci stránky* 'towards the end of the page';
	TWHEN.AFT *k večeru* 'towards the evening', *k 15. říjnu* 'towards the 15th of October';

	EXT.APPX *Bylo nás ke stu* 'there was towards one hundred of us';
	EXT.NIL *Zmučená k nepoznání* 'devastated beyond recognition'
	PAT *k jízdě nezpůsobilý* 'unfit for driving', *povolat/ chystat se k boji ´* call up s.o./be ready to fight´, *náklonnost* 'propensity to(wards)', *odpor* 'aversion to', *vůle k* 'will to'
	AIM *voda k pití* 'water for drinking', *probouzí* 'rouses'/*podněcuje* 'instigates'/*přispívá* 'contributes'/*volá* 'calls' *k odpovědnosti* 'to responsibility';
	ATT *k mé radosti* 'to my pleasure', *k lítosti* 'to regret', *k plné spokojenosti* 'to full satisfaction';
	REG *právo k nemovitostem* 'right to immovebles'
kolem	p: LOC.*kolem Jsou všude kolem nás* 'They are all over around us.'
	DIR2.k*olem Prošli kolem nás* 'They passed by us';
	TWHEN.APPX *kolem poledne* 'around noon';
	ACT.APPX *přišlo kolem třiceti lidí* 'about thirty people came';
	PAT.APPX *přinesli kolem dvaceti lahví* 'They have brought some/about twenty bottles', *váží to* 'it weighs', *stojí to kolem* 'costs about'; *bylo jich*.ACT *kolem padesáti*.PAT.APPX 'There were about fifty of them';
kromě	RESTR *všichni kromě tebe* 'all but you';
kvůli	CAUS;
mezi co	DIR3.*mezi*1 *Dal to mezi ty drobnosti* 'he put it among the trinkets';
	DIR3.*mezi*2 *Dal to mezi stůl a gauč* 'he put it between the table and the settee';
mezi čím	LOC.*mezi*1 *byli mezi sebou* 'they were amongst themselves'
	DIR2.*mezi*1 *prošel mezi lidmi* 'he passed among the people'
	LOC.*mezi*2 *byl mezi dvěma ohni* 'he was between two fires'
	DIR2.*mezi*2 *prošel mezi Scyllou a Charybdou* 'he passed between Scylla and Charybdis'
	TWHEN.INTV *stalo se to mezi pátou a šestou* 'it happened between five and six (o´clock)';
mimo	p: LOC.*mimo působí mimo Prahu* 'he operates outside P.'
	DIR2.*mimo prošel mimo tebe* 'he passed by you';
	RESTR *všichni mimo tebe* 'all but you';
místo	SUBST *místo tří přišli jen dva* 'in place of three only two have arrived';
na co	p: DIR3.*v na Spořilov* 'to S.', *na schůzi* 'to the meeting', *orientovat/zaměřit koho*.PAT *nač* 'direct s.o. to st.', *orientovat_se*2 *na co* 'to orientate o.s. to st.';
	DIR3 *na dal to na stůl* 'he put it on the table'
	TWHEN *na Štědrý večer* 'on Christmas Eve'
	TFHL *přijel na tři dni* 'he came for three days', *práce na rok* 'work for a year';
	MANN *zavřít na klíč* 'close with a key', *nabarvit na růžovo* 'paint pink';
	PAT *zapomenout nač* 'forget s.t.', *čekat* 'wait for', *ptát se* 'ask for', *odpovědět* 'reply', *chystat_se nač* 'be about to', *návrh* 'proposal', *reakce na* 'reaction to', *to je na rozvod* 'that amounts to a divorce';
	EFF *rozsypat se na prach* 'disintegrate into powder', *rozbít na kousky* 'break into pieces', *učesat vlasy na drdol* 'to comb one's hair into a knot';
	RSTR *dům na spadnutí* 'a tumbledown house', *opravy na počkání* 'repairs while you wait';
	ACT.APPX *bylo nás na sto* 'we were about a hundred'
	EXT.APPX *přivedli jich na dvě stě* 'they have fetched about two hundred of

	them';
	EXT.NIL *zaplatit na halíř* 'pay to the last penny';
	INTT *šel na jahody* 'he went picking strawberries'*, na nákup* 'shopping'*, na procházku* 'for a walk'*, na trávu* 'mowing grass', *zůstat na večeři* 'stay for dinner';
	REG *úrodný na ovoce* 'productive of fruits'*, na počet málo* 'few in number'*, štědrý na sliby* 'lavish with promises', *zemřít na* 'die of'*, zpívat na povel* 'sing at command'*, otevřít na žádost* 'open at request', *na zaklepání* 'at a knock'*, na návrh* 'at the suggestion'*, na popud* 'at an impulse';
	AIM *kapky na trávení* 'drops for digestion';
	TOWH *přeložil výuku na pátek* 'he transferred the classes to Friday'; *demonstrace je svolána na šestou hodinu* 'the demonstration has been called together for six o'clock';
	TWHEN.NIL *na Štědrý den* 'on Xmas Day';
na čem	p: LOC.*na leží na stole* 'it lies on the table'
	LOC.*v na Spořilově* 'in Spořilov'*, na schůzi* 'at the meeting';
	TWHEN.NIL *na úsvitě* 'at daybreak', *na konci* 'at the end';
	MEANS *jet na žebřiňáku* 'to be transported on a rack waggon', *objasnit na příkladu* 'to explain by an example'
	REG.NIL *prospívat na duchu* 'prosper on one's mind', *přibývalo mu na lačnosti* 'his greed increased', *ochromnout na všech údech* 'become paralysed in all limbs', *pomatený na smyslech* 'deranged in his senses', *klesat na mysli* 'lose heart', *získat na vážnosti* 'gain in respect';
	PAT *lpět* 'stick to'*, záviset* 'depend on'*, trvat na názoru* 'insist on an opinion';
	INTT *byl na jahodách* 'he was picking strawberries'*, na kurzu* 'in a course of lectures'*, na prázdninách* 'on holiday' (LOC can stand along with INTT, they cannot be coordinated: *byla na prázdninách u babičky* 'she spent her holidays at her grandmother's';
nad co 'above'	p. DIR3.*nad*
	CPR.DFR *nad slunce jasnější* 'brighter than sun' (lit. 'above sun brighter', with comparative);
	EXT.MORE *nad jiné milý člověk* 'a gentleman nicer than any others';
nad čím	p.LOC.*nad* 'over, above';
	DIR2.*nad Trať se vine nad řekou* 'the railway line winds over the river';
	CAUS *naříkat* 'lament for'*, zoufat nad něčím* 'despair of', *zaradovat se nad* 'rejoice over';
	PAT *vládnout* 'reign over'*, vítězit* 'prevail over'*, zamyslet se* 'spare a thought for';
naproti	LOC.*naproti* 'opposite' *stojí naproti nádraží* 'stands opposite the railway station';
	DIR3.*naproti dej to naproti* 'put it to the opposite side';
	DIR2.*naproti cesta se vine naproti svahu* 'the path is winding opposite the slope';
napříč čeho	DIR2.*nad* 'across';
na rozdíl od	CPR.DFR 'in contradistinction to';
následkem čeho	CAUS 'as a result of';
na úkor čeho	BEN.AGST 'to the detriment of';
na vrub koho	BEN.AGST 'to someone´s debit';

navzdory čemu	CNCS 'in spite of';
na základě	NORM: *na základě dohody* 'on the basis of an agreement';
na způsob	CPR 'in the manner/fashion of';
nedaleko čeho	see *blízko* 'not far from';
nehledě k	CNCS 'regardless of';
nezávisle na čem	REG.WOUT 'independently of';
o co	p.DIFF *větší o* 'taller by'; *opírat_se oč* PAT 'lean on';
	MANN *běžet o závod* 'run for all one is worth';
	MEANS *brousit nůž o kámen* 'to grind a knife on a stone'*, zazvonit lžičkou o sklenici* 'to ring a spoon against a glass', *mrštit talířem o zem* 'to fling a plate on the ground'*, spálit se o kamna* 'to burn o.s. on the stove';
o čem	TWHEN.NIL *o Velikonocích* 'at Easter';
	PAT *mluvit o* 'talk about'*, dohadovat se o* 'argue about'; *říkat o* 'to say about' (*že…*EFF 'that…')*, vypravovat (psát) o kom* (PAT) *povídku* (EFF) 'to narrate (write) a short story about s.o.', *říkat o kom, že…* (EFF) 'say about s.o. that…' *mluvili o tom, že pršelo*.PAT Lit. '(they) narrated about (it) that (it) rained';
	MANN *jít o holi* 'walk with a stick', *žít o bramborách* 'live on potatoes';
	RSTR *báseň o čtyřech slokách* 'a poem in four stanzas'*, kružnice o poloměru* 'a circle with/having radius of';
od	p: DIR1.*u odvrátit* 'avert from'*, upustit* 'refrain from'*, ustoupit od* 'retreat/recede from';
	TSIN *od soboty* 'since Saturday';
	MANN *od oka* 'by rule of thumb', *dům od domu* 'from house to house';
	AIM *od toho jsou jiní* 'for that purpose others are competent';
	CAUS *dát něco od cesty* 'give s.o. st. for the trouble', *ruka bolí od…* 'my hand hurts from…';
	ORIG *odvozovat původ od…* 'derive one's descent from', *očistit* 'clear of', *rozlišit* 'differentiate st. from', *rozeznat od…*'distinguish from', *pozdrav od* 'greetings from';
	ACT with passive voice, resultative aspect and nouns of action – *namalováno od* 'painted by'*, malba od* 'a painting by'*, má uvařeno od mámy* 'he has got his cooking done by mother';
	TFRWH *zbylo od Vánoc cukroví* 'there are some sweets left from X-mas'*, cukroví od Vánoc* 'sweets from X-mas';
od ... do	DISJ with lemma "*od_do*": *od hlavních bodů do nejmenších detailů* 'from the main points to the slightest details';
od ... po	DISJ with lemma "*od_do*" *od hlavních bodů po nejmenší detaily* 'from the main points down to the slightest details';
od ... přes ... do	DISJ with lemma "*od_přes_do*" *od hlavních bodů přes příklady do nejmenších detailů* 'from the main points through examples to the minutest details';
ohledně čeho	REG.NIL 'as regards';
okolo	p: LOC.*kolem Jsou všude okolo nás* 'they are all over around us';
	DIR2.*kolem chodit okolo domu* 'walk around the house';
	TWHEN.APPX *okolo druhé hodiny* 'about two o'clock';
	EXT.APPX *okolo dvou set* 'about two hundred';
oproti	see *proti*;
po co	DIR3.*u voda sahá po kolena* 'the water reaches up to the knees', *bořit se po* 'to

sink in up to';

TPAR *po tu chvíli* 'for the while'

TTILL *až po tu chvíli* 'until then';

LOC.*na po mnoho kilometrů* 'for many kilometres';

EXT *zamilovat se po uši* 'to fall in love head over ears';

po čem — p: TWHEN.BEF *po jeho příchodu* 'after his arrival';

TWHEN.NIL *po večerech* 'in the evenings', *po druhé* 'at the second time'

THO.BEF *po každém dešti* 'after every rain';

LOC.*na viset po boku* 'hang at one's side', *kantorovat po mnoha městech* 'schoolmastering about many towns';

ACT.DISTR *Každou hodinu přišlo po dvou koledníc*ích 'each hour carolsingers kept arriving two at a time';

PAT.DISTR *dal dětem po dvou jablíčkách* 'he gave the children two apples each';

PAT.NIL *touha* 'desire for', *žízeň* 'thirst for', *pátrat* 'search for', *dychtit* 'yearn for', *ohlížet_se po* 'look back at';

DIR2.*přes jezdí po Čechách* 'he is travelling about Bohemia';

DIR3.*na házet po kom kamením* 'throw stones at s.o.';

MANN *po špičkách* 'on tiptoe', *po dobrém* 'amicably', *po hlavě* 'headfirst', *po očku* 'with furtive glances at', *po venkovsku ustrojená* 'dressed in rustic fashion', *po domácku* 'domestically', *po anglicku* 'English-like', *po vojensku* 'in a military way';

ORIG *tvář poďobaná po neštovicích* 'face covered with pockmarks', *jmenovat se po* 'be named after', *jméno po* 'name after', *peníze po manželovi* 'money left by her husband', *vdova po řediteli* 'a director's widow';

HER *jmění po otci* 'a fortune inherited from father', *dědit po* 'inherit from';

ATT *po pravdě* 'according to facts';

poblíž — see *blízko*;

pod co — p: DIR3.*pod* 'under';

EXT.NIL *pod obraz* 'drunk as a lord', *pod psa* 'bloody awful' (phrasemes);

EXT.LESS *pod padesát* 'under fifty'

pod čím — p: LOC.*pod*;

DIR2 *tunel vede pod…* 'the tunnel leads under…';

MANN *prodat pod rukou* 'sell underhand', *pod trestem* 'under punishment';

podél — see *podle*

podle — p: NORM *podle pravidel* 'observing the rules', *pravítka* 'agreeing with a ruler', *rady* 'as advised', *jména* 'as to the name', *barvy pleti* 'according to the colour of one's skin', *módy* 'in keeping with the fashion', *abecedy* 'alphabetically' (i.e., in accordance with, with the aid of);

CRIT *podle něho* 'according to him', *podle jeho názoru* 'in his opinion', *poznat* 'recognize by', *posuzovat* 'judge by', *rozlišovat podle obleku* 'distinguish by one's dress', *podle všeho* 'by all appearances' (i.e., according to, origins of the opinion);

COND *podle možnosti* 'if possible';

LOC.*kolem podle/podél potoka leží vesnice* 'a village is situated along the stream';

DIR2.*kolem podle/podél potoka vede cesta* 'a road leads along the stream';

po dobu čeho	THL 'for the time of';
pomocí	MEANS 'with the aid of';
po stránce	REG.NIL *po stránce právních předpisů* 'in respect of legal regulations';
postupem	TWHEN.NIL *postupem času* 'in the course of time';
po vzoru	NORM *po vzoru Norů* 'after the fashion of Norwegians';
pro	p: with INAN noun: AIM - *natáhnout se pro* 'reach out for', *byl pro smích* 'he was a laughing-stock', but also *jít pro lékaře* 'to fetch a physician';
	p: with ANIM (also with names of institutions, collectives, etc., sometimes with abstracts: "for the benefit of"): BEN.NIL *udělat co pro koho* 'do st. for s.o.';
	TFHL *pro tu chvíli* 'for that moment';
	REG *pro koně má oči* 'he has a (good) eye for horses';
	CAUS *zatknout pro* 'to arrest for', *pro co zapomeno nač* 'to forget st. for st.';
pro případ čeho	AIM *pro případ potřeby* 'for a case of need';
prostřed čeho	see *uprostřed* 'in the middle';
prostřednictvím	MEANS 'through the mediation of';
proti	p: BEN.AGST *mluvit proti* 'to speak against';
	PAT *protestovat* 'to protest', *bouřit se* 'revolt', *vystupovat* 'stand against';
	LOC.*naproti leží proti oknu* '(it) is situated opposite the window';
	DIR2.*naproti projít proti radnici* 'walk opposite the town hall';
	DIR3.*naproti postavit proti oknu* 'to place st. opposite the window';
	CPR.DFR *proti tobě je starý* 'as compared with you he is old';
	REG.NIL *imunní proti* 'immune to';
	CNCS *proti očekávání* 'contrary to expectations', *proti všem zásadám* 'against all principles';
průběhem čeho	TPAR *průběhem času (léta) nastaly změny* 'in the course of time (summer) changes have set in';
před co	p: DIR3.*před* 'to the front of';
před čím	p: LOC.*před tykat mu před lidmi* 'to be on intimate terms in public';
	PAT *varovat* 'warn of', *prchat* 'flee from', *ostýchat se* 'be shy in s.o.'s presence';
	DIR2.*před projít přede dveřmi* 'to pass in front of the door';
	TWHEN.AFT *před hodinou* 'an hour ago';
	THO.AFT *před každým deštěm* 'before any rain';
přes	p: DIR2.*v* 'over';
	TPAR *přes celé léto se tam nedostal* 'all of the summer he failed to get there';
	CPR.DFR *přes dvacet* 'over twenty';
	CNCS *přes své dobré vychování* 'in spite of his good breeding', *přese všechno* 'for all the';
	LOC.*přes bydlí přes dvůr* 'he lives across the yard';
přes	in the configuration *od … přes … do* 'from … over … to', see the preposition *od*;
při	p: COND *při nedostatku* 'under the lack of';
	LOC.*u při dveřích* 'at the door';
	CNCS *při vší opatrnosti* 'for all the caution';
	TWHEN.NIL *při práci* 'at work', *při nemoci* 'under illness';
přihlížeje k	REG.NIL 'paying regard to';

přiměřeně čemu	NORM 'adequately to';
při příležitosti	THO.NIL 'at the occasion';
s čeho	DIR1.na 'down from';
s čím	p: ACMP *chodit s klukem* 'go out with a boy', *chodit s holí* 'walk with a stick', *s úctou* 'with respect', *s povzdechem* 'with a sigh', *se zdravou kůží* '(get off) unscathed',
	ADDR *dohodnout se s kým*.ADDR *na čem*.PAT 'to settle with s.o. on st.', *souhlasit s* 'agree with';
	PAT with the reflexive verbs not having any other PAT (*seznámit_se* 'make oneself acquainted with') see Sect. 2.6.B; *jednat*2 *s kým*.PAT *o čem*.EFF 'negotiate with s.o. about st.', *jednat*3 *s kým*.PAT *jak*.MANN 'to treat s.o. somehow'; also: *obchod s něčím*.PAT 'a shop with';
	MEANS *dělat s lopatou* 'work with a shovel';
	REG *s tím neměl štěstí* 'he was out of luck with it';
	CPR *s léty přibývají zkušenosti* 'with years the experience keeps accumulating';
	ATT *s radostí* 'with pleasure', *s odporem* 'with disgust', *s úžasem* 'with astonishment';
se zřetelem k, na	REG.NIL 'with a view to';
skrz co	DIR2.uvnitř 'through what';
s ohledem k, na	REG.NIL 'with respect to';
soudě podle čeho	CRIT *soudě podle táty* 'judging by my father('s opinion)', *podle šatů* 'by the clothes';
stran	REG.NIL *jednat stran dodávek* 'negotiate re deliveries';
s výjimkou čeho	RESTR 'with the exception of';
u čeho	p: LOC.*u*;
	COND *u vědomí čeho* 'being aware of';
úměrně k čemu	NORM ´proportionately to´;
uprostřed	p: LOC.uprostřed 'in the middle of' *uprostřed náměstí* 'in the centre of the square';
	TWHEN.*uprostřed uprostřed léta* 'in the middle of summer';
u příležitosti	TWHEN 'at the occasion';
uvnitř čeho	LOC.*uvnitř* 'inside';
v co	TWHEN.NIL *v neděli* 'on Sunday';
	DIR3.*v uvést v pohyb* 'set in motion';
	EFF *změnit v* 'change into', *vyčesat vlasy v drdol* 'comb the hair up into a knot', *rozdělit co ve skupiny* 'to divide st. into groups';
	PAT *věřit ve vítězství* 'to believe in victory', *propuknout v pláč* 'to break into tears';
v čele čeho	ACMP *v čele společnosti* 'at the head of the company';
v čele s	ACMP *v čele s předsedou* 'being led by the chairperson';
v čem	p: LOC.*v orientovat_se1* 'to get orientation in'
	TWHEN.NIL *v minulém století* 'in the last century';
	MANN *žít v souladu* 'live in harmony', *nadávat ve všech řečech* 'swear in all languages', *účtovat výlohy ve výši* 'bill expenses in the level of';
	COND *v tom případě* 'in that case';
	REG *písmo v své úhlednosti vynikající* 'a script outstanding in its neatness';

	PAT *utvrzovat koho* (ADDR) *v čem* (PAT) 'to confirm s.o. in st.', *vynikat v čem* 'to excel in st.';
	ATT *ve skutečnosti* 'in reality';
včetně čeho	ACMP.NIL 'including';
v duchu čeho	NORM 'in the spirit of';
v důsledku čeho	CAUS 'as a result of';
vedle	p: LOC.*vedle* 'by the side of';
	DIR3.*vedle položit co vedle čeho* 'to put st. beside st.';
	DIR2.*vedle vedle našeho domu vede cesta* 'along our house there is a path', *leží vedle mě* 'she lies beside me';
	RESTR *vedle historie studovala češtinu* 'besides history she was studying Czech', *vedle nenávisti byly i sympatie* 'along with hatred there were sympathies as well';
ve prospěch čeho	BEN.NIL 'to the benefit of';
ve shodě s	NORM 'in keeping with';
ve smyslu čeho	NORM 'in accordance with';
koho	CRIT 'according to';
ve srovnání s	CPR.NIL 'in comparison with';
ve světle	NORM 'in the light of';
ve věci čeho	REG.NIL 'on the issue of';
vinou koho, čeho	CAUS 'owing/due to';
vlivem	CAUS 'under the influence of';
vně čeho	LOC.*mimo* 'outside';
v období čeho	TWHEN.NIL 'in the period of';
v oblasti čeho	LOC.*v* 'in the sphere of';
v oboru čeho	LOC.*v* 'in the field of';
v osobě koho	MEANS 'in the person of';
v podobě čeho	MANN 'in the form of';
v poměru k	REG.NIL 'in relation to';
v porovnání s	CPR.NIL 'in comparison with';
vprostřed	see *uprostřed* 'in the middle of';
v procesu čeho	TWHEN.NIL 'in the course of';
v protikladu k	CPR.AGST 'in contrast with';
v průběhu čeho	WHEN.NIL *stalo se to v průběhu cesty* 'it happened in the course of the journey';
	TPAR ('during', i.e. in the whole course) *dělo se to v průběhu cesty* 'it was going on in the course of the journey';
v rozporu s	BEN.AGST 'at variance with';
v souhlase s	NORM 'in agreement with';
v souladu s	NORM 'in harmony with';
v souvislosti s	ACMP.NIL 'in connection with';
vůči komu, čemu	REG.NIL 'in the face of';
vyjma co, čeho	RESTR 'except for';
v zájmu čeho	AIM 'in the interests of st.';
v zájmu koho	BEN.NIL 'in a person's interests';

v zastoupení	SUBS 'on behalf of';
v závislosti na	CRIT 'in dependence on';
vzdor čemu	CNCS 'in spite of';
vzhledem k	REG.NIL 'with regard';
z	p: DIR1.*v žít z práce* 'make a living from work'*, mluvit ze zkušenosti* 'speak from experience'*, probudit se ze spánku* 'to wake up from a sleep', *obvinit* 'blame for'*, zkoušet* 'examine in'*, usvědčit* 'convict';
	DIR1.*na z povrchu* 'from the surface', *ze zdi* 'from the wall';
	DIR1.*v jeden z chlapců* 'one of the boys', *včera zvolili tři členy výboru z pěti, příští týden zvolí zbývající dva* 'yesterday they elected three members of the committee out of five, next week they will elect the other two'
	DIR1.*mezi*1 *nejlepší z lidí* 'the best among the people', *zvolili pět členů výboru z deseti kandidátů* 'they elected five members of the committee from ten candidates'
	PAT - *sestávat z* 'to consist of', *skládat se z* 'to be composed of';
	ORIG *vyrobit* 'produce from';
	TFRWH *z pátku* 'from Friday'*, z dětství* 'from childhood';
	CAUS *ze sympatie* 'out of sympathy'*, z přinucení* 'under constraint'*, strach z* 'fear of';
	MANN *ze všech sil* 'with all one's might'*, z doslechu* 'from hearsay';
za co	p: DIR3.za *položit*/*dát za skříň* 'place st. behind the cupboard';
	CAUS *za zásluhy* 'for merits', *trest za* 'punishment for';
	MANN *vzít za kliku* 'press the handle', *tahat za vlasy* 'pull by the hair';
	PAT *přimlouvat se za* 'speak up for s.o.'*, trpět* 'suffer for', *vyznamenat koho* (ADDR) *zač* 'decorate s.o. for', *mstít se komu* (ADDR) *zač* 'have one's revenge on s.o. for st.'; *cena za* 'price for';
	SUBST *dát* 'give', *prodat za* 'sell for', *za otce jednal strýc* 'it was the uncle who acted for father', *platit zač* 'pay for', *letenka za* 'air ticket for';
	EFF *zvolit koho* (PAT) *za* 'to elect s.o. as', *vydávat* 'pass s.o. off for', *pokládat* 'consider s.o.´, *považovat co za* 'regard st. as', *vyměnit co za* 'exchange st.', *uložit za úkol* 'to charge s.o. with a task';
	COMPL *jít za svědka* 'to come as a witness';
	TWHEN.BEF *přijde za týden* 'he will come in a week', *bude tam za hodinu* 'he will be there in an hour';
	THL *napsal to za dvě hodiny* 'he wrote it down in two hours' (only with perfective verbs);
	TPAR *za celou hodinu* 'for the whole hour';
za čeho	TPAR *za deště* 'in the rain' (with imperfective verbs);
	TWHEN *psal*/*napsal to za minulého ředitele* 'he was writing/he has written it down under the last director', *za krále Jiřího* 'in the reign of king George';
za čím	LOC.*za ležet za* 'to lie behind';
	DIR2.*za procházet za plotem* 'to walk behind the fence', *chodit za děvčetem* 'to court a girl';
	DIR3.*za odjet za manželem* 'to leave to meet her husband', *dívat se za* 'to look behind';
za příčinou	CAUS 'because of';
zásluhou	CAUS 'thanks to';

za účelem	AIM 'with the aim of';
ze stanoviska	REG.NIL 'from the standpoint of';
ze strany	ACT with passive: *ze strany dodavatele bylo účtováno* 'from the part of the supplier it was charged';
	ORIG with active voice: *ze strany dodavatele jsme dostali* 'from the part of the supplier we obtained';
z důvodu	CAUS 'for the reason of';
z hlediska	REG.NIL 'from the point of view';
znad	DIR1.*nad* 'from above';
zpod	DIR1.*pod* 'from below';
zpoza	DIR1.*za* 'from behind';
zpřed	DIR1.*před* 'from the front of'.

2.2.2.3 Subordinating conjunctions and connecting expressions

Let us remind that *ať* 'let', *co* 'what', *kdy* 'when' do not always represent conjunctions. Neither *protože, když, jenže* etc. are regarded as conjunctions if they introduce a whole sentence, i.e., if they do not connect two clauses; in such cases they are taken as nodes which depend on the verb, with the functor PREC.

We have included also conjunctions often considered as subordinating, which, however, we conceive as an expression of the functor CSQ (since they cannot stand at the beginning of a complex sentence - which can be regarded as the most reliable criterion of coordination).

a_proto, a_tak, a_tedy	CSQ (consequence, functor for coordination)
aby	p. AIM *šel tam, aby* 'he went there to';
	PAT *její snaha, aby se líbila* 'her endeavour to to please the eyes', *chtít, aby* 'to want to', *možnost, aby* 'possibility of';
	ACT *je směšné, aby* 'it is ridiculous to', *je možné, aby* 'it is possible to';
	COMPL only with negation: *Sotva kdy kolem toho jde, aby to neviděl* 'he hardly ever passes by it without seeing it'; this holds in cases where *aby ne = aniž* 'without -ing' (i.e., if it is not a purpose);
ač(koli/v)	CNCS 'although';
aniž	COMPL and Neg with the verb; *šel, aniž (by to) viděl = šel, přičemž neviděl* 'he went without seeing it = he went while not seeing'
ať	p.CNCS *ať jsou sebemenší, jsou dobré* 'however small they may be, they are good';
	PAT *dej pozor, ať tě neuslyší* 'mind out that they do not hear you', *řekni, ať jde pryč* 'tell him to go away';
ať_ať, ať_nebo	DISJ *ať mladí nebo staří* 'be they young or old';
až	TWHEN.NIL with imperfective verbs: *až tu poroste tráva* 'by the time grass will be growing here';
	TWHEN.BEF with perfective verbs: *až vyjedeme* 'after we have set out';
	RESL *bylo tam plno, až omdlévali* 'it was so full there that they began fainting';
	DISJ *tři až čtyři* 'three, up to four';
až_na	RESTR 'but for';
byť	CNCS 'even though';
co	TSIN 'what';

	in relative clauses with the 'absolute' *co* 'what', if the clause contains a coreferential pronoun, *co* is treated like a conjunction, i.e. with no corresponding node in the TGTS (*žena, co jsem ji potkal* 'the woman whom I met', *Schránku, co jsem do ní vhodil dopis, už vybrali* 'They have already emptied the box into which I threw my letter'); without such a corefential pronoun, *co* itself is understood as a pronoun, i.e. with a corresponding node in the TGTS (as a case of grammatical coreference); this pronoun is in Nominative or in Accusative (*Žena, co potkal na schodech, má krátkou sukni* 'The woman he met on the staircase, has a short skirt', *Člověk, co jde po chodníku, je můj přítel* 'The man that walks on the pavement is my friend'); the coreferential pronoun can occur also with Acc: *Ten kluk, co ho Jirka potkal* 'that boy that George met' (with *on* the value of COREF:*kluk* will be supplied):
	`kluk`
	`  \`
	`  potkal.RSTR`
	`  /  \`
	`on.PAT  Jirka.ACT`

co nej-	EXT 'as … as possible'; *co největší* 'as big as possible'

což	'which' - also with indirect cases, as well as with a preposition: Comma.CONJ will join both the clauses and *což* will obtain a functor according to its function in the second of them
	dali mu bonbon, bez čehož by dál nešel 'they gave him a chocolate, without which he would not have come in'
	`Comma.CONJ`
	`/  \`
	`dali  nešel`
	`  /`
	`což.ACMP.WOUT`

dokud ne	TTILL (following that functor no negation is introduced in the subordinate clause) *nepůjdu, dokud neslíbíš* 'I won't go until you promise';
dokud	THL 'as long as';
dřív(e) než	TWHEN.AFT 'before';
hned jak/když	TWHEN.JBEF 'at the moment (that)';
i kdyby	CNCS 'even if';
i když	CNCS 'even when';
jak	p. MANN 'as' *píše, jak mluví* 'he writes as he speaks';
	PAT *radit se, jak* 'to confer on how';
	COND *jak nepoví, budu…* 'if s/he doesn´t tell, I shall…';
	COMPL *viděl tě, jak jdeš* 'he saw you walking';
	PAR *jak se dá předpokládat* 'as can be presumed', *jak lze soudit* 'as can be concluded';
jak	is also an adverb of MANN, 'how';
jakmile	TWHEN.JBEF 'as soon as';
jako (by)	CPR.NIL 'as (if)';
jakožto	COMPL 'as';
jelikož	CAUS 'inasmuch, since';
jen co	TWHEN.JBEF 'as soon as';
jestli(že)	COND 'if';
ježto	CAUS 'since';
kdežto	CONFE 'while';

kdy	TWHEN.NIL 'when(ever)' *kdy chtěl* 'any time he wanted', *kdy se mu zachtělo* 'whenever he felt like';
kdyby	COND 'if';
kdykoli(v)	THO.NIL 'whenever', provided the nominalized form of the clasue is *(vždy) při* 'any time at';
	THO.BEF if the nominalized form corresponds to *(vždy) po* 'any time after';
když	TWHEN.NIL 'when';
	TWHEN.BEF see the comment on *kdykoli*;
ledaže	RESTR 'unless';
-li	COND 'if';
načež	TWHEN.BEF
než	p. TWHEN.AFT 'before';
	TTILL
	CPR.DFR *jiný* 'other than', *jinak* 'otherwise', *větší než já* 'taller than I';
	RHEM *nezbývá* (*nic*.ACT.ELID), *než*.RHEM *odejít*.RESTR Lit. 'remains (nothing.ACT.RHEM) than.RHEM (to) leave.RESTR';
nežli = než 'than';	
pokaždé (po každé), když	THO.NIL or BEF 'any time when';
pokud	p. COND 'if';
	THL - can be recognized by the meaning: 'as long as';
	RESTR *(jen) pokud* 'only as long as';
	REG *pokud jde o* 'as far as sth. is concerned';
poněvadž	CAUS 'because';
pročež	CSQ 'thus';
protože	CAUS 'because';
před tím (předtím), než	TWHEN.AFT 'before when', see *než*; a secondary conjunction: *předtím_než*;
přestože	CNCS 'in spite of';
přičemž	CONJ (as distinct from *což* it doesn't have any other node than a conjunction);
sotva (že)	TWHEN.JBEF 'as soon as';
takže	CSQ 'so that' classed with coordination, since it can stand between clauses only, never at the beginning of a compound sentence;
třeba(že)	CNCS '(al)though';
zatímco	TPAR 'while';
	CONFR *zatímco Bristol je v Anglii, Glasgow je ve Skotsku* 'while B. is in England, G. is in Scotland';
zda(li)	PAT 'whether' *otázka, zda* 'a question, whether';
	ACT zda přijdou.ACT všichni, nebylo jasné.PAT 'whether all would come was not clear'
že	p. PAT 'that' *řekl, že* 'he said that', *výčitky* 'scruples about', *námitka* 'objection that', *poznámka* 'remark that', *tvrzení* 'statement that';
	ACT *zdá se, že* 'it seems that';
	MANN *dělat tak, že* 'work so, that';
	RESL *tolik, že* 'so much that', *mám ruce zmrzlé (tak), že prsty nenatáhnu* 'my hands are frozen up (so that) I cannot stretch my fingers'.

2.2.2.4 Agreement with the noun

With a congruent ajective (or relative clause) it must be decided (mostly on the basis of semantics) also in the LC whether the case is

p. **RSTR** a typical attribute, such as *velký dům* 'large house', *lidé, kteří mají podobné názory* 'people who have similar opinions', *tři kluci* 'three boys', *(byly tam černé a modré šaty) vybrala si (ty) modré*.RSTR '(there was a black and a blue dress there), she chose the blue one' (see 2.5); *ten, který spí*.RSTR 'the one that is sleeping' (with the verb depending on *ten* ´the one´)

criterion: should the attribute be omitted, the scope of the semantic validity of the sentence would change - not every house is large, not all people have similar opinions (but all differ from animals, see below on DES)

DES (descriptive, i.e., non-restrictive attribute) - *zlatá Praha* 'Golden Prague', *stověžatá Praha* 'hundred-towered P.', *lidé, kteří se liší od zvířat* 'people, who differ from animals'

> Note: The functor DES is assigned to all pseudo-relative clauses (Cz. 'nepravé'); for the time being they are regarded as "genuine" relative clauses even in the MC

APP (appurtenance) - *můj sešit* 'my copybook'*, mámina sukně* 'mother's skirt'

COMPL (predicative complement) - *přišel bos* 'he arrived barefooted', *přivezli je zraněné* 'they have brought them wounded'

PAT (with copula: *byl nemocný* 'he was ill'*).*

Agreement between **two nouns:**

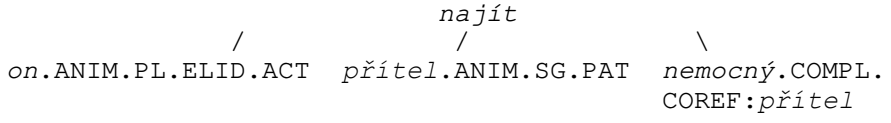
In the LC as well as in the MC:

The type *předseda Zeman* 'chairman Z.' (i.e. a proper noun preceded by another noun agreeing in number and case, while one of them depends on the other in ATS), similarly as e.g. *Miloš Zeman*, is represented as mother node *Zeman* and daughters *předseda*, *Miloš* (with the functor RSTR); other pairs of nouns in a similar way: *pan*.RSTR *starosta* 'Mr. Mayor', *řeka*.RSTR *Vltava* 'river V.'; a different treatment concerns an indeclinable nominative: *stanice Vltava*.ID 'Broadcasting station V.', also without agreement in number: *město Litoměřice*.ID 'town L.', *okres Litoměřice*.ID 'district L.', *vesnice Petrovice*.ID 'village P.', *jméno Petrovice*. ID 'name P.'*;* with a reverse word-order: mother - daughter: *Máme Zemana předsedu*.RSTR *a Zemana ministra*.RSTR 'we have Z. the chairman and Z. the minister'.

> Note: Sometimes, with inanimate nouns, the boundaries between RSTR and ID are not clear; if disagreement in case is possible, e.g., *parník* (*-u*) *Hradčany* 'steamboat H., of the steamboat H.', *pojem shody* 'concept of agreement', the daughter (placed to the right) will have the functor ID; therefore also *parník Zvíkov*.ID 'steamboat Z.'*, pojem shoda*.ID 'concept agreement', *řeka*.RSTR *Vltava* 'the river V.', but: e.g., *řekou Ohio*.ID '(down) the river (Instrumental) Ohio' (Nominative). Further provisional criteria are: with animate nouns (functions, etc.) RSTR stands to the left, with inanimate nouns (concepts, foreign rivers, stations named after sth., etc.) ID stands to the right - indeclinable nominative.

Complement (predicative complement)**:**

In the LC as well as in the MC the functor COMPL dependent on the verb and gets in its attribute COREF the lemma of the second "governing" word (from which Atv 'depends' in the analytic tree): *Našli přítele nemocného*.COMPL.COREF:přítel 'they have found (their) friend ill'; in the MC also CORNUM is filled: it obtains the value ORD (the serial number of the antecedent: with the numbers following the decimal point if the antecedent is a restored deleted node) and so is ANTEC, which obtains the functor of the antecedent.



The relation of a complement to the item with which it stands in agreement can be regarded as one kind of grammatical coreference.

2.3 Lexical parts of the tags

(a) Gen(eral participant)

The contents of Sect. 2.3 hold for the LC as well as the MC, unless with individual points an explicit statement to the contrary is found.

Even in the LC it is necessary to distinguish general participant from the absence of the participant, although with all participants except for ACT (see below) Gen is always expressed by zero and is marked out, therefore, with ELID; this doesn´t refer to cases where *někdo* 'someone'*, nevím kdo* 'I don't know who'*, cokoli* 'anything', etc. could fit, but to such instances as: "for what it usually holds good", "what is typical of the given action", "those (that), who (what), as a rule, occupy(-ies) such a position", etc.

The lexical value of the node for general participant is represented by Gen, i.e., in the tectogrammatical tree for the sentence *knihy se tisknou* 'the books are printed' Gen stands as the lexical value with the functor ACT transformed from the original analytic (ATS) *se* 'self' AuxR; for the sentence *Jirka už dobře čte* 'George reads well already' the supplied node has lexical value Gen and functor PAT. It is not easy to distinguish between the absence of an optional participant and the presence of Gen. One of the criteria for this difference is the dialogue test: e.g., with *Co dělá Honza? Píše dopis. - Komu?* 'What's Johny doing? He is writing a letter. - To whom?' the response could be: *To nevím* 'That I do not know' - an absent ADDR. On the other hand with *Jirka už dobře čte* the answer to a possible question *Co*? 'What (does he read?)' could rather be: *No to, co tak děti obyčejně čtou* 'Well, what children would normally read' (with PAT having the lemma Gen).

(i) Gen with ACT:

With ACT Gen is expressed by reflexive passive or by a compound passive without the Actor, e.g., *Domy se stavějí z cihel* 'Houses are built from bricks', *Dům byl postaven ve dvacátých letech* 'The house was built in the twenties' (reflexive passive is treated automatically).

However, e.g., *Milan se zabil* is ambiguous - (1) 'Milan killed himself', or (2) 'Milan was killed': either

(1) *Milan*.ACT *zabít se*.PAT.COREF:*Milan* (in the MC CORNUM gets the serial number of the word *Milan* and ANTEC is filled in), or

(2) a derived intransitive (middle) verb *zabít_se* (*Milan upadl a zabil se* 'Milan fell down and was killed'): Milan.ACT zabít_se…

Gen is not the case with *Domníváme se* 'We assume'; *Tady to dělají takhle* 'Here they do it in the following way'*; Než se ohlídneš, uteče* 'Before you have a look around he'll run away'… - there the pertinent personal pronoun remains as lemma with ACT (*já* 'I'*, on*.PL 'he', *ty*.SG 'you')*;*

(ii) Gen with PAT:

Typical examples: *Ta trouba dobře peče* Gen.PAT 'That oven bakes well'; *Jirka už umí číst* Gen.PAT. 'George can read already';

However, PAT.Gen is not the case with: *Neruš ho, on ještě čte* 'Don't disturb him, he is still reading': this is not Gen but the absence of optional PAT; he is reading sth., anything, I do not know what - there is neither PAT nor ADDR there; in the verbal frame they are both optional; *číst* 'to read' has the frame ACT (PAT) (ADDR) (EFF);

(iii) Gen with ADDR:

Dědeček dobře vypravuje Gen.ADDR Gen.PAT 'The grandfather narrates well': he narrates to those that listen what may be ever normally narrated; it is general validity and habituality that argue in favour of Gen;

Seděli jsme u stolu a babička četla pohádky 'We were sitting at the table and granny was reading fairy tales' - this is ambiguous: either Gen.ADDR (i.e., to those at the table), or she was reading for herself, ADDR is absent (being optional in the frame).

> Note: *Babička si četla pohádky* 'Granny was reading fairy tales for herself' would be as follows: *babička*.ACT *číst si*.ADDR.COREF:*babička*; in the MC *si* will have ANTEC and CORNUM filled in as well.

(iv) Gen with EFF and ORIG:

Předělal tu loutku 'he remade that marionette' - Gen.EFF does not mean 'into anything', but again into a marionette or what marionettes usually may be remade into;

with *předělat* 'remake'*, změnit* 'change' the frame is ACT, PAT, EFF, (ORIG);

e.g., *vyrůst* 'grow' represents at least three tectogrammatical verbs, with three frames:

*vyrůst*1 ACT PAT ORIG: *Chlapec*.ACT Gen.ELID.ORIG *vyrostl v muže*.PAT 'The boy grew into a man'; *Syn*.ACT *vyrostl z dítěte*.ORIG *v muže*.PAT 'The son grew up from a child into a man'

Similarly *Ze semene vyrostl dub* 'From the seed an oak grew', however: **Semeno vyrostlo v dub* '*The seed grew into an oak' will not do which is why we have to distinguish *vyrůst*2 from *vyrůst*1:

*vyrůst*2 ACT PAT, similarly as *stát se* 'become', *vyplývat* 'follow' (*z čeho* 'from' is PAT):

Z chlapce.PAT *vyrostl muž*.ACT. lit.: 'From the boy grew a man'

Z městečka.PAT *vyrostlo velké centrum*.ACT 'From a town a large centre grew'; however:

malé městečko.ACT *vyrostlo ve velké centrum*.PAT - the verb is *vyrůst*1, there is Gen.ELID.ORIG;

*vyrůst*3 ACT: *Ten kluk už hodně vyrostl* 'That boy has grown up considerably', *Vyrostl z dětských kalhotek*.DIR1 'He has grown out of his baby's shorts'

vyvinout se 'develop' has meanings 1, 2, 3, with similar frames;

vyspět 'mature' has only 2, 3;

vyplývat 'follow' has just 2;

(b) on, ona, ono, oni, ony, ona 'he, she, it, they'

They have the single lemma *on* with the grammatemes of gender and number; for other pronouns see above in 2.2.1.

(c) Neg

is a lexeme (with the functor RHEM) for *ne, nikoli,* prefix *ne-* with verbs, and *ani* unless it is used as a coordinating conjunction); in the case of negative verbs the node with the lexeme Neg depends on V;

In the LC, Neg is primarily placed before V; the decision depends on TFA:

(a) if the node V has the value F in the attribute TFA, Neg depends on it from the left and has F;

(b) if V has T, Neg can depend either from the left and then it has T (*Dnes Jirka nepřišel proto, že je nemocen* 'Today George has not come because he is ill'), or from the right (with F); the latter case holds if the verb stands outside the scope of negation: (*nepřišel s omluvou, ale s vysvětlením* 'he has not come with an excuse but with an explanation').

> Note that e.g. *nevelký* 'rather small, lit. not-large', *nepřítel* 'enemy, lit. not-friend' represent derived words, they are not divided.

(d) Focussing operators

The words (adverbs, particles) *jen(om)* 'only'*, taky/é* 'also'*, též* 'too'*, právě* 'just'*, dokonce* 'even'*, i* 'and also' (unless it has a coordinating function) etc. (see the list in Appendix A) obtain the functor RHEM; *ani* obtains RHEM and also Neg.RHEM as a sister (depending on the same verb), unless it represents a coordinating conjunction.

The rhematizer depends prototypically on the verb; only in such examples as *byli tam jen ČEŠTÍ studenti* 'there were CZECH students only there' it depends on some other member (on the noun in this case), since there is only an adjective within its scope, which, therefore, should be its sister; that often is different from the ATSs.

(e) Particles

They differ functionally (if not in their morphemic shape) from conjunctions in that they do not link up clauses (they follow a full stop, etc.). This concerns e.g.*, tedy, tudíž, totiž* 'then, therefore, that is to say'*, a* 'and', *protože* 'because'*, když* 'when', but '*jenže*'; they have the functor PREC ("referring to the preceding text"); also the apposition particles, such as *tj.* 'i.e.'*, totiž* 'that is to say'*, a to* 'namely', *jako* 'as' obtain the functor PREC, if they introduce a whole clause, rather than an apposition.

(f) "Subjectivizing" (emotional) pronouns

The free pronominal Dative forms *ti, vám, nám* lit. 'to you, to us' (unless they represent ADDR or BEN) are automatically assigned ETHD, i.e. the abbreviation for the functor "ethical dative";

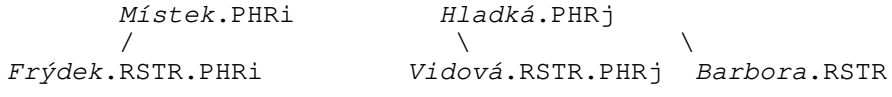
ten, on ("parallel to the subject") obtain the functor INTF (intensifier), e.g., *Ono (To) prší* 'Look, it is raining'*; On Jirka myslí…* Lit. 'He George thinks…'.

(g) Compound proper names

of the type of *České Budějovice, Hradec Králové,* but also *Frýdek-Místek* are structurally analyzed and, in addition to this, they obtain the grammateme PHRi.

A hyphen or a space inside a compound proper name are regarded as surface conventions of their spelling and the diffrences are not reflected in TGTS.

Such cases as *Frýdek-Místek* and *Barbora Vidová Hladká* are treated as follows:



The same holds for the spelling *Vidová-Hladká*.

(h) For the technical lexical value Cor, see Sect. 2.5.B.3 and 2.6.B; for *se* and *se*_Recp, see the section on grammatical coreference, 2.6.B. For the value Emp, see Sect. 2.2.2 (Sentences without the governing verb).

2.4 Coordination, apposition and parenthesis

The lemma of the conjunction is preserved; there can appear also further conjunctions, other than those contained in the list above, e.g.:

with the functor CONJ: *jakož i*, *jak_tak, jednak_jednak* 'as well as'*,*

with DISJ: *či* 'or'*, buďto_nebo/či* 'either - or'*, ani* 'neither'*, až* 'till'*, ať_ať* 'be - or',

with ADVS: *nýbrž* 'but'*, avšak, jenže* 'but, however', *sice_ale* 'though - but', *kdežto* 'whereas',

with CSQ: *tudíž* 'therefore'*, takže* 'thus'*, pročež* 'therefore',

with GRAD: *i*, *a_také*, *ani* 'even'

Statements about intervals are handled as a specific type DISJ with lemma *až*: *od - do* 'from - to'*, mezi* 'between'*, až* 'up to'; a hyphen or dash in such a function is replaced by lemma *až*:

V období mezi *1995 a 1999 (1995-99)* 'In the period between 1995 and 1999'*,*

Přinesl 5 až 10 knih 'He brought 5 up to 10 books',

*Přišlo tam 100 až 200 člen*ů '100 up to 200 members turned up there';

a comma without any conjunction can stand here, too, with the lemma being Comma: *dvacet, třicet jich bylo* 'there were twenty, thirty of them'.

To be distinguished:

nebo, či 'or':

každý tam byl buď nemocný, *nebo raněný* - DISJ 'everyone there was either sick or wounded'

posílají tam nemocné nebo raněné - ambiguous between DISJ and CONJ - 'they are sending there the sick or wounded' and 'they are sending there the sick and the wounded'.

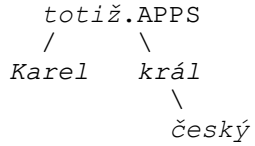
For the time being, such configurations as "Nováček s Kordou" (Attr in ATS) 'N. with K.' are not interpreted as coordination. The attribute obtains the functor ACMP (Accompaniment).

With the syntactic grammateme Reltype the value CO is assigned automatically to each member of coordination construction (to distinguish it from a modification of the whole coordinated construction).

Apposition

If in ATS apposition includes an AuxY (with such words as *totiž* 'that is to say'*, tj.* 'i.e.'*, a_to* 'viz., namely', *jako* 'as', etc., the role of the "apposition conjunction" will be played by the lemma of this particle, which obtains the functor APPS and on which both (all) apposited parts will depend; if there is no such word the lemma with APPS will be Comma (or, as the case may be, Dash, Colon, Brackp).

E.g.: *Karel, totiž král český* 'Karel, that is to say the Czech king'



An example of a coordination of appositions: *Václav, král český, nebo Zikmund, císař římský* 'Václav, the Czech King, or Zikmund, the Roman Emperor'.

As for apposition, we assume that it does not have a common modification: in *český král, Karel, který byl synem Elišky* 'the Czech King, Karel, who was the son of Eliška' the relative clause depends on *Karel;* on the other hand, in *král, který vládl mnoho let*, *Karel* 'the King, who ruled many years, Karel' the relative clause depends on *král* 'King'. To put it generally: the modification depends on the nearest member of the apposition; if both are "the nearest", then it is the left-hand one.

Parenthesis

Two situations may obtain:

(i) If in the ATS the inserted member has the analytic function AuxY_PA or PRED_PA (i.e., it is not syntactically integrated in the sentence), it obtains the functor PAR;

e.g.: *přijdu, myslím.*PAR, *pozdě* 'I will come, I think, late'*;*

(ii) If in the ATS the inserted member has some other analytic function before _PA, then it gets a corresponding functor; in the MC it automatically acquires (in the syntactic grammateme Reltype) the value PA to mark it off as parenthesis;

e.g.: *Předpokládám, že podmět (jestli vyjadřuje*.COND.PA *činnost*.PAT.PA),… 'I assume that the subject (if it expresses an activity),…'

(iii) The grammateme PA is assigned to all the nodes within the parentheses.

2.5 Deletions

The contents of Sect. 2.5 hold for the LC as well as the MC, unless with individual points an explicit statement to the contrary is found.

This section deals with restoration of nodes deleted in the surface structure of a sentence and it is divided into three subsections:

A. when to restore a deleted member and where to put it,

B. what to restore (a pronoun, a concrete lemma of the respective item); and under which conditions,

C. when to supply also something dependent on the restored node.

In this step of the conversion of analytic structures to the tectogrammatical ones the point in question lies in constructing nodes which have no counterparts in the corresponding analytic trees; such nodes are indispensable in the tectogrammatical, underlying structure of the sentence, yet they have been deleted in the surface structure.

I**n the LC**, each of such restored nodes is denoted by the value ELID in the attribute DEL.

In **the MC**, the values ELID and ELEX must be distinguished: ELEX is assigned when it is appropriate - according to the meaning of the given sentence - to add its dependent members to it, which, however, we do not specify;

e.g.: in the TGTS of *Včera navštívil Jirka Marii a Milan Alenu* 'Yesterday George visited Mary and Milan Alena', the nodes for *Milan*.ACT and *Alena*.PAT depend on the restored second occurrence of the verb *navštívit*, but for *včera* no further dependent node is supplied, and the existence of further modification which could be transferred from the first clause gets merely indicated by adding *navštívit*.CO.ELEX as the rightmost member of the coordination structure.

If such dependent members belong to a node which had not been deleted, this node gets the value EXPN in the attribute DEL, e.g. *Včera Jirka navštívil Marii a Milan potkal*.EXPN *Alenu* 'Yesterday George visited Mary and Milan met Alena'.

This means that we do not take pains over finding exactly which free modifications of the second coordinated member have to be repeated and which do not.

A. When and whereto restore elements

A.1 When to restore a node

We have adopted the following general principle:

In the **LC** as well as in the **MC** the following nodes get restored:

(i) the governing node to the node with the Afun ExD in ATS,

(ii) nodes for deleted obligatory modifications with the vervs and the verbal nouns ending with *–ní* or *–tí*, and

(iii) in coordination structures etc. nodes for the deleted repetitions of the governing word.

The cases under (iii) are supposed to be covered to a large extent already by (i).

> Note: Thus, no optional free modifications are supplied unless they represent governing words added to an attribute with ExD, or, as the case may be, the modifications follow from some specific context directly.

In the LC, the functor is supplied and so is (see B) the lemma as well as the gender and number with nouns and with the pronoun *on*; with *já, ty, my, vy* 'I, you, we, you-Pl' the number will be assigned automatically; with the node Cor (see B.3) also COREF is added (the lemma of the antecedent; if the antecedent is coordinated, it is the lemma of the conjunction that is supplied into the COREF); in the MC lemma is filled in as in the LC, and so are the values of CORNUM and ANTEC; with the type *potkal* (see B.4.3) the value PREV is assigned to the attribute CORSTN.

The following more detailed instructions are applied:

(a) A governing node to the node having in the ATS the function ExD is introduced unless the node with ExD depends directly on a node with AuxS (or, as the case may be, on some other Aux that is placed close below AuxS).

(b) With verbs, a node is introduced for an obligatory modifying member (in case it can be found within the valency frame in the dictionary), e.g., DIR3 with the verb *přijít* 'come'; a new node is also introduced in the case of a missing obligatory inomissible participant, as in *Potkal on*.PAT, see B.4.3.; for an omissible member see Sect. B.4.2 below; also a subject to a verb is supplied if it fails to be expressed in the surface structure (this step follows from a property of Czech as a language that can elide a pronoun playing the part of subject, a so-called pro-drop language), see Sect. B.4.1 below; a node having the lexical value of a personal pronoun (*on, já,*

my, ty, vy with gender and (with *on*) number according to the form of the verb and to what has been understood from the context):

[*My*.ANIM.PL] *Byli jsme tam všichni.* '(We) were there all.'

Marie a Jana přišly a [*on*.FEM.PL] *posadily se*. 'Mary and Jane came and sat down.'

Děti.FEM.PL *přišly a* [*on*.FEM.PL] *posadily se*. 'Children came and sat down.'

Number with *já, my, ty, vy* will be supplied in the second phase of the automatic procedure.

For LC, the same instructions hold for the restoration of nodes for obligatory participants of nouns derived from verbs in the most productive way (by the suffix *-ní* or *-tí*: e.g. *zpracovávání* '(repeated) processing', *zpracování* 'processing', *dobývání* 'the process of conquest', *dobytí* 'conquest'), but not with such nouns as *výběr* 'selection', *vývoj* 'development', *náhrada* 'replacement' etc. With other nouns only those participants are represented in the TGTS that appear also in the surface shape of the sentence. Thus e.g. in the TGTS of *Musí zůstat pod dohledem* 'He must be kept under supervision' no ACT is restored as dependent on the noun *dohled* 'supervision', and in the TGTS of *První skupina už odtud odešla* 'The first group has already left' there is no MAT restored as dependent on *skupina* 'group'.

In the MC, all obligatory participants are restored with all nouns to the frame of which they belong, if they are deleted in the surface shape of the sentence. ACT is restored also with postverbal adjectives derived from passive participles: e.g. *odvozený* 'derived', *přijatý* 'accepted', *darovaný* 'donated', *adresovaný* 'addressed', PAT is restored with the active counterparts, such as *píšící* 'writing'.

In the MC, the functor of the antecedent is inserted into COREF, and also CORNUM and CORSNT are filled in.

(c) Nodes for deleted elements in coordination constructions are supplied, see point B.1; a new node is introduced only if another node depends on it or in the case of an obligatory modification.

(d) Nodes for general participants are introduced with the lexical value Gen (i.e., Gen.ACT, Gen.PAT, Gen.ADDR, Gen.EFF, Gen.ORIG), see Sect. 2.3 (a) above and B.2 below.

(e) A new node is introduced in cases of grammatical coreference (the relation of control with infinitives and other forms: a node for the controlled element is added), see B.3 below.

(f) A new node is introduced for a noun that governs a congruent attribute expressed by an adjective (*Přišli jen mladší* 'Only the younger (ones) came'; *Našli jen modré* 'They found the blue (ones) only';

this does not concern those adjectival words with which we assume a substantival function as well, such as:

pronouns as *ten(to)* 'this'*, některý* 'some'*, všichni* ´all' (with the exclusion of possessive pronouns other than *naši, vaši* 'our/your people, family, parents'; only the latter function also as nouns),

cardinal numerals,

superlatives,

substantivized adjectives as *nemocný* 'ill'*, raněný* 'wounded'; for a list see above in the Sect. 2.1 (iv).

Examples:

*Z kluků přišli tři.*ACT. 'From the boys three have arrived.'

*Zbylí dva.*ACT (*kluci* - not restored) *jsou…* 'The remaining two (boys) are…'

*Zvolili tři.*PAT *z pěti místopředsedů.* 'They elected three from the five vice-presidents.'

A noun is not restored even with adjectival words in constructions with a copula (*Kluci byli úspěšní*.PAT 'the boys were successful') or with the functors EFF and COMPL: e.g., *pokládat za své.*EFF 'to regard as (one´s) own', *našli je nemocné.*COMPL 'they found them ill', etc.

(g) With a negative verb a node with the lemma Neg and the functor RHEM is inserted as one dependent on the verb from the left (whose lemma then will be positive): *nenapíše* 's/he will not write' is represented as Neg.RHEM *psát*.

This does not concern adjectives (*nepsaný* 'unwritten', *nepodařený* 'unsuccessful, miscarried') or nouns (*neurvalec* 'rowdy'), or any of other non-verbal parts of speech (*nemnoho* 'not many/much', etc.); with them just one node remains with a negative lemma.

A.2 Where to place a restored node

The restored node is always placed to the left of its governing node. Should more sister nodes be restored, then their order must conform to the systemic ordering of their functors, i.e., ACT first, followed by most of the free modifications, then ADDR, PAT and EFF in this order.

In case of rewriting of AuxR as Gen.ACT, the node carrying Gen.ACT remains in the place it occupied in the ATS.

The nodes governing coordination constructions are placed between the parts of this construction (for a more precise specification, see Sect. 3).

B. What lemma is assigned to the restored node

General principles:

(1) Where filling in a pronoun (or, as the case may be, a pronominal adverb) would suffice and the meaning of the sentence is not changed, we confine ourselves to supplying that pronoun only, see B.4.

(2) A lexical unit for a noun or adjective is supplied only if required by syntax or semantics, see B.1; with verbs, the lexical value is added always (this will be done in the second phase of the automatic procedure), since there is no pronominal verb in Czech (such as the English *do*).

(3) For special lemmas, see the respective sections: Sect. B.2 for the lemma *Gen*, Sect. B.3 for *Cor*, Sect. 2.2.2 for *Emp*, Sect. A.1 (g) for *Neg*, and Sect. 2.4 for *Comma*.

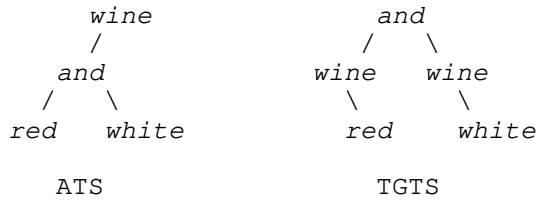
B.1 Filling in lexical units

The value of the lemma is filled in by the lexical value in case it is systemically obvious from the sentences itself not only in which syntactic position something has been deleted, but also it can be recognized which particular lexical unit should occupy this position. In particular, this is the case of an elided member in a coordination construction provided there is something that depends on it; for a reconstruction of obligatory modifications, see B.4.2-3 below.

In the MC the insertion of the lexical value is done by a human annotator; in the LC, only the node as such is restored, with the lemma ´???´; the respective concrete lexical unit will be added in the 2nd phase of the automatic analysis.

Examples:

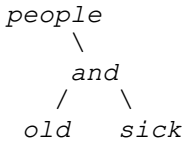
červené a bílé víno 'red and white wine'



Such a TGTS is assigned only if it is semantically clear that a common modification of both the conjuncts is not the case. If not, then the following two situations may obtain:

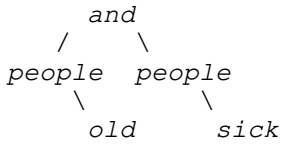
staří a nemocní lidé 'old and sick people':

(a)



This structure is assigned if there is a close conjunction in terms of meaning or if the semantics is not clear; if it is obvious that the conjunction means *lidé staří a lidé nemocní* (i mladí) 'old people and sick people (even young ones)', the structure is:

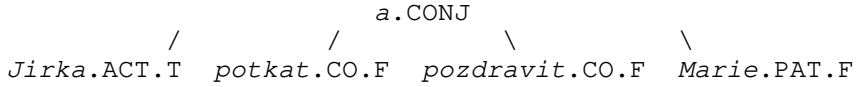
(b)



A similar ambiguity is present, e.g., with *nové knihy a časopisy* 'new books and magazines'; on the other hand, it is clear that for such examples as *ministerstva práce a sociálních věcí a zdravotnictví souhlasila, že…* 'The Ministeries of Labour and Social Affairs and of Health agreed that…' under the second conjunction *a* 'and' a node for *ministerstvo* will be restored, while in the first coordinated structure the nodes for *práce* 'labour' and *věc* 'affair' will both depend on the first node for *ministerstvo* 'ministery' (the change of PL to SG takes place in the MC only).

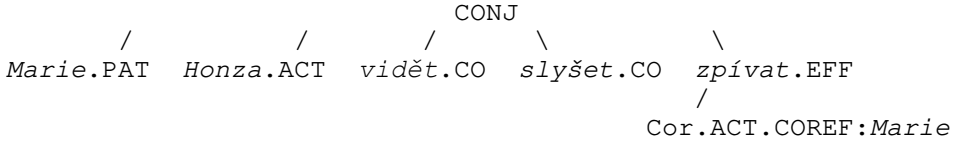
The general principle is that coordination is to be placed as 'low' as possible in the structure. The consequence of this general principle is that often nothing is restored: the simplest structure is preferred, i.e., the coordination of participants is preferred to the sentential coordination, so that, e.g., two nouns depend on the coordination of verbs:

Jirka potkal a pozdravil Marii 'George met and greeted Mary'



However, sometimes the restoration of a node is inevitable: this is the case with grammatical coreference, where the lemma of the antecedent is placed in the COREF in the LC, too (again by a single annotator in the second pass):

Marii Honza viděl a slyšel zpívat. Lit. 'Mary I saw and heard sing.'



Common modification of a coordinated predicate is not involved in such constructions as: *Jirka potkal Marii a pozdravil (ji)* 'George met Mary and greeted (her)'; here ACT (George) is common for both verbs, but PAT under *pozdravil* 'greeted' gets the lemma *on* (if it is understood that George greeted Mary, or someone else known from context), or the lemma *Gen*.

B.2 The general participant Gen expressed by zero

The notion of 'general participant' (*Gen*) was already discussed in Sect. 2.3 (a) above. In the present Section, we present some illustrative examples, in the connection with restoration of nodes.

> Note: If an obligatory participant is not expressed, it, as a rule, appears as Gen, unless it ought to be restored on the basis of context.

Examples of zero realizations of Gen with different participants:

Actor:

Honza tam slyšel Gen.ACT *střílet* 'John heard shooting there': with the infinitive (in Czech; verbal noun in English);

Ty domy Gen.ACT *byly vystavěny už dávno* 'Those houses were built already a long time ago': with the passive construction.

Patient:

Jirka hodně čte Gen.PAT 'G. reads a lot.'

Matka uklidila Gen.PAT, *uvařila* Gen.PAT *a lehla si.* 'Mother tidied up, cooked and went to bed.'

Za deset minut budu mít uklizeno Gen.PAT 'In ten minutes I shall have done with cleaning.'

Náš chlapec už čte Gen.PAT (in the MC the grammateme of modality FAC is added to the verb) 'Our boy can read already.'

Ta trouba dobře peče Gen.PAT 'That oven bakes well' (the same as with the preceding example).

However, compare the following example:

Neruš tátu, on ještě čte. 'Don't disturb father, he is still reading.'

In the TGTS for this sentence, PAT is absent; in the frame of the verb *číst* 'to read' PAT is optional, the verbs *číst* 'read', *psát* 'write' etc. have the frame ACT (ADDR) (PAT) (EFF); thus e.g. the sentence *Napsal mu o té schůzi* 'He wrote him about that meeting' has two interpretations:

(i) "*že bude*" 'that it would take place' - Gen.PAT
(ii) "*jaká byla*" 'what was it like' - no PAT, only EFF

On the other hand, the frame of the verb *říkat* 'to tell' is ACT (ADDR) PAT EFF: to tell smt. (EFF) about smt. (PAT), cf. the following example:

Gen.EFF *Řekli mu.*ADDR *o té schůzi*.PAT. 'They told him about the meeting.'

With the verb *mluvit* 'to speak (about smt.PAT) (with s.o.ADDR)' both PAT and ADDR are optional, so that in the following sentences there is just Gen.PAT and no ADDR:

To dítě už Gen.PAT *mluví*. 'That child can speak already.'

Gen.PAT *Mluví anglicky*. 'S/he speaks English.'

Mluví o té oslavě.PAT *už dva měsíce*. 'He has been speaking about that celebration for two months already.'

Když s ním.ADDR Gen.PAT *mluvíš, máš pocit vstřícnosti*. 'Whenever you speak with him you feel something like communicative atmosphere.'

Musíš s ním.ADDR Gen.PAT *promluvit.* 'You have to speak with him.'

There is no PAT nor ADDR in the following sentence:

*Když Petr.*ACT *mluví, všichni ztichnou*. 'When Petr is speaking all become silent.'

Addressee:

Dědeček dobře Gen.ADDR *vypráví své zážitky*.PAT. 'Grandfather relates well his experiences.'

Jana Gen.ADDR Gen.PAT *prodává u Bati*. 'Joan sells at Bata´s.'

In both these sentences, the restoration of a General participant is supported by a habitual or customary, non-topical character of the given action (the grandfather is a good narrator, she is a shop-assistant).

Effect:

The verbs such as *předělat* 'remake, reshape, reconstruct, rebuild, transform', have the frame ACT PAT (ORIG) EFF, cf. the following examples:

Gen.EFF *Předělali nám byt* (*teď se nám toho tam víc vejde*). 'They have rebuilt our flat (now there will be more space for our things to go in)'.

Gen.EFF *Předělali nám topení (a máme to teď dražší)*. 'They have reconstructed our heating system (and we have it more expensive now).'

Note: A support for such a treatment can be found in the dialogue test, cf. the incoherence of the following sequence: *Na co? *Nevím (Na nějaké modernější…)*. 'Into what? *I don't know. Into some more modern one.'

Further examples of a similar type:

Otec předělal loutku. 'Father remade the puppet'.

Matka přešila šaty. ´Mother remade the dress.'

Přestavěla pokoj/dům. ´She has rebuilt the room/house.'

Cf. also Sect. 2.3.(a) for the frames of the verbs *vyrůst* 'grow', *vyvinout se* 'develop', *vyspět* 'mature', *vyplývat* 'follow'.

B.3 Verbs of 'control'

The lemma Cor is assigned to the restored node (which can have the functor ACT, PAT or ADDR in the TGTS) representing the zero subject with infinitive; this node is coreferential to the Actor, Patient or Addressee or even a free modification BEN (as in *chodit pozdě je pro Janu typické* 'to come late is typical for Jane' where *chodit* is the governor of Cor.ACT.ELID.COREF: *Jana*) of the governing verb:

In the LC a node dependent on the infinitive is established (placed to the left of the infinitive) and assigned the lexical value Cor and the functor ACT (with passive infinitive PAT or ADDR). (In the second pass, another annotator fills in the attribute COREF with the lemma of the word with which the subject of the infinitive is coreferentially related.)

In the MC also the values of the attributes CORNUM and ANTEC are added.

Examples:

With the frame of the verb *hodlat* 'intend', ACT is the (functor of the) 'controller', with the verb *radit* 'advise' it is the ADDR.

Podnik plánoval zvýšit výrobu. 'The firm planned to step up its production.'
In the LC: the restored node depending on the verb *zvýšit* has the complex label Cor.ACT.ELID.COREF:*podnik*.
In the MC: the restored node depending on the verb *zvýšit* has the complex label Cor.ACT.ELID.COREF:*podnik*.CORNUM:number of the word-form *Podnik.*ANTEC:ACT.

Rodiče Jirkovi radili jít tam. 'The parents advised George to go there.'
In the LC: the restored node depending on the verb *jít* has the complex label Cor.ACT.ELID.COREF:*Jirka*.
In the MC: the restored node depending on the verb *jít* has the complex label Cor.ACT.ELID.COREF:*Jirka*.CORNUM:number of the word-form *Jirka*.ANTEC:ADDR.

Note: We distinguish between (i) *Jirka slíbil přijít* 'George promised to come', where a node is established depending on the infinitive and having the lemma *Cor* and the functor ACT (since the alternative that someone else would be coming is out of question) and (ii) a structure with textual coreference *Jirka slíbil, že přijde* 'George promised that (he) would come', where the restored node (with the functor ACT) depending on the verb of the embedded clause has as its lemma the personal pronoun *on* 'he' with gender assigned according to the context.

Further examples for the LC:

Honza se chystal tam jít. 'John was prepared to go there.' the restored node depending on the verb *jít* has the complex label Cor.ELID.ACT.COREF:*Honza*

Honza se bál být spatřen. 'John was afraid lest he should be seen (lit. 'to be seen').'
the restored node depending on the verb *spatřit* has the complex label Cor.PAT.ELID.COREF:*Honza*; in addition, with *spatřit* either Gen.ACT.ELID or *on*.ACT.ANIM.PL.ELID is restored, according to the context

Note: Passive is not distinguished from active in the TGTSs; a possible solution of cases in which the truth conditions differ (e.g. *Převezli ho dobrovolně*. '(They) transported him voluntarily.' vs. *Byl převezen dobrovolně*. 'He was transported voluntarily.') has to be found in the future.

Nouns of action as objects of verbs of control are treated in a similar vein:

Radili Jirkovi k odchodu. Lit.: 'They advised Jirka to departure.'
the restored node depending on the noun *odchod* has the complex label Cor.ACT.ELID.COREF:*Jirka*;

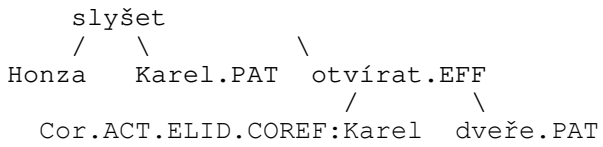
Podezírají ho z útěku před zodpovědností. 'They suspect him of flight from responsibility.'
the restored node depending on the noun *útěk* has the complex label Cor.ACT.ELID.COREF:*on*

Očekávali ode mne trvalý dohled nad pacientem. Lit. 'They expected from me a constant supervision over the patient.'
the restored node depending on the noun *dohled* has the complex label Cor.ACT.ELID.COREF:*já*.

In the MC again the values of CORNUM as well as those of ANTEC will be specified.

The so-called Slavonic infinitive with accusative is also treated in a similar way:

Honza slyšel Karla otvírat dveře. 'John heard Charles open the door.'
(*slyšet* 'hear' has the frame ACT PAT (EFF)):



Honza slyšel otvírat dveře. 'John heard the door open.'
two nodes are restored: one depending on the verb *slyšet*, with Gen.PAT, and another depending on the verb *otvírat*, with Cor.ACT.COREF:Gen

In the MC, with all these examples also CORNUM and ANTEC will be filled in.

Note: The verb *slyšet*, however, has also another frame, namely ACT PAT (where PAT is in Accusative or it is a dependent clause), so that, e.g., with *Honza slyšel, že se dveře otvírají* 'John heard that the door was opening' or *Honza slyšel otvírání dveří* 'John heard the opening of the door', no node will be restored depending on the verb *slyšet*.

B.4 Restoration of a pronominal, anaphoric element

B.4.1 Zero subject (pro-drop)

Examples:

on.ACT.ELID *Udělal to*.PAT. '(He) did it.'

Note: In *Pršelo* '(It) rained' there is no ACT.

Byla předběhnuta několika jinými. '(She) was left behind by some others.'

Here, the reference of the zero subject is indistinct and is to be decided on the basis of the context; the sentence can be interpreted either as:

on.FEM.SG.PAT.ELID *Byla předběhnuta několika jinými*. 'She was left behind by some others.'

or (in the context e.g. of *Naše děvčata běžela dobře, ale přesto neměla úspěch*. 'Our girls were running well, yet they failed to succeed.') as:

on.NEUT.PL.PAT.ELID *Byla předběhnuta několika jinými*. 'They were left behind by some others.'

In the LC: only the personal pronoun *on* is added. The number with *já, ty, my, vy* will be supplied automatically, but with *on* the number and the gender is assigned manually, in the second pass, by a single annotator, according to the form of the verb and corresponding to what has been understood from the context. The functor is assigned manually everywhere.

Examples:

my.ANIM.ELID.ACT *Byli jsme tam všichni*. 'We (boys) were there all.'

my.FEM.ELID.ACT *Byly jsme tam všechny*. 'We (girls) were there all.'

Marie a Jana přišly a on.PL.FEM.ELID.ACT *posadily se*. 'Mary and Jane arrived and (they) sat down.'

Děti přišly a on.PL.FEM.ELID.ACT *posadily se*. 'Children arrived and (they) sat down' (Note: *děti* 'children' is FEM in plural, the same as with *oči* 'eyes', *uši* 'ears'.

my.ANIM.ELID.ACT *Vstali jsme a my*.ANIM.ELID.ACT *odešli*. 'We got up and left.'

Včera Jirka přijel a dnes on.ANIM.SG.ELID.ACT *odjel*. 'Yesterday George arrived and today (he) left.' (Note: This is a case of coordination, but only a pronoun is assigned, because the node in question is a dependent node and not a member of the coordination construction. See Sect. B.1 above.)

B.4.2 Deleted omissible obligatory participant

In cases of deletion of an omissible obligatory participant in the ATSs, a node with a pronoun (pronominal adverb) is restored in the TGTSs in both the LC and MC.

Examples:

Jirka do Černošic napsal a Jarda přišel. 'George wrote to Černošice and Jarda came.'

It should be decided according to the context whether *sem* ´here´ or *tam* ´there´ with ELID.DIR3 will be restored. In case of *sem* 'here' nothing is filled into COREF, CORNUM and CORSNT, with *tam* 'there' such attributes are filled in according to the antecedent, in the MC only.

Similarly:

Jirka odsud/odtamtud.ELID.DIR3 *odešel*. 'Jirka left from here/there.'

B.4.3 Inomissible obligatory participant

If in the surface structure an inomissible obligatory participant is deleted, a node with a pronoun (pronominal adverb) is restored in the TGTS.

Examples:

(*Potkal Jarda Jirku?* 'Did Jarda meet George?') *on*.ANIM.SG.ELID.ACT *on*.ANIM.SG.ELID.PAT *Potkal*. 'Met.'
It is evident that it was Jarda who met George, but the point in question is a textual coreference. Therefore *on* is assigned as the lexical value not only to ACT (see above in B.4.1) but also to PAT; only in the MC the values of COREF:*Jarda* (together with the corresponding CORNUM and CORSTN) appear with the Actor, and COREF:*Jirka* (together with the corresponding CORNUM and CORSTN) with the Patient. ANTEC is not filled in because this is done with grammatical coreference only, see Sect. 2.6 above.

(*Potkals našeho Jirku?* 'Did you meet our George?') já.ANIM.ELID.ACT *Jirka*.ELEX.PAT *Potkal*. 'Met.'
The restored node for *Jirka* gets ELEX rather than ELID, because its modification ('our') is not copied from the context. (But cf. D below for *Jirka potkal a pozdravil Marii.* 'George met and greeted Mary.'*)*

Jirka potkal Marii a pozdravil. 'George met Mary and greeted.'
The assigned TGTSs should distinguish between the possible interpretations for the deleted Patient: either *ji* 'her' (= Mary; then the restored node has the complex label *on*.FEM.SG.ELID.PAT), or *ho*/*je* 'him/them' with he corresponding values of the gender and number, according to the context (maybe in connexion with a bet, etc.); in another case, the restored node has the label Gen.ELID.PAT. (But cf. the end of Sect. B.1 above, for the example *Jirka potkal a pozdravil Marii*. 'George met and greeted Mary.'). COREF and other attributes will be filled in the MC only, because this is again a case of textual coreference.

B.4.4 Restoration of a pronominal node

In case the deleted noun governs a congruent attribute expressed by an adjective.

If such a situation obtains in a coordination construction and the particular lexical value is obvious, this lexical value is assigned to the restored node; see the example *červené a bílé víno* 'red and white wine' above in B.1. Else, a pronominal lexeme is filled in.

In the MC, the restored node is assigned also a functor as well as gender and number; sometimes *ten* instead of *on* is filled in as the lemma, e.g.:

Ty *kulaté dá Jirka doprava, ty hranaté dá doleva*. 'Those rounded Jirka will put to the right, the angular (ones) (he will) put to the left.'

*ten.*INAM.PL.ELID.PAT *kulaté dá Jirka doprava, ten*.INAM.PL.ELID.PAT *hranaté on*.ANIM.SG.ELID.ACT *dá doleva*.

The governing node is not restored with such words as *nemocný* 'sick', *raněný* 'wounded', *ten(to)* 'this, that', *některý* 'some', *všichni* 'all', etc., also with superlatives, etc., i.e. with those adjectival forms with which we presume a nominal function as well, see above in Sect. A.1 (f) and the list in 2.1 (iv).

B.4.5 Deleted pronouns of laziness

Example:

Můj bratr políbil svou ženu a já taky. 'My brother kissed his wife and so did I.'

From this sentence standing alone, it is not clear whether I kissed my own wife or his wife. if this ambiguity can be solved on the basis of context, it will be distinguished by the assignment of two TGTSs in the MC:

(i) for the interpretation 'I kissed my wife, too.':
Můj bratr políbil svou ženu a já já.ANIM.SG.ELID.APP *on*.FEM.SG.ELID.PAT.COREF:*žena políbil taky.*

(ii) for the interpretation 'I kissed his wife, too.':
Můj bratr políbil svou ženu a já on.ANIM.SG.ELID.APP.COREF:*bratr*
on.FEM.SG.ELID.PAT.COREF:*žena políbil taky.*

In the LC, only the nodes for the deleted verb *políbit* and for the Patient of the second coordinated clause are restored: *on*.FEM.SG.PAT.ELID *políbit*.

> Note: A pronoun of laziness that has not been left out remains as a pronoun, and COREF is filled in only in the MC.

C. Notes concerning more complex examples

(i) Wherever the lexical unit can be restored in a non-systematic way, yet univocally (it may be carried over from the preceding sentences, crossing the full stop), the lemma of the restored unit is added to the node with the supplied pronominal element as the value of the attribute COREF; this is done in the MC.

Example:

Jirka přišel k Jiřině a dal jí kytku. 'George came to Jiřina and gave her a flower.'
If it is clear from the context that George did not give a bunch of flowers to someone about whom something in this sense had been said before, then the value of COREF with the node *on*.FEM.SG.ADDR will be *Jiřina*, with CORNUM equal to the serial number of the node with the lemma *Jiřina* in the first clause of the coordinated structure.

(ii) If the the antecedent is a pronoun, a pronoun is also assigned as the lemma of the restored node.

(iii) Generality (Gen) applies also with adjectives:

Jan je Gen.ADDR.ELID *věrný manžel*. 'John is (a) faithful husband.'

Jirka je Gen.PAT.ELID *hrdý/pyšný člověk*. 'George is (a) proud/vain man.' (i.e. proud of himself, of the circumstances around himself, of his merits,…).

For further examples see Panevová (1998b).

(iv) As has been already said, lexical values must be supplied when restoring nodes for verbs, since Czech does not have a pronominal verb such as the English *do*. Therefore, the lemma '???' rather than a lexical verb such as *dělat* 'to do' is assigned in the LC, since an insertiosn of dělat 'to do' might change the sense of the sentence. The lemma '???' will be replaced by a concrete lexical value in the 2nd phase of the automatic processing.

2.6 Coreference

Two types of coreference can be distinguished: textual (see the Section A below) and grammatical (reflexive and relative means, cases of control, complement, see the Section B below).

In case of textual coreference, the restored node is assigned a pronominal lemma in the LC (this node gets a functor, but COREF is not filled in).

In case of grammatical coreference, in the MC the value of COREF is filled in (it is the lemma of the antecedent or, as the case may be, of the 'controller'). In the LC this is done in the second pass by a single annotator.

As coreference also wider anaphoric relations are understood, which need not always concern a full referential identity: e.g., *Jirka zase nepřišel. Oni všichni…* 'George failed to turn up again. They all…' - *oni* 'they' is interpreted as referring to George.

A. Textual coreference

In the MC: the attributes COREF (lemma of the antecedent), CORNUM (serial number of the antecedent), and CORSTN are filled in, while ANTEC (which is for grammatical coreference only in the MC) is not; the attribute CORSTN is automatically assigned the value NIL, which is changed into PREV by the human annotators of the MC if the antecedent is in the preceding sentence.

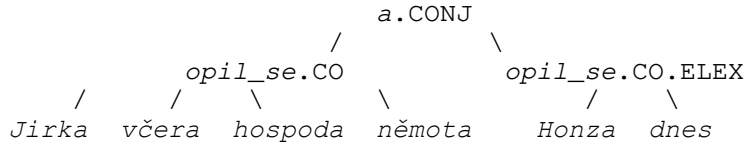
Examples:

Jirka potkal Marii a pozdravil. 'George met Mary and greeted.'
If it is clear that George greeted Mary, the node *on*.FEM.SG.PAT.ELID is restored, with the attributes COREF:Marie and CORNUM:Nr.Marie; else Gen.PAT depending on the verb *pozdravil* is restored.

Políbil Marii včera a já dnes. 'He kissed Mary yesterday and I today.'
The second occurrence of the verb *políbit* is restored, together with its PAT; the attribute COREF of PAT is filled in by the respective value.

(*Vsadili jsme se, že do týdne Jirka políbí Marii a já Janu.*) *Jirka ji políbil včera a já dnes*. '(We made a bet that in a week George would kiss Mary and I Jane.) George kissed her yesterday and I today.'
The reference of the restored participants is again obvious from the value of COREF for the restored PAT (the speaker kissed Jane).

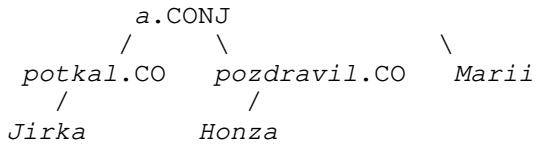
Jirka se včera v hospodě opil do němoty a Honza dneska. 'Yesterday in the pub George drank himself under the table and John today.'



NB: If there are more dependents on the restored node that might be copied from the antecedent structure and it is not clear which of them are to be really restored (did Johny drink himself also under table and again in the pub?), then ELEX rather than ELID is assigned to the restored governing member and the dependents that are free modifications are not copied; however, obligatory modifications are always copied and restored. The specific symbol ELEX serves the purposes of a further linguistic study. In the LC only ELID is used, never ELEX.

No restoration occurs with:

Jirka potkal a Honza pozdravil Marii. 'George met and John greeted Mary.'



In the LC: a node with pronominal lemma is restored (this node obtains a functor, but its COREF is not filled in), which, however, concerns the following cases only (cf. the conditions listed in Sect. 2.5.B above):

(a) zero subject:

Spali. '(They) slept.'
on.ANIM.PL.ACT.ELID *Spali*.

but: (*Lehli jsme si a*) *spali*. '(We lay down and) slept.'
my.ANIM.PL.ACT.ELID *Spali*. (see Sect. 2.5.B.1 above);

(b) some other deleted obligatory participant:
omissible (*přišel sem/tam* '(he) came here/there', see Sect. 2.5.B.4.2 above),
inomissible (e.g. PAT with *potkat*, see Sect. 2.5.B.4.4).

(c) a noun on which an adjective (with ExD in the ATS) depends, see Sect. 2.5.A (f) and 2.1.(iv) above.

B. Grammatical coreference

Grammatical coreference concerns such means of expression as reflexive pronouns (e.g. *svůj* 'his', *se* '-self'), relative pronouns (e.g. *který* 'which'), the subject of an infinitival clause (control) and the predicative complement.

In the LC, the lemma of the restored node and the value of COREF (the lemma of the antecedent) are filled in.

In the MC, also the values CORNUM and ANTEC are assigned; in the attribute CORSTN the value NIL (assigned automatically) remains unchanged because with grammatical coreference the antecedent occurs in the same sentence).

(i) With *se/si* and *svůj*:

The forms *se, si, sebou, svůj, sebe, sobě…* (forms of reflexives) have a common lemma *se*; in the LC for *se, si, sebou* the lemma *se* as well as the value of COREF (i.e., the lemma of the Actor) will be added in the second phase of the automatic procedure; with *svůj, sebe* and with passive both the lemma and COREF will be supplied manually in the second pass, by a single human annotator.

In the MC, in addition to the lemma and the value of COREF the following values will be assigned:

(a) Where reciprocal use is not the case:

In the MC also the values of CORNUM (the number of the antecedent) and ANTEC (the functor of the antecedent) and the syntactic grammateme are assigned.

The values of the functor (and of the syntactic grammateme) are assigned according to the values of the Afun in the ATSs:

with *se*: *se*.Obj is changed to *se*.PAT (see also the point (b) below)
se.AuxR is changed to Gen.ACT
se.AuxT is attached to the lexeme of the verb (e.g. *smát_se* 'laugh', *šířit_se* 'extend')

with *si*: *si*.Obj is changed to *se*.ADDR (or *se*.PAT, according to the valency)
si.Adv is changed to *se*.BEN
si.AuxT is attached to the lexeme of the verb (e.g. *pospíšit_si* 'hurry up')
si.AuxO is changed into *se*.ETHD (e.g., *postupte si* 'get forward', *přijďte si* 'come')

with *svůj*: se.APP

(b) Reciprocal use of *se, si, sebou…*:

In the MC the pronoun occurring in the ATS is changed into the lemma *se*_Recp (reciprocal, 'mutual(ly)'), and the participants in the relation of reciprocity are analyzed as a coordination. The values of the attributes CORNUM and ANTEC are assigned as well.

Examples:

*Honza.*ACT *a.*CONJ *Jirka.*ACT *se*_Recp.PAT *střídali*. 'John and George took turns.'

*Střídali se*_Recp.PAT *Honza*.ACT *a*.CONJ *Jirka*.ACT. Lit.: 'Took turns John and George.'

Honza.ACT *a*.CONJ *Marie*.ACT *se*_Recp.PAT *líbali*. 'John and Mary kissed.'

Jan.ACT *a*.CONJ *Marie*.ACT *se*_Recp.ADDR *dopisovali* Gen.PAT. 'John and Mary were in corrrespondence.'

Chlapci.ACT *se*_Recp.PAT *střídali*. 'The boys took turns.'

This solution applies only where the reciprocal participants are coordinated or in plural (and mutuality is implied, e.g. John is taking turns with George and at the same time George is taking turns with John, i.e., between themselves); since the antecedent is coordinated, it is the lemma of the conjunction that appears in COREF. Wherever the participants are connected by means of the preposition *s* 'with' and a mutual relation is involved, the analysis proceeds in a different way:

Honza.ACT *střídá_se s Jirkou*.PAT. 'John takes turns with George.'

Honza.ACT *s Jirkou*.PAT *střídají_se*. Lit.: 'John with George take turns.'

Honza.ACT *s Marií*.PAT *líbá_se rád*.CMPL. Lit. 'John with Mary likes to kiss.'

Honza.ACT *s Marií*.PAT *líbají se rádi*.CMPL. Lit. 'John with Mary like to kiss each other.'

Jan.ACT *dopisuje_si s Marií*.ADDR *o termínech*.PAT. 'John corresponds with Mary about terms.'

Jan.ACT *s Marií*.ADDR *dopisují_si o termínech*.PAT. 'John with Mary correspond about terms.'

This is different from *Přijeli Lendl s Kordou.*ACMP (Lit. 'Arrived Lendl with Korda'); the criterion is the word-order position of the (mutual) participant. If such words as *spolu* 'together', *navzájem* 'mutually' occur here, they obtain the functor MANN.

In the LC, reciprocity is disregarded (i.e. _Recp is not attached to *se*, *si*) and the pair *Honza a Jirka* 'John and George' is treated as a coordination of typical (i.e., not mutual) subjects. Such verbs as *střídat* 'take turns'*, líbat* 'kiss' have the lemma *střídat_se, líbat_se*, respectively, and their participants are assigned either the values ACT and PAT as in: *Honza.*ACT *se_střídali s Jirkou.*PAT. 'John took turns with George.'(and, in a similar vein, *Honza s Jirkou se střídali*. 'John with George took turns.' is represented as *Honza.*ACT *se_střídali s Jirkou.*PAT), or these participants are represented as a coordination of Actors (*Honza a Jirka se střídali*. 'John and George took turns.').

(iii) With relative clauses: see Sect. 2.2.2.4 (congruent attribute, functor RSTR or DES, both in LC and MC).

(iv) With verbs of control: see Sect. 2.5.B.3.

(v) A certain kind of grammatical coreference may be seen in the relation of the predicative complement to the subject (or to some other sentence part); see Sect. 2.2.2.4.

3. TOPIC-FOCUS ARTICULATION

Topic-focus articulation (TFA) will be treated in the MC and (separately, by specialized annotators) also for the LC. The assignment is done after the structure of TGTSs has been built, but when making the decisions about the values of the attribute of TFA, sometimes the surface shape of the sentence and its ATS (which is hidden on the screen, but always accessible for the annotators) has to be taken into account.

The automatic preprocessing procedure, preceding the manual tagging, assigns F to every node; this value will be changed manually into T or C if necessary, according to the instructions below.

If the verb has a complex form in the ATS, then in the TGTS its node is typically placed in the position that in the ATS is occupied by the lexical, rather than by the auxiliary, verb (i.e. the position of the infinitive or the participle is decisive).

Further general principles:

(1) (a) whatever depends on the verb from the left has, as a rule, T;

the value F remains only where it can be clearly recognized that "new information" or a new relation is involved; sentential stress would be placed there in the spoken form of the sentence (falling pitch, intonation centre), e.g. *TÁTA přišel* 'FATHER came'; in the written Czech texts (especially intellectualized, technical ones, although not e.g. in written represenations of speech), the rightmost position gets F quite regularly;

the value C is asigned to the contrastive part of the topic (which expresses an element taken from a set of alternatives), e.g.:

Jedině s úspěšnými vzory.C *se můžeme poměřovat*. Lit. 'Only with successful models.C (we) can compare oureselves';

Jirkovi.C *to Martin nedal*. Lit. 'To George it Martin did-not-give'; *Janu*.C *Marie neviděla*. Lit. 'Jane(Acc.) Mary(Nom.) did-not-see';

(b) in the prototypical case, the node that depends on the verb from the right and occupies the rightmost position, gets the value F;

(c) any verb and what is between it and the rightmost position (see (b) above) gets, in principle, F (this also concerns are semantically "poor" words, such as *být* 'be', *mít* 'have', *činit* 'do' etc.); if the node represents a repeated (not necessarily verbatim) unit from the preceding text (within the sentence as well as from a previous context), it obtains T;

(d) the nodes that are more deeply dependent (such as an attribute, etc.) are, as a rule, assigned F (for exceptions, see (c) above);

(e) as a rule it holds, in the TGTSs, that every node having T or C depends on the left from its head node, and every node having F depends on the right from its head, with the following exceptions:

a focus sensitive particle with F precedes its head node if the latter has F (see (3) 11. below);

a proxy focus has T, but is placed to the right of its head in the TGTS, see (2)(ii) below; a further similar point is specified in (3) 7.

> Note: The node of a coordinating conjunction is not assigned any value in its attribute TFA, i.e. it is assigned NIL by the automatic preprocessing; the preprocessing inserts F as the TFA values of all the other nodes, which then are changed to T or C, in accordance with the instructions specified here.

(2) Treatment of structures that are non-projective at the analytic level (the following suggestions are tentative, for both the MC and LC):

(i) the node occurring at the non-projective leftmost position will be assigned the symbol C(ontrast) instead of T (for exceptions, see point (iii) below), that is to say, its contrastive use is assumed, and it will be placed in the projectively leftmost position (all nodes that depend on it follow and obtain T or F according to the instructions concerning the remaining nodes); clitics will also be placed in a projective way, yet they obtain index T (rather than C);

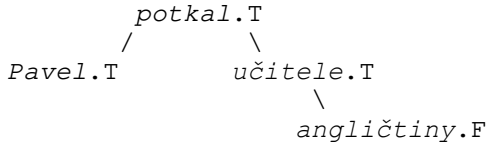
An example:

K jásotu.C *není nejmenší důvod*. 'For jubilation, there is not the slightest reason'

A contrastive node depending to the left need not be in a non-projective position, and yet it should get C: e.g.:

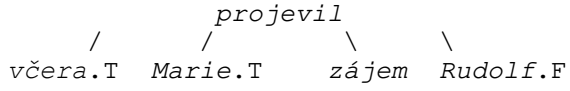
Jeho.C *jsem neznal; ji*.C *jsem poznal hned.* 'Him I didn't know; as for her, I recognized her at once'.

(ii) If to the right of the verb the first F node is deeply embedded (i.e. it does not depend directly on the verb), its governing node obtains index T and this governing node, the 'proxy focus' (see Hajičová, Partee and Sgall 1998), is placed (as an exception) to the right of its own governing node (an example: *Pavel*-T *potkal*-T *učitele*-T *angličtiny*-F 'Paul met the English teacher'):



(iii) the node placed in a non-projective position to the left is usually assigned C, but it may get T if it is not placed at the beginning of a clause (or if other factors show that it is not the least dynamic item) and at the same time, the governing word is 'quasimodal', be it a simple word (*plánovat* 'to plan', *rozhodnout_se* 'to decide to'), or a phraseme, with a quasimodal meaning (*mít čest* 'to have the honour to', *pokládat si za čest* 'to consider it a honour to'), or else it is a phraseme of the kind of *projevit zájem o* 'to express interest in';

the list of such phrasemes and quasimodal verbs is being compiled, step by step; the dependent words dependent on a phraseme are placed as dependent on its head, e.g. *včera*.T *o Marii*.T *projevil zájem Rudolf*.F Lit. 'yesterday about Mary expressed interest Rudolf'



(iv) a predicate complement at the beginning of a clause gets C, e.g. *Jako prozřetelné*.C *prověřil*.F *čas*.F *rozhodnutí*.F *Milenino*.F. Lit. 'As the time(Nom) decision(Acc) of-Milena'

(3) The order of the items within the topic:

(a) the leftmost position in the TR is occupied by the node of topic proper (typically a noun group); if in the uppermost subtree of the TGTS there is a node with C, this node is placed leftmost;

(b) then similar contextually bound nodes follow, i.e. those nodes that do not belong to (c) - (f);

(c) next come nodes with the functor PREC;

(d) then: restored deleted nodes, e.g. ACT.ELID;

(e) the TGTS nodes for clitics;

(f) temporal and local settings, i.e. adverbials of these two kinds that are neither topics proper, nor contrastive.

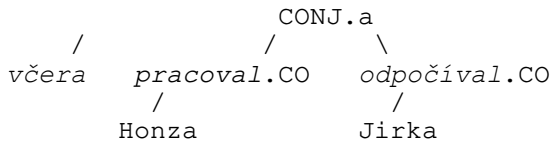
Note 1: Expressions such as *u nás* 'with us'*, dnes* 'today', *minulou sobotu* 'last Saturday' typically are settings, but they can also function as topics proper, cf.:

Dnes (topic proper) *tam* (setting) *nepůjdu* (focus); also in *Včera Jirka odjel* 'Yesterday George left' *včera* can be a setting and *Jirka* can be topic proper and thus placed in the first position; the decision on the positions of *Jirka* and *včera* should be made in accordance with the context (what the sentence 'is about' is topic proper).

If an adverbial of the given kind occurs in the focus, we do not call it a setting (e.g. *Přišla sem včera* 'She came here yesterday').

Note 2: If there are more than one node with the characteristics under one of the points (a) - (f), they should be ordered in accordance with systemic ordering (roughly: ACT - ADDR - most of the free complementations - PAT - EFF).

Note 3: A node that depends on a whole coordination precedes all nodes depending on the individual parts of the coordination, e.g.: *Honza včera pracoval a Jirka odpočíval*, lit. 'John yesterday worked and George rested' has a tree of this shape:



(4) The assignment of the values T and F is further guided by the following principles:

1. The value T is assigned to the nodes representing repeated (not necessarily verbatim) units from the preceding text (within the sentence as well as form the previous context); this concerns above all the cases of coreference, etc. (even with verbs). It is to this group where also the nodes supplied in the tectogrammatical tree belong (exceptions concern coordination, e.g. in such a construction as *černý a modrý inkoust* 'black and blue ink' the first occurrence of the noun is elided; this, however, is F, while the second, undeleted, occurrence is T).

2. Indexical expressions - *já, my, teď, můj, náš, tady, letos, zítra* 'I, we, now, my, our, here, this year, tomorrow' etc. - and also expressions with the functor PREC (*proto, tedy* 'therefore, then' etc.), obtain, as a rule, the value T; it is only with contrast, as bearers of intonation centre (sentential stress, mostly at the end of the sentence) that they are assigned F.

3. General pronominal words such as *někdo, něco, jednou, nějaký* 'someone, something, once, some' get the value F (with the same exception as with the verb, see (1)(c) above);

4. An attribute expressed by an adjective, noun or pronoun (except for indexical expressions etc.) and dependent on a noun is, under normal circumstances, assigned F (though it may stand to the left of the N in the surface word order); the value T is assigned only if this attribute is repeated or obvious from the preceding context:

starý dům 'old house' --> *dům starý*.*F*, but:

dům starý.*F a starý*.*T park* 'old house and old park'.

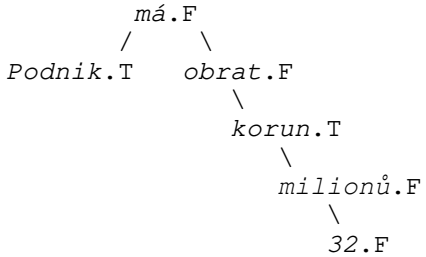
The value T is also assigned to typical counted nouns in the constructions with numerals (the governing noun is proxy focus in the examples below) unless it again follows from the context that the given noun is in contrast with some other noun in the sentence (e.g. crowns with dollars, years with months etc.); as for the placement of the governing noun with T in such constructions, see the note in (ii) above.

Examples:

Utratil za tu učebnici sto korun. 'He spent for the textbook (one) hundred crowns' --> *Utratil*.T *za tu učebnici*.T *korun*.T *sto*.F

Bylo mu osmnáct let. 'He was eighteen years old' --> *Bylo*.F *mu*.T *let*.T *osmnáct*.F

Podnik má obrat 32 milionů korun. 'The enterprise has a turnover of 32 million crowns' --> *Podnik*.T *má*.F *obrat*.F *korun*.T *milionů*.F *32*.F



If there is one or more attributes to the left of a governing noun, they are 'moved' from the left to the right in a mirror-like way, e.g.:

Koupila hezký malý obrázek --> '*Koupila obrázek*.F *malý*.F *hezký*.F '(She) bought a nice small picture'.

Prozradíte DALŠÍ elitní jména 'You will disclose FURTHER distinguished names' --> *Prozradíte*.F *elitní*.T *jména*.T *DALŠÍ*.F (here again *jména*.T stands to the right of the verb, as a proxy focus, see above and point (ii)).

If a noun is modified by attributes both from the left and from the right (in the ATS), the position of the attributes depending from the right relative to those moved there from the left is determined by the semantic closeness; e.g.:

žlutý motýl, kterého jsme viděli 'yellow butterfly we saw' --> *motýl žlutý… viděli*

malá kočka domácí lit. 'little cat domestic' --> *kočka domácí malá*

stará pasta na parkety lit. 'old wax for floor' --> *pasta parkety stará*

všechny malé sešity ze školy, které má v tašce 'all the small school copybooks which s/he has in (her/his) satchel' --> *sešity ze školy*.F *malé*.F, *které*.T *má*.F *v tašce*.F *všechny*.F (here the restrictvie reading of the relative clause is distinguished from its non-restrictive reading: in the latter case this clause would occupy the rightmost position).

5. The value T is assigned to the weak forms of pronouns (*tě, ti, ho, mu, mi,* mostly also *mě* 'of you, to you, him, to him, to me')*;* the same holds for *mně, ji, jí, jim, jej, nás*, etc. 'to me, her, to her, to thme, him, us' unless sentential stress is placed on them, and for other clitics.

6. The strong forms of pronouns (*tebe* 'you'*, jemu* 'him'*,* etc.) obtain the value F if they (or their prepositions, as the case may be) carry sentential stress (falling, i.e., the intonation centre, not only an optional rising, contrastive, stress - this would be assigned C); pronouns following a preposition have always strong forms, so that the form is not decisive in these cases for the assignment of the TFA value and they can get the value T, e.g.:

Pro něj.C *to přinesu ZÍTRA.* 'For him.C I´ll bring it tomorrow.'

Zítra to přinesu PRO něj.F. 'Tomorrow I´ll bring it FOR him.F.'

Zítra to pro něj.T *PŘINESU.* Lit. 'Tomorrow it for him.T I´ll BRING.'

ZÍTRA to pro něj.T *přinesu.* 'TOMORROW I´ll bring it for him.T.'

7. The symbols governing coordination structures have neither T nor F and they are written between both parts of the coordination connexion; here again, typically the coordinated nodes with T precede the conjunction and those with F follow it; only in a case of restored deletion it can happen that the left part of the pair has F while the right one T (see above in the example (1) in Sect. 2.5.B.1: *červené*-F [*víno*-F] a *víno*-T *bílé*-F lit. 'red.F [wine.F] and wine.T white.F'; in such a case, both the nodes (with F and T) are placed to the right of the conjunction, F preceding T.

8. The reconstructed node is always placed to the left of its governing node and it has index T (contextually bound), if possible; an exception is, e.g., the just mentioned (above, sub 7) situation: *červené*-F [*víno*-F] a *víno*-T *bílé*-F lit. 'red.F [wine.F] and wine.T white.F.'

9. As regards complex sentences, the following instructions hold:

(a) coordinated/paratactic compound sentence - see point 7 above; each coordinated clause has a TF articulation of its own:

*Tom.*T *přinesl*.F *knihy*.F *a pak*.T *Jirka*.T *odnesl*.F *noviny*.F 'Tom brought books and then George took-away newspapers.'

Knihy.C *odnesl*.F *a noviny*.C *přinesl*.F. 'The books and newspapers he did bring'.

(b) direct speech is, as a rule, more dynamic than the introductory sentence; the verb in the introductory sentence thus primarily gets T and the verb in the direct speech gets F; this holds true also in cases when the introductory clause follows the direct speech:

Jirka řekl.F (or T): *"Je*.F *dobře."* 'Jirka said.F (or T): "It's.F fine."'

"Je dobře," řekl Jirka. '"It's.F fine," Jirka said.'

In a secondary case, which is probably more rare, the direct speech moves to the left (and its verb gets T), as in e.g. *Taky JIRKA řekl: "Je dobře."* 'Also JIRKA said: "It is fine"' (with capitals denoting the placement of the intonation center), which would be represented in the same way as if it had the following shape: "*Je.*T *dobře" řekl.*T *taky JIRKA*.F. '"It is.T fine," said.T also JIRKA.F.'

(c) In a subordinate connexion of clauses the dependent clause usually preserves its position to the left (with T) or to the right (with F) according to the surface ordering unless its intonation or stress are marked (i.e., unless one of the verbs is repeated, etc.).

If the verb of an adverbial clause is placed before the governing verb of the main clause and is in a contrastive position (i.e. it has at least a rising intonation contour), the dependent verb gets C and remains to the left, as in e.g.:

Protože se program.T *osvědčil*.C*, použili*.F *ho…* 'Because the programme.T proved-to-be-good.C, they used.F it for…'.

(d) If some node in a relative clause obtains C, then this node should be placed to the left of the relative word introducing the relative clause.

10. As regards free modifications, MOD and MANN have as a rule F (even if they are placed left on the surface, the so-called adverbial attribute: *rychle řekl* lit. lit. 'quickly he-said'*; snad řekl* 'perhaps he-said'*; dobře řekl…* lit. 'well he-said'*)*; ATT has usually T (*Naštěstí přišli* 'Fortunately they arrived').

11. Focus sensitive particles ('rhematizers'):

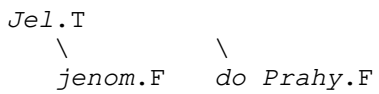
In the LC the position of a rhematizer, including Neg, and the value of its TFA attribute are decided as follows:

(i) if the TFA attribute of its governing node has the value F, then the rhematizer is assigned F and depends on its head from the left;

(ii) if its head has T, the rhematizer can depend from the left and then it has T (*Dnes Jirka nepřišel proto, že je nemocen* 'Today Jirka has not turned up because he is ill'), or from the right (with F), when the head is outside the scope ('focus') of the rhematizer (*… nepřišel s omluvou, ale s vysvětlením* '…he has not turned up with an excuse, but with an explanation'), or when, as the case may be, there is nothing else than the rhematizer in F.

Some further examples:

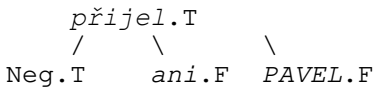
Jel.T *jenom*.F *do Prahy*.F '(He) travelled only.F to Prague.F':



Jenom.T *jel*.T *do Prahy*. '(He) only travelled to Prague.'

Nepřijel ani PAVEL. Lit. 'Has arrived neither PAUL. ' and

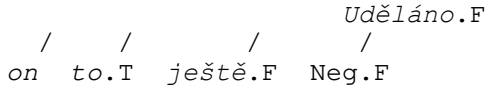
Ani PAVEL nepřijel. 'Not even PAUL has arrived' would be represented in the same way: *Neg*.T *přijet*.T *ani*.F *PAVEL*.F



12. In the remaining cases it holds that the node which depends from the left is assigned the value T and the node dependent from the right obtains the value F.

13. If in a verb complex an auxiliary verb has evidently a rhematic character (incl. the modal verbs proper), the lexical verb is placed in the focus part and is assigned F, even if its lexical value is contextually bound:

e.g. *Uděláno to ještě NEMÁ*. Lit. 'Done it yet he-has-not.'



The situation is similar in the following example, where the surface order cannot express the distinction between the thematic verb and the rhematic future:

Ještě to neudělal, ale určitě.F *to*.T *udělá*.F. Lit. 'Yet it he-has-not-done, but certainly.F it.T he-will-do.F.'

Concluding remark

A number of questions requires further examination, be it the need for more refined articulation, etc. and, therefore, further empirical research, or technical adjustments of problems whose more complete solution is known, but has not been applied at this stage owing to their exacting nature. An overview of such points will be published separately.

We do hope that even though achieved in this preliminary way, the syntactic (as well as the morphemic) annotations of parts of the Czech National Corpus will be useful in that the present form of the Prague Dependency Treebank makes it possible for interested researchers to collect large sets of data relevant for their research in a much more easy and rapid way than was possible using manual excerption. Any monographic research oriented at an issue from grammar, style or present-day development of Czech can then bring proposals how to amend the classification of the given set of data and how to bring the annotation procedure to a higher level, perhaps also to raise the degree of its automation.

Appendix: A list of adverbs (incomplete)

We are presenting here provisionally some adverbs (and particles). The selection is not systematic, adverbs of productive types (e.g., *česky, právnicky, lexikálně* 'Czech, juristically, lexically') cannot be listed; their functor is MANN. Ambiguity ATT/MANN is frequent - then it is typical that an adverbial of ATT ("bylo prosté, spravedlivé etc., že to udělal" 'It was plain, just, etc. that he has done it') would stand at the beginning of the sentence (clause): *prostě nepřišel* 'He plainly failed to come', while MANN ("bylo prosté, spravedlivé. etc., jakým způsobem to udělal" 'It was plain, just etc., in which way he has done it') usually follows the verb *oblékl se prostě* 'He's dressed plainly'; as ambiguous will be interpreted *Prostě se oblékl a šel tam:* '(He) simply dressed and went there' or 'Plainly, he dressed and went there': simply - ATT, plainly - MANN.

Adverbs derived in a productive way from adjectives have in the MC (and in the dictionary) the lemma of the adjective: e.g., *samozvaně --> samozvaný, česky --> český, snadno --> snadný*: 'self-appointedly, selfappointed; (in) Czech, Czech; easily, easy'.

ADVERB		**FUNCTOR**
absurdně	absurdly	ATT/MANN
alternativně	alternatively	MANN
ani	neither, nor	RHEM (or conjunction CONJ/GRAD)
apropó	'a propos'	ATT
asi	perhaps, maybe, about	MOD (pravděpodobně 'probably') or EXT.APPX (zhruba 'roughly')
bezpochyby	doubtless, indubitably	MOD
biologicky	biologically	MANN
bláznivě	foolishly	MANN
bohužel	regrettably, unfortunately	ATT
celkem	altogether, in all, on the whole	EXT
částečně	partly, partially	EXT
často	often	THO
česky	Czech	MANN
čestně	honestly	MANN
dál, dále	further, farther, next	PREC
díkybohu	thank goodness	ATT
dobře	well	MANN
dodatečně	additionally	MANN
dokonce	even	RHEM
doslovně	literally	ATT
doufejme	hopefully	MOD (without commas, else:PAR)
duchovně	spiritually	MANN
duševně	spiritually, mentally, intellectually	MANN
důvěrně	confidentially	MANN
ekonomicky	economically	MANN
emocionálně	emotionally	MANN
fakticky	factually	ATT
formálně	formally	ATT/MANN

fyzicky	physically	MANN
hlavně	mainly, chiefly	RHEM
hloupě	stupidly	ATT/MANN
hodně	plenty, a lot	EXT
hmotně	materially	MANN
hypoteticky	hypothetically	ATT
charakteristicky	characteristically	ATT/MANN
chybně	mistakenly, erroneously	ATT/MANN
i	as well as, also, too	RHEM or conjunction CONJ/GRAD
ideálně	ideally	ATT
intrikánsky	in a plotter's way	ATT
iracionálně	irrationally	MOD/ATT
jak	how	MANN
jasně	clearly	MOD/MANN
jednoduše	simply	MANN
jen	only	RHEM
jenom	only, just	RHEM
ještě	yet, still	RHEM
jistě	certainly, surely	MOD
již	already	RHEM
jmenovitě	nominally	ATT
klasicky	classically	MAN/ATT
koneckonců	after all	ATT
konečně	finally	TWHEN
konkrétně	concretely	APPS (conjunction) or MANN
krásně	beautifully	ATT/MANN
krátce	shortly	MANN (but *krátce po, před* '…before, after' -->EXT)
v krátkosti	in short	MANN
kupříkladu	e.g.	APPS (conjunction)
laskavě	kindly	MANN
legračně	funnily	MANN
lexikálně	lexically	MANN
lingvisticky	linguistically	MANN
logicky	logically	ATT/MANN
lstivě	artfully	MANN
málem	almost	RHEM
matematicky	mathematically	MANN
materiálně	materially	MANN
mazaně	cunningly	ATT
metaforicky	metaphorically	MANN
milosrdně	mercifully	ATT
mimochodem	by the way, incidentally	MANN
mimoto	besides	PREC

morálně	morally	MANN
moudře	wisely	ATT/MANN
možná	possibly	MOD
mylně	mistakenly	ATT
náhodně	accidentally	MANN
náhodou	by chance	ATT/CAUSE
nakonec	finally	TWHEN
náležitě	appropriately	ATT
nanejvýš	at (the) most	RHEM
naopak	conversely	PREC
napovrch	superficially	MANN
např.	e.g.	RHEM
například	e.g.	RHEM
národnostně	ethnically, nationally	REG
následovně	conseqently	MANN
naštěstí	fortunately	ATT
navíc	moreover	PREC
ne (Neg)	not	RHEM
necharakteristicky	uncharacteristically	MANN/ATT
nejapně	ineptly	ATT/MANN
nejen	not only	RHEM
nelogicky	illogically	MANN/ATT
nemilosrdně	unmercifully	ATT
nemoudře	unwisely	ATT/MOD
neočekávaně	unexpectedly	ATT
neoddiskutovatelně	indisputably	ATT
nepochybně	undoubtably	MOD
nepopiratelně	undeniably	MOD
nesporně	incontestably	MOD
nespravedlivě	unjustly	ATT/MANN
nesprávně	incorrectly	ATT/MANN
nešťastně	unluckily	ATT/MANN
netypicky	untypically	ATT/MANN
neuvěřitelně	incredibly	MANN/EXT
nevhodně	inappropriately	MANN/ATT
nevyhnutelně	inevitably	MOD
nevysvětlitelně	unaccountably	ATT
nutně	by necessity	MOD
obecně	generally	EXT
obrazně	figuratively	MANN
obzvlášť	especially	RHEM
odedávna	long since	TWH
odporně	disgustingly	ATT
odtud	from here	DIR1

oficiálně	officially	MANN
omylem	by mistake	CAUS
opět	again	RHEM
opravdu	indeed	ATT
osobně	personally	MANN
but on osobně	he pers.	RSTR
pak	then	PREC
podezřele	suspiciously	MANN/ATT
podobně	similarly	CPR (followed by *jako* 'as' or dative; else PREC)
politicky	politically	MANN
poměrně	relatively	EXT
popřípadě	as the case may be	APPS (conjunction)
potenciálně	potentially	MOD
potěšitelně	gratifyingly	ATT
pouze	only	RHEM
povětšinou	largely	RHEM
prakticky	practically	MANN
pravděpodobně	probably	MOD
právnicky	juridically	MANN
právě	just, exactly	RHEM
právem	by right	ATT
proboha	for goodness' sake	ATT
proklatě	zounds	MANN
prosím	please	ATT
prostě	plainly	ATT/MANN
proto	therefore	CAUS
prvořadě	firstly	RHEM
především	above all	RHEM
překvapivě	amazingly, astonishingly	ATT/RESL
přesně	exactly, precisely	MANN
převážně	largely	RHEM
přibližně	approximately	EXT.APPX
příjmově	as for income	REG
příkladně	for example	APPS (conjunction 'například', e.g.) or MANN
přinejmenším	at least	ATT
případně	as the case may be	APPS (spojka)
přirozeně	naturally	ATT/MANN
racionálně	rationally	ATT/MANN
raději	had better, rather	ATT
radostně	joyfully, cheerfully	ATT
realisticky	realistically	ATT/MANN
relativně	relatively	EXT
riskantně	riskilly, hazardously	ATT/MANN

rovněž	too, also	RHEM
rozhodně	decidedly	MOD/MANN
rozumně	reasonably	ATT/MANN
sem	here	DIR3
schválně	on purpose	CAUS
skoro	almost, nearly	RHEM
skutečně	as a matter of fact	ATT
skandálně	scandalously	RESL
slavně	gloriously	RESL
sotva	hardly, barely	ATT
spíš	rather	PREC
spíš_než	rather_than	CPR.DFR
spravedlivě	justly	MANN/ATT
správně	correctly	MANN/ATT
stejně	equally	CPR (but *přesto* 'all the same'-->PREC
stěží	hardly, scarcely	ATT
striktně	strictly	MANN
stručně	briefly, in brief	MANN
střízlivě	soberly	MANN
šťastně	happily	MANN/ATT
štěstím	of happiness	CAUS
tady	here	LOC
tak	so	MANN/PREC
také, taky	also, too	RHEM
takřka	practically	RHEM
tam	there	LOC/DIR3
teď	now	TWH
technicky	technically	MANN
téměř	almost	RHEM
teoreticky	theoretically	MANN
tj.	i.e.	APPS (conjunction)
to_je(st)	i.e.	APPS (conjunction)
toliko	solely, merely	RHEM
totiž	that is to say, namely	APPS (conjunction)
trochu	a little	RHEM
tudy	this way	DIR2
typicky	typically	MANN/ATT
údajně	allegedly	ATT
upřímně	candidly	MANN
určitě	certainly, definitely	MOD
úskočně	insidiously, treacherously	ATT
už	already	RHEM
vážně	seriously	MANN but: *vážně se to stalo* 'it really happ*ened*' ATT

vcelku	altogether, on the whole	EXT
včetně	inclusive	RHEM
velmi	very	EXT
většinou	mostly, largely	RHEM
vhodně	aptly, conveniently	ATT/MANN
viditelně	visibly	RESL
vlastně	actually	ATT
vnějškově	outwardly	MANN
vnitřně	inwardly	MANN
vskutku	in effect	ATT
všeobecně	broadly	MANN
vůbec	in general, at all	RHEM
výlučně	exclusively	RHEM
výslovně	expressly	MANN
zákeřně	insidiously	ATT
zároveň	at the same time	RHEM
zároveň_i	along with	RHEM
zásadně	on a matter of principle	MANN/MOD
zase	again	RHEM
zatraceně	damn	EXT
zázračně	miraculously	CAUS/EXT
zázrakem	by a miracle	CAUS/EXT
zdánlivě	seemingly	ATT
zejména	particularly	RHEM
zhruba	roughly	MANN
zjevně	avowedly, apparently	ATT/MANN
zpropadeně	damn	EXT
zřetelně	evidently, distinctly	RESL
zvlášť	apart, separately	RHEM

References

Bémová A., Buráňová E., Hajič J., Kárník J., Pajas P., Panevová J., Štěpánek J. and Urešová Z. (1997). *Anotace na analytické rovině: návod pro anotátory* [Annotations on the analytic level: instructions for the annotators.] Technical Report ÚFAL TR-1997-03. Praha: Karlova Univerzita.

Dokulil M. and Daneš Fr. (1958). K tzv. mluvnické a významové stavbě věty. In: *O vědeckém poznání soudobých jazyků*. Praha, 231-246.

Hajič J. (1998). Building a syntactically annotated corpus: The Prague Dependency Treebank. In: *Issues of Valency and Meaning. Studies in Honour of Jarmila Panevová* (ed. E. Hajičová) (pp. 106-132). Praha: Karolinum.

Hajič J. and Hajičová E. (1997). Syntactic Tagging in the Prague Tree Bank. In: *Proceedings of the Second European Seminar "Language Applications for a Multilingual Europe"* (eds. R. Marcinkeviciene and N. Volz) (pp. 55-68). Kaunas.

Hajič J., Hajičová E., Panevová J. and Sgall P. (1998). Syntax v českém národním korpusu. In: *Slovo a slovesnost 59*, 168-177.

Hajič J., Hajičová E., Panevová J. and Sgall P. (1998). Specifikace stromových struktur na tektogramatické úrovni značkování a základní kroky překladu z analytické na tektogramatickou úroveň. Technická zpráva projektu GAČR 405-96-0198 "Formální reprezentace jazykových struktur".

Hajič J., Hajičová E. and Rosen A. (1996). Formal Representations of Language Structures. *TELRI Newsletter 3*, 12-19.

Hajič J. and Hladká B. (1997). Probabilistic and rule-based tagger of an inflective language - a comparison. In: *Proceedings of the Fifth Conference on Applied Natural Language Processing.* Washington, D.C., 111-118.

Hajič J. and Hladká B. (1998). Czech language processing - POS tagging. In: *Proceedings of the First International Conference on Language Resources & Evaluation.* Granada, Spain.

Hajičová E. (1993), *Issues of sentence structure and discourse patterns.* Praha: Karlova Univerzita.

Hajičová E. (1998), Prague Dependency Treebank: From analytic to tectogrammatical annotations. In: *Proceedings of the Conference TSD 98.* Brno.

Hajičová E., Partee B.H. and Sgall P. (1998). *Topic-Focus Articulation, Tripartite Structures, and Semantic Content*. Dordrecht: Kluwer.

Hajičová E., Panevová J. and Sgall P. (1998). Language Resources Need Annotations To Make Them Really Reusable: The Prague Dependency Treebank. In: *Proceedings of the First International Conference on Language Resources & Evaluation.* Granada, Spain, pp. 713-718.

Hajičová E., Sgall P. and Skoumalová H. (1995). An automatic procedure for topic-focus identification. In: *Computational Linguistics 21*, 81-94.

Panevová, J. (1973). Věty s všeobecným konatelem. In: *Studia Slavica Pragensia.* UK Praha, 133-144.

Panevová J. (1980). *Formy a funkce ve stavbě české věty*. Praha: Academia.

Panevová, J. (1996). More Remarks on Control. In: *Prague Linguistic Circle Papers, Vol. 2* (eds. E. Hajičová, O. Leška, P. Sgall, Z. Skoumalová). J.Benjamins Publ. House: Amsterdam- Philadelphia, 101-120.

Panevová, J. (1998a). Ellipsis and zero elements in the structure of the sentence. In: *Tipologija, grammatika, semantika. K 65-letiju Viktora Samuiloviťa Chrakovskogo.* Sankt-Peterburg: Nauka, 67-76.

Panevová, J. (1998b). Ještě k teorii valence. In: *Slovo a slovesnost 59*, 1998, 1-14.

Panevová, J., Benešová, E. and Sgall, P. (1971). Čas a modalita v češtině. In: *AUC - Philologica Monographia 34*, Praha.

Panevová, J. and Skoumalová, H. (1992). Surface and Deep Cases. In: *Proceedings of COLING '92.* Nantes, 885-888.

Panevová, J. - Petkevič, Vl. (1997). Agreement in Czech and Its Formal Account. In: *Formale Slavistik* (eds. U. Junghanns and G. Zybatow), Frankfurt am Main: Vervuert Verlag, 321-333.

Petkevič V. (1995). A new formal specification of underlying representations. In: *Theoretical Linguistics 21*: 7-61.

Petkevič V. (in press). *Underlying structure of sentence based on dependency*. Praha: Karlova Univerzita.

Sgall P. (1967). *Generativní popis jazyka a česká deklinace*. Praha: Academia.

Sgall P. (1992). Underlying structure of sentences and its relations to semantics. In: *Wiener Slawistischer Almanach, Sonderband 33. Festschrift für V. Ju. Rozencvejg*, ed. Tilmann Reuther, 273-282.

Sgall P., Hajičová E. and E. Buráňová (1980). *Aktuální členění věty v češtině*. Praha: Academia.

Sgall P., Hajičová E. and Panevová J. (1986). *The Meaning of the Sentence in Its Semantic and Pragmatic Aspects*. Dordrecht: Reidel and Praha: Academia.

Šmilauer V. (1947). *Novočeská skladba*. Praha: Mikota.