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

P R A H A

MANUAL FOR ANNOTATION OF DISCOURSE RELATIONS IN THE PRAGUE DEPENDENCY TREEBANK

Lucie Poláková, Pavlína Jínová, Šárka Zikánová, Zuzanna Bedřichová, Jiří Mírovský,
Magdaléna Rysová, Jana Zdeňková, Veronika Pavlíková, Eva Hajičová

ÚFAL Technical Report
TR-2012-47

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

Copies of ÚFAL Technical Reports can be ordered from:

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

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

Technical Report

MANUAL FOR ANNOTATION OF DISCOURSE RELATIONS IN THE PRAGUE DEPENDENCY TREEBANK

LUCIE POLÁKOVÁ
PAVLÍNA JÍNOVÁ
ŠÁRKA ZIKÁNOVÁ
ZUZANNA BEDŘICHOVÁ
JIŘÍ MÍROVSKÝ
MAGDALÉNA RYSOVÁ
JANA ZDEŇKOVÁ
VERONIKA PAVLÍKOVÁ
EVA HAJIČOVÁ

Institute of Formal and Applied Linguistics
Faculty of Mathematics and Physics
Charles University in Prague

December 2012

Contents

This work was carried out thanks to the support from the following grant projects: ME10018 *Towards a computational analysis of text structure* under the KONTAKT program, GA ČR P406/12/0658 *Coreference, discourse relations and information structure in a contrastive perspective*, P406/2010/0875 *Computational Linguistics: Explicit description of language and annotated data focused on Czech*, LINDAT-Clarin project of the Ministry of Education of the Czech Republic (project LM2010013) and GA UK 103609 *Textual (Inter-sentential) Relations and their Representation in a Language Corpus*.

1. Introduction

This report serves as an annotation manual for the portrayal of interclausal textual relations (or discourse relations) on the material of the Prague Dependency Treebank (PDT) version 2.5. Within the framework of the tectogrammatical representation (TR), the underlying syntactic structure of sentences including topic-focus articulation and basic coreference relations, has been described in detail (see the "large" manual – *Mikulová et al. 2006, Annotation on the tectogrammatical layer in the Prague Dependency Treebank. Annotation manual.*). The annotation of discourse relations is based on the tectogrammatical representation (tree structures and nodes, syntactico-semantic annotation) and, in some aspects, TR is adopted for the portrayal of discourse relations. This manual maintains the terminology describing the tectogrammatical representation and presupposes at least a basic knowledge of annotation on this layer. There are some newly introduced and explained terms from the field of analysis of discourse relations, especially those that are inspired by a partnership project *Penn Discourse Treebank 2.0*.

The purpose of the annotation of discourse relations is to mark the semantic interconnection of utterances in a textual document and thus allow observing, on a large amount of data, by which language means a sequence of utterances is linked in one coherent complex. In this way, the already existing detailed linguistic analysis "within the sentence" will be supplemented by "inter-sentential" information. In addition to this semantic text structure, another project dealing with text linking is in progress: annotation of extended coreference and bridging relations (*The extended textual coreference and bridging relations, Nedoluzhko and Mírovský 2011*). The output data will combine the annotations of both projects.

The manual is divided into several sections: after the introductory chapters (1 and 2), a part of the manual (Chapter 3) deals with the relationship between the tectogrammatical representation and discourse annotation, or, more precisely, it describes the extent to which it was possible to take the existing tectogrammatical annotation and apply it to the textual annotation. The key chapters of the manual are chapters 4 and 5, which focus on the actual instructions for the annotation of discourse relations in PDT. Chapter 6 discusses various problematic structures and presents the principles adopted for their annotation. Chapter 7 is devoted to evaluation of the manually annotated data.

2. Basic terms

The terms *discourse* and *text* are used synonymously or with little variation concerning the use of *discourse* for language usage in general. *Text* is then used in the sense of written document, artefact, mostly meaning a specific text from PDT on which we carry out the analysis. At the same time, we keep to the traditional use of the word *text* for terms such as *textual coherence*, *textual coreference*, etc. We use expressions such as *segment of the text/discourse*, *text/discourse unit*, *text/discourse argument*, *textual/discourse relations*, etc., synonymously.

The terms *clause* and *sentence* will be used in accordance with the manual for annotation on the tectogrammatical representation in PDT; which means that a *clause* contains one predication, *sentence* is a hyperonym for both clause, and complex and compound sentences as well as for utterance. It is often relevant for our purposes to characterise the sentence as a unit "from full stop to full stop".

Conjunction is understood as a traditional word class category; *linking element* as any language expression with the connecting function at the level of sentence description; (*discourse* or *textual*) *connective* (Czech: *textový konektor*) as connecting elements that have this function at the level of discourse description. This term is based on both the traditional linguistic description in Czech grammars (but we do not define the opposition junctor – connector; on the contrary, we understand junctors, i.e. linking elements between clauses, as coming under discourse connectives) and the translation of the English word *discourse connective*. The term *discourse marker* is one of those linguistic expressions that signal a certain discourse function but are not necessarily limited to the connective function only (i.e. to link two text segments).

The term *discourse/textual relations* (we prefer *textual relations* in Czech, *discourse relations* in English) is problematic. Linguistically in general, it denotes all types of relations that occur in the text; i.e. coreference relations, topic-focus relations and other relations as well. In this manual, however, this term is used in a narrower sense – it is understood as referring to discourse (textual) relations based on syntax, i.e. those discourse relations that usually signal the interconnectedness of the neighbouring clauses or sentences by a specific operator, a discourse connective.

In text linguistics, the terms *coherence* and *cohesion* are often used inconsistently. In this manual, we follow the principle that *coherence* is semantic continuity and consistency of the text and one of the basic preconditions for the intelligibility of the text, while *cohesion* is understood as a demonstration of *coherence* on the surface layer.

The terms *predicate* and its *arguments* are used in Functional Generative Description (FGD, the approach serving as the theoretical basis for the Prague Dependency Treebank) for the characterization of a verb and its complementations. At the level of discourse, by the predicate of a (usually binary) relation is considered the discourse connective that accepts arguments in the form of certain text units, ordinarily clauses.

Note: the English translations of (not only) real-data Czech examples often deal with translation limits. Due to the language differences, some of the translations are the nearest possible approximation to the original Czech text/expression.

3. The relation of the tectogrammatical representation and the annotation of discourse

3.1. Tree structures and tectogrammatical nodes

The tectogrammatical structure represents a sentence as a dependency tree whose root is usually a verb in the predicate function. The tectogrammatical nodes represent autosemantic lexical units; the edges between them usually express dependency, i.e. they represent the relation between the governing and the dependent node. Although the semantic type of this relation belongs to the edge, it is portrayed with the dependent node as its functor. A crucial role in the annotation of discourse is what the tectogrammatical structure (the tree) looks like and which functors belong to the tree nodes.

3.1.1. Coordination and "coap" nodes

Coordination and apposition are realized by "coap" nodes [nodetype = coap] on the tectogrammatical representation. These nodes represent coordinating conjunctions and other linking elements (including some punctuation marks) that link the members of coordination – see Figure 1, where the conjunction *a* (*and*) is represented by a *coap* node with the functor CONJ (conjunction). The members of the coordination are clauses with the governing verbs *potřebovat* (*to need*) and *být* (*to be*).

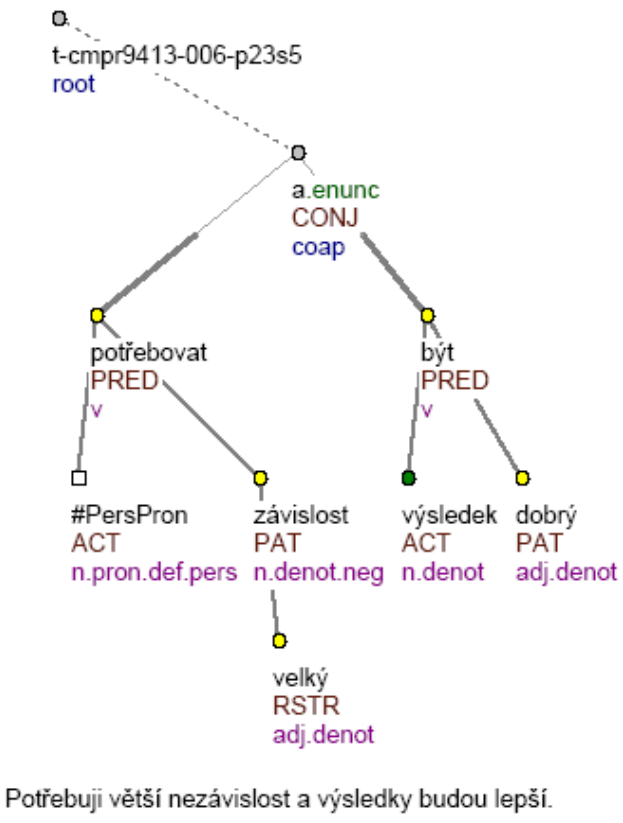


Figure 1: A tectogrammatical tree with a coordination of predicate verbs

(*I need more independence and the results will be better.*)

The t-representation distinguishes between coordination of clauses, coordination of clause elements and mixed coordination (see Chapter 6.1.2 in the large manual). However, the only type relevant for the discourse relations is clausal coordination. The semantic type of coordinate relation (i. e. the functor in the *coap* node) may have the following values (for more details, see Chapters 6.12.1, 6.12.2 a 6.12.3 in the large manual):

ADVS – adversative relation

CONFR – confrontation

CONJ – conjunction

CONTRA – contradiction

CSQ – consequence

DISJ – disjunction

GRAD – gradation

REAS – reason

APPS – apposition

OPER – operations and mathematical intervals

We must decide whether the coordination is really clausal in all nodes with these functors and then check the type of semantic relation (functor) according to the principles of the annotation of textual semantic annotation tags. For further details, see section 3.2, Adoption of a part of the tectogrammatical annotation for the annotation of discourse relations, and Chapter 5, System of textual semantic relations.

3.1.2. Atomic nodes

Atomic nodes on TR [nodetype = atom] represent such expressions that are not incorporated into the sentence by syntactic dependencies but rather modify the meaning of the sentence or its part "from above", i.e. outside the basic syntactic structure. The edge of the atomic node does not represent dependency (intermittently marked). The atomic nodes may have seven functors, of which the following are relevant for the annotation of discourse relations: PREC (reference to the previous context), CM (modifier of coordination), RHEM (rhematizer), ATT (speaker's attitude to the content of the utterance). The expressions with these tectogrammatical functors usually operate (PREC, CM), or may operate, in certain contexts (RHEM, ATT) as discourse connectives.

3.1.2.1. The functor PREC

The expressions with the functor PREC are a basis for defining a group of discourse connectives (Figure 2). This functor is given to those linguistic expressions that connect the clause in which they occur with the previous context. The definition from the large manual: "The PREC functor (reference to the PREceding Context) is a functor for such an atomic node (= without a dependency edge) that represents an expression signaling the linkage of the clause to the preceding context" (p. 534).

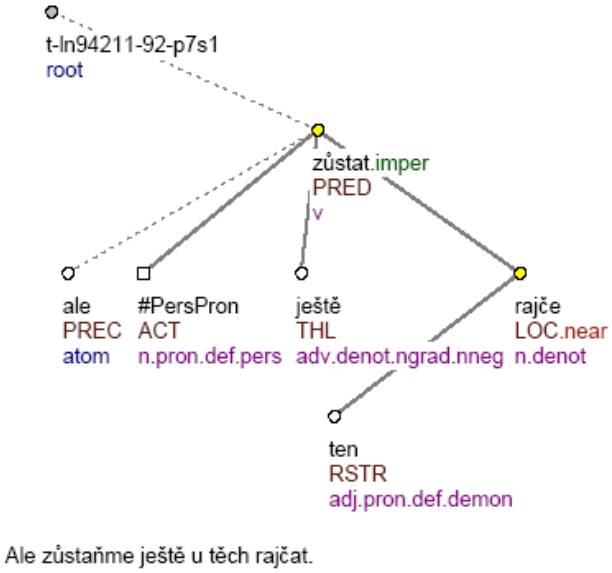


Figure 2: A tectogrammatical tree with a PREC node

(*But let's stay yet with those tomatoes!)*

These expressions comprise a formally heterogeneous class that includes coordinating linking elements, some adverbial expressions and some particles. In most cases, these expressions have also another meaning that is overridden by the meaning of PREC. In other words, the expressions with the functor PREC may be divided (with a few exceptions) into their "original" semantic classes (i.e. they may be assigned another functor). The conjunctions used as PREC still have their original specific syntactic function to link clauses or clause elements. The decision to evaluate them as PREC is made only on the basis of the absence of the first clause in the sentence where the expression occurs. Therefore, the coordinating conjunctions in the initial position always have the functor PREC.[1] The adverbs used as PREC keep also their adverbial characteristics, usually temporal, etc.

The expressions with the functor PREC and the context to which they refer may occur in a single sentence complex although those expressions in most cases/usually go across the sentence boundary. So, PREC combines the reference both within and across the sentence (i.e. "over the full stop") even though the tectogrammatical representation does not describe this explicitly.

1 Coordinating conjunctions have a fixed position between the connected clauses, unlike subordinating conjunctions, which can introduce the whole expression. Thus, coordinating conjunctions in the initial position must refer across the sentence boundaries in which they occur.

If a subordinating conjunction occupies the initial position in a simple sentence, it does not have the functor PREC because subordinating expressions are "hidden" on the tectogrammatical representation, i.e. they do not have their own tectogrammatical node. These cases of an additionally attached dependent clause, i.e. if the sentence is parceled out, are evaluated as the ellipsis of the main clause.

Nevysílají české Události právě pro ty banality. Protože právě jejich znalost by na Slovensku mohla dělat neplechu. – They do not broadcast Czech Události (Events) just for those banalities. Because it is precisely their knowledge that could bring about/do mischief in Slovakia. PDT

PREC is thus the basic functor for the set of expressions which are discourse connectives. The annotation of discourse relations notices mainly these nodes but is not limited to them!

3.1.2.2. The functor CM

The tectogrammatical representation assigns the functor CM (conjunction modifier) to the expressions that modify coordinating linking elements. The tree graph captures the atomic nodes of these expressions with the non-dependency edge as direct descendants of the given coordinating linking expression (see Figure 3). Formally, part of coordinating linking elements may be expressed by most particles and adverbs with the primary function of rhematizers, various structuring particles, some adverbs with the primary function of adverbials and other expressions (see the large manual p. 594). The functor CM is usually not assigned to conjunctions.

The category of CM thus overlaps lexically also with some expressions that occur mainly in the function of PREC. The inclusion of all these expressions into the category of CM is based on whether they are part of multiword linking expressions or whether they modify some linking element in the given context (*nejen*_CM, *ale*_GRAD *i*_CM – *not only*_CM, *but*_GRAD *also*_CM). They also have a specific position in the sentence – they usually stand between two coordinated nodes after the first part of the linking element expressed by the coordinating node (coordinating nodes are mainly basic conjunctions or punctuation marks, mostly the comma (cf. t-lemma = #Comma)).

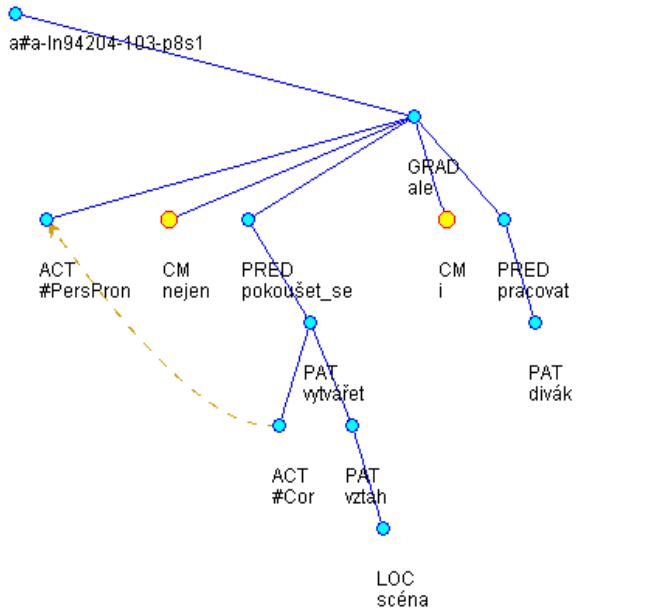


Figure 3: A tectogrammatical tree with two CM nodes

(*We try **not only** to create relationships on stage [but] we **also** work with the audience.*)

One sentence or clause may contain several modifiers of coordination and they (may) occur in various combinations. The typical examples are:

nejen CM *ale* GRAD *třeba* CM *i* CM – *not only* CM *but* GRAD *perhaps* CM *also* CM,
#Comma CONJ *případně* CM *i* CM – #Comma CONJ *possibly* CM *also* CM

If we focus on the relation between the expressions with the functor CM and those with PREC, we find that a certain subset of expressions labelled as CM is actually formed by expressions referring to the "preceding context", which only stand between clauses in one sentence, usually after the comma or the conjunction. The tectogrammatical annotation is interpreted in the way that these expressions lose their linking function, which is adopted by the node of coordination. The cases where this node is represented by a comma are given only by convention, cf.:

Snažím se projet co nejméně, [#Comma] CSQ *proto* CM *jezdím na žižkovské nákladové nádraží. – I try to spend as little as possible*, [#Comma] CSQ *therefore* CM, *I go to the goods station in Žižkov.* PDT

Snažím se projet co nejméně. Proto PREC *jezdím na žižkovské nákladové nádraží. – I try to spend as little as possible.* Therefore *PREC, I go to the goods station in Žižkov.*

If the expression with the functor CM is functionally homonymous with the expression of PREC (i.e. it can occur as PREC as well) and, at the same time, it helps the coordinate structure to select the functor of the node for the coordinate connection, the node with the functor CM along with the coordinate node should be analysed as a multiword discourse connective. Typical representatives of this group are expressions of result, such as *a proto, a tedy, a tudíž, a tak* (all meaning *and therefore*, *and thus*) or adversative expressions *a přece, a přesto* (meaning *and yet*). In these cases, the conjunction may be replaced by a comma.

When sorting expressions with the functor CM into those that operate as a part of the discourse connective and those that do not, we must first of all take into account that a large group of coordinate structures with the modifier CM is coordination of clause elements (not of whole clauses):

V *nákladním autě byly nalezeny náboje, benzín, zásobníky s plynem a* také *mléko. – We have found cartridges, petrol, gas tanks and* also *milk in the truck.* PDT

These coordinations are not evaluated as a matter of discourse with one exception: the node with the functor CM is a negative particle. In such cases, it is necessary to explore the meaning of the sentence and the scope of negation without regard to the tectogrammatical annotation.

The functor CM is closely bound to the linking elements (coap nodes). This means that, theoretically, it appears as a discourse connective very often. The annotation should monitor and capture carefully the cases when the given coordinate relation (and thus also the expression with the functor CM) is to be evaluated in a different way in terms of discourse semantics, i.e. with a tag other than the one it has on the tectogrammatical representation. If so, the node with the functor CM is annotated as a part of a multiple connective; if the *coap* node is a comma, the comma is not annotated as a connective and the CM node itself is usually the connective.

3.1.2.3. The functor RHEM

The functor RHEM stands for rhematizers (or focalizers) on the tectrogrammatical layer. The function of these expressions is to signal the categories of the topic-focus articulation. Usually, they indicate the new, context independent information (the rheme). Their position in the sentence is usually before the element to which they are related (which they rhematize); on the tectogrammatical representation, the edge above the RHEM node determines the position of the rhematizer in the deep structure and so determines its scope (more in the large manual, p. 1102).

A list of Czech expressions which can have the function of rhematizers was created for the purposes of the tectogrammatical annotation in PDT. However, these expressions may often have other functions, mainly the function of adverbials, modal and attitude modifiers (ATT and MOD), conjunction modifiers (CM) or expressions referring to the preceding context (PREC).

The last and the most important homonymy between the functors RHEM and PREC applies mainly to the expressions *také, též, i, rovněž* (all meaning *as well*, *also*)*, zároveň* (*at the same time*)*, spíše* (*rather*)*, nejspíš* (*most likely*), *zase* (*again*), *jen* (*only*) and *naopak* (*on the contrary*). Concerning these expressions, it is necessary to focus on their scope.

Petr uklidil <u>také</u> RHEM *v kuchyni. – Peter <u>also</u>* RHEM *cleaned up the kitchen.*

If the scope of the rhematizer (mainly from the group mentioned above) is wide, i.e. it covers the finite verb, this expression has usually the function of a connective as well.

Petr vyluxoval celý byt. <u>Také</u> RHEM (PREC?) *vytřel podlahy. – Peter has done the vacuum cleaning in the whole flat. He wiped the floor <u>as well</u>* RHEM (PREC?).

In such cases, the interpretation of the rhematizer (marked with the functor RHEM on the tectogrammatical representation) as a connective depends on the context, the right semantic interpretation and, finally, on the annotator's decision in difficult and ambiguous cases.

The rhematizers with the narrow scope, i.e. with the range only to one sentence element (which is not the finite verb), must be examined also in terms of their position in the text. Usually, they are not connectives but some such expressions in the initial position clearly have the text structuring function. The research of these rhematizing particles is not yet completed. This issue is generally wider and is related to the categories of the topic-focus articulation in the Czech sentence. It is necessary to deal with the potentially rhematizing and text structuring expressions in more detail on the basis of the annotated data in the near future.

3.1.2.4. The functors MOD and ATT

The functor ATT (attitude) applies to words expressing certain attitude (evaluative or emotional opinion) of the speaker to the content of the clause or its part (more in the large manual, p. 529). The functor MOD (modality) applies to expressions with the modal nature (p. 533). The elements with the functors ATT and MOD show similar behaviour as rhematizers: they express a degree of probability or some attitude to the content and they thus refer to the whole utterance or to its part (to a certain subtree).

The expressions with the functor MOD do not usually participate in the discourse structuring, and it is not necessary to pay any special attention to them. Nevertheless, the semantic category of ATT, as defined in the annotation scheme of PDT, contains expressions whose attitudinal (evaluative) semantics is strong in a variable degree: the weaker it is, the higher the probability that they will also have other functions – e.g., the ability of rhematization or the ability to refer to the previous context. The functor ATT in the tectogrammatical annotation overlaps lexically with the functor PREC, especially in the case of the expressions *vždyť* (*indeed*), *stejně* (*equally/anyway*), *ovšem* (*certainly*), *ostatně* (*after all*), etc. These expressions may be evaluated as connectives adding a supplemental explanation or explication with other semantic differences of emphasis, affirmation or concession, see the following examples:

Naší oporou by mělo být i fantastické domácí publikum. Vždyť ATT *máme kapacitu stadionu 5000 míst a dva týdny před ligou už jsme prodali 3000 permanentek. – Our support should be also a fantastic home audience. Indeed* ATT*, we have the capacity of 5000 seats and we have already sold 3000 season tickets two weeks before the league.*

This example demonstrates that the expression ATT is here replaceable by the expression *totiž* (*because*) – the relation to the previous sentence remains the same but the evaluative meaning is lost.

The particles structuring the text are also expressions such as *krátce* (*shortly*), *prostě* (*simply*), *vlastně* (*actually*), *zkrátka* (*in short*), which are annotated as ATT or MANN in PDT, i.e. never as PREC:

Spadl z hradby a srazil si vaz nebo co, zkrátka už nevstal. – He fell from the wall and broke his neck or something like that; in short, he did not get up any more.

The cases in which these expressions are sentence elements may be relatively easily recognized from their functions of utterance content modifiers. Except for the function of sentence elements, these expressions may be evaluated as connectives with a secondary meaning of speaker's attitude.
We take the following aspects into account:
1. The core of the expressions ATT is relatively stable – it expresses the speaker's attitude "strongly" and it does not usually have another function (e.g., *bohužel – unfortunatelly*, *jaksi – somehow*, *naštěstí – fortunately*, *pochopitelně – of course*, *evidentně – obviously*).
2. Discourse connectives themselves clearly indicate the presence of the "first" argument, i.e. they connect two or more arguments that enter into a discourse relation. The core of the expressions ATT does not meet this requirement while the example above does.

3.1.3. Ellipses in tectogrammatics and in the annotation of discourse relations

We have prepared detailed instructions for capturing different types of ellipses. In terms of discourse structure, this issue is closely linked to the definition of text units. The annotation of discourse relations uses the analysis of elliptical expressions on the tectogrammatical representation and usually follows this analysis. The detailed rules for annotation of ellipses can be found in the chapter on problematic structures.

3.2. Adoption of part of the tectogrammatical annotation for the annotation of discourse relations

Some tectogrammatical functions (functors) represent the same meanings as the textual semantic tags. From the theoretical point of view, this means that we admit that some relations between the clauses within a sentence are the same as between the (variously large) text units within the text. An example is conjunction (conj) or condition (cond).

Some tectogrammatical functors are therefore the same as the textual semantic tags. These functors are taken to the textual annotation without modifications. Some meanings had to be revised, more subtly distinguished or regrouped to fit better to the description of discourse. Thus, new tags arose, unused in tectogrammatics (e.g., specification, correction, restrictive opposition) and some were defined differently (opposition, concession, reason – result, etc.).

We have adopted, i.e. automatically looked up and copied, the following items from the tectogrammatical annotation to the annotation of discourse:

1. a number of dependency relations between clauses within sentences
2. a number of coordinate relations with the *coap* nodes (only coordination of clauses)

Ad 1: Dependency relations in the tectogrammatical tree are not re-annotated. The only exceptions are cases with the pragmatic semantic relation of reason, condition or contrast. In these cases, we re-annotate the dependency edge by a discourse arrow with the appropriate type of pragmatic relation.

Ad 2: Some functors of the *coap* type (i.e. coordination and apposition) are (i) taken into the annotation of discourse relations, i.e. they are not re-annotated by different tags; (ii) some are not taken into account at all (e.g. the functor CONTRA is associated only with the coordination of sentence elements); (iii) some need to be revised. The relation between the members that are linked by the given linking element may be different on TR and in the textual annotation. Moreover, the scope of the arguments (members of coordinate relation) may be different as well.

Ad (iii) The syntactic relations of the coordinate relations on the tectogrammatical representation (represented by *coap* nodes) are re-annotated by the discourse relations, i.e. we draw the discourse arrow between the coordinated members in the following cases:

- CONJ – if the relation is not pure conjunction

- ADVS – if it has another contrastive meaning than opposition (opp) and mainly if the meaning is unambiguously concessive (conc)
- clausal APPS – if possible, we divide them more subtly, i.e. they are always assigned a new type of arrow (spec, equiv, gener, etc.)
- we carefully check the meaning of relations with the linking elements *což* (*which*) in all its forms, *aniž* (*without*) and *pročež* (*consequently*), and, if needed, we re-annotate them

This applies in principle also to other relations with *coap* nodes: DISJ, CONFR, CSQ, GRAD, REAS, OPER

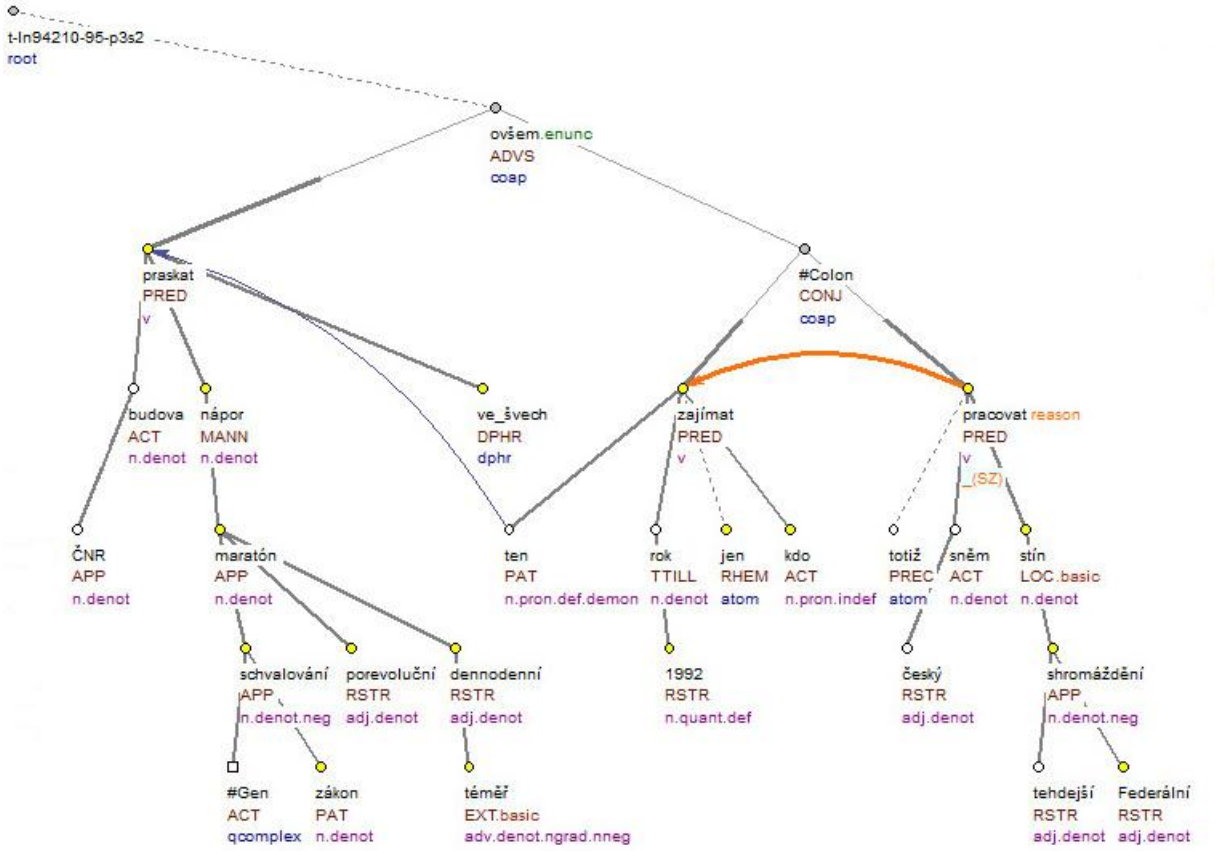


Figure 4: A tectogrammatical tree with a re-annotated coordination (CONJ) to discourse reason – result

Pod náporem téměř dennodenního porevolučního maratónu schvalování zákonů budova ČNR praskala ve švech, ovšem do roku 1992 to zajímalo jen málokoho: český sněm totiž pracoval ve stínu tehdejšího Federálního shromáždění.
Under the onslaught of almost everyday post-revolutionary marathon of passing laws, the building of ČNR bursted at the seams, but only a few people were interested until 1992: because the Czech Parliament had worked in the shadow of the former Federal Assembly.

4. The principles of annotation

4.1. Discourse arguments and connectives

As **argument** in the discourse annotation is inderstood (in this phase of the project) a syntactic structure with a finite verb. An argument may be realized by a syntactically independent clause (whether in terms of a sentence between two final punctuation marks or in terms of a coordinated part of a compound sentence) but also by a dependent clause, which is not, however, a valency complementation of the predicate of its main clause (except for nominal content clauses).

An infinitive structure is not a discourse argument. The only exception is when it has the functor PRED. (The t-representation allows such cases, e.g., *Proč o tom uvažovat? – Why to think about it?*)

The extent of an argument is derived from the principle of minimality – the argument includes only the minimum number of clauses that carry the semantics of the relation. The principle is not related so much to dependent clauses in a single tree but rather to the number of trees included in a single argument. Attributive clauses are considered a part of the argument. Removing an attributive clause must be justified. Other types of dependent clauses are subject to the same principle.

Connectives of discourse relations are primarily the nodes with the functor PREC and the nodes coap; however, annotations are not limited only to them. Generally, an expression is a discourse connective if it opens two positions that are occupied by two valency independent syntactic structures containing a finite verb (i.e. the arguments). The language means with the connective function necessarily express the semantic relation of the arguments. In the first stage, the annotation is limited only to the discourse relations with connectives.

Whether the expression in question is a discourse connective or not is always dependent on the particular context. Some connectives are typical of discourse relations (e.g., *totiž – because, však – however*), some of them become connectives only in certain contexts (*jinak –*

otherwise, *podobně – similarly*, *naproti tomu – on the contrary*, etc.). Discourse connectives are the expressions of the following classes:

a) coordinating conjunctions: *a* (*and*), *ale* (*but*)*, však* (*but*), *nebo* (*or*), *proto* (*therefore*)*…*
b) subordinating conjunctions: *ačkoliv* (*although*)*; místo, aby* (*instead*)*; s tím, že* (*with the fact that*) *…*
c) particle expressions (including rhematizers): *totiž* (*because), ovšem (however), zkrátka (shortly), dokonce (even), také (too), například (for example)…*
d) adverbs: *potom* (*then*)*, následně* (*as follows*)*, stejně* (*equally/alike*)*, současně* (*at the same time*), *tak* (*so*) *…*
e) certain uses of pronouns: *kromě toho* (*except for this*), *k tomu* (*in addition to this*), *naproti tomu* (*on the other hand*), *tím* (*by this*) *…*
f) idiomatic multiple-word connective means formed by linking of different expressions: *na jedné straně* (*on the one hand*)*, stručně řečeno* (*in short*)*, jinými slovy* (*in other words*)*…*
g) elements formed by letters or numbers expressing enumeration: *a), b), 1., 2.…*
h) two punctuation marks: colon and dash.

4.2. Discourse relation between two arguments; the list structure

The annotation of discourse relations in PDT differentiates discourse relations between two arguments, and the list structure. By a discourse relation between two arguments is meant the semantic relation of conjunction, gradation, instantiation, etc. (for more details, see Chapter 5). In the first stage, we only annotate discourse relations between two arguments with an expressed (explicit) connective.

A list structure is an enumeration which can be marked (e.g., by numbers or letters) or unmarked, in most cases preceded by an opening statement with a colon (in our terminology the hypertheme). The hypertheme is not considered a list item. The annotation includes only evident list structures – the characteristic features of which are the enumeration. Clearly definable items which are similarly structured have an approximately equal length (a maximum of 3–5 sentences) and they contain a finite verb. Other structures are provided with an annotator's comment "a candidate for a list". The list may appear both with and without explicit connectives (asterisk, bullets, etc. may be also evaluated as connectives if there is a strong reason to do so). The list is introduced by a hypertheme – if it contains a verb, the

arrow leads from the first item to this verb although the list specifies only the nominal phrase in such a case. If the hypertheme contains only DENOM, the arrow leads to the node with this functor. (We are leaving aside the possibility of a list without a hypertheme for now because such an example was not yet found in the actual annotations.)

4.3. The annotation in TrEd

The annotation of discourse relations in PDT consists in plotting the arrows between the arguments of these relations in TrEd and in marking some elements relevant for the structuring of a text. The arrows depicting discourse relations are described in section 4.3.1; some elements relevant for structuring of a text are introduced in section 4.3.2.

4.3.1. The discourse arrow

The discourse arrow always leads from the second argument to the first; it links the highest nodes of arguments.

All attributes are assigned to the initial node, from which the arrow leads (usually the node PRED or coap). The order of arguments is chosen in a way which enables the semantic label of relation – e.g., REASON (reason – result), PRECED (precedence – succession) – to be described by an argument expressing the given semantics (the second argument in REASON expresses the reason of action; the second argument in PRECED expresses the action preceding the action of the first argument). The order is irrelevant in case of some relations and it is chosen in a way so that the first argument is on the left; e.g., in the relation CONJ (conjunction) and EQUIV (equivalence). The table of all relations along with the order of arguments is presented in Chapter 5 (Table 3).

Some attributes of the arrow (type, discourse type, connectors.rf, start_range, target_range, src and comment) are directly visible on the initial node of the arrow or when moving the mouse on the arrow: after a while, a frame with the type of the arrow and connective appears. All attributes are then displayed when clicking on the initial node of the arrow.

4.3.1.1. Attributes and their values

4.3.1.1.1. Type

The attribute type expresses a type of the arrow. It may have two values – *discourse* (a thick orange arrow in the annotation window) and *list* (a thin orange arrow). The value *list* belongs to the arrows among the items of the list and the hypertheme; the value *discourse* to all other arrows (see discourse type below).

4.3.1.1.2. Target_node.rf

The attribute target_node.rf contains the identifier of the target node of the arrow. If the target node is missing (from the first item of the list in the case of the arrow *list*), it remains undefined.

4.3.1.1.3. Start_range a target_range

The attribute start_range captures the extent of the second argument of the relation that is marked by the arrow; the attribute target_range expresses the scope of the first argument. The extent of the argument is always indicated in the number of trees (or subtree and trees).
The attribute start_range (target_range) may have these values:
"0" means that the argument is a subtree of the node inclusive, i.e. a subtree of the node from/to which the arrow leads. This value of the attribute start/target_range is set automatically to the arrows of the *discourse* type and specifically to the arrows of the *list* type (the arrows of the *list* type have only start_range, which is set automatically according to the number of trees in the item – see 4.3.1.3).
"n" where n>0 means that the argument of the relation is a subtree of the node from/to which the arrow leads + n of the following trees – this means that the arrow always begins and ends in the tree/subtree being most on the left that belongs to the argument. See the schematic representation for 2 situations:

trees	1	2	3	1	2	3
arguments	arg 1			arg 2		
the arrow	the tree to which the arrow leads			the tree from which the arrow leads		
attributes				start_range 2 target_range2		

Table 1: Annotation of the attributes start_range and target_range: order Arg1 – Arg2

trees	1	1	2	3
arguments	arg 2	arg 1		
the arrow	the tree from which the arrow leads	the tree to which the arrow leads		
attributes	start_range 0 target_range2			

Table 2: Annotation of the attributes start_range and target_range: order Arg2 – Arg1

"**group**" expresses that the argument is a group of nodes which has a clear boundary and does not consist exclusively of the whole trees. Groups participate in the annotation of the relation between the main and the dependent clause if we disagree with the semantic interpretation captured already on the t-layer. The identifier of a given group of nodes (a positive integer) is in the attribute start_group_id, or in target_group_id.

"**forward**" means that the argument of the relation is the subtree of the node from/to which the arrow leads + an indefinite number of sentences toward the end of the text (i.e. if it is impossible to identify a clear boundary).

"**backward**", on the contrary, means that the argument is the subtree of the node from/to which the arrow leads + an indefinite number of sentences toward the beginning of the text (again if it is impossible to identify a clear boundary).

4.3.1.1.4. Start_group_id and target_group_id

The attributes start_group_id and target_group_id contain an identifier of a group of nodes (a positive integer) if the attribute start/target_range has the value *group* (i.e. if the first and/or the second argument is formed by a group).

4.3.1.1.5. Discourse type

The attribute discourse type expresses a semantic type of a discourse relation. This attribute may have a total of 23 values, e.g., CONJ (conjunction), SPEC (specification), CONC (concession), etc. All these values are specified in detail in Chapter 5.

4.3.1.1.6. Connectors.rf

This attribute contains the id of all nodes from the tectogrammatical and the analytical layers that form a connective of the given relation.

4.3.1.1.7. Src

The attribute src contains the abbreviation of the annotator who annotated the given relation. All arrows are assigned this attribute when the annotator submits his or her data.

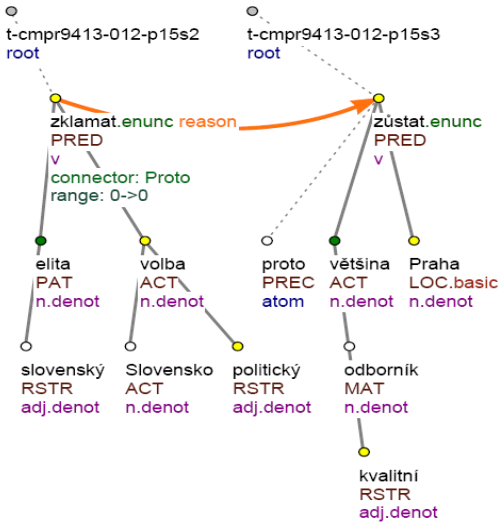


Figure 5: An example of annotation with the discourse arrow

Slovenská elita byla zklamána politickou volbou Slovenska. Proto (PREC) *většina kvalitních odborníků zůstala v Praze.*
The Slovak elite was disappointed by the political choice of Slovakia. This is why (PREC), most (of the) quality experts remained in Prague.

An example of the annotation of a list structure:

(9) *K tomu, aby zaměstnavatel pracovníkovi za škodu opravdu odpovídal, musí být splněny tyto podmínky* ***(hypertéma)****: – The following conditions must be met so that the employer can be really responsible for the damage caused to his worker* ***(hypertheme)****:*
(10) *1.* ***(PREC)*** *Zaměstnanci musí vzniknout škoda, tj. musí dojít k určitému snížení hodnoty jeho majetku (v některých případech mu vzniká i právo na náhradu ušlého zisku). – There must be damage caused to the employee, i.e. the value of his assets must be reduced (in some cases, he has even the right to compensation for loss of profits).*
(11) *2.* ***(PREC)*** *Zaměstnavatel nebo jiná fyzická či právnická osoba, která jedná jeho jménem, musí porušit své právní povinnosti. – The employer or another physical or legal person acting on his behalf must violate his or her legal obligations.*

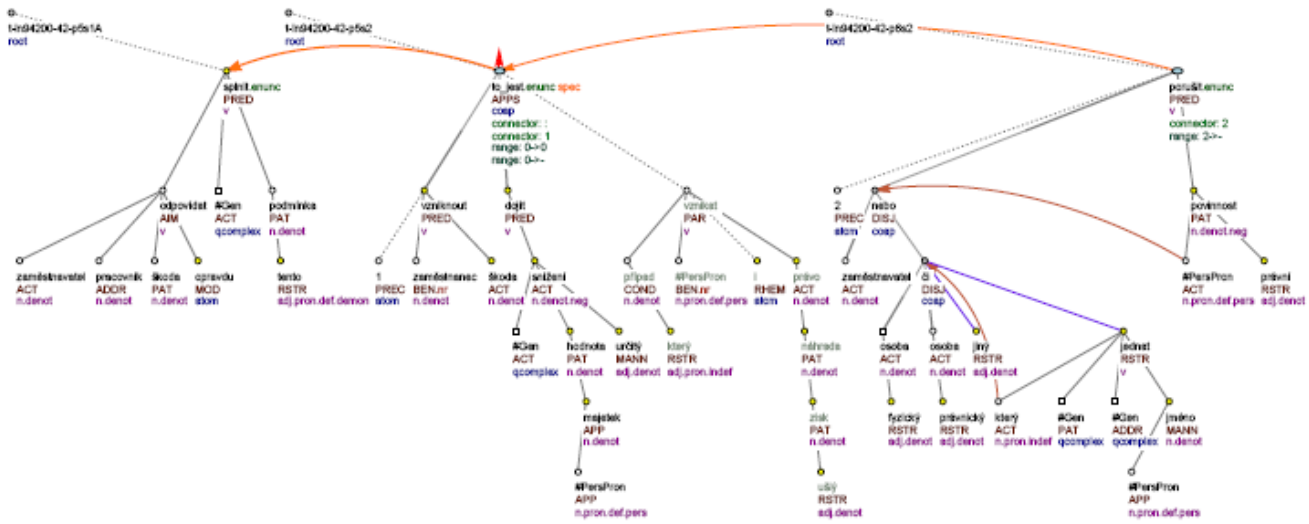


Figure 6: An example of annotation of a list structure

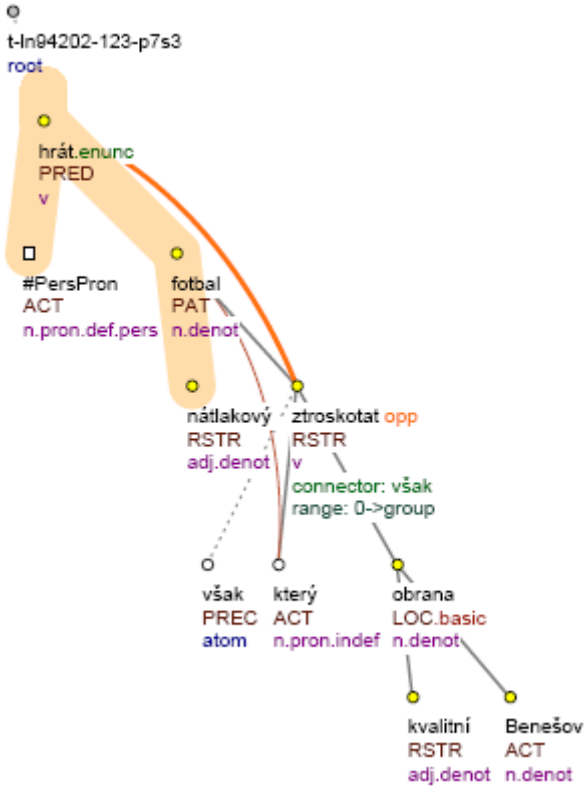


Figure 7: An example of annotation using a group – a structure with a pragmatic relative clause

Chtěli jsme hrát nátlakový fotbal, ***který*** *<u>však</u> ztroskotal na kvalitní obraně Benešova.*
We wanted to play coercive football, which, however, failed due to the quality defence of Benešov.

4.3.2. Other annotated issues

4.3.2.1. The attribute is_heading – the annotation of headings

The attribute **is_heading** may have the values 0 and 1. If there is no value in the attribute, 0 is assumed. The value 1 is assigned to the root of the (sub)tree that represents the heading or subheading in the text. Both headings and subheadings are annotated without distinction. On the contrary, the authors' names, their abbreviations, the location and the source of the article or other information about the text (e.g., about the place of the event, the year of publication of a reviewed book, the cast of a theatrical performance, etc.) are not annotated at all, see the examples:

Parlament rozšiřuje (is_heading = 1) *své sídlo – Parliament extends* (is_heading = 1) *its headquarters*

V bloku malostranských paláců a měšťanských domů vznikne (is_heading = 1) *přes sto poslaneckých kanceláří se 180 pracovními místy – Over a hundred of MP offices with 180 jobs will arise* (is_heading = 1) *in a block of palaces and town houses in the Lesser Town.*

Jan (is_heading = 0) *Štětka*

Když se koncem osmdesátých let rekonstruovala historická budova českého sněmu na Malé Straně v Praze, nikdo z architektů asi nepředpokládal (is_heading = 0)*, že už za pět let bude svým uživatelům malá... – When the historic building of the Czech Parliament in the Lesser Town in Prague was reconstructed in the late eighties, none of the architects expected* (is_heading = 0) *that it would be small for its users in five years ...*

5. System of discourse semantic relations

Text units are interconnected by semantic relations, which were classified for annotation in PDT both on the basis of functors on the tectogrammatical representation and on the hierarchy of semantic tags used in the Philadelphia Penn Discourse Treebank project. Some types of semantic relations appear also in the syntactic annotation of PDT, i.e. "within the sentence"; however, others do not and, therefore, we introduce them newly. At the same time, we maintain the Philadelphia concept in the way that relations are divided according to their semantics into four basic categories: *temporal*, *contingency* (causal relations), *comparison* (contrastive relations) and *expansion* (broadly conceived conjunction, elaborative relations). The system of relations also includes the so called pragmatic relations, i.e. relations that are formally similar to semantic relations (they may have even the relevant connective of the given semantic relation) but their meaning is different – very often, they express a relation of presuppositions or another pragmatic phenomenon.

A complete list of annotated relations with examples is given in the following Table 3. These textual semantic relations are then individually analysed in this chapter. The second column of the table includes the English term of the particular relation (mostly used also in the Penn Discourse Treebank if the same or a similar relation is included there – cf. *PEDT Annotation Manual 2.0*). The third column contains the abbreviation used for the annotation of the relation in PDT. The fourth column contains information about the order of the arguments (and about the direction of the arrow for practical annotation) – in some relations, the order is important because each of the arguments is different in nature (e.g., the argument with condition, the argument with the result of condition); in other relations, the order does not matter and the direction of the arrow is given conventionally.

Intersentential semantic relations do not have to be always signalled by a textual connective. At the current first stage, however, we annotate only those relations where the connective is present. Nevertheless, the examples used in the table of relations and further below do not sometimes include it – we want to demonstrate that the connective does not have to be necessarily present in the text and yet the relation in the given context remains unchanged.

Table 3: System of discourse semantic relations for PDT

The name of the relation		**abbreviation**	**orientation – <u>the arrow always leads to A1</u>**	**Example**
TEMPORAL	(A & B)			
precedence – succession		Preced	A2 happens first, the arrow usually leads to the right	*The lamp sputtered for a while. Finally, it extinguished.* *The lamp extinguished. Before that, it only sputtered for a while.*
synchronous		Synchr	A1 is on the left	*The tenth hour struck and the lamp was still shining.*
CONTINGENCY	(A -> B)			
reason - result		Reason	A2 is reason, A1 result	*He was dismissed because he worked irresponsibly.* *He pulled the rope hard. It snapped.*
pragmatic reason – result		*f_reason*		*Grandmother is home because the lights are on in the kitchen.*
condition – result of the condition		Cond	A2 is condition, A1 result of the condition	*I will make pancakes. But first you must buy eggs.*
pragmatic condition – result of the condition		*f_cond*		*Are you thirsty? There is juice in the fridge.* *If you do understand it, so I do not.*
purpose		Purp	A1 is action, A2 purpose	*She goes to train regularly. She wants to lose weight.*
explication		Explicat	A2 is explication	*He is a thief. He was shop lifting.*
COMPARISON	(A & B)			
confrontation		confr (previously juxt)	A1 is on the left	*The worker is mortal, the work is alive, Anthony is dying, the bulb is singing.*
opposition		Opp	A1 is on the left	*He heard everything. But he saw nothing.*
pragmatic opposition		*f_opp*	A1 is on the left	*It is going to rain this weekend. But Czechs will block the highways anyway.*
restrictive opposition + exception		Restr	A2 is restrictive opposition or exception	*I will come. I only do not know when.*
concession		Conc	A2 is what is usually in the dependent concessive clause	*They died. And yet they still speak.*
correction (or replacement) + chosen alternative (substitution)		Corr	A1 is negated or corrected by the second sentence	*He did not wait at home. He followed her to work.*
gradation		Grad	A1 is lower degree	*He was running. What is more, he was speeding.*
EXPANSION	(A & B)			
conjunction		Conj	A1 is on the left	*He went straight. He did not look left nor right.*
instantiation		Exempl	A1 is more general, A2 is an example	*She never spent evenings at home. For example, she went for walks with friends.*
specification		Spec	A1 is more general, A2 is detail, not example	*He tries to reduce debt. He earns more money.*

equivalence		equiv	A1 is on the left	*The method is up to you. Just do it by yourself.*
generalization		Gener	A1 is more specific, A2 more general	*They lent him some money. In short, they helped him.*
conjunctive alternative		Conjalt	A1 is on the left	*We may go to the cinema. Or we may go for a coffee (or both).*
disjunctive alternative		Disjalt	A1 is on the left	*Behave decently. Or do not come here!*

5.1. Temporal relations (TEMPORAL)

The basic semantics of temporal relations may be described as A & B, where A is valid and B is valid; within this group, we speak about their time correlations. Most of the relations that express a certain period of time are reflected already on the TR.

In the annotation of discourse relations, we mark only the relations **preced** (precedence and succession) and **synchr** (simultaneity), namely in those cases where the given relation is expressed by a textual connective (e.g., *zároveň – at the same time, zatímco – while, pak – then, předtím – before, nakonec – finally*). The temporal expressions such as *now*, *today*, etc., which refer to some extralinguistic temporal situation and not to another proposition, are not annotated; nor do we mark constructions with two temporal expressions such as *dosud* (*so far*) – *nyní* (*now*), *dříve* (*earlier*) – *dnes* (*today*) because they are not primarily temporal but contrastive expressions that have been already annotated under the annotation of the topic-focus articulation.

Mainly within the temporal relations, it happens frequently that there are two relations between the arguments – very often opposition together with some temporal connective (*pak ale – but then, nejdřív ovšem – but firstly, a zároveň – and at the same time*...). For two relations between the same arguments, we have established a standardized comment *second_rel*. In the trees, we annotate only the relation considered to be primary; the second relation is only mentioned in the comment as *second_rel*.

This situation is typical especially of OPP and PRECED; within the potential coexistence of CONJ and PRECED, PRECED is always more important (and in this case, we do not write CONJ as the second relation in the comment).

Temporal connectives are in most cases: *potom, pak, poté, posléze, vzápětí, následně* (all synonyms for *then*, *thereafter*). Rather temporal connectives are: *mezitím* (*meanwhile*)*, dále* (*further*). *Později* (*later*) is sometimes a temporal connective. In most cases, connectives are

not (always dependent on the context): *už* (*already*), *ještě* (*still*), *okamžitě* (*immediately*), *tehdy* (*at that time*), *snad* (*perhaps*), *nakonec* (*finally*), *konečně* (*finally*), *nadále* (*hereafter*), *dosud* (*so far*), *opět* (*again*), *znovu* (*again*), *zatím* (*for now*). If there are multiple connectives within the sentences such as *nejdříve – potom* (*firstly – then*), *nejprve – pak* (*firstly – then*), we add both parts to the arrow.

5.1.1. Precedence – succession (asynchronous)

preced

Argument 1 is the proposition happening later in time and the arrow always leads to this position. Therefore, this relation may have two realizations: either the order of the arguments in the text corresponds to the time behaviour (as in the first example) or the later proposition in real time (second example) occurs in the text firstly. These two realizations differ only in the direction of the arrow.

Lampa chvíli prskala. Nakonec zhasla. – The lamp sputtered for a while. Finally, it extinguished. (the arrow leads to the right, i.e. to the word *zhasnout – extinguish*)

Lampa zhasla. Předtím chvíli jen prskala. – The lamp extinguished. Before that, it only sputtered for a while.
(the arrow leads to the left, i.e. to the word *zhasnout – extinguish*)

Note to the connective *později (later)*:
Although the expression *později (later)* carries more lexical meaning than *potom* (*then*) or *vzápětí* (*soon, immediately*), we consider it a connective just like *potom* (*then*) and *vzápětí* (*soon, immediately*). See the following example.

*Po činu z místa utekl **a později** se přihlásil na policii. – He escaped after the crime **and later** he owned up to the police.*

There are exceptions for cases when this expression has absolutely clearly the function of adverbial of time. Most probably, it depends on the fact that it is in rhematic position (the rheme should not be the connective) – see the example:

V půl jedné přišla Jana. Eva přišla později. – Jane came at half past one. Eve came later.

5.1.2. Simultaneity (synchronous)

synchr

Both propositions are happening simultaneously in time, argument 1 is the first proposition in the sequence of the text. The arrow thus leads always back (to the left).

Město postihla krize a nezaměstnanost. Zároveň začala nová éra svobodných celních zón. – The city was affected by crisis and unemployment. At the same time, a new era of free customs zones started.

In this relation, we must also pay attention to another meaning of the temporal connective than the original. In the given context, it does not have to mean simultaneity but it may express either precedence or the structuring of the text (not the structuring of the temporal event) – see the example (annotated as *preced*):

P. Dvorský zahájí program áriemi od B. Smetany a A. Dvořáka. K této literatuře se hlásím jako k vlastní, řekl Dvorský. Zároveň připomněl, že v Čechách se mu vždy dostávalo velké pozornosti. – Mr. Dvorský will start the program by arias by Smetana and Dvořák. I accept this literature as my own, said Dvorský. He also noted that, in Bohemia, he always received great attention.

5.2 Causal relations (CONTINGENCY)

This group contains extended causal relations, most of which are expressible both at the level of the sentence and between individual sentences or larger text units (we may argue about the purpose relation that is expressed intersententially almost exclusively by means of the modal verb *chtít – to want* or its synonyms). The basic semantics of this group may be expressed in the sense A -> B, A implies B or A is (causally) related to B. Partially, this group includes the relation of textual concession but it also belongs to the group of *comparison* (contrast), whereit is left for our needs. See more details on the semantics of concession in a separate chapter.

5.2.1 Reason – result

reason

Reason – result is a very common textual relation, the most common way of expressing causality in the text. Similarly to precedence – succession, it is usually realized "on both sides", the order of arguments is arbitrary. The proposition expressing reason is always A2. The arrow always leads to the proposition expressing result (A1).

Dostal výpověď. Pracoval totiž nezodpovědně. – He was dismissed because he worked irresponsibly.
Tahal za lano silně. Utrhlo se. – He pulled the rope hard. It snapped.

In some cases, a purely semantic relation of reason and result is difficult to distinguish from explication. See further details on this issue in the chapter about explicative relations. Tectogrammatical functors CSQ, REAS and CAUS are left as they are. They all express the relation of cause and effect. Although it is possible that in some cases they have a rather explicative meaning, there is still no reliable rule to distinguish *reason* and *explication*. Therefore, we do not pay attention to them for now. However, if there is a case that would not seem to be reason – result, it will be re-annotated as *explicat*.

5.2.2 Pragmatic reason – result

f_reason

Semantic relations of reason and result are not seemingly causal structures (e.g., expressed by the conjunction *protože – because*), where causality is understood by means of some inference or unexpressed content that the reader/listener may simply infer. In such cases, we speak about the relation of the so called pragmatic reason – result that is annotated as *f_reason*. Cf. the examples:

Petr je doma, protože čeká na opraváře. – Peter is at home because he is waiting for a repairman. reason

Petr je doma, protože se svítí v oknech. – Peter is at home because the lights are on. f_reason

The second example could be paraphrased as follows: I suppose that Peter is at home because I see the lights in his windows. Thus, the relation of reason and result exists only between the fact that the speaker thinks something and why he thinks so, not in the content of his thought. The arrow leads, as well as in the real reason – result, always to the proposition expressing the pragmatic result (A1).

5.2.3 Condition – result of the condition

cond

The conditional relation is usually expressed between clauses within one sentence, namely by a clear and limited repertoire of subordinate conjunctions (*jestliže, kdyby, když, -li, pokud*... all meaning *if*). Sometimes, however, it may be expressed asyndetically or by means of coordinating linking elements, modal verbs or the interrogative and the imperative.

For the practical annotation, A2 is condition, A1 the result of the condition. It means that the arrow leads from the proposition expressing condition to the proposition expressing result of the condition.

The main types of condition are the following:

1. with the connective *pokud – if* or *pokud ale – but if*, *pokud ovšem – however if*:

Půjdu na výstavu. Pokud ovšem stihnu napsat aspoň kousek diplomové práce. – I will go to an exhibition. But only if I manage to write at least a piece of my thesis.

2. with the verbs *muset* (*must*) + connectives *ale*/*ovšem* (*but/however*) + connectives *předtím/nejdřív* (*before/firstly*):
Nakreslím ti formuli. Ty mi ale nejdřív musíš přinést pastelky. – I will draw you formula. You must, however, firstly bring the crayons.

3. with the verb in the imperative + connective *ale* (*but*), *ovšem* (*however*):

Nakreslím ti auto. Přines mi ale pastelky. – I will draw you a car. But bring me the crayons.

Pay attention: However, if A1 is without the connective *ale* (*but*) or *ovšem* (*however*), we do not evaluate it as condition. It is only a command:
Nakreslím ti formuli. Přines mi pastelky. – I will draw you a formula. Bring me the crayons.

4. with the verbs in the imperative or with one in the imperative and one in the future tense:

Sáhni na to a jsi synem smrti. – Touch it and you are a goner.
Sáhni na to a já tě zabiju. – Touch it and I will kill you.

This subgroup of condition was found only within the sentence (tree) till now. In the case that such a relation occurs in the tree and is not annotated as condition on TR, it must be re-annotated.

5. infinitive + the verb in the conditional
Být na tvém místě, tak bych tam šel. –If I were in your shoes [literally in Czech *To be in your shoes*], *(so) I would go there.*

5.2.4. Pragmatic condition – result of a condition

f_cond

There are two types of pragmatic condition:

1. These are the cases where A1 is not expressed but implied as a positive answer to the question in the sense of the inferred *if*. A2 may (but does not have to) contain the connective

tak (*so*) or *pak* (*then*) + the imperative or the conditional. True semantic condition exists here between the presupposition and the second argument. These cases are marked as *f_cond*.

Máš žízeň? [if so] *Tak se napij. – Are you thirsty?* [if so] *So drink.*
Nemůžeš najít práci? [if so] *Pak bych se přihlásila na jeden z těch velkých internetových serverů, kde ji nabízejí. – You cannot find a job?* [if so] *Then I would sign up for one of those big internet servers where it is offered.*

2. This group includes the cases of discourse relations that express condition formally but not semantically (the fulfillment of the condition is not necessary for the validity of the content of the second argument, see the example). It is also considered *f_cond*.

Kdybys měl žízeň, tak je pivo v ledničce. – If you were thirsty, so beer is in the fridge.

The arrow leads (just like in the case of true semantic condition) from the proposition expressing pragmatic condition to the proposition expressing the result of the condition.

5.2.5 Purpose

purp

In the textual relation of purpose, A2 expresses the purpose for which A1 is carried out. The arrow always leads from A2 (expressing purpose) to A1. As already indicated in the introduction to causal relations, purpose is primarily expressed within the sentence as a clause element or, even more often, as a dependent clause. Two independent sentences with the meaning of purpose are basically in the relation of reason – result with the added intentional component expressed mainly by the verb *chtít* (*to want*) or its equivalents.

At the same time, Czech does not have any intersententially applicable connective of purpose. For these reasons, it is possible not to consider purpose to be a type of a textual semantic relation (as in the Penn Discourse Treebank). For the annotation of PDT, we still maintain this relation in order to confirm or exclude our speculations.[2]

[2] The relation of pragmatic purpose is not introduced because it is always interpretable by another semantic relation.

Věra chodí pravidelně cvičit. Chce totiž zhubnout. – Věra goes to train regularly. (The purpose is that) She wants to lose weight. purp
Protože chce Věra zhubnout, chodí pravidelně cvičit. – Since Věra wants to lose weight, she goes to train regularly. reason

5.2.6 Explication

explicat

In explicative relation, the second argument (A2) usually gives an explanation of the content of the first argument (A1). The arrow always leads to the argument that is being explained, i.e. (in most cases) to the left. From the semantic point of view, it is not a purely causal relation but a relation of synonymy/similarity and closeness of given propositions.

Hráli dobře, dali totiž pět gólů. – They played well, they scored five goals.

In some cases, it is difficult to distinguish explication from reason and result, sometimes it is even impossible. We have three criteria for the identification of these two relations for the time being, but they may not always be valid and sometimes are difficult to apply:
1. in the reason-result relation, we state the real motivation of action and we never formulate the same thing differently. On the contrary, from the semantic point of view, explication occurs in case of content synonymy/similarity and closeness (the meaning of the first utterance is more general/wider than the content of the second), cf. the sentences:

Je zloděj, protože má hlad. – He is a thief because he is hungry. reason
Je zloděj, protože krade. – He is a thief because he is shoplifting. explicat
Je zloděj, krade totiž. – He is a thief, he is shoplifting. explicat
Je zloděj, vždyť krade. – He is a thief, indeed, he is shoplifting. explicat

2. the order of arguments in the text: the explicative relation usually does not allow the reverse order of arguments, the explanation follows the content that is supposed to be explained. The arguments in a/the reason-result relation may be in an arbitrary order.

3. the importance of utterances: in explicative relation, the first utterance is usually more important than its following explanation; concerning reason and result, this distinction (importance vs. minor importance) is not evident.

We must be careful mainly as regards the discourse connectives *totiž*, *vždyť*, *přece* (these Czech connectives do not have their exact English counterparts, the utterances are usually translated with the connectives such as *because, as, indeed* or asyndetically). They are ambiguous, as they may express both explication and causality.

5.3 Contrastive relations (COMPARISON)

The basic semantics of contrastive relations puts two propositions into contrast. It may be described as A & B, where the contents of the propositions A and B are different, dissimilar, contradictory or compared.

The group includes mainly these syntactic relations from the tectogrammatical annotation: adversative (ADVS) and confrontational (CONTR, CONFR) relations, concession (CNCS) but also, for example, gradation (GRAD). In addition to those, new relations have been established (*correction – replacement, restrictive opposition and exception, pragmatic opposition*). Their introduction probably contributes to the fact that semantics of contrastive relations is the most highly elaborated – see the Table 4 below. If some types of opposition may specify the annotation on TR, the relations are annotated. The basic textual relation identical with ADVS is opposition (*opp*). The tectogrammatical functor ADVS is not replaced by *opp*.

Unlike tectogrammatical relations, textual semantic relations do not reflect hypotaxis and parataxis, which is demonstrated most in the group of contrastive relations. Therefore, we suppose that e.g. concession may be expressed in the text also paratactically, by devices expressing adversative relations. However, the important thing is not the connective but the real meaning.

Note to the extent of arguments for contrastive relations: If we are looking for an argument of a relation following a contrastive relation (i.e. the structure contains opposition or another contrast), it is likely that only the second part of the contrastive complex will be A1 because this complex opens a new theme in its second part. A1 of the following relation does not go beyond this thematic turning point. See A1 for the relation signalled by the connective *tedy* (*thus*) in the following example:

To je sice dobré pro základní orientaci, ***ale v konkrétní situaci potřebuje zákazník i konkrétní konzultaci.*** **(A1)** *Bude tedy asociace vyvíjet i poradenskou činnost? – It is maybe good for basic orientation* ***but in a concrete situation the customer needs also a concrete consultation.*** **(A1)** *Will the association develop also the consulting services, then?*

Table 4: The overview of contrastive discourse relations

The name of the relation	Abbreviation	Description	Connectives	Example
opposition	**Opp**	A1 is in the unspecified opposition to A2	ale, však, avšak, jenže, jenomže... (all meaning but, however)	*He asked for postponement of exams, which was not allowed to him.*
confrontation	**Confr**	A1 and A2 confront certain properties of two components	zatímco (while), zato (while), ale zato (but while), zato však (however while), naproti tomu (on the contrary), přitom (nevertheless), a zatím (menawhile)...	*While wages are falling, prices are increasing.*
restrictive opposition + exception	**Restr**	The content of A2 restricts the validity or scope of the content of A1 (partial unfulfilled expectations)	jen, jenom, jedině, toliko, pouze, jenže, jenomže, ledaže... (all meaning only, but only...)	*I will come. I only do not know when.*
correction (or replacement) + chosen alternative (substitution)	**Corr**	The content of A2 is replacement or substitution for an invalid content of A1, possibly the replacement is not valid either (unfulfilled expectations with or without replacement: not A1, but A2; not A1, but also not A2)	totiž, nýbrž, ale (all meaning but, however), naopak (on the contrary)...	*It was not freezing. On the contrary, the weather became nice.*
concession	**Conc**	A2 is in contradiction to A1 against the expectations about the validity of A1 retrospectively implied by A2	přestože, přesto, a přesto, i když... (all meaning yet, although)	*He has been on the dole for two months. Yet, he is not looking for a job.*
gradation	**Grad**	A2 expresses a higher or lower degree of quality than A1	ale i (but also), dokonce i (even also), dokonce (even)...	*Not only was he uneducated. He was even completely illiterate.*
pragmatic opposition	**f_opp**	A1 and A2 are formally in opposition but the semantics is usually unclear or vague	various connectives for contrast	*It is going to rain this weekend. But Czechs will block the highways anyway.*

5.3.1 Confrontation (juxtaposition)

confr

The relation of confrontation (within the sentence as well as between higher text units) indicates that two phenomena, situations, etc., have two different properties or a different degree of one property. A simple scheme of confrontation is: "Component A has the property X, while component B has the property Y", where components A and B are from a certain set (e.g., people) and properties X and Y are somehow related – they are often two opposite poles on one scale. The direction of the arrow is given by convention to the first argument in the text.

It is sometimes difficult to distinguish confrontation from opposition. However, opposition does not correspond to the scheme mentioned above – mostly it does not contain two components but only one.

The relation of confrontation usually includes the conjunctions *zatímco* (*while*), *kdežto* (*whereas*):

Na jihu je teplo, zatímco na severu zima. – The south is warm while the north is cold.
Ty máš čas, ale já už musím jít. – You have time but I must go.
Pepík pracuje dobře, ale Toník výborně. – Pepík works well but Toník excellently.

Some cases of double temporal expressions of the type *dříve – nyní* (*previously – now*) implicitly express confrontation. In these examples, confrontation is included in the annotation of the topic-focus articulation on TR (contrastive theme, tag *c* in the tectogrammatical attribute *tfa*). Such cases may be found automatically and, therefore, it is not necessary to mark either confrontation or temporal relations in the textual layer again:

Dosud totiž ozbrojení muži hlídali jen veřejnosti nepřístupné prostory. Nyní se budou přímo pod pracovnami poslanců pohybovat desítky cizích lidí. – The armed men have guarded only the off limits spaces so far. Now tens of strangers will be walking directly under the offices of politicians.

5.3.2 Opposition

opp

The textual relation of opposition corresponds to the tectogrammatical adversative relation (ADVS). The annotations should carefully distinguish it from other contrastive relations, especially from concession and confrontation (see the Table 4; also, see the chapter about confrontation and concession). The direction of the arrow is given by convention to the first argument in the text.

Chtěla bych pracovat jako doktorka. Nemám na to ale vzdělání. – I would like to work as a doctor. But I do not have the education.

Note to the annotation:
The following connectives and the relation they signal are also considered a case of OPP: *ale současně* (*but simultaneously*)*, zároveň však* (*but at the same time*), etc. (see the example (X)). Within a sentence, we maintain ADVS; intersententially, we annotate it as OPP and add both connectives. (If *zároveň* /*at the same time*/ or *současně* /*simultaneously*/ clearly have a strong temporal meaning, we annotate second_rel SYNCHR and the connective).

Respektuji skutečnost, řekl včera Doležal, že majitel novin má právo rozhodovat o personálním obsazení redakce, ale současně platí, že glajšaltovat noviny z ideologických důvodů je nemravnost za všech okolností. – I respect the fact, said Doležal yesterday, that the newspaper owner has the right to decide about the editorial staff, but at the same time, to manipulate the newspaper for ideological reasons is immoral in all circumstances.

5.3.3 Pragmatic opposition

f_opp

Pragmatic relations were defined as marks for cases when we formally express a certain relation (by means of a relevant connective) but we cannot interpret it semantically. However, in case of these relations (where the form and semantics do not correspond to each other), we always prefer the semantic criterion.

The relation of pragmatic opposition is such a relation that (similarly to pragmatic reason – result and pragmatic condition) does not express the real opposition and, at the same time, it is not possible to interpret it as another semantic relation. Usually, this case is also a relation of one argument and some presupposition or inference; the form (as well as the connective) is

adversative but the meaning is not clearly describable. We may also say that the arguments are vaguely connected or their meaning is, due to a large degree of inference, "at a great distance":

O víkendu má pršet. Ale Češi stejně ucpou dálnice. – It is going to rain this weekend. But Czechs will block the highways anyway.

This type of relation also includes stylistically inappropriate text units. For example, concessive conjunctions are ordinarily used in sentences that do not express concession:

I když má česká měna velmi dobrý zvuk a ČR je proslulá i dobrými odborníky, přesto jsou akcie českých podniků ve srovnání se zahraničím stále vysoce nadhodnoceny. – Although the Czech currency has a very good sound and the Czech Republic is known also for good professionals, the shares of Czech companies are still highly overvalued in comparison with other countries.

The category of pragmatic opposition contains both pragmatic concession and pragmatic contrast; these two types are not further distinguished due to the opacity of their meaning. The direction of the arrow is given by convention to the first argument in the text.

5.3.4 Restrictive opposition + exception

restr

Restrictive opposition is a relation in which the validity of the first argument is limited by the content of the second argument or the second argument expresses an exception to the first. The arrow leads from the exception or restrictive opposition to the more general (usually first) argument. On TR, this textual relation corresponds to the functor RESTR.

Sport vůbec nepěstuju. Jen si jdu občas zaplavat. – Sport is not my cup of tea. Only occasionally I go swimming.

Some cases of restrictive opposition are close to the condition or the implicit condition is included there. We prefer restrictive opposition if the significance of the restriction or the exception is still primary and the whole expression is primarily not interpreted as condition:

Kup mi hrušky, ale jenom ty zralé. – Buy me some pears, but only the ripe ones.
Každá krajina má svou krásu. Jenom ji musíte umět vidět. – Every landscape has its beauty. You must only be able to see it.

The basic connectives of restrictive opposition are the connectives of opposition (*ale* /*but*/, *však* /*however*/) and restrictive rhematizers (*jen, jenom, pouze*, etc., all meaning *only*; on TR, they have the functor RHEM).

Nenajdeme v nich ani stopy po orchestrálních partech či akordeonových melodiích, jen v závěrečných Summer Eyes zazní varhany. – You will not find any trace of the orchestral parts or accordion melodies there, only in the final Summer Eyes, there will be a sound of the organ.

5.3.5 Concession

conc

Concession expresses a specific type of opposition. A2 implies certain expectations associated with A1 (the causal component of concession, cf. *it is raining* implies: *people do not go out*) and, at the same time, it is negatively defined against it (a contrasting component of opposition, cf. *yet I will go out*). It is precisely this fact of regressive implication of certain expectations that differentiates concession from ordinary opposition (*opp*). However, it is still difficult sometimes to distinguish these two types of discourse relations, and there are cases where the relation may be interpreted in both ways. Both types of these relations may be distinguished by the fact that the arguments in opposition are in contrast to each other without the existence of implicated expectations (accepted by all authors and addressees of the text) that would be denied:

Když o deset let později obrátil ke gumě pozornost louisvilleský lékárník John Colgan, existovala již řada žvýkačkových milionářů.
Přesto však byly dveře pro zlepšovatele otevřeny dokořán, většina gumy byla stále ještě jen povrchově oslazený či ochucený kousek chicle.
– When gum began to interest Louisville pharmacist John Colgan ten years later, there hadalready been many gum millionaires.
However, in spite of that, innovators had the door wide open; most of the gum was still only a little sweetened or flavoured piece of chicle. conc

Petr má kolo, ale nemá auto. opp – *Peter has a bicycle but he does not have a car*. Opp

Přestože má Petr kolo, nemá auto. opp – *Although Peter has a bicycle, he does not have a car*. opp (without a wider context, there are no denied expectations)

The arrow in concession always leads from what is usually in the dependent concessive clause, i.e. from the argument to which we may add the concessive conjunction (A2).

5.3.6 Correction, replacement

corr

Correction is a relation in which the content of the second argument corrects or replaces the content of the first argument. One of them is always (at least implicitly) negated – in the vast majority of cases, it is the first. A typical connective of correction is *nýbrž* (*rather*); we annotate also each node with negation as a part of the connective:

Vláda se rozhodla vnést více systémovosti do právního systému tím, že nepůjde o nějakou kampaň, ale spíše o důraz na celý další legislativní proces. – The government decided to introduce more systematicity to the legal system in the way that it will not be a campaign but rather a focus on the further legislative process.

Pure negation can also serve as a connective:

Žádná aktivita státu, vědců či ekologických iniciativ nemůže tento systém nahradit, může ho jen rozbít. – No activity of the state, scientists or environmental initiatives can replace this system, it can only break it.

On TR, most cases of correction within a single tree are marked with the functor ADVS (the negated or corrected part is the first) but this relation occurs sometimes also within the parts of structures marked CONJ. These cases are re-annotated. From the semantic point of view, this textual relation includes also the structures within the sentence with the tectogrammatical functor SUBS (substitution, replacement). However, these cases are not re-annotated.
The arrow leads to the negated or replaced part, i.e. in most cases, to the left – in the following case, however, it leads to the right:

Chytrý bankéř si klienty přece vytváří, a ne se jich zbavuje. – A smart banker rather creates his clients and does not dispose of/get rid of/want to see the back of them.

5.3.7 Gradation

grad

The textual relation of gradation corresponds to the relation of gradation within the sentence (with the tectogrammatical functor GRAD). It compares a different degree of one property or two different actions where the deciding factor of the gradational character is mainly the context. Sometimes it is difficult to define the gradational nature of semantic relations against the pure conjunction. It may also depend on the reader's interpretation. In the annotation of PDT, we mark only indisputably gradational connections. If we hesitate, we write an annotator's comment. The first argument always expresses a lower degree of the property and the arrow always leads to it, i.e. it may be in both directions.

Státní úředníci nemají dostatečný respekt. Nemají možná snad ani představu o požadavcích Listiny základních práv a svobod. – State officials do not have enough respect. They have perhaps not even an idea about the requirements of the Charter of Fundamental Rights and Basic Freedoms.

5.4 Broad conjunction, elaborative relations (EXPANSION)

Discourse relations in the group of *expansion* correspond approximately to the relations of conjunction (with the functor CONJ) and apposition (APPS) in the tectogrammatical representation. From the annotation on the tectogrammatical representation, i.e. within the sentence, we have adopted the relations where conjunction or apposition is constructed by a clause, i.e. the verbal nodes with their subtrees (not nominal phrases). This distinction may be problematic if one of the verbs is elided – more details can be found in the chapter on ellipses (6.1).

The nature of the whole group of expansion, the so called elaborative relations, is different from the previous three groups. Unlike them, the semantics of expansion is not always motivated syntactically but rather by a compositional structure of the text. We may call them text structuring relations; they are called rhetorical in some textual linguistic studies. They mainly determine how the content of the utterance is related to the content of the whole previous section – whether it expands the content, brings a summary, gives examples, etc. These relations are also often identifiable among the larger text units such as paragraphs.

Apposition within the sentence, as one of those text structuring relations captured in PDT, only indicates the content parallelism of its neighbouring propositions. For this reason, all verbal appositions (with the functor APPS) are re-annotated as some of the *expansion* relations, usually specification, generalization and equivalence.

5.4.1 Conjunction

conj

On TR, all clauses are connected within the graphic sentence (i.e. from full stop to full stop); there is more or less no graphic sentence divided into independent trees. If there is no semantic connection between the clauses, they are linked only by technical coordination with the functor CONJ. This means that the compound sentences with CONJ must be checked – we must control whether some semantic relation actually exists there. If it turns out that it is not so, we comment upon it on the coordinating node CONJ. The direction of the arrow is given by convention to the first argument in the text.

Expressions *což* (*which*), *přičemž* (*and, at the same time*), *čímž* (*thereby*), etc., are annotated as CONJ on TR. Although CONJ is used appropriately in most cases, it is always necessary to assess whether there is not any different relation.

Regarding the conjunction *and*, we have decided to abandon the implicit meaning of succession. It would be very laborious and it seems that it is trivial. Interesting cases, if they arise, may be certainly commented upon:

*Kandidoval jsem kdysi v městských volbách v Janově za liberální stranu **a** drobní hokynáři si u mne houfně stěžovali na bandity. – I have once stood in municipal elections for the Liberal Party in Genoa **and** small grocers massively complained to me about the bandits.*

In these cases, we annotate ***a*** (***and***) **+ synchr**.
If the conjunction *and* appears with another connective, we annotate the relation indicated by this connective in the vast majority of cases; *and* is only added as the second connective – see, e.g., *a pak* (*and then*) (preced), *a tedy* (*and so*), *a tak* (*and so*) (reason, explicat...).

Note to the annotation:
A) The expressions ***což* (*which*), *přičemž* (*and, at the same time*), *čímž* (*thereby*)**, etc., are annotated as CONJ on TR – the cases where this annotation is not adequate for the given context were re-annotated.
B) There is a context in which the connective **však/ale** (**however/but**) combines the arguments in a copulative relation (sometimes with the trace of gradation). This is the context that only introduces a new theme and does not express a real contrast (see the examples X–Z). These cases were annotated as CONJ or GRAD.

(4) Stále nikdo neví, ze které strany do naší obce dorazí civilizace v podobě vodovodu. – Nobody knows yet from which side of our village civilization in the form of water supply will arrive.
(5) Ale již nyní jsou pohromadě peníze, za které bude naší vesnicí rozvedena síť trubek od domu k domu. – But we have already collected the money to distribute the network of pipes from house to house through our village.
(6) I kolem mých oken jsou již zabodnuty kolíky, usnadňující orientaci bagristovi, který brzy zjizví terén výkopem. – Even around my windows, there are pins stuck facilitating the excavator operator's orientation who will scar the land early by excavation.
/10 more sentences elaborating on the topic water supply/

Hudbou blízké budoucnosti ovšem není pouze vodovod. – However, the near future is not only the water supply.
Ten je pro obec povinně aktuální od okamžiku, kdy rozhodla svolit k vybudování mamutí skládky odpadu, vzdálené od nás pouze několik kilometrů. – It is a current topic from the moment when the village agreed to build a huge landfill waste just a few kilometres distant from us.

Strážníci se včera v centru Prahy v ulici Na příkopě zaměřili na řidiče vozů taxislužby. – Yesterday the policemen focused on taxi drivers in Prague in the street Na příkopě.
Ti totiž na stanovišti porušují dopravní předpisy a parkují kolmo k chodníku i přes to, že byli několikrát napomenuti. – They violate the traffic rules and park perpendicularly to the pavement even though they have been warned several times.
Takové stání je pro ně výhodnější, protože tak mohou svým "vyvoleným kolegům" držet místa na "štaflu". – Such parking is more profitable for them, because, in this way, they can hold places for their "chosen colleagues".
Za špatné stání strážníci účtovali blokovou pokutu až 500 korun. – The policemen charged them up to 500 crowns/CZK for this bad parking.
Kromě toho se však zaměřili i na kontrolu dodržování podmínek pro provozování taxislužby, které předepisuje koncesní listina. – In addition, however, they focused also on the control of compliance with the conditions for taxi services that are prescribed in the concession deed. (nak)
FOTO: MAFA - MICHAL RŮŽIČKA

Pink Floyd pozdravili publikum, nadšeně reagující zejména na starší písničky, v průběhu koncertu několika českými větami. – Pink Floyd greeted the audience, responding enthusiastically especially to the older songs, with several Czech sentences during the concert.
Ještě potěšitelnější však pod deštivým pražským nebem byla perfektní práce zvukařů. – But even more heartwarming was the perfect job of sound engineers under the rainy skies in Prague.

5.4.2 Instantiation (exemplification)

exempl

In the relation of instantiation, the first argument contains a set (e.g., activities, behaviours, etc.) and the second selects its subset as an example. The arrow leads from A2 (example) to A1 (superordinate term). It is important that both groups cannot be identical; the example must represent a selection of the total, i.e. one group is not viewed from different perspectives.
A typical connective is *například, třeba* (*for exampl*e). (On the tectogrammatical representation, these expressions are usually evaluated as rhematizers but it does not forbid them to have also another function – to participate in the construction of the text if they are opening the position for two arguments.)

Inovovali jsme také receptury pracích prášků, zvýšili podíl účinných látek a parfémů. U detergentu "Toto" jsme <u>například</u> řešili problém s udržením stálé kvality, protože jednotlivé partie byly nevyvážené. – We have also upgraded the laundry detergent recipes and increased the proportion of active ingredients and fragrances. <u>For example</u>, we have solved the problem of maintaining the consistent quality for the detergent "Toto" because the individual parts were unbalanced. exempl

U silnic se zřejmě vyplatí podnikat vždy, u hlavních tahů se pravděpodobnost úspěchu ještě zvyšuje. Jezdí-li <u>například</u> po hlavním tahu mezi Berounem a Rokycany podle úseků 9000 až 12000 vozidel denně, pohybuje se tu při obsazenosti vozidla jen dvěma lidmi 18000 až 24000 potenciálních zákazníků. – It seems that business is always worth doing at (the) roads; trunk routes even increase the likelihood of success. <u>For example</u>, if 9000 to 12000 vehicles drive on the main route between Beroun and Rokycany a day, there are (with the occupation of the car only by two people) from 18000 to 24000 potential customers. exempl

Discourse instantiation vs. sentence element (coreferential) instantiation

The relation of instantiation may link sentences (= textual relation, annotated) but also nominal phrases (= coreferential relation, not annotated).

Co tedy dělat v situaci, když se ke kolapsu takového podniku přidává propouštění lidí i z jiných odvětví, <u>třeba</u> ze zemědělství? – So what to do in the situation when the collapse of such a company is added to by dismissals of people even from other sectors, <u>such as</u> agriculture? (coreferential relation, not annotated)

Many tokens are apparently mixed (nominal phrase – sentence); sometimes, it is not easy to decide whether the first argument is just a noun phrase or the whole sentence.

Svou váhu má též datování. Zejména u autorů, u nichž se speciálně cení některé období tvorby. Znalec však většinou pozná, kdy byl obraz namalován, podle motivu či způsobu malby. Raný Kaván je například cennější než pozdní. – Dating is also of importance. In particular for authors who are evaluated especially for some period of their work. An expert, however, usually knows when the picture was painted – on the basis of the motive or the method of painting. For example, the early Kaván is more valuable than the late.
(A1: *authors who are evaluated especially for some period of their work* – not annotated)

Průpravu jsem měl všeho druhu. Třeba při rozvozu jsem denně přenesl pěkných pár tun na zádech. – I had training of all sorts. For example, during delivery, I transferred quite a few tons on my back a day.
(A1: *the training of all sorts / I had training of all sorts* – not annotated)

Summary: textual instantiations are only those cases where we are absolutely sure that the left part contains the expression of the whole action, not only a nominal phrase.

5.4.3 Specification

spec

In the relation of specification, the second argument expresses a detail or more concrete information about the statement in the first argument. Again, as with instantiation, there is a subset related to the content of the first argument but it is not a case of giving an example!!
Specification also includes the relation between hypertheme (title) of the list structure and a group of items in the list, see the chapter on list structures (4.2). A typical connective is a colon (unless it has the functor PRED – in that case, it represents the verb). Specification occurs very often without any connective (such cases, however, are not yet annotated). Within the sentence, many appositions are specifications. We must beware of overusing this relation in cases that rather express instantiation or equivalence.
The arrow leads from the detail to more general argument, i.e. to the left in the vast majority of cases.

Snaží se omezit zadlužení. Vydělává víc peněz. – He tries to reduce debt. He earns more money. spec
Zdeňku, piš, říkával jsem autorovi, ***kdykoliv jsme se potkali,*** *a* ***bylo to dost často. –*** *Zdeněk, write, I used to tell the author* ***whenever we met*** *and* ***that was fairly often***. spec

5.4.4 Equivalence

equiv

The relation of equivalence connects two arguments, in which we speak twice about the same denotation or action but "in other words" in each case. However, the second argument is neither more specific (specification) nor more general (generalization) in relation to the first; it is not an example (exemplification) or explication, either . The difference may be described by the examples that follow (but very often it is impossible to keep a clear boundary between these relations in the real text, and the semantics is determined by the connective). These relations are similar because they introduce some claim and then bring its elaboration in the following text unit. The direction of the arrow is given by convention to the first argument in the text.

Je zloděj, krade totiž. – He is a thief, as he is shoplifting. explicat
Je zloděj, jinými slovy krade. – He is a thief, in other words, he is shoplifting. equiv

Dnes nebo zítra se v dolní komoře polského parlamentu - v Sejmu - očekává hlasování, které bude mít vážné politické důsledky, ***ať už dopadne jakkoliv****,* *tj.* ***bude-li zákon odmítnut či přijat.*** *– Today or tomorrow the lower chamber of the Polish Parliament – the Sejm – expects voting that will have serious political consequences* ***whatever the outcome will be****,* *i.e.* ***whether the law will be rejected or accepted***. equiv

5.4.5 Generalization

gener

The relation of generalization expresses generalization or summarization – the second argument contains a summary of the content of the first argument. The arrow always leads from A2 to A1, i.e. in most cases, to the left. Typical connectives are non-conjunction expressions and expressions such as: *stručně řečeno* (*shortly*)*, zkrátka* (*simply*)*, krátce* (*in short*)*, prostě* (*simply*)*, vlastně* (*actually*)...

Byl nesnášenlivý, netrpělivý a panovačný. Prostě to byl takový malý tyran. – He was intolerant, impatient and imperious. Simply, he was a little tyrant. gener

5.4.6 Conjunctive alternative

conjalt

Conjunctive alternative refers to a relation where A1 and A2 represent alternatives that may be understood in both a copulative and a disjunctive relation. In the following example, it is possible to interpret the relation between A1 and A2 as two alternatives that may but donot have to exclude each other – we may either go to the cinema or dinner or both. The arrow always leads from A2 to A1 (A1 is on the left).

Můžeme jít do kina. Nebo bychom mohli jít na večeři. – We may go to the cinema. Or we may go to dinner. conjalt

As for the conjunction *nebo* (*or*), we do not re-annotate the functor CONJ to *conjalt*. We suppose that all such *nebo's* (*or's*) are *conjalt* and the annotation will be done later automatically.

5.4.7 Disjunctive alternative

disjalt

Disjunctive alternative expresses a relation where A1 excludes A2 and vice versa. The arrow leads from A2 to A1 (A1 is on the left). This relation corresponds to the functor DISJ on the tectogrammatical representation. However, it is also necessary to check the annotation of the conjunctive alternative.

Chovej se tu slušně. Nebo sem nechoď! – Behave decently. Or do not come here! disjalt
Ještě nevím, co budeme dělat o víkendu. Asi pojedeme do Českého Krumlova. Anebo možná zůstaneme v Praze. – I do not know yet what we will do this weekend. I guess we will visit Český Krumlov. Or maybe we will stay in Prague. disjalt

6. Various problematic structures

6.1. Ellipsis in discourse

The annotation of discourse relations must deal with several types of ellipses, namely:
a, Ellipsis of the governing clause
b, Contextual ellipsis of the governing verb (mainly, but not only, the ellipsis of the second predicate in coordination)
c, Grammatical ellipsis of the governing predicate verb
d, A special case: contextual ellipsis of a modal predicate

a, Ellipsis of the governing clause

(Neodešla.) Protože by to nestihla. – (She has not gone.) Because she would not have caught it.
(Stát se nesmí spoléhat na placené školství.) Ledaže by si pronajal místa na soukromých ústavech. – (The State must not rely on paid education.) Unless it rented places at private institutions.

The tectogrammatical representation captures ellipsis of the governing clause in the way that its governing verbal node is usually copied from the preceding sentence to the dependent clause. If we need to lead the arrow to/from this governing clause, we lead it to/from the original one, not to/from the copied one. (The principle is that we would like to eliminate the added elided clause in the future and to "join" the clauses that are parceled out to the governing clauses as if these sentences were not parceled out.) However, we must pay be careful in case of "lower" relations, where we need to include only the dependent clause – in these cases, it does not matter where the the governing clause occurs.
If there is a mistake on the tectogrammatical representation and the missing governing clause is not added to the tree with the dependent clause, we add the discourse arrow to the governing verb in the preceding tree.

b, Contextual ellipsis of the governing verb

There are several types of contextual ellipsis of the governing verb. The first and the second examples are not problematic. They are annotated in the same way as if the given verbs were present in the surface realization, i.e. as discourse relations:

Pavel nepřinesl nic. Hanka ovšem čokoládu. – Paul did not bring anything. But Hanka a bar of chocolate.
V sobotu pracovali všichni, ale v neděli nikdo. – Everyone has worked on Saturday, but no one on Sunday.

Problematic contextual ellipses of verbs are the cases that elide the second verb in coordination. Firstly, we must decide whether we deal with clausal coordination (between clauses, verbs) or "lower" coordination (between non-verbal sentence elements). Mostly, this has already been solved in the tectogrammatical trees – the verb is either copied or not, so the decision is not up to us. However, it may happen that we will not agree with the capturing of the tree. In such cases, there are two basic criteria according to which we may make a decision:

the criterion of the same function (or of semantic compatibility), i.e. the same or a semantically compatible functor (e.g., two complementations of manner – although of a different type) for coordinated members. If the members of a potential coordination of sentence elements have a different functor and are semantically incompatible, it is a clausal coordination.
the criterion of "modification": if any member of the coordination is modified by a rhematizer or the absent predicate is modified, it is not a member coordination (i.e. on the contrary, it is a discourse relation). Analogously, if two sentences are linked:

Koupím chleba. A mléko. – I will buy bread. And milk.

it is a member coordination, i.e. "non-discourse", while in the sentences:

Koupím chleba. A asi i mléko. – I will buy bread. And perhaps also milk.

there is a coordination with the copied predicate, i.e. verbal and a matter of discourse.
If the coordination is modified by negation, it is verbal coordination, i.e. a sentence such as:

Koupím chleba. A ne mléko. – I will buy bread. And not milk.

is paraphrased in this way: *I will do two things – I will buy bread and I will not buy milk*; not as: *I will do one thing – I will buy bread and "not-milk"*. If there is negation in the joined sentence with the contextual ellipsis of the governing verb and if the negation is not present in the preceding sentence, we evaluate this case as clausal joining, i.e. a matter of discourse.
We understand modification as a modification of a verb, not as a modification of a noun phrase (as attributes). This means that a modification of an elided verb also includes examples of the following type:

Koupím chleba. A mléko večer. – I will buy bread. And milk in the evening. (the modification of a verb – clausal coordination)[3]

However, not of this type:

Koupím chleba. A plnotučné mléko. – I will buy bread. And full-cream milk. (the modification of a noun – member coordination)
Byl to hezký, ale únavný večer. – It was a nice but exhausting evening. (a relation between the attributes; only 1 clause)

Another example of this type follows; the ellipsis is considered clausal and the relation is annotated as discourse:

Státní úředníci nemají dostatečný respekt, možná snad ani představu o požadavcích Listiny. – State officials do not have enough respect, maybe not even an idea about the requirements of the Charter.
Trámy mají být po rekonstrukci znovu instalovány, avšak [mají být instalovány] již pouze jako okrasná atrapa zavěšená na nových stropech. – The beams are to be installed again after the

[3] Other examples of clausal coordination:
Přinesl to včas, ovšem Marii. – He brought it in time, but to Mary.
Přinesl to včas a Marii. – He brought it in time and to Mary.

renovation but [are to be installed] only as an ornamental imitation hanging on the new ceilings.

In this case, the complementation of manner ("as imitation") does not have its counterpart (adjunct of manner) in the first clause. These constructions are always annotated as discourse.

c, Grammatical ellipsis of the governing predicate verb

Grammatical ellipsis of the governing predicate contains an elided member that has a generated node with the substitutional t-lemma #EmpVerb. This node and its subtree are considered a proposition in principle. If another clause is linked to it by a connective, we annotate such a case as a discourse relation.

Nač [#EmpVerb] *ten spěch? A proč se tak mračíte? – Why* [#EmpVerb] *the rush? And why are you frowning?*

d, A special case: contextual ellipsis of a modal predicate

Ale přes den se nedá dělat nic, snad jen instalovat snímací kamery a služba je bude muset nějak ohlídat, říká ing. Dastych. – But you cannot do anything during the day, perhaps only install sensing cameras and the service will have to guard them somehow, says ing. Dastych.

The second member of coordination lacks a modal verb and, therefore, the coordination of two infinitives arises. These examples (the elided member *can* is a modal verb that is hidden in the node of the verb *to do*) are annotated as discourse relations (RESTR) for now. We face here a deeper problem of the relation between the content of the proposition and its modal modification. This solution is temporary and the issue will be further examined. All such cases have been collected (in addition to the annotation) and provided with an annotator's comment.

6.2. Semantically "underspecified" constructions

Semantically "underspecified" constructions like *děje se to, že..., stalo se to, že...* (*it happens that..., it happened that…*) have the discourse arrow in the lower level (i.e. where the concrete content is expressed) only if there is a clear reason to do so.

6.3. The arrow to coordinated clauses

If coordinated clauses occur within one argument, the arrow is drawn to/from the coordinating node (coap) (however, if one of the coordinated clauses is not a part of the argument, the arrow is drawn to the root of the clause that belongs to the argument).

6.4. Pragmatic relative clauses

Pragmatic relative clauses are re-annotated only if they contain both a relative and another linking element:

Norský trenér Olsen vzal pro zápas s Běloruskem za základ kádr z MS, v němž <u>však</u> už schází libero Bratseth, který ukončil kariéru. – Norwegian coach Olsen took a cadre of MS as the basis for the match with Belarus, which, <u>however</u>, lacks sweeper Bratseth, who stopped his career.

If we re-draw the edge, the connective is not a relative but only other connectives (*however* in this case). Relative clauses without other connectives are left without discourse annotation because it is often impossible to decide whether they are "proper" or pragmatic relative clauses.

Other pragmatic relative clauses are re-annotated:

Nakonec vyhráli, když ještě v polovině třetí třetiny prohrávali 3:0. – Finally, they won when they were trailing in the middle of the third period 3:0. conc
Ulehla, aby už nevstala. – She laid down not to get up any more. conj or opp

If we are re-drawing the edges in all the pragmatic relative clauses, it is important to ensure that the

rest of the structure except for the subordinate clause (i.e. the part of the tree corresponding to the main clause) is grouped together. This is prevention for cases when one part of the tree would refer to the whole tree, including that part itself.

6.5. Annotation of verbs introducing an assertion (first level of the tree) and the assertion content (lower layers)

- the type ***it speaks for A and it speaks for B*** (a pure conjunction of introductory verbs) – the connective belongs to the this relation; the lower level contains either nothing (the contents of the introduced speeches are in no relation – *řekl, že sněží, a řekl, že maminka bude večer doma – he said that it was snowing and he said that his mother would be at home in the evening*) or the implicit conjuctions (*Svědčí to o tom, že Česko má daleko do Evropy. Svědčí to i o tom, že v Čechách existuje diskriminace žen. – This suggests that the Czech Republic is far from Europe. It also suggests that there is discrimination against women in the Czech Republic*). The implicit conjunction will be annotated during the annotation of implicit relations.
- the type ***he said A but he (also) said B*** – the contents of saying are clearly in contrast; the connective belongs to these dependent contents (the lower level) and is annotated within them. The higher level contains (implicit) conjunction
- the type ***he said A but he did not say B*** – *but* is in the higher level (*Řekl, že přijel včera, ale neřekl, že dorazil o půlnoci. – He said he had arrived yesterday but he did not say that he had arrived at midnight.*)
- **it follows that the lower level is connected only if the types of the the relations in the higher and lower levels are different**

6.6. Structures with comparison

a, comparative structures with the functor CPR – these structures have always an inserted verb; therefore, we must comment on them within the discourse annotation. Because of the enormous complexity of the issue in PDT and due to the fact of how these cases are annotated on the t-representation (the verb is always inserted within one branch of a tree), we leave comparative structures with the functor CPR completely beyond the scope of our annotation.

Therefore, we do not annotate anything in the examples like:

Děčínskou odvetu zahájily hymny z koktavého gramofonu a rychlá branka z hole Alinče už po 16 vteřinách hry.
Stejně jako v Liberci převládala (znovu i díky špatné ledové ploše) bojovnost nad hokejovou krásou.
– The retaliation of Děčín was launched by anthems of stuttering phonograph and a quick goal by Alinč already after 16 seconds of the game.
As in Liberec, the fighting spirit prevailed (again also thanks to bad ice) over hockey beauty.

b, comparative structures without the functor CPR – the following examples demonstrate clearly that the expressions *stejně tak* (*equally*), *podobně* (*similarly*), *obdobně* (*likewise*), *stejně jako* (*the same as*) connect or structure the text in some way. Some usages of these expressions have even the functor PREC on the t-representation. As for these expressions, we always consider whether their function is rather discoursive in our sense (e.g., whether they are replaceable by a more typical connective) or different. According to this, we either annotate a discourse relation or not.

Chce snad někdo namítat, že žádná Hašprtánie neexistuje?
Stejně tak dnes neexistuje územněprávně ani Morava.
– Does someone want to argue that there is no Hašprtánie?
Equally, there is no Moravia as a legal administrative district today.

Stačí je jen vzájemně propojit, spustit Mascu, označit v diáři to, co si přejete přenést, a po chvíli se vám data objeví v počítači.
Obdobně to probíhá i v opačném směru.
– It is enough to link them together, run the Masca, mark in your diary what you want to transfer and your data will appear on your computer after a while.
Similarly, it works in this way also in the opposite direction.

The expression *stejně* (*equally* but also *anyway*) appears sometimes as one_arg:

Tento transformační "polštář" (podobně jako další polštář v podhodnoceném kurzu měny proti paritě její kupní síly) má ještě značné rezervy, které při existujícím tempu "vypouštění"

obou polštářů vydrží zhruba až do roku 2000 (přitom v roce 1999 podle všech proroctví, která se zatím naplňují, bude už stejně konec světa). – This transformational "pillow" (like another pillow in the undervalued exchange rate against its purchasing power parity) has still considerable reserves that will last in existing rate of "discharge" of the two pillows approximately until 2000 (according to all the prophecies not yet fulfilled, there will be the end of the world in 1999 anyway).

6.7. Structures with apposition

These structures are annotated only in the case of clausal apposition, i.e. the apposition of two syntactic structures with the finite verbs, one of which may be elided and inserted. Clausal appositions require a closer semantic marking (the mark APPS itself does not say much about the semantics). Thus, the constructions with APPS must be always annotated.
Appositions of sentence elements and mixed appositions are not annotated.

The structures like the following example do not have an inserted verb within the apposition on the t-representation. The annotators of the t-representation have decided to consider these examples **non-verbal structures**. We respect their decision.

*Mezi návštěvníky zatím údajně převládají Němci, **vesměs znalci či příznivci Mahlerovy hudby, lidé toužící po poznání.** – Reputedly, the dominant group among the visitors are Germans, **mostly experts or fans of Mahler's music, people eager for knowledge**.*

The annotated data contained cases of mixed APPS, which were, however, divided into two trees (theverb was copied in the second of them). These cases are annotated. We also add a comment that it is a case of **mixed APPS in two trees**. These cases are likely to be very rare.

Zdeněk Matějček: Předkládáme i problémy, na které se zapomíná.
Tak například úmrtí dítěte nebo narození postiženého dítěte.
– Zdeněk Matějček: We present also the issues that are usually neglected.
For example, an infant's death or a birth of a handicapped child.

6.8. Structures with untypically introduced direct speech

PDT contains sentences with direct speech of this kind:

Trenér Uh. Hradiště Juran byl naopak spokojen: Teplice hrály v prvních dvaceti minutách excelentně. – The coach of Uh. Hradiště, Juran, was, on the contrary, satisfied: Teplice played excellently in the first twenty minutes.

Teplický trenér Bičovský neskrýval zklamání: Byli jsme před utkáním přesvědčeni, že vyhrajeme. – The coach of Teplice, Bičovský, did not hide his disappointment: we were confident before the game that we will win.

These sentences are annotated according to the semantic relation between the two parts regardless of how the structure is captured on the t-representation. There is either CONJ between the two parts or the second part is linked to the first one with EmpVerb represented by a participle (*řka – saying*). If we annotate a structure with an inserted participle, we group the upper part of the tree (the introductory clause) without the inserted participle and the arrow leads from/to the node of the verb at the bottom of the tree (the content of direct speech). The inserted participle thus remains outside the annotation.
The same structures without a colon (only with a comma) follow the rule that a comma is not considered a discourse connective and they are left aside the annotation.

6.9. The relation of question and answer

The relation of question and answer was left aside the annotation of discourse relations because it represents a different type of discourse structure.
For a better idea about this issue, two examples are presented here. The second sentence is an answer to a question; it actually introduces the contents of the pronoun and, therefore, it is not a discourse relation between two syntactic structures with a finite verb.

***Co** to způsobilo?*
Například to může být škola, sport, hudba atd.
*– **What** caused it?*
For example, it may have been school, sports, music, etc.

*Na negativní věci se nabalují další a další a už ani nevíme, **co** bylo na začátku.*

Mohl to být třeba nedostatek lásky rodičů, především matky.
*– More and more things add to the negative ones and we no longer know **what** was in the beginning.*
It could have been, for example, a lack of love of parents, especially the mother's.

6.10. Shared modifiers of coordinated structure

When we are grouping an argument containing shared modifiers, it is not necessary to include the shared modifiers into the group (or to create a group just because of them) – see Figure 8 below. The nodes of shared modifiers are understood as a connected subtree with their governing node (i.e. effective but not coordinating).

Kořeny tohoto zvratu měly ovšem hlubší příčiny, především velkou hospodářskou krizi.
Její důsledky se projevily nejprve v drastickém dopadu na sociální postavení velkých skupin obyvatelstva, odrazily se i v radikalizaci jejich politických názorů a promítaly se pak nepříznivě do všech oblastí fungování politického systému.
– However, the roots of this reversal had deeper grounds, especially the Great Depression.
Firstly, its results had a drastic impact on the social status of the large population groups; they were also reflected in the radicalization of people's political views and then they were reflected adversely in all areas of the operation of the political system.

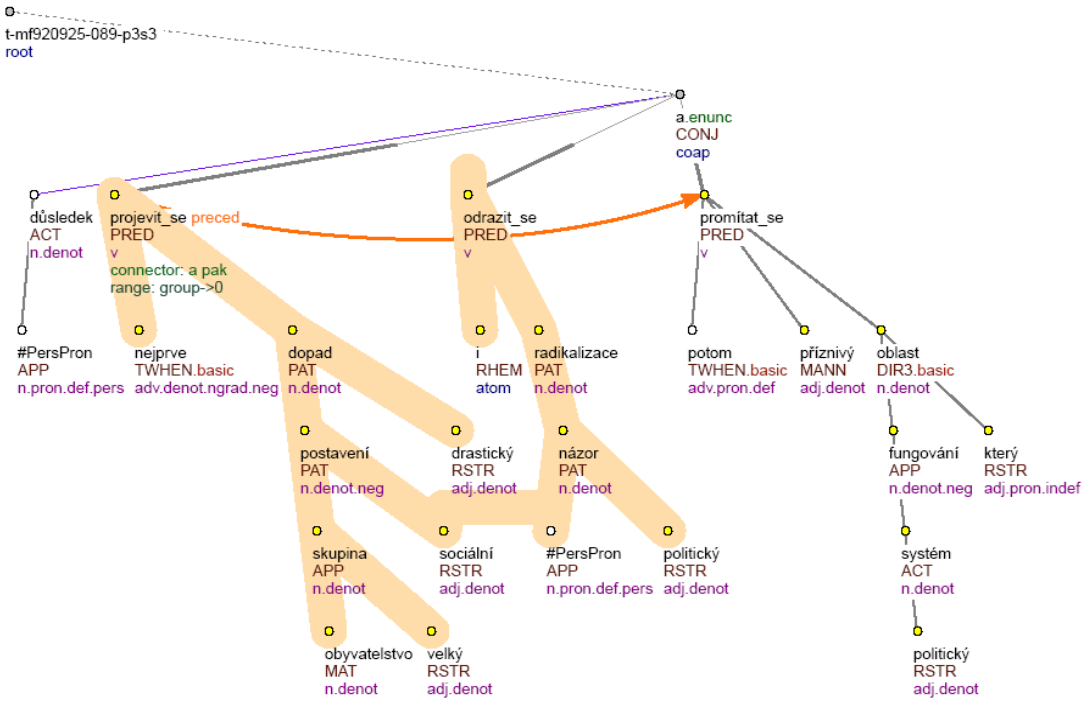


Figure 8: Annotation of shared modifiers

6.11. Structures with the connective *a to* (*and that*)

TR captures structures with ***a to*** (***and that***) within a single tree as APPS; the part after ***a to*** (***and that***) often contains the same verb copied from the first part. Even though we may consider the potential transferability of the whole structure to a simple sentence in many cases, it will never be a completely synonymous construction. Since the verb is copied, the sentence elements added to it are different and, as already mentioned, the structure is not transferable to a simple sentence without a change of meaning, we annotate the corresponding discourse relation between the parts of APPS (mostly SPEC or CONJ).

A to (*and that*) is sometimes followed by another connective. Since EmpVerb stands instead of the main clause, the connective belongs to the subordinate clause and it is annotated on TR in this way as well (see Figure 9). Between the branches of APPS, there is only the relation of CONJ. On the basis of this, we also annotate only the relation of CONJ between the independent clauses linked by *a to* (*and that*) even if there is another connective (see the following Figure 10).

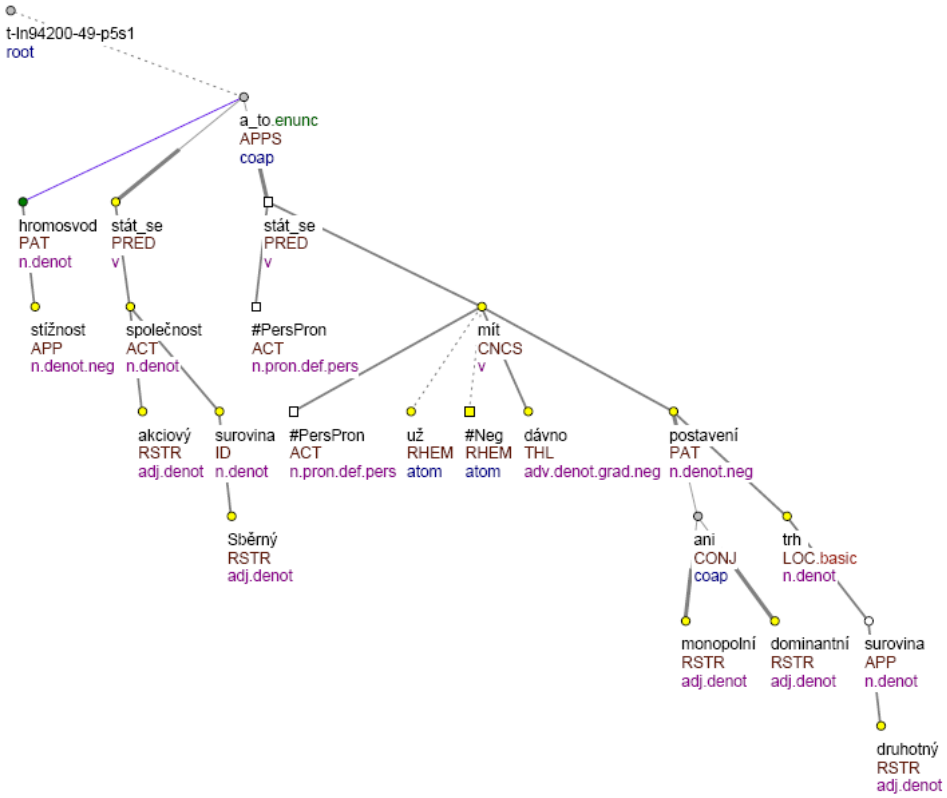


Figure 9: Annotation of *a to* (*and that*)

Hromosvodem stížností se stala akciová společnost Sběrné suroviny, a to přesto, že už dávno nemá ani monopolní, ani dominantní postavení na trhu druhotných surovin. – All complaints were oriented to the joint-stock company Sběrné suroviny and that even though it has no longer the monopoly or dominant position on the market of secondary raw materials.

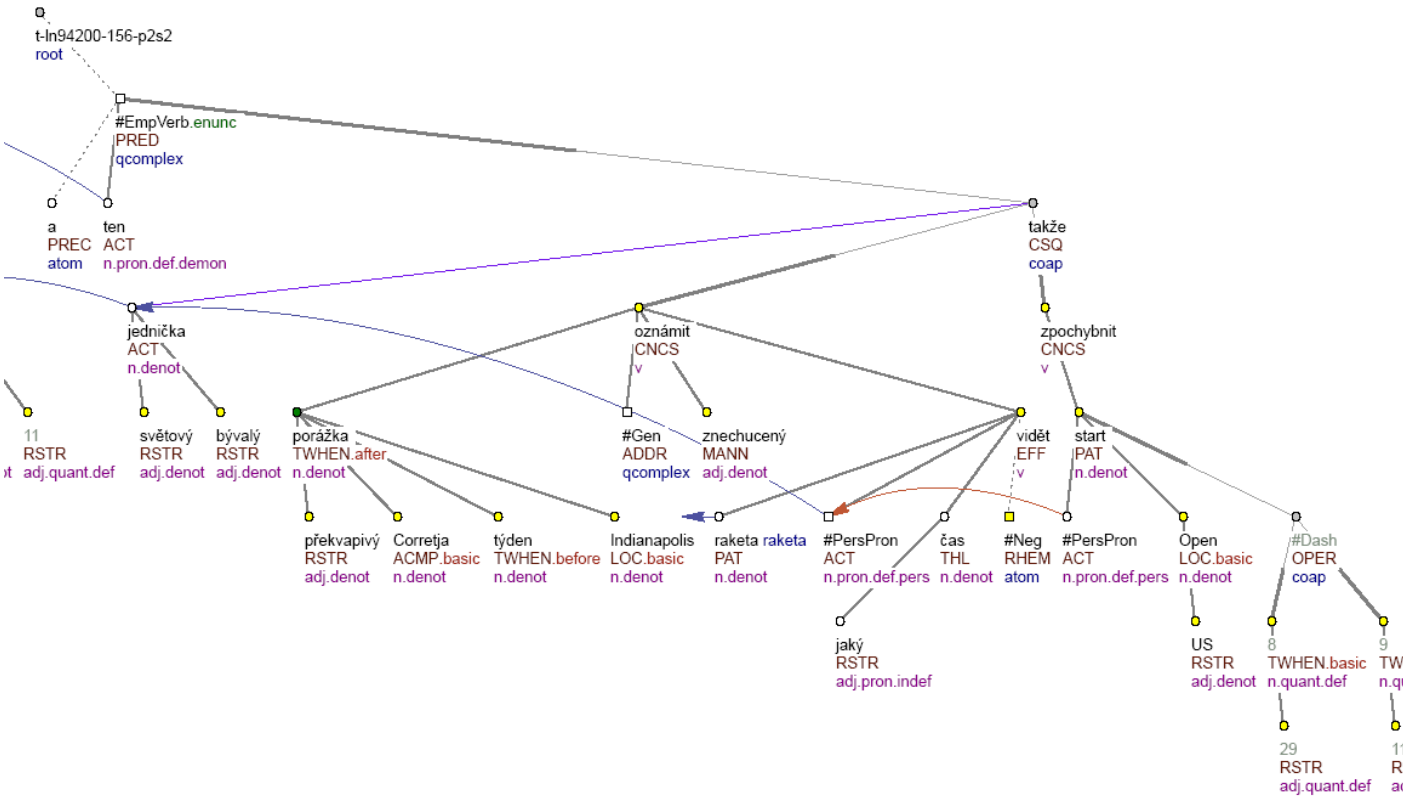


Figure 10: Annotation of *a to* (*and so*) with a PREC

Ve Flushing Meadows už dva dny pilně trénuje Jim Courier (11. tenista ATP).

A to navzdory tomu, že po překvapivé porážce s Corretjou před týdnem v Indianapolisu bývalá světová jednička znechuceně oznámila, že rakety nechce nějaký čas vidět, takže zpochybnila svůj start na US Open (29. 8. - [11.9]11[11.9].[11.9]9.).

– Jim Courier (the 11th ATP tennis player) has been training diligently in Flushing Meadows for two days.

And (he has been doing) so despite the fact that, after a surprising defeat by Corretja a week ago in Indianapolis, the former world number one said with disgust that he does not want to see rackets for some time and questioned his start at the US Open (29.8. - [11.9]11[11.9].[11.9]9.).

6.12. Structures with deictic connectives

We understand deictic connectives as a combination of a preposition and a demonstrative pronoun that has clearly a connective function in the text. In most cases, a deictic connective may be replaced by a typical connective (e.g., *díky tomu* > *proto* – *due to this fact* > *therefore*); its demonstrative element typically refers to a whole clause. Deictic connectives

are the following expressions: ***místo toho* (*instead of that*), *k tomu* (*thereto*), *naproti tomu* (*besides*), *kromě toho* (*apart from that*), *mimo to* (*except for that*), *navzdory tomu* (*despite that*), *díky tomu* (*thanks to this fact*), *kvůli tomu* (*due to this fact*), *vzhledem k tomu* (*because of that*), *vedle toho* (*besides that*), *spolu s tím* (*together with this fact*)** in contexts such as:

Máme určité kontaktní možnosti ve všech státech, odkud pocházeli zahraniční studenti v bývalém Československu.
Díky tomu bychom tam mohli hledat uplatnění pro naše lidi, a naše licence je proto pojata dosti široce.
Jenomže v těchto zemích sami nevědí, co s nezaměstnanými, odpověděl Otto Brabec z agentury SERVUS na otázku, jak se daří zprostředkovávat práci v zahraničí.
Vzhledem k tomu dominuje v činnosti agentury zajišťování studentských pracovních pobytů především [] v Německu.
– We have certain contact possibilities in all states from which foreign students came in the former Czechoslovakia.Thanks to this fact, we could seek opportunities for our people there, and our license is thus conceived quite broadly.
However, these countries themselves do not know what to do with the unemployed, said Otto Brabec from the agency SERVUS when he was asked how successful the mediation of the work abroad is.
Because of that the dominating activity of the agency is ensuring student work stays mainly [] in Germany.

Note: The expressions ***tím/a tím* (*this/and this*)** are more likely not to be connectives but it is necessary to bear them in mind.

6.13. Structures with the connective *s tím, že* (*with the fact that*)

A subordinate clause linked to the main clause by a connective ***s tím, že*** (***with the fact that***) expresses an unclear or "blurred" condition. It has a functor ACMP on TR that is given by the form; the semantics may vary (is "blurred") – it may be concession, purpose, condition, conjunction, etc. These structures are re-annotated as a relation that really occurs between the parts of the multiple sentence in the given case. The connective is the whole expression ***s tím, že*** (***with the fact that***) (see the large manual, p. 500).
The real meaning may be tested by substitution of a different connective.

6.14. Diversions from the topic in large arguments

Short diversions in large arguments were included in the argument in cases such as (A) here. In this example, the first argument marked with the arrow from sentence 134 contains sentences 112–132. However, sentences 121–123 are a diversion from the topic and, from a certain point of view, it would be reasonable not to include them in the argument. However, the thematic continuity is not interrupted significantly, so this diversion may be lost also from the readers' point of view. Therefore, these sentences are retained as part of the argument.
(A similar diversion is sentence 133 but it is easy not to include it, as it stands in the end.)

(A)

(111) Dublin- město hospod a knih – Dublin – a city of pubs and books

(112) V porovnání s Corkem je Dublin městem kultury a zábavy. – In comparison with the city of Cork, Dublin is a city of culture and entertainment.

(113) Hospody(je jich tu na sedm set) se střídají s knihkupectvími. – Pubs (there are about seven hundred of them) take turns with bookshops.

(114) Zatímco v prvních je k dostání především guinness, irská káva nebo horká whiska, v těch druhých dominují knížky Jamese Joyce, Williama Butlera Yeatse, George Bernarda Shawa, Oscara Wilda, Samuela Becketta a dalších známých irských literátů. – While the first offer mainly Guinness, Irish coffee or hot whiskey, the second are predominated by books by James Joyce, William Butler Yeats, George Bernard Shaw, Oscar Wilde, Samuel Beckett and other famous Irish writers.

(115) Žádné jiné město se nemůže pochlubit třemi nositeli Nobelovy ceny za literaturu. – No other city can boast about three Nobel Prizes for Literature.

(116) Dubliňané a Irové vůbec jsou na své velikány pyšní. – Dubliners and all Irish are proud of their greats.

(117) Dokazují to nejen nesčetná vydání jejich děl, ale také sochy, fotografie, plakáty, citáty na suvenýrech a pohlednicích, a dokonce i názvy hospod nesoucích jejich jména. – This is demonstrated not only by frequent publishing of their works, but also by their sculptures, photographs, posters, souvenirs and postcards with their quotes, and even by the names of pubs bearing their names.

(118) Největším literárním fenoménem Dublinu je samozřejmě James Joyce. – The biggest literary phenomenon of Dublin is, of course, James Joyce.

(119) Podle Irů je jeho Odysseus, zachycující jeden den života jednoho Dubliňana, největším románem dvacátého století vůbec. – According to the Irish, his Odysseus describing one day of a Dubliner is the biggest novel of the entire 20th century.

(120) Uvedení spisovatelé jsou díky tomu, že publikovali převážně v angličtině, často omylem považováni za Angličany. – The above mentioned writers are often mistakenly regarded as English because they published their books mostly in English.

(121) Irština, patřící ke starým galským jazykům, ostatně není ani dnes mezi lidmi zdaleka tak rozšířená, i když úřední nápisy a oznámení jsou dvojjazyčná. – Irish, belonging to the old Gaelic languages, is not much widespread among people even today although official signs and announcement are bilingual.

(122) Její úpadek pokračuje, přestože ve školách se povinně vyučuje. – Its decline continues even though it is compulsory at schools.

(123) Mimochodem Irsko se řekne Éire a Dublin je Baile Átha Cliath. – By the way, Ireland is Éire and Dublin Baile Átha Cliath.

(124) Irská hospoda se od české poměrně dost liší. – The Irish and Czech pubs differ significantly.

(125) Připomíná velký bar. – The Irish resembles a big bar.

(126) Uprostřed místnosti stojí většinou čtverhranný pult, kde se obsluhuje ze všech stran. – In the middle of the room, there is usually a rectangular desk with service given from all sides.

(127) Roznášení pití ke stolům je zcela výjimečné. – Distributing drinks to the tables is absolutely exceptional.

(128) V nabídce tradiční irské hospody je kromě několika druhů piva(přestože je Irsko vyhlášené svým černým pivem Guinness, Irové dávají přednost spíš světlému) studená i horká whiska, irské likéry a samozřejmě i pověstná irská káva. – The traditional Irish pub offers, apart from a few kinds of beer (although Ireland is famous for its black beer Guinness, Irish people rather prefer light beer), both cold and hot whiskey, Irish liqueurs and of course the famous Irish Coffee.

(129) Její příprava zabere barmanovi zhruba deset minut. – The barman spends about ten minutes to make it.

(130) Podává se do skleničky na víno a voní z ní whiska. – It is served in a wine glass and smells of whiskey.

(131) V pořádné zdejší hospodě by neměla chybět ani tradiční živá hudba a zpěv. – The proper local pub should not lack the traditional life music and singing.

(132) Nejčastěji se hraje na tahací harmoniku a kytaru, někdy se objeví i keltská píšťala. – The people play mainly the accordion and guitar, sometimes also the Celtic whistle.
(133) Kromě tradiční hudby se Irové prosadili významně i v moderní muzice, stačí připomenout rockovou skupinu U2 či zpěvačku Sinéad O' Connorovou. – Except for the traditional music, the Irish are very successful also in modern music, just remind the rock group U2 or the singer Sinead O 'Connor.
(134) Ale Dublin nejsou jen hospody a knihy. – But Dublin is not only the pubs and books.

6.15. Untypical CONTINGENCY structures

Some CONTINGENCY structures require restatement of one of the arguments for their interpretation: e.g., in the following example , it is necessary to reformulate the part before the hyphen in the sense *if the security is handed over ... to the vault ...* in order to get the sense of the relation with the part after the hyphen (*it is no longer necessary to pass it to the lien agent*).

Zastavený cenný papír může být také předán spolu se zástavní smlouvou či její úředně ověřenou kopií do úschovy třetí osobě – pak už není třeba ho předat zástavnímu věřiteli. – A pledged security along with the mortgage agreement or its certified copy may be also submitted to the custody of a third person – then it is no longer necessary to pass it to the lien agent.

In some cases, there are more possibilities of interpretation/such reformulating. E.g., in the first two examples in this section, the connective (*potom – then, to – that*) may be replaced by "in that case/in such case" or by the wole subordinate clause. The relation may be then interpreted as rather temporal or causal; in the case of the second example, we may also think about explication.

Londýn by se stáhl zpět za La Manche. – London would have pulled back behind La Manche.
A Německo, zoufale se snažící udržet upadající jednotu, by náhle zjistilo, že je kontinentálním suverénem. – And Germany, desperately trying to keep the declining union, would have suddenly discovered that it is the continental sovereign.
Výsledkem by nebylo opakování třicátých let s neodvratitelným náběhem ke konfrontaci, s čím však by bylo nutno počítat, by byl pocit frustrace, vlna obezřetnosti a vyvedení Evropy z rovnováhy, jež by další Maastricht odsunuly na neurčito. – The result would not be a

repetition of the thirties with an inevitable rise of confrontation; however, it would be necessary to expect a feeling of frustration, a wave of prudence and pulling Europe out of balance, which would postpone further Maastricht indefinitely.

Otázkou je, zda by potom (= v takovém případě) USA chtěly v rozdělující se Evropě zůstat a nést všechna případná rizika spolu s ní. – The question is whether the U.S. would have wanted then (= in that case) to remain in the dividing Europe and carry any risks with it.

Schopnost mít v hlavě šikovně uspořádanou a dostatečně rozsáhlou kartotéku slov a umět z ní vybírat přiléhavé výrazy nemůže být výsadou spisovatelů. – The ability to have arranged and sufficiently extensive files of words in the head and be able to choose the fitting expressions from it cannot be a privilege of writers.
A nelze očekávat, že nás k ní dovede dlouholetá zkušenost, ***to*** **(= v tom případě)** *by se dobře vyjadřovali jenom geronti. – And we cannot expect that long experience will lead us to it;* ***in such case****, only geronts would express their ideas properly.*

These cases are solved in this way: If the first argument may be reformulated by adding a subordinate clause with the connective *jestliže/pokud/když...* (all meaning *if, whether*) without difficulty, we annotate it as a common condition with a comment. In ambiguous contexts, we make a decision according to the prevailing semantic aspect; the second semantic aspect may be captured by a comment.

7. Verification Experiments

7.1. Inter-Annotator Agreement Measurement

Throughout the work on annotation of the interclausal semantic relations, we have measured the inter-annotator agreement. The whole volume of data, i.e. 49,431 sentences, was divided into 10 parts for the annotation purposes – 8 of them serve as training data for automated processes in the future (train-1 – train-8), 2 of them as checking parts (dtest and etest). Each part was further divided into five roughly equal subparts (of approximately 1000 sentences) that were given to the annotators, respectively. In each of these parts, we have selected an overlap of approx. 200 sentences. This overlap was annotated in parallel by all annotators who worked on the given part and it formed the basis for the measurement of the inter-annotator agreement. (The annotators did not know which files were selected for these purposes.)

In total, the data contained 22 annotated types of interclausal semantic relations (e.g., conjunction, condition, precedence – succession – see Table 3 in Chapter 5). At a higher level of generalization, we may talk about four classes – temporal relations, contingency, comparison and expansion.

For inter-annotator agreement, we have used three different methods: F1 based on recognition of connectives, agreement on types and Cohen's kappa. F1 based on connectives proved to be the optimal method for assessing the agreement in annotations of such complex textual phenomenon as the interclausal semantic relations (more details including comparison with other methods see Mírovský et al. 2010).[4] In this approach, agreement is understood as such situation in which connectives identified in both/multiple annotations have a non-empty intersection. The more F1 verges to one, the better is the inter-annotator agreement on connectives. The agreement on types demonstrates the proportion of relations that annotators evaluated uniformly from the semantic point of view (they assigned them the same type) among the relations that were captured in both annotations. Cohen's kappa is considered one of the most reliable methods for the assessment of the inter-annotator agreement – it takes into account also the cases when the annotators were in agreement only by chance. Also for this

[4] Mírovský, Jiří, Mladová, Lucie, Zikánová, Šárka. 2010. *Connective-Based Measuring of the Inter Annotator Agreement in the Annotation of Discourse in PDT*. In Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010), Beijing, China. s. 775–781.

measure is valid, the more it verges to one, the better the agreement among the individual annotators.

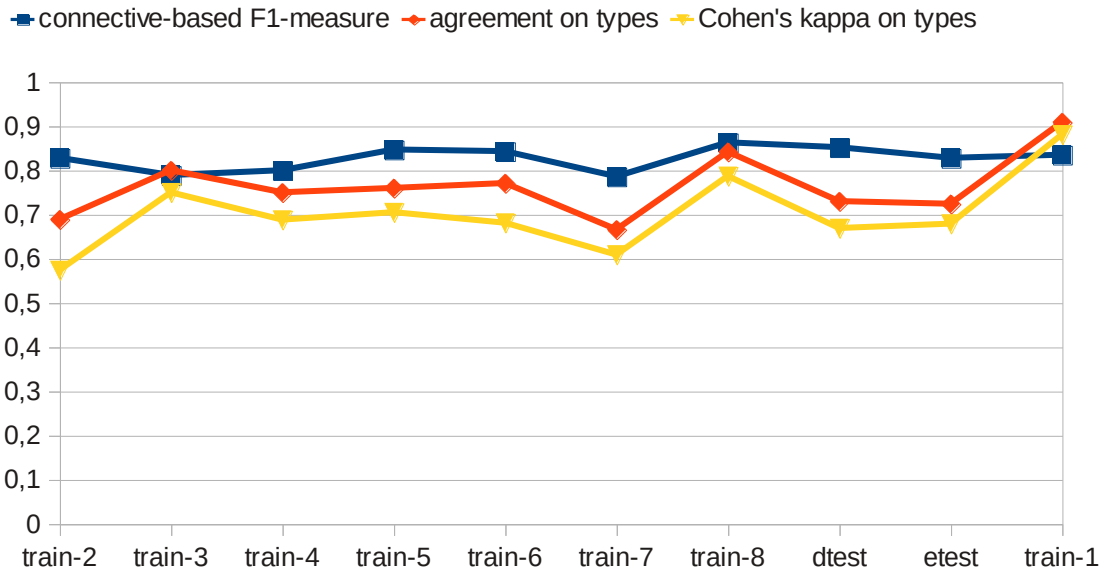
In addition to the regular measurements of agreement throughout the whole project, we have carried out a comprehensive cross-sectional measurement on all parts of the corpus in fall 2011. For clarity, we compared data from two annotators who as the only ones annotated all parts of the treebank. The measurement was limited to the inter-sentential relations. The results of the measurements are presented in Table 5.

measurement	F1	agreement on types	kappa on types
train-2	0,83	0,69	0,57
train-3	0,79	0,8	0,75
train-4	0,8	0,75	0,69
train-5	0,85	0,76	0,71
train-6	0,84	0,77	0,68
train-7	0,79	0,67	0,61
train-8	0,86	0,84	0,79
dtest	0,85	0,73	0,67
etest	0,83	0,72	0,68
train-1	0,84	0,91	0,88

Table 5: Inter-annotator agreement gradually measured on the parallel data in all parts of the treebank

The order of the individual parts of data in Table 5 corresponds to the time sequence of annotation. The part train-1 is included as the last one because this part was completely re-annotated in the end. This section was the first the annotators worked on, and therefore it was possible to expect that train-1 will contain inaccuracies resulting from the lack of experience with annotations.

For better clarity, the data from Table 5 are transferred to Graph 1.



Graph 1: Inter-annotator agreement gradually measured on the parallel data in all parts of the treebank

In both Table 5 and Graph 1, we can observe a slightly upward tendency in terms of agreement on connectives (F1 measure) and the highest agreement on types in the most recent annotated part. These data, in our opinion, reflect the gradual acquisition of experience with the texts. They are also the result of a gradual fine-tuning of the annotation concept.

Table 6 presents the results of measurements of all types on all parallel data simultaneously. There were altogether 44 documents and 2,087 sentences.

measurement	F1	agreement on types	kappa on types
all parallel data	0,83	0,77	0,71

Table 6: Inter-annotator agreement measured on all parallel data

Generally, these results can be considered satisfactory for the given type of annotation. For example, the agreement on types 0.77 corresponds well with the results in the Penn Discourse Treebank based on a similar theoretical approach of a research group led by prof. Aravind Joshi – in this corpus, the measured value was 0.8 (for details see Prasad, R. et al. 2008. The Penn Discourse Treebank 2.0. In Proceedings of the 6th International Conference on Language Resources and Evaluation, CD-ROM). We should also keep in mind that we have measured the inter-annotator agreement only for inter-sentential relations for now. Relations within a sentence are already adequately captured on the tectogrammatical representation of the Prague Dependency Treebank in most cases. When these relations are annotated automatically in the next phase of annotation, the agreement will increase even more.

The measurement of the inter-annotator agreement included also the analysis of disagreements among annotators. The results of this analysis were the basis for checking procedures for the data.

Our measurement method demonstrates that there may be altogether two types of disagreement: disagreement on the connective and disagreement on the semantic type. In both cases it is necessary to distinguish situations with an error from situations with two acceptable solutions (the expression may be or need not be a connective in the given context, i.e. the given context allows both types – both semantic interpretations of the relation). As for connectives, the analysis demonstrated that the vast majority of cases (99%) fall into errors – the annotators overlooked the given connective. Most of these cases (75%) contained expressions that may be or need not be connectives in different contexts and it is, therefore, easier to overlook them in the text. In case of disagreement on types, the errors of annotators accounted for 26%, other cases allowed a double interpretation: 33% were cases in which the annotators disagreed on the semantic type, but they chose types within the same class, and 41% were cases in which the annotators chose types belonging to different class. Situations in which the annotators agree on class were understood as agreement in the approach of Penn Discourse Treebank. Therefore, we can generally state that differences in semantic types are an issue of an ambiguous context in 75%.

Agreement and disagreement on the class level is clearly presented in Table 7. Each class is provided both with a number of cases in which the annotators agreed on the class (cells with a grey background) and all variants of confusion.

	con	**caus**	**add**	**time**	**total**
con	137	2	5	1	145
caus	1	49	5		55
add	4	8	60	3	75
time		1	1	7	9
total	142	60	71	11	284

Table 7: Contingency table of agreement on four classes: contrastive relations, causal relations, additive relations, temporal relations

In our opinion, this table relatively clearly demonstrates that the annotation is quite consistent in terms of classes.

Within the individual classes, there is quite often a disagreement among the individual types of the contrast class and among some of types from other classes (generalization and equivalence, explication and reason). This information offers a valuable feedback for checking and adjusting the interpretation of the given relations on the basis of real-text data.

7.2. Automatic checking procedures

During the manual annotation of discourse relations, we have formed proposals on automatic checking procedures that allow either directly find and correct errors of certain types or at least suggest where an error probably occurs. Some of these checking scripts have been already programmed and implemented in December 2011 (or they are made continuously with every part of processed data), other are in progress and they will be activated on all data at once. We expect that these procedures will detect also the types of errors that will require manual or semi-automatic data revision.
There are some examples of checking scripts that are running / will be activated:

The rules that are always valid:

 (`_find_errors_1.ntred`) every relation (arrow) is provided with a connective – on condition that it is the type 'discourse' and it does not have a discourse type 'spec' (some other cases may have exceptions; it is also included in the script)

 (`_find_errors_2.ntred`) in every relation (arrow), the attribute src is non-empty

 (`_find_errors_3.ntred`) nodes from/to which discourse arrows lead are either complex nodes (nodetype = complex) with the value of "v" in the grammateme sempos or roots of the coordinate structures (nodetype = coap) or they are quasi-complex (nodetype = qcomplex) and have substitutional t-lemma # EmpVerb.

 (`_find_errors_4.ntred`) at least one arrow leads from group/to group

 (`_find_errors_5.ntred`) a group consists of fewer nodes than one tree, or on the contrary more nodes than one tree (in the future stronger – boundaries of the group do not coincide with the boundaries of the trees, i.e. the group consists of fragments of trees, never exactly of whole trees)

 (`_find_errors_6.ntred`) a list structure includes more than one item

 (`_find_errors_7.ntred`) if the attribute start_range is assigned to some node in the last t-tree of the document it can only have values 0 or group.

Rules with possible exceptions:

- the arrow of a 'list' type usually has a connective; the arrow of a 'discourse' type with a discourse type 'spec' usually has a connective
- attribute is_heading usually belongs to the effective root of the tree
- every node with the functor PREC is either a connective or it is provided with a comment (an exception is *tak – so*, *pak – then* in pairs like *jestliže-pak – if-then, pokud-tak – if-so, když-tak – when-so* etc.)
- every file usually contains at least one attribute is_heading = 1

The ÚFAL/CKL Technical Report Series

ÚFAL

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

CKL

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

Technical Reports

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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