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

Topic–Focus Articulation in English Texts on the Basis of Functional Generative Description

Kateřina Rysová, Magdaléna Rysová, Eva Hajičová

ÚFAL Technical Report
TR-2015-59

ISSN 1214-5521

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

Copies of ÚFAL Technical Reports can be ordered from:

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

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

Topic–Focus Articulation in English Texts on the Basis of Functional Generative Description

KATEŘINA RYSOVÁ

MAGDALÉNA RYSOVÁ

EVA HAJIČOVÁ

Charles University in Prague

Faculty of Mathematics and Physics

Institute of Formal and Applied Linguistics

Prague 2015

Citation: Rysová, Kateřina, Magdaléna Rysová and Eva Hajičová. 2015. *Topic–Focus Articulation in English Texts on the Basis of Functional Generative Description*. Technical report TR-2015-59. Prague: Charles University in Prague, Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics. ISSN 1214-5521. Available from: <http://ufal.mff.cuni.cz/techrep>

Contents

Acknowledgements

We gratefully acknowledge support from the Ministry of Education of the Czech Republic (project LM2010013 – LINDAT/CLARIN *Establishing and operating the Czech node of pan-European infrastructure for research*) and the Czech Science Foundation (GA CR, project P406/12/0658 *Coreference, discourse relations and information structure in a contrastive perspective*) and the Charles University in Prague (project P46 PRVOUK – *Programs of development of scientific areas at the Charles University in Prague: Informatics*). This work has been using language resources developed, stored and distributed by the LINDAT/CLARIN project.

Abstract

In the report, we introduce the concept of topic–focus articulation on the basis of Functional Generative Description. Firstly, we present the crucial terms connected with topic–focus articulation – mainly contextual boundness and communicative dynamism and we describe operational criteria how to detect topic and focus: the so-called question test and test by negation. In the next part, we present the annotation principles for annotation of topic–focus articulation in the Prague Czech-English Dependency Treebank.

List of Abbreviations

c – contrastive contextually bound nodes
CB – contextual boundness
CD – communicative dynamism
f – contextually non-bound nodes
PCEDT – Prague Czech-English Dependency Treebank
PDT – Prague Dependency Treebank
t – non-contrastive contextually bound nodes
TFA – topic–focus articulation

Abbreviations for functors used in the PCEDT and in the PDT are in a separate table at the end of the document.

1 What Is Topic–Focus Articulation

One way to look at discourse is to view it as a sequence of utterances, taking into account the so-called information structure of the sentence (topic–focus articulation). This aspect of sentence structure is a good "bridge" towards a study of (at least one aspect of) the dynamic development of discourse. This, of course, is not a new idea: To our knowledge, its first comprehensive treatment, though taken from a psychological rather than linguistic perspective, was provided by Weil (1844). According to Weil (1978, p. 11), "Words are the signs of ideas; to treat of the order of words is, then, in a measure, to treat of the order of ideas." Weil recognized two types of "movement of ideas," namely *marche parallèle* and *progression*: "If the initial notion is related to the united notion of the preceding sentence, the march of the two sentences is to some extent parallel; if it is related to the goal of the sentence which precedes, there is a progression in the march of the discourse" (ibidem, p. 41). It should not be overlooked that Weil (ibidem, p. 45) also noticed the possibility of a reverse order which he calls 'pathetic': "When the imagination is vividly impressed, or when the sensibilities of the soul are deeply stirred, the speaker enters into the matter of his discourse at the goal."

In more modern terms, one can say that two adjacent utterances may either be linked by their topics or the topic[1] of one utterance may be linked to the focus of the preceding one (see the two basic types of thematic progressions in Daneš, 1974).

The readers or hearers of a text are accustomed of being informed from a particular perspective. They expect to receive a certain anchor, i.e. to start with what they have already known and on the basis of this "old" knowledge they accept "new" concepts or new relations among previously mentioned elements. These new concepts or new relations then fit into the previous (con)text and become known. And again, through the information that was just obtained, people can accept more new information. The same principle is usually reflected in the build-up of a text and on lower layer, in the formulation of individual sentences. In this way, topic–focus articulation performs the communicative function of the text.

[1] In different approaches to this domain of study different terminology is used: topic – focus, theme – rheme, background – focus, etc. The underlying ideas are very close to each other, though there are, of course, differences in their interpretation.

2 The Importance of Topic–Focus Articulation – Language Comic and Misinterpretation

Besides the above mentioned communicative function, topic–focus articulation is also a language phenomenon that significantly affects the sentence semantics, cf. Example (1).

(1) *Entry with dogs on leash only.*

The sentence in Example (1) can be interpreted in two ways: (i) the entry of dogs is allowed only if they are on a leash, or (ii) the entry is allowed only if you have a dog (on a leash). The intonation center is put on the word *leash* in both cases: *Entry with dogs on LEASH*[2] *only*. The two interpretations vary in the scope of the focus particle *only* (called focalizer or rhematizer).[3] In the first case, the focus particle *only* concerns the participant *on leash* – while in the second case, only pertains to the whole prepositional group with *dogs on leash*.

The misinterpretation of the topic and focus of a sentence may cause misunderstandings between the speaker and the addressee and may also be a source of language comic, see e.g. Example (2).

(2) *Why do we dress baby girls in pink and baby boys in blue? Because they do not know how to dress themselves.*

In the most common interpretation of the sentence, the pronoun *we* stays in the background of our attention; the emphasis is put on the colors of girls' and boys' clothing. However, the answer deals with the pronoun *we* as if it were emphasized: It says why the baby girls and boys are dressed exactly by us (not why they are dressed in pink and in blue as we would probably expect). It should be noted that the position of the intonation center again plays an important role here. Both examples illustrate the importance of the distinction between the information the addressees understand as the topic of the sentence, and the information newly introduced and non-identifiable.

In the first sections, we describe the theoretical basis and fundamental notions of the theory of topic–focus articulation that we subscribe to, such as contextual boundness,

[2] The intonation center is henceforth marked in capitals.
[3] For the interpretation of rhematizer, see Hajičová (1995). A detailed analysis of this category based on the PDT material is given by Štěpánková (2014).

communicative dynamism and topic and focus. In the next parts, we outline how topic–focus articulation is captured in the Prague Czech-English Dependency Treebank.

3 The Theoretical Basis

The original formulations of what is now more generally referred to as the information structure of the sentence were based on a dichotomy, be it a distinction between psychological subject and psychological predicate, theme – rheme, topic – comment, topic – focus, presupposition and focus, given and new information etc. In structural linguistics, the pioneer of the study of these topics was Mathesius, who refers to Weil (1844) quoted above, and to linguists around Zeitschrift für Völkerpsychologie, von der Gabelentz (1868), Paul (1886), and esp. Wegener (1885), though criticizing their terms psychological subject and psychological predicate (Mathesius, 1907). Mathesius himself refers to this articulation by the Czech term *aktuální členění* (literally translated as "actual articulation") because it is determined (guided) by the "actual," that is "topical" situation of the speaker and concerns the way, in which the sentence is incorporated into the factual relation to the situation from which it originated (Mathesius, 1939). Mathesius distinguishes between *východiště výpovědi* (initial starting point of the utterance, its basis), which he specifies as "what is known or at least evident in the given situation and from where the speaker starts" on the one hand and *jádro výpovědi* (nucleus of the utterance), that is "what the speaker utters about with respect to the starting point of the utterance." Mathesius prefers the above specification rather than using known and unknown. However, already in Mathesius' writings we see a certain inclination to recognize a more articulated scale rather than a mere dichotomy, when he says that the starting point may contain more than a single element so that it is possible to speak about the center of the starting point and the accompanying elements which "lead from the center to the nucleus." Referring to the position of the sentence predicate, Mathesius writes that the predicate is a part of the nucleus but on its edge rather than in its center and represents a transition between the two parts of the utterance.

Mathesius' observations inspired the fundamental work of Firbas and his team. As Mathesius' original Czech term *aktuální členění větné* is not directly translatable into English and apparently inspired by Mathesius' use (Mathesius, 1929) of the German term *Satzperspektive* Firbas used the term functional sentence perspective (FSP). Very early in the development of the FSP approach, the binary articulation into theme and rheme was complemented – also in line with Mathesius' ideas mentioned above – by a more structured

approach introducing the notions of transition and even a more scalar notion of communicative dynamism (CD). From this point of view, theme was specified by Firbas (1964) as being constituted by an element or elements carrying the lowest degree(s) of communicative dynamism within a sentence (which was later modified by Firbas (1992) in the sense that theme need not be implemented in every sentence, while in every sentence there must be rheme proper and transition proper). The concept of communicative dynamism was characterized by Firbas (1971) as a hierarchy of degrees carried by a linguistic element of the sentence, i.e. "the extent to which the element contributes towards the development of communication." The basic distribution of communicative dynamism would then reflect what Weil (1844) called the "movement of the mind."

Almost in parallel with FSP, but also partly as a reaction to it, Sgall and his collaborators in Prague developed the theory of topic–focus articulation (TFA) (see e.g. Sgall, 1967; Sgall, Hajičová and Benešová, 1973; Sgall, Hajičová and Buráňová, 1980; Sgall, Hajičová and Panevová, 1986; Hajičová, Partee and Sgall, 1998). The theory of topic–focus articulation is an integral part of the formal model of Functional Generative Description of language, namely of the representation of sentences on the underlying (tectogrammatical) sentence structure. These tectogrammatical representations are viewed as dependency trees, with the main verb being the root of the tree. Every node of the tree carries – in addition to other characteristics such as the type of dependency – an index of contextual boundness: a node can be either contextual bound or non-bound. This feature, however, does not necessarily mean that the entity is known from the previous context or new but rather how it is structured as for the information structure of the sentence.

With the help of the bound/non-bound primary opposition, the distinction between the topic and the focus of the sentence can be defined depending on the status of the main verb (i.e. the root) of the sentence. If the verb is contextually bound then the verb and all the nodes depending (immediately or not) on the verb constitute the topic, the rest of the sentence belongs to its focus; if the verb is contextually non-bound, then the verb and all the nodes depending on it to the right constitute the focus, while the rest of the sentence belongs to its topic (see the definition of topic and focus in Sgall, 1979).

The left-to-right dimension of the tree serves as the basis for the specification of the scale of communicative dynamism: Communicative dynamism is specified as the deep word order, with the least dynamic element standing in the leftmost position and the most dynamic element (the focus proper of the sentence) being the rightmost element of the dependency tree.

In spoken language, the most important means of expressing the difference in topic–focus articulation is the sentence prosody including the placement of the intonation center; in our more recent work with spoken language corpora, the characteristics of the curve were considered as a marker of a contrastive topic (Veselá, Peterek and Hajičová, 2003).

Currently, the phenomenon of topic–focus articulation is included essentially in most formal (and empirical) language descriptions under different names, such as information structure (the term used by a number of authors, e.g. by Steedman, 1991 or Lambrecht, 1996); see also the treatment of communicative structure in the Meaning – Text Theory as developed by Mel'čuk (1981).

In our analysis, we use the Functional Generative Description as the main theoretical basis for our linguistic approach and also as the basis for annotating topic–focus articulation in the Prague Czech-English Dependency Treebank; we also utilize the term topic–focus articulation.[4]

4 Basic Terms of Topic–Focus Articulation

The description of topic–focus articulation is based on three main features: (i) contextual boundness; (ii) communicative dynamism and (iii) sentence division into topic and focus. Topic and focus are defined on the basis of the first two characteristics. Therefore, we introduce the contextual boundness and communicative dynamism phenomena first and then describe the conception of topic and focus within Functional Generative Description approach.

4.1 Context and contextual boundness

Sentences in a coherent text are interconnected by various types of relationships (explicitly marked or implicitly present) – the relationship of contextual boundness between sentence items and the context is one of them. The context can be provided by the previous sentences (i.e. by the previous text or texts) or by the broader setting of situation in which the text is created or perceived. The situational context is not fixed and its setting can influence the text perception (e.g. Shakespeare's dramas were understood differently in 17th century than now because the situational context has changed). The situational context includes any shared or

[4] For comparison of the FGD approach with the further approaches to topic–focus articulation, see Hajičová (1972); Sgall, Hajičová and Benešová (1973); Sgall (1975); Hajičová, Partee and Sgall (1998) or Hajičová (2012).

generally known information, which may be determined by the immediate situation or longer experience, senses, culture or other factors.

Depending on the context, we can decide for every sentence item (that is relevant for topic–focus articulation) whether it is contextually bound or non-bound. In the Functional Generative Description, the contextual boundness is a property of an element of the sentence (expressed or absent in the surface sentence structure) which determines whether the author uses the sentence element as given (for the recipient), i.e. uniquely determined by the context, see Hajičová, Partee and Sgall (1998). It means that contextually bound sentence items are deducible from the broader context, see Example (3).[5]

(3) *(Jane is my best friend.) She is very NICE.*

The pronoun *she* is contextually bound because it is deducible from the previous context. On the contrary, all other sentence items are contextually non-bound in this case because they bring information that cannot be deduced from the (previous) context.[6]

The relationship of contextual boundness may seem similar to coreferential and anaphoric relations. Nevertheless, they do not necessarily coincide since they describe data from different points of view. In Example (4), items *her* and *him* have a coreferential relation to some previous sentence items. However, they are contextually non-bound, as they present the items from the context in a new, indeducible relation.[7]

[5] The sentence in parentheses denotes the context, be it immediately preceding or distant, in which the example sentence is supposed to be uttered.

[6] For determining the contextual boundness of a particular node, it is not so important whether the item itself is known from the context, but whether it is used as known by the author – cf.: *Romney and Obama are presidential candidates. OBAMA$_f$ is the favorite$_t$. / Obama$_t$ is the FAVORITE$_f$.* ("t" is a label for contextually bound nodes, "f" for contextually non-bound nodes).

The sentence *Romney and Obama are presidential candidates* may be followed by a sentence in which the node *Obama* is presented as contextually non-bound (see the first case) or bound (see the second case).
In the first case, the author says who is the favorite (it is Obama): *Romney and Obama are presidential candidates. (And you probably want to know who is the favorite.) OBAMA$_f$ is the favorite$_t$.*

In the second case, the author speaks about Obama – he or she says about him that he is the favorite (in this case, the communication could continue by mentioning the second candidate): *Romney and Obama are presidential candidates. (And you probably want to know something about them.) Obama$_t$ is the FAVORITE$_f$. and Romney is supposed to LOSE.*

[7] Contextually non-bound may be also a sentence item that is already known to the addressee: *Do you think it's better to finish the business or continue for a higher score? – Definitely TO FINISH$_f$.*

Repetition of a previously used item (*to finish*) clearly brings new information to the addressee in this case; therefore, this item is contextually non-bound.

In this context, we may mention some interesting cases in terms of communication perspective, see the following sentences: *Business$_t$* ***is$_f$ BUSINESS$_f$.*** / *Promise$_t$* ***is$_f$ PROMISE$_f$.*** / *The situation$_t$* ***is$_f$ what$_f$ it$_t$ IS$_f$.***

In these sentences, the expression that appears in the part that is spoken about (topic) is the same as the expression in the part that says something (focus). Although it seems that similar sentences have no function in

(4) *(For some Catholics, Mary$_1$ is more important than Christ$_2$.) They go to HER$_1$ not to HIM$_2$.* (Perspective Digest)

4.1.1 Contrast and contextual boundness

A common way of how information can be formulated is to express it in contrast with the known context. This contrastivity is reflected also in the topic–focus articulation structure. Namely, contextually bound sentence elements can stand in contrast as in Example (5):

(5) *(We have two children.) John$_c$ is the YOUNGER, Mary$_c$ is the OLDER.*[8]

In this example, *John* and *Mary* are presented by the author as contextually bound – they were introduced in the first sentence and now they are referred to as a starting point for the flow of the text in which information about their age is presented. On the other hand, they are presented in contrast to each other, with the background formed by the word *children*. In the Functional Generative Description, this case is discerned as a special subtype of contextual boundness – the contrastive contextual boundness.

There are typical ways to formally express the feature of contrastivity for contextually bound items. One of them is contrastive stress, as in Examples (5). In Czech, also specific (long) forms of pronouns are used to express contrastivity, while non-contrastive contextually bound pronouns have short forms, cf. Examples (6) and (7) with a long stressed contrastive form *tebe* [*you*] and a short clitic non-contrastive form *tě* [*you*].

(6) *Tebe$_c$ já NEZNÁM.*
lit. *You$_c$ I DO_NOT_KNOW.*
Concerning you$_c$, I DO NOT KNOW you.

a literal level (it is not effective to inform someone that something is something; it would be a vicious circle), they are functional in authentic communication.

The aim of similar sentences is not to give a definition of the expression in the role of subject, but they have a function as a whole structure. In most cases, these examples are idioms that gain the real meaning from the context, e.g., 'business is about profit, anything else is secondary' or 'if we promise something, we should fulfil it'.

[8] Here and in further examples, contrastive contextually bound items are labeled with c, non-contrastive contextually bound items bear a mark t, the contextually non-bound nodes are marked as f.

(7) *Já tě$_t$ NEZNÁM.*
lit. I you$_t$ DO_NOT_KNOW.
I DO NOT KNOW you$_t$.

So far we have dealt with contrastivity for contextually bound items. For the contextually non-bound sentence members, such a distinction is not considered to be relevant, because as a matter of fact, the newly presented items always concern a choice of alternatives and to some extent stand in contrast to the previous context. Unlike contextual boundness, contextual non-boundness has no special formal means for discerning the feature of contrastivity, see Example (8).

(8) *While you*$_{\mathbf{c}}$ *prefer TEA*$_{\mathbf{f}}$, *we*$_{\mathbf{c}}$ *prefer COFFEE*$_{\mathbf{f}}$

In the example, there are two semantic contrasts: 1) *you – we*, 2) *tea – coffee*. Whereas contrast expressed by the pronouns (*you – we*) is a part of topic (and it is thus annotated: both pronouns are assigned the value of contrastive contextually bound expressions, i.e. the value "c"), the second contrast (*tea – coffee*) is not captured in the annotation because the given expressions are in focus (where the contrastivity is not marked – both items are assigned the value for contextually non-bound items, i.e. "f").

Commonly, when annotating topic–focus articulation, we take into account especially the preceding context. Nevertheless, contrast is such phenomenon that may recognized (in some cases) from the context of the following sentences. At the same time, contrast may concern even longer chains of items, see Example (9).

(9) *On Monday*$_{\mathbf{c}}$, *I usually clean my ROOM.*
On Tuesday$_{\mathbf{c}}$, *I usually go for a walk in the PARK.*
On Wednesday$_{\mathbf{c}}$, *I usually do SHOPPING.*

From the first occurrence of the temporal modification *on Monday*, the reader does not know whether it is used contrastively to the following expressions (as in Example (9)) or not. Only when reading the second sentence with the item *on Tuesday*, we realize that the expression *on Monday* stands in contrast to the following number of days. At the same time, the backward reinterpretation of text is common in communication, see Example (10).

(10) *The children had very nice bows. However, they did not have any arrows.* / *They tied them themselves.*

In Example (10), the reader may understand the right meaning of the word *bow* only from the following context (the reader has to reinterpret it when he or she understood it incorrectly during the first reading). Similarly, the annotator may re-annotate the value of contextual boundness when he or she reads the following context.

On the contrary, the author of text knows his or her communication perspective from the beginning. Therefore, the author knows already during uttering the given expression whether the word *bow* is used in the meaning of a tool or piece of clothing. Similarly, the author knows whether he or she uses some contextually bound expression contrastively – and if so, he or she puts a contrastive stress on it (in the spoken form).

To sum up, the theory discerns two basic categories: contextual non-boundness, contextual boundness and a subcategory of the contrastive contextual boundness. A distribution of sentence items with various values of contextual boundness is presented in Example (11).

(11) *Across the river$_c$ Magda$_t$ and Kovarik$_t$ could now$_t$ see$_f$ a FIRE$_f$ with two$_f$ figures$_f$ beside it$_t$. When they$_t$ moved$_f$ closer$_f$, they$_t$ could make$_f$ out two$_f$ white$_f$ HORSES$_f$ against the background$_f$ of the dark$_f$ bushes$_f$. Then$_t$ he$_t$ [Kovarik] RECOGNIZED$_f$ them$_t$.*
(Škvorecký, 1986)

We can observe that mainly the temporal and circumstantial adjuncts in the role of scene setting (e.g. *beside it, now*) and subjects presented as given (e.g. *Magda and Kovarik, they*) are contextually bound. On the contrary, most of predicates (e.g. *could see, could make out, recognize*) are contextually non-bound because they are not deducible from the context. Contrastive contextually bound sentence items are rather rare in authentic texts. In Example (11) there is only one item marked as contrastive contextually bound sentence element, namely the local setting *across the river*. The location is given by the broader context of the situation but it offers a choice of one alternative out of several others (*on this side of the river, at distance, …*) given within that context.

At the same time, we can see that contextually bound sentence items can be modified also by contextually non-bound sentence elements (e.g. *two figures beside it*) and on the

contrary, contextually non-bound sentence items can be modified by dependent contextually bound elements, see Figure 1 in Section *8.4 Dependency Trees*.

4.2 Communicative Dynamism

When observing the sentence and its contextually bound and non-bound parts, we can see that the individual sentence items mutually differ in degrees of their relative importance. Firbas (1971) characterized this phenomenon as communicative dynamism and postulated the concept of information hierarchy in the sentence. Firbas likened communicative dynamism to information flow. He claimed that the degree of communicative dynamism is specified as relative importance with which the given element contributes to the development of communication, i.e. to what extent the sentence element moves the communication forward.

The Functional Generative Description took over this concept and applied it in formal description. According to Hajičová, Partee and Sgall (1998), communicative dynamism is a property of a sentence element that reflects its relative degree of communicative importance attributed to it by the author – compared with other sentence elements in the sentence; contextually non-bound sentence elements are considered to be more dynamic than sentence elements contextually bound (be they non-contrastive or contrastive). Communicative dynamism is not seen as a dichotomy but as a scale with more degrees. Such a scale is reflected in the so-called deep word order. Deep word order describes the organization of elements in a sentence structure according to their increasing communicative dynamism. In some cases, deep word order can be directly related to the surface word order,[9] see Example (12).

(12) *He*$_{t}$ *looked*$_{f}$ *at MAGDA*$_{f}$. (Škvorecký, 1986)

In Example (12), there is one contextually bound item (*he*) and two contextually non-bound items (*to look*, *Magda*). The contextually bound item carries the lowest degree of communicative dynamism, i.e. the lowest relative degree of importance, and it is followed by contextually non-bound items that carry a higher degree of communicative dynamism. At the

[9] The surface word order is the ordering of sentence elements in the surface structure, i.e. the word order in sentences realized in authentic texts (for more details, see Rysová and Mírovský, 2014). The difference between the deep and surface word order occurs more frequently in languages with a grammatically fixed word order (such as English), while with languages such as Czech the surface word order is typically governed by topic–focus articulation and as such corresponds to the deep word order (see Rysová et al., 2015).

same time, the predicate (*to look*) carries a lower degree of communicative dynamism than the element *Magda*, despite the fact that both of them are contextually non-bound.

Empirical investigations of topic–focus articulation in Czech have indicated that the individual values of communicative dynamism are connected with contextual boundness. However, it is supposed that the individual values of communicative dynamism function differently among contextually bound sentence items in comparison with contextually non-bound items (directly dependent on their governing verb). The order of contextually bound modifications directly depending on the verb is determined in the scale of communicative dynamism by the choice of the author and it may be affected by various factors – the language factors (e.g. Actor may be chosen as the least dynamic item more easily than other participants), the situation factors (e.g. whether the entity was mentioned in the immediately preceding context or whether it is not really activated in the consciousness of the author and addressee) or by factors related to the text composition (e.g. use of contrast).

On the other hand, the contextually non-bound verb modifications directly depending on the verb are supposed to follow the so-called systemic ordering (Sgall, Hajičová and Buráňová, 1980, see also Zikánová, 2006; Rysová, 2011; Rysová, 2014), i.e. a scale of communicative dynamism for contextually non-bound sentence items directly dependent on their governing verb. Systemic ordering presumes e.g. that contextually non-bound Patient carries a higher degree of communicative dynamism than e.g. contextually non-bound Temporal modification in English sentences. The existence of systemic ordering in languages is considered to be language independent but the individual degrees of it are language specific (i.e. systemic ordering in Czech is different than systemic ordering in English).

In English, systemic ordering is only rarely reflected in the surface word order. However, in Czech we can study its systemic ordering particularly from the surface word order. In most cases, the contextually non-bound sentence items (directly dependent on their governing verb) also follow the systemic ordering in surface word order because Czech is a language with free word order and its surface word order is affected by communicative dynamism to a large extent (unlike English).

While in English e.g. the order of the members carrying the highest degree of communicative dynamism is mostly grammatically fixed, in Czech they are usually placed at the very end of the sentence, cf. Example (13).

(13) *Potom*$_{t}$ [*on*$_{t}$] *je*$_{t}$ *POZNAL*$_{f}$. (Škvorecký, 1991)
lit. *Then*$_{t}$ [*he*$_{t}$] *them*$_{t}$ *RECOGNIZED*$_{f}$.
Then$_{t}$ *he*$_{t}$ *RECOGNIZED*$_{f}$ *them*$_{t}$. (Škvorecký, 1986)

In Example (13), the most dynamic element is *poznal* [*recognized*]. All other sentence items carry a lower degree of communicative dynamism. In Czech, this fact is captured also in surface word order – the most dynamic sentence element is placed in the last position whereas the object *je* [*them*] stands before the predicate. On the contrary, in English, the last item is the word form *them* that is contextually bound and therefore also less dynamic than the contextually non-bound predicate *recognized*. On the basis of this example, we can see that Czech surface word order is much more influenced by communicative dynamism than the word order in English. In English, surface word order is affected more by grammatical factors than by topic–focus articulation.

4.3 Topic and Focus

On the basis of the previously described phenomena (contextual boundness and communicative dynamism), it is possible to distinguish two parts of the sentence – topic and focus. These terms no longer concern the individual sentence elements as contextual boundness and communicative dynamism but are related to the larger parts of sentences.

Generally speaking, between topic and focus, there is a relation of aboutness – focus says something about the topic (cf. Hajičová, Partee and Sgall, 1998). A simple example of topic and focus can be demonstrated as follows (topic is in plain text, focus is printed in bold):

(14) *He*$_{\text{t}}$ ***looked*$_{\text{f}}$ *at MAGDA*$_{\text{f}}$**. (Škvorecký, 1986)

The sentence is about *him*, hence this is the sentence topic. The other part of the sentence (*looked at MAGDA*) is a statement *about him*, i.e. sentence focus.

In a more detailed description, we can characterize topic as the part of a sentence that consists of all contextually bound sentence items directly dependent on their main governing verb. These items can also be further modified by other sentence members (e.g. by attributes) that can be contextually bound or non-bound – all such modifiers are also a part of topic.

At the same time, focus consists of all contextually non-bound sentence items directly dependent on their main governing verb. Also these items can be further modified by other sentence elements (like by attributes) that can be contextually non-bound or bound – all such modifiers are also a part of focus, see Example (15):

(15) (*I have two cats.*) *The black*$_c$ *one*$_t$ ***is*$_f$ *my*$_t$ *FAVORITE*$_f$**.

Example (15) demonstrates that the element *my* is a part of focus, though it is contextually bound.

The governing verb itself can be contextually bound or non-bound. If it is contextually bound, it is a part of topic; if it is contextually non-bound, it is a part of focus, see Example (16). For more details about the algorithm for detection of topic and focus, see, in particular, Sgall, Hajičová and Panevová (1986), Zikánová, Týnovský and Havelka (2007) and Rysová, Mírovský and Hajičová (2015).

(16) *He* ***looked*$_{\text{f}}$ *at MAGDA*** *while Magda* *looked*$_{\text{t}}$ ***at someone ELSE***.

The first occurrence of the governing verb *to look* is contextually non-bound and it is a part of focus. On the contrary, its other occurrence is contextually bound (deducible from the context) and therefore it is a part of topic.

In terms of communicative dynamism, topic is (as a whole) less dynamic than focus. At the same time, the individual items of topic have different degrees of communicative dynamism. The least dynamic item (i.e. the item with the lowest relative degree of importance) is called **topic proper**. Also the individual parts of sentence focus carry different degrees of communicative dynamism and the most dynamic item is called **focus proper** (in the spoken variant of the sentence, focus proper also carries the intonation centre), see Example (17):

(17) *Then*$_{t}$ *he*$_{t}$ ***RECOGNIZED***$_{\mathbf{f}}$ *them*$_{t}$. (Škvorecký, 1986)

In Example (17), topic proper is the sentence element *he* (Actor is very often the item that is spoken about) and focus proper is the predicate because it carries the most important information.

In the next example, we present the sentences from previously used Example (11) once more – this time not only with values of contextual boundness of individual sentence items but also with marking of topic and focus in each sentence. Topics are written as plain text and focuses are printed in bold.

(18) *Across the river*$_{c}$ *Magda*$_{t}$ *and Kovarik*$_{t}$ *could now*$_{t}$ ***see***$_{\mathbf{f}}$ **a** ***FIRE***$_{\mathbf{f}}$ ***with two***$_{\mathbf{f}}$ ***figures***$_{\mathbf{f}}$ ***beside it***$_{\mathbf{t}}$. *When they*$_{t}$ *moved*$_{t}$ *closer*$_{f}$, *they*$_{t}$ ***could make***$_{\mathbf{f}}$ ***out two***$_{\mathbf{f}}$ ***white***$_{\mathbf{f}}$ ***HORSES***$_{\mathbf{f}}$ ***against the background***$_{\mathbf{f}}$ ***of the dark***$_{\mathbf{f}}$ ***bushes***$_{\mathbf{f}}$. *Then*$_{t}$ *he*$_{t}$ [Kovarik] ***RECOGNIZED***$_{\mathbf{f}}$ *them*$_{t}$.
(Škvorecký, 1986)

All sentences contain focus but not all of them also have topic (there are e.g. some sentences that are formed only by focus proper). The focus proper is an obligatory part of every sentence. It brings the most important information – the main message. Without the main message, it would not make sense to use the sentence in authentic communication. On the other hand, the topicless sentences (sometimes called hot news) are not rare. Such sentences are typically headlines or first sentences of the text presenting some new objects on the scene or very short sentences, see Examples (19)–(22).

(19) ***How*$_{\mathbf{f}}$ *Colorado*$_{\mathbf{f}}$ *State*$_{\mathbf{f}}$ *Won*$_{\mathbf{f}}$ *by Losing*$_{\mathbf{f}}$ *Jim*$_{\mathbf{f}}$ *McElwain*$_{\mathbf{f}}$ *to FLORIDA*$_{\mathbf{f}}$.**

(20) ***Once upon a time*$_{\mathbf{f}}$ *there were*$_{\mathbf{f}}$ *three*$_{\mathbf{f}}$ *FROGS*$_{\mathbf{f}}$.**

(21) ***ATTENTION*$_{\mathbf{f}}$*!***

(22) ***Page*$_{\mathbf{f}}$ *45*$_{\mathbf{f}}$.**

5 Detection of Topic and Focus

As we have seen in all previously mentioned examples, the main issue in recognizing topic and focus in sentences is an appropriate identification of the contextual boundness of individual sentence items. At the same time, topic and focus can also be detected by using operational criteria, the following two being most useful: the so-called **question test** and **test by negation**.

5.1 Question Test

The range of the focus can be reliably detected by the question test. Its formulation assumes that for every sentence it is possible to determine a set of questions which can be appropriately answered by the given sentence (with its given surface word order and given realization of intonation), see Example (23).

(23) *Tomorrow, I will read a MAGAZINE.*

For the sentence realization with the intonation centre placed at the item magazine, examples of appropriate questions are *What will you do tomorrow?* or *What will you read tomorrow?* On the contrary, an example of an inappropriate question is *When will you read a magazine?*

Each appropriate question must fully represent the relevant features of the context in which the sentence may be used. However, it should be noted that it is an artificial pair of question and answer and that it is not a natural dialogue.

The aim of the test is to identify to which part of the sentence (topic or focus) the given elements belong. In the test only the appropriate questions are used. Those sentence elements that are contained in each of the appropriate questions belong to the sentence topic; those of its elements that are not found in any given set of appropriate questions belong to its focus; those elements of the sentence which only occur in some of the appropriate questions (but not in all of them) create the potential range of the sentence topic or focus.

The application of the question test is illustrated in Example (24). We also formulated a set of possible questions – for each question, we indicated which elements from the response to the question are not included in the question itself.

(24) *Kids are playing with SNOW.*
(24a) *What are the kids playing with? (… With SNOW.)*
(24b) *What are the kids doing? (… They are playing with SNOW.)*

The member *snow* is not present in any of the created sentences – this member is thus determined as the sentence focus proper. On the other hand, the item *kids* occurs in both of them – this item is therefore the topic proper. Other sentence elements stand between the two terminal points (on a scale of communicative dynamism) and they are the potential range of the sentence topic or focus depending on the appropriate questions. It is also possible to imagine a context indicated in question below where none of the sentence elements are included in the question. If we accept this possibility, Example (24) would be understood as a sentence without topic, i.e. as hot news.

(24c) *What is going on? (... Kids are playing with SNOW.)*

While assembling the set of possible questions, we may see that according to the context (represented in questions), the tested sentence can have three possible meanings, i.e. three possible interpretations[10] of topic–focus articulation (focus is printed in bold in every example).

(25a) *What are the kids playing with?*
Kids are playing ***with SNOW***.

(25b) *What are the kids doing?*
Kids ***are playing with SNOW***.

(25c) *What is going on?*
Kids are playing with SNOW.

[10] with the given intonation

5.2 Test with Negation

Besides the question test, we can also use tests associated with negation as an operational criterion for determining certain aspects of topic–focus articulation. Hajičová (1973) consistently deals with this relationship of topic–focus articulation and negation (see e.g. also Hajičová, 1975). She concludes that in the primary case, the scope of negation is consistent with the focus of the sentence; the relation of the focus to the topic is thus negated (the sentence says that the focus is not true in relation to the topic).

The test with negation can be complemented by the notion of possible continuations as introduced by Chomsky (1969). His approach is based on the fact that in a natural continuation of the sentence, focus may contain parts of sentences that may be replaced by other parts, standing in a similar position (e.g. after the conjunctions *but*, *rather*). Chomsky particularly exemplifies this idea for questions and negative sentences but it is possible to also use it for the affirmative or negative form of Example (24) with its natural continuations, see Example (26).

(26) *Kids are* (*not*) *playing with SNOW.*

(26a) *Kids are not playing* ***with SNOW*** *but* *with sand.*

(26b) *Kids* ***are not playing with SNOW*** *but* (*they*) *are watching TV at home.*

(26c) ***Kids are not playing with SNOW*** *but* *parents are buying sweets.*

The results obtained by Chomsky's method are the same as those from the question test. According to the context, there are also three possible interpretations of topic–focus articulation (with the given intonation).

6 Topic–Focus Articulation in Interrogative Sentences

In the previous sections, we have focused on the description and analysis of TFA in declarative sentences. TFA of other types of sentences is basically similar (i.e. we can apply question test or test with negation on them), thus we do not concentrate on them separately. However, more attention is paid to the interrogative sentences because the use of operational

criteria mentioned above for distinguishing their topic and focus is more problematic by them. Therefore, in this section, we focus on the specifics of interrogative sentences (both yes/no questions and wh-questions) in terms of TFA, see also Hajičová (1984).

On the basis of PCEDT material, we may say that also yes/no questions exhibit a similar behavior (described in previous sections) to other sentence types – see Examples (27)–(30) (focus proper marked in capitals).

(27) *Do you go to the SEA this year?*

(28) *Do you go to the sea THIS YEAR?*

(29) *DO you GO to the sea this year?*

(30) *Do YOU go to the sea this year?*

The boundary between topic and focus in these questions can be traced by test with negation. The part of the interrogative sentence in scope of possible negation is its focus, the other part its topic.

More complex is the analysis and determination of TFA in wh-questions. These sentences are specific because of the presence of interrogative words like *why, who, when* etc. In some approaches (see Mathesius, 1941), the interrogative word is supposed to be focus because it represents unknown information and all other sentence items belong to topic. However, Daneš (1949) demonstrates that focus proper of the question may be the interrogative word as well as another item of the sentence, which is illustrated in the following examples. In Example (31), focus proper is the element *when*, in Example (32) the prepositional phrase *to Berlin* (contrasting with the expression *to Prague*).

(31) *We are going to visit PRAGUE.*
WHEN do you go there?

(32) *I will go to PRAGUE tomorrow.*
When will you go to BERLIN?

In this context, Firbas' study (1976) highlights the need to take into account two perspectives – the perspective of the speaker and the perspective of the hearer. The speaker does not want to say any new information but on the contrary, he or she wants to get this information from the hearer by using an interrogative pronoun. Other items of the interrogative sentence are known to the speaker. However, not all of these items are equally important for the hearer and the speaker must indicate from which perspective he or she should approach the question. The most important indicator of this perspective is intonation – if intonation center is placed on another question element than interrogative pronoun, this element is (according to Firbas) focus proper.

The boundary between the topic and focus part of the wh-questions is thus possible to be determined on the basis of the contextual boundness opposition but in this case, the main indicators of contextual boundness are detection of the perspective from which the question is formulated and intonation (placement of intonation centre).

7 Local and Global Topic–Focus Articulation

In texts, we may distinguish the so-called global and local topic–focus articulation, see Hajičová et al. (1998) that correspond to topic–focus articulation of sentences as a whole and individual clauses (see Section 8.6 *How Are Large Units of Text Annotated for Topic–Focus Articulation*). **The global topic–focus articulation** concerns all syntactic units that form a sentence as a whole. These may be simple sentences, individual sentences within a compound (paratactic) sentence, sentence equivalents (see e.g. *Oops!*), nominal clauses (*Scandal!*) or syntactically unintegrated parentheses.

At the same time, sentences as a whole may contain some individual (dependent) clauses (i.e. clauses that modify their governing and hierarchically higher unit). Individual clauses are mostly a part of focus or topic of their governing unit (on the level of global topic–focus articulation).[11]

However, we may distinguish **the so-called local topic–focus** articulation, i.e. topic and focus within individual clauses. Individual clauses are subordinate clauses, participles and gerunds (functioning as subordinate clauses, mostly conditional) or syntactically integrated parentheses whose basis is verb.

[11] However, in some cases, the boundary between global topic and global focus may be also inside a dependent clause, cf.: *What do you think that John will give me tomorrow? – I think that John will give you*$_{topic\_1}$ *a* ***new BOOK***$_{focus}$ *tomorrow*$_{topic\_1}$. More details in Hajičová et al. (1998).

8 Representation of TFA in the Prague Czech-English Dependency Treebank

The phenomenon of topic–focus articulation (under different names) is captured in several annotated corpora, e.g. the Potsdam Commentary Corpus (Stede and Neumann, 2014); the ANNIS Database (Annotation of Information Structure; Dipper, Götze and Skopeteas, 2007); the Muli corpus (Baumann et al., 2004); the Switchboard Corpus (Calhoun et al., 2005); the DannPASS (Danish Phonetically Annotated Spontaneous Speech; Paggio, 2006) or the Penn TreeBank (Bohnet, Burga and Wanner, 2013). Every corpus representation of topic–focus articulation is unique and often differs significantly from other ones. Both the technical approaches and the annotated features are different.

There are two Prague treebanks with annotated topic–focus articulation: the Prague Dependency Treebank (the latest version in Bejček et al., 2013, PDT) containing 50 thousand of annotated sentences in Czech (it is a corpus of written newspaper texts) and the Prague Czech-English Dependency Treebank (Hajič et al., 2012, PCEDT) containing original English texts from the Wall Street Journal (from the corpus Penn TreeBank) and their parallel translations to Czech (for each language, the corpus contains almost 50 thousand sentences).[12]

In this section, we introduce practical issues connected with annotating topic–focus articulation in the Prague Czech-English Dependency Treebank using the theory of the Functional Generative Description described above (see also Mírovský, Rysová, Rysová and Hajičová, 2013).

The Prague Czech-English Dependency Treebank is annotated mostly manually on several layers – on the layer of words, morphological layer, analytical layer (i.e. surface syntax) and the so-called tectogrammatical layer (i.e. deep syntactico-semantic layer). The tectogrammatical layer contains e.g. annotation of topic–focus articulation and annotation of sentence participants and free modifications like Actor, Patient, Addressee, Locative, Manner or Temporal modification etc. Moreover, the corpus also contains annotation of coreference and anaphora. The annotations are carried out on the dependency trees.[13]

[12] The annotation of topic–focus articulation is carried out on 1/10 of the PCEDT (i.e. on 5 thousand Czech and 5 thousand English parallel sentences) and has not been published yet.

[13] For detailed description of Prague treebanks and syntactico-semantic annotation of English texts see Cinková et. al, 2006.

8.1 A Layer of Deep Syntax – Tectogrammatical Layer

As we have mentioned above, the annotation of topic–focus articulation proceeds on the tectogrammatical sentence layer, i.e. on the layer of deep syntax, and it is done on the dependency trees (which is unique within the other corpus annotations of topic–focus articulation).

The tectogrammatical trees also contain reconstructed sentence items, i.e. items (nodes) that are deleted in the surface sentence structure. Therefore, it is also possible to annotate nodes present only in the deep sentence structure (e.g. elided subjects, obligatory participants from the valency frame of verbs or actual ellipses) but clearly participating in topic–focus articulation – elided sentence participants are usually contextually bound, see Example (33), but not necessarily in all cases, see Example (34).

(33) Surface structure: *Some mobile devices are able to carry classified information.*
Deep structure: *Some mobile devices are able to carry classified information* [*from one person*]$_{t}$ [*to another person*]$_{t}$.

(34) Surface structure: *primary and secondary schools*
Deep structure: *primary* [*schools*]$_{f}$ *and secondary schools*$_{t}$

When "reconstructing" the deep layer, it is not necessary to know which particular lexemes were omitted in the surface structure (here e.g. *from one person* or *from one side* etc.). The annotation is carried out on the supplemented types of sentence participants (i.e. e.g. "verb modification *where from*", "verb modification *where to*" etc.], see Example (35):

(35) Surface structure: *Some mobile devices are able to carry classified information.*

Deep structure: *Some mobile devices are able to carry classified information* [directional verb modification: where from]$_{t}$ [directional verb modification: where to]$_{t}$.

However, we may find also opposite cases when an item present in the surface structure does not appear separately in the deep structure – cf. e.g. the so-called anticipatory *it*[14] and existential *there* as in the following examples.

(36) Surface structure: *It is difficult to make predictions especially about the future.*
Deep structure: *Ø is difficult to make predictions especially about the future*

(37) Surface structure: *There are many ways to change your life.*
Deep structure: *Ø are many ways to change your life.*

Expressions *it* and *there* are located in the place that is grammatically common for sentence subject (in its surface realization). However, they do not have any syntactico-semantic role (e.g. they are not an Actor etc.) and therefore, they are not a part of a deep (semantic) representation of a sentence (see also Section 9.10 *Phenomena Appearing on the Scene*). As seen, the deep structure of a sentence thus does not have to be "richer" than the surface structure in all cases.

The information about topic–focus articulation in the tectogrammatical layer is based on the following two characteristics: (i) value of contextual boundness and (ii) value of communicative dynamism. The sentence division into topic and focus is not explicitly annotated but is well deducible from the two annotated phenomena.

8.2 Annotation of Contextual Boundness in the PCEDT

In the first step of annotation, we evaluate each node (relevant for topic–focus articulation)[15] of a tree in terms of contextual boundness. In this respect, we distinguish sentence elements that are (i) contrastive contextually bound (marked as c and highlighted in green in our figures) (ii) non-contrastive contextually bound (marked as t and highlighted in white) and (iii) contextually non-bound (marked as f and highlighted in yellow), see Figure 1 in Section *8.4 Dependency Trees*.

[14] Different situation is with the so-called false subjects like *It is freezing*. See Section *9.11 Semantic Type of a Sentence Item (Node)*.
[15] The tfa value is not assigned e.g. to the technical root of the sentence or to the nodes representing coordinating conjunctions.

8.3 Annotation of Communicative Dynamism in the PCEDT

The second step of annotation of topic–focus articulation in the PCEDT concerns communicative dynamism, i.e. ordering of nodes in the tree with respect to their communicative dynamism that grows from the left to the right. The deep ordering of the sentence elements in the PCEDT may be thus different from the surface word order of the given sentence.

The main rule for the communicative dynamism annotation in the deep word order is that in the dependency tree, the contextually bound nodes are placed to the left of the governing node, whereas the contextually non-bound nodes are placed to the right. The node that is placed in the rightmost position is the focus proper (i.e. the most dynamic part of the given sentence carrying the intonation centre). The annotation of communicative dynamism in the PCEDT is demonstrated in Figure 1.

8.4 Dependency Trees

The annotation of topic–focus articulation is carried out on dependency trees, see Figure 1.

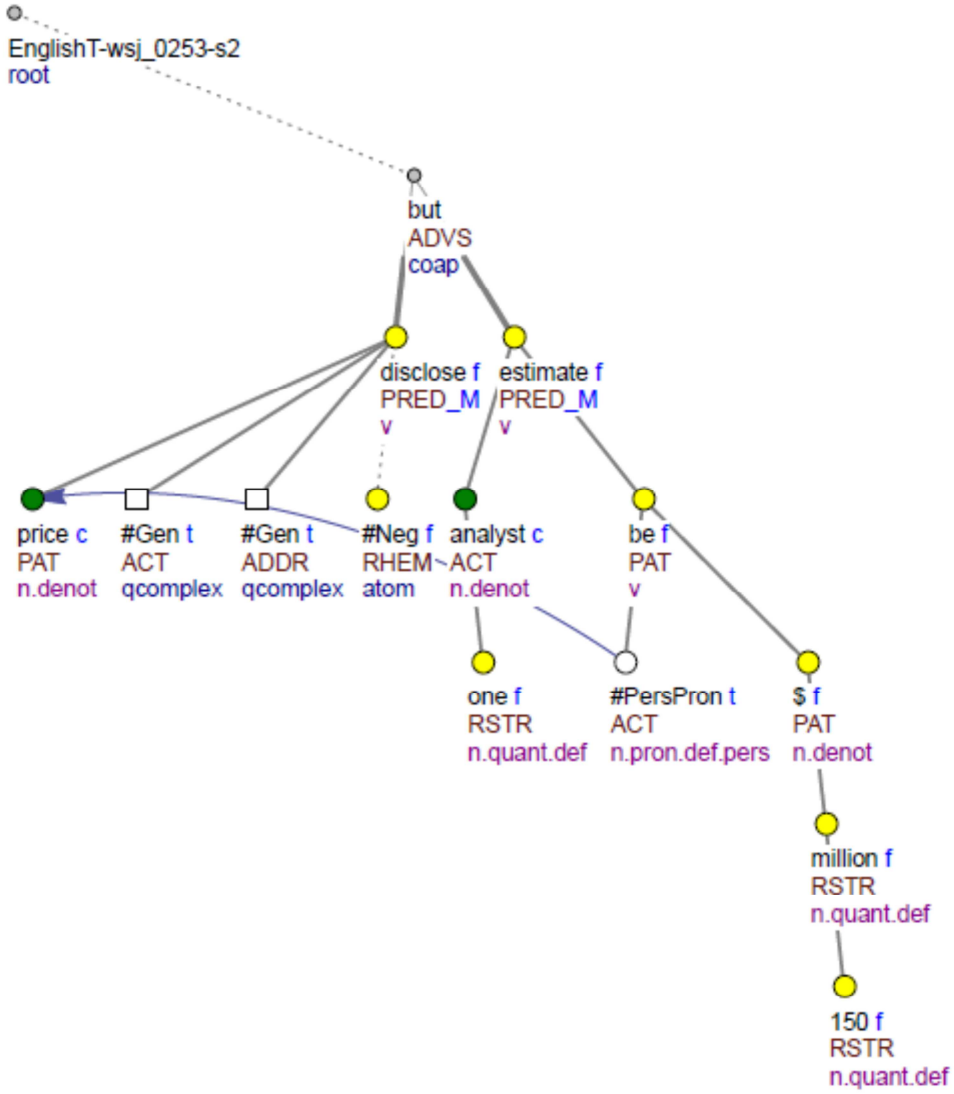


Figure 1: Dependency tree from the Prague Czech-English Dependency Treebank with the annotation of topic–focus articulation for the Example (38) ***The price wasn't disclosed but one analyst estimated that it was $150 million.***

Dependency tree in Figure 1 captures the following sentence (the preceding context is given in the square brackets; the text is from the *Wall Street Journal*), see Example (38).

(38) [*Armstrong World Industries Inc. agreed in principle to sell its carpet operations to Shaw Industries Inc.*] *The price wasn't DISCLOSED but one analyst estimated that it was $150 million.*

The individual nodes of the dependency tree are labeled with wheels (for the items that are explicitly expressed in the surface structure of the sentence) and squares (for the items that are omitted from the surface but are present in the deep structure). The lines between the individual nodes are the so-called edges. Each node is assigned a lemma and the syntactico-semantic function called functor (e.g. PRED: Predicate of the main clause, ACT: Actor; PAT: Patient; ADDR: Addressee; RHEM: Rhematizer; RSTR: Restrictive Attribute; ADVS: root of the coordination expressing Adversative Relation). Furthermore, the figure captures some other information from the morphological layer and an arrow expressing coreference relation (*price* ← *it*).

Topic–focus articulation, precisely contextual boundness (see Section 4.1 *Context and Contextual Boundness*) is assigned to each node (relevant for topic–focus articulation) by one of the indices "c", "t" and "f". The index "c" is given to contrastive contextually bound nodes (marked in green in the figure), the index "t" to non-contrastive contextually bound nodes (marked in white in the figure) and the index "f" to contextually non-bound nodes (marked in yellow in the figure).

Contextually bound nodes (both contrastive and non-contrastive) directly dependent on the governing verb (having the functor PRED) and all nodes dependent on them form the topic of the sentence. Contextually non-bound nodes directly dependent on the governing verb and all nodes dependent on them constitute the focus of sentences. The governing verb is either a part of the topic (if it is contextually bound) or focus (if it is contextually non-bound). For more details see Section 5 *Detection of Topic and Focus*.

The tree in Figure 1 contains two utterances (it is a coordination of two structures in adversative relation). The first is *The price wasn't disclosed* and the second is *One analyst estimated that it was $150 million*. Each of them has its own topic–focus articulation.

The second utterance contains two clauses: the first one is governing (*One analyst estimated*) and the second is dependent (*that it was $150 million*). Dependent clauses are treated as a part of the governing clause (cf. *One analyst estimated something*).

In our case, we can see from the tree that the topic of the first utterance is *The price*, the focus part is *wasn't disclosed*. The topic of the second utterance is *One analyst*, focus is *estimated that it was $150 million*. The boundary between topic and focus is not explicitly marked during the annotation (but it is well deducible from the tree).

Another phenomenon that is annotated within topic–focus articulation in the dependency trees is communicative dynamism. This is captured through the so-called deep word order, i.e. by the order of the individual nodes, see Section 4.2 *Communicative*

dynamism. It is obvious from the example above that contextually bound nodes directly depending on the governing verb are located to the left of the verb and contextually non-bound to the right. This suggests that contextually bound nodes carry a lower degree of communicative dynamism (they are less informative or less communicatively important) than contextually non-bound nodes. Similarly, contextually non-bound attributes are always to the right of their governing nouns because they have a higher degree of communicative dynamism. The dependency tree is projective.

Example (39) demonstrates the division into topic and focus and contextual boundness of the individual nodes without the tree structure (i.e. the global topic–focus articulation). The example (unlike the tree structure) does not capture the communicative dynamism of the individual nodes. In the square brackets, there are the participants that are not expressed in the surface structure (the verb *to disclose* has 3 obligatory inner participants in its valency frame: Actor – *someone*, Patient – *something* and Addressee – *to somebody*).

(39) {*The price*}$_{.c}$ [Actor: *someone*]$_{.t}$ [Addressee: *to somebody*]$_{.t\ topic\_1}$ {***was DISCLOSED***}$_{.f}$ ***not***$_{.f\ focus\_1}$ *but one*$_{.f}$ *analyst*$_{.c\ topic\_2}$ ***estimated***$_{.f}$ ***that it***$_{.t}$ ***was***$_{.f}$ ***$***$_{.f}$ ***150***$_{.f}$ ***million***$_{.f\ focus\_2}$.

Example (40) demonstrates the local topic–focus articulation of the dependent clause without the tree structure.

(40) *The price wasn't DISCLOSED but one analyst estimated that it*$_{.t\ topic\_local}$ ***was***$_{.f}$ ***$***$_{.f}$ ***150***$_{.f}$ ***million***$_{.f\ focus\_local}$.

8.5 Which Sentence Items (Nodes) are Annotated in terms of Contextual Boundness and Communicative Dynamism

Topic–focus articulation concerns all **sentence elements** (including **attributes**) as well as structures integrated into clause loosely, e.g. **salutations**, **parenthesis**, sentence adverbials expressing **authorial comment** or (evaluating or emotional) **attitude** to the content like *unfortunately*, *really*, *hopefully*, **modal** characteristics like *probably*, *sure*, *perhaps*, but also the so-called **rhematizers** like *only*, *particularly*, *for example*, *also*, *not* or expressions connected to the **previous context** like *however*, *therefore*, *thus*, *hence* (in examples like: *He said I was unpleasant. However, it is not true.*), **interrogative** and **relative** expressions like *who*, *that*, *which*, **interjections** like *oops*, **answering particles** like *yes, no* (*Yes, that is true*). In terms of topic–focus articulation, we annotate also **measurement units** expressed by words, abbreviations or symbols (e.g. *million, dollar, tonne, kg, %, $, EUR*). Also each member of **apposition** (e.g. *Shakespeare, a writer*) as well as **coordination** (e.g. *plays and poems*) has its own value of contextual boundness.

Topic–focus articulation is **not annotated separately** when the sentence items do not have their own node in the deep structure, i.e. prepositions (e.g. *from*, *to*), articles (*a*, *an*, *the*), anticipatory *it* (*It is nice to meet you*)[16], existential *there* (*There is a meeting today*), inherent reflexive pronouns, e.g. *-self* in constructions like *They pride themselves*, particles in phrasal verbs like *make up*, auxiliary verbs as parts of an analytical predicate (*I have been dancing*), modal verbs (*can*, *could*, *may*, *might*, *shall*, *should*, *must*, *ought to*, *will*) in combination with infinitives (*I can dance*).[17]

We do not annotate separately connective expressions of coordinations and appositions like *either – or*, *neither – nor*, *versus* etc. (cf. *I wanted to go home and sleep but I could not*), expressions modifying these items (*It is not big but even huge*) and expressions representing punctuation marks like *:* (colon); *(* (parentheses); */* (slash); – (dash) or mathematical signs or intervals (like *3 plus 4*, *5 – 2*, *four times five*, *starting Monday*, *from January to June* etc.). Foreign expressions are treated as a whole, i.e. topic–focus articulation

[16] However, we do annotate topic–focus articulation in cases when *it* is in the function of the so-called false subject (*It is raining*). We consider it to be contextually bound is such cases.

[17] However, quasi-modal verbs (e.g. *to have to*, *to be able to*, *to be bound to*) as well as phrasal verbs (e.g. *to start*, *to stop*) and copular verbs (*He is a student*) are considered to be separate items (nodes) in the deep structure of the sentence and therefore, topic–focus articulation is annotated **separately** in these cases (*I started to play chess*).

is not annotated for the individual items separately (<*Česká investiční společnost*> *went bankrupt*).

However, it does not mean that we understand the listed items (e.g. auxiliary verbs like *have been*, reflexive pronouns like *themselves* etc.) as excluded from the topic–focus articulation. From the perspective of the deep structure, they are always a part of another node (e.g. the individual parts of *have been dancing* on the surface layer act together as *dance* on the deep layer; surface form *can swim* is represented by a single node *swim* etc.). The mentioned items are thus treated in topic–focus articulation as one complex node together with other sentence item to which they are grammatically related.

Of course, we may argue that in some cases, also the mentioned items (that are a part of complex node on the deep layer) may be the only contextually non-bound member in the sentences (i.e. the focus proper), cf. *You cannot sing? – I CAN sing!* or *Are you going to Prague? – I am going FROM Prague.*). Similarly, the focus proper may be even one grammatical category, e.g. verb tense (*Has he got money? – He HAD money. But he divorced.*).

Thus it may happen that in a single node on the deep layer, there are connected two items with different value of contextual boundness. In this case, we apply the rule that **contextual non-boundness takes precedence over contextual boundness**, i.e. if at least one subpart of a particular node (e.g. a modal verb that is connected with the lexical verb, but also a single characteristics of a word like grammatical category) is contextually non-bound, the whole node is considered to be contextually non-bound (*I am going* {*FROM Prague*}**.f**; *I* {*CAN sing*}**.f**; *He HAD***.f** *money*).

Contextually bound expressions (nodes) are labeled as “t” in the examples and contextually non-bound as “f”. In the braces, there are sentence items that constitute a single node with a complex value in the deep structure. The focus part of the sentences is marked in bold, the topic remains unmarked. The topic part of the sentences does not always have to coincide with the contextually bound nodes and the focus part with the contextually non-bound nodes (for details on contextual boundness and topic and focus, see Sections 5 *Detection of Topic and Focus* and 4.2 *Communicative Dynamism*). Communicative dynamism is not marked in these examples due to better clarity. In the square brackets, there is the preceding context.

(41) [*What happened to you?*] *I*.t *unfortunately*.t ***lost*.f *all*.f *phone*.f *NUMBERS*.f.**

(42) [*What did he say?*] *He*.t *said*.t ***I*.t *was*.f *UNPLEASANT*.f.** *However*.t, *it*.t ***is*.f *not*.f *TRUE*.f.**

(43) [*Who do you like?*] *I*.t *like*.t ***IMPRESSIONISTS*.f, *especially*.f *Paul*.f *CÉZANNE*.f.**

(44) [*Did you read something?*] *I*.t *read*.t ***many*.f *BOOKS*.f *({e.g.}*.f *Robinson*.f *CRUSOE*.f *or Gulliver's*.f *TRAVELS*.f*).***

(45) [*Where have you been?*] *I*.t {*have been*}.t **{*to LONDON*}.f.**

(46) [*Who did you meet yesterday?*] *Yesterday*.t *I*.t *met*.t **{*the woman*}.f *who*.t *CALLED*.f *you*.t.**

(47) [*What is this?*] ***Is*.f** *it*.t ***your*.t *SUITCASE*.f?**

(48) [*What is the weather?*] *It*.t **{*is RAINING*}.f** – *I*.t ***THINK*.f**

(49) *It is nice to MEET you.*
Deep structure: ***is*.f *nice*.f {*It to MEET*}.f *you*.t.**

(50) [*What is the programme today?*] {*There is*}.t **{*a MEETING*}.f** *today*.t.

(51) [*What do they do?*] *They*.t **{*PRIDE themselves*}.f.**

(52) [*What have you been doing?*] *I*.t **{*have been DANCING*}.f.**

(53) [*What can you do?*] *I*.t **{*can SWIM*}.f.**

(54) [*What do you have to?*] *I*.t {*have to*}.t ***go*.f *HOME*.f.**

(55) [*What did you do?*] *I*.t ***stopped*.f *SMOKING*.f.**

(56) [*What does he do?*] *He*.t ***is*.f {*a STUDENT*}.f.**

(57) [*What do you know about Česká investiční společnost?*] <*Česká investiční společnost*>$_{.t}$[18] ***{has been SOLD}*$_{.f}$**

(58) [*What do you know about this book?*] *This book contains all the known plays and poems of William SHAKESPEARE, a writer and an actor.*

Deep structure: *This*$_{.t}$ *book*$_{.t}$ ***contains*$_{.f}$ *all*$_{.f}$ *known*$_{.f}$ *plays*$_{.f}$ *and poems*$_{.f}$ *William*$_{.f}$ {*of Shakespeare*}$_{.f}$, {*a writer*}$_{.f}$ *and* {*an actor*}$_{.f}$.**

(59) [*Few details offered on quarantines over Ebola*

NEW YORK –] *A nurse who treated Ebola patients is the test of quarantine POLICIES.*

Deep structure: {*A nurse*}$_{.t}$ *who*$_{.t}$ *treated*$_{.t}$ *Ebola*$_{.t}$ *patients*$_{.t}$ ***is*$_{.f}$ {*the test*}$_{.f}$ *quarantine*$_{.f}$ {*of POLICIES*}$_{.f}$.**

[18] The foreign expression *Česká investiční společnost* is divided into the individual nodes in the dependency tree, but the topic–focus articulation is not annotated for each node separately but for the expression as a whole.

8.6 How Large Units of Text Are Annotated for Topic–Focus Articulation

In a simple sentence, we can find one set of topic and focus, see Examples (60) and (61). In compound sentences, topic and focus are recognized in every sentence (clause) separately, see Examples (62), (63) and (64). Finally, in complex sentences, only one set of topic and focus is present, see Examples (65) and (66).

Topic–focus articulation is annotated also in all sentence equivalents (like headings, captions etc.), see Examples (67) and (68). If there are more sentence equivalents with the function of individual statements in one unit (bounded at the beginning of a sentence with a capital letter and full stop at the end), we consider them to be separate independent clauses and we annotate the topic–focus articulation separately for each of them (like in a compound sentence), see Examples (71), (72), (74).

We proceed similarly if there are any types of independent clauses functioning as individual statements within one sentence unit (topic–focus articulation is thus annotated for each of the independent clause separately), see Examples (69), (70), (73).

Gerunds and participial constructions are considered to be dependent clauses (they have a function of subordinate clauses, often conditional), and therefore, we do not annotate topic–focus articulation for them separately (we consider them, in most cases, to be a part of a governing clause, similarly as subordinate clauses), see Examples (81) and (82).

Parentheses are divided into two groups: syntactically integrated into the sentence structure and syntactically unintegrated into the sentence structure. Parentheses that are syntactically integrated are perceived as a part of the sentence structure to which they belong and topic–focus articulation is not annotated for them separately, see Examples (75) and (76). On the contrary, parentheses that are syntactically unintegrated are viewed as independent sentence units and in these cases, topic–focus articulation is thus annotated separately, see Examples (77), (78), (79) and (80).

To sum up, topic–focus articulation is annotated separately in **each independent clause (i.e. in verbal, nominal, interjectional clauses as well as parentheses).**

In the following examples, all members of the focus/focuses are marked in bold (regardless contextual (non)boundness of their individual nodes), all members of topic/topics are not marked (again regardless context (non)boundness of their individual nodes). Contextual boundness of the individual nodes is here not indicated because of better clarity. However, it may happen that some of the focus items (such as items indirectly dependent on the governing verb of the main clause) are contextually bound or that some of the topic items

(like elements indirectly dependent on the governing verb of the main clause) are contextually non-bound.

(60) *The restaurant*$_{\textbf{topic}}$ ***is located on the Destiny, a boat owned by New York Cruise LINES***$_{\textbf{focus}}$

(61) *I* $_{\textbf{topic\_1}}$ ***DON'T think***$_{\textbf{focus\_1}}$ *so*$_{\textbf{topic\_1}}$.

(62) *Summer*$_{\textbf{topic\_1}}$ ***may be FADING***$_{\textbf{focus\_1}}$, *but we*$_{\textbf{topic\_2}}$ ***can still have some seasonal FUN***$_{\textbf{focus\_2}}$

(63) *Customers*$_{\textbf{topic\_1}}$ ***order at the counter on the upper DECK***$_{\textbf{focus\_1}}$, and *food*$_{\textbf{topic\_2}}$ ***is brought to tables in an open-air dining AREA***$_{\textbf{focus\_2}}$

(64) *Grandmother*$_{\textbf{topic\_1}}$ ***is SLEEPING***$_{\textbf{focus\_1}}$, *Grandfather* [*is sleeping* = contextual ellipsis of predicate] $_{\textbf{topic\_2}}$ ***TOO***$_{\textbf{focus\_2}}$

(65) *The Wall Street Journal*$_{\textbf{topic}}$ ***has selected 100 legacies from World War I that continue to shape our LIVES today***$_{\textbf{focus}}$

(66) *Two separate attacks on men from India's northeast*$_{\textbf{topic}}$ ***cannot be prosecuted as racist crimes because India doesn't have a law to protect northeasterners against racist VIOLENCE***$_{\textbf{focus}}$

(67) ***Page 45***$_{\textbf{focus}}$

(68) ***Total Hurt by Oil PRICE***$_{\textbf{focus}}$

(69) ***Yes***$_{\textbf{focus\_1}}$, *you* $_{\textbf{topic\_2}}$ ***are RIGHT***$_{\textbf{focus\_2}}$

(70) ***OOPS***$_{\textbf{focus\_1}}$, *something* $_{\textbf{topic\_2}}$ ***is WRONG***$_{\textbf{focus\_2}}$*!*

(71) ***John SMITH***$_{\textbf{focus\_1}}$, ***New YORK***$_{\textbf{focus\_2}}$

(72) ***POOH***$_{\text{focus\_1}}$, ***NO***$_{\text{focus\_2}}$*!*

(73) ***Wedding SCANDAL***$_{\text{focus\_1}}$***: Whose Bride Lost Her MEMORY***$_{\text{focus\_2}}$*?*

(74) ***OH***$_{\text{focus\_1}}$, ***my TEETH***$_{\text{focus\_2}}$*!*

(75) *It*$_{\text{topic}}$ ***was a clever (and nice) GESTURE of you***$_{\text{focus}}$.

(76) *I*$_{\text{topic}}$***'d like to go home now (if you don't MIND)***$_{\text{focus}}$.

(77) *After their start in 2010, the Giants*$_{\text{topic\_1}}$ ***won their final six GAMES***$_{\text{focus\_1}}$ *(the following year, they*$_{\text{topic\_2}}$ ***lost the championship to the COLTS***$_{\text{focus\_2}}$*).*

(78) ***Jane MERRICK***$_{\text{focus\_1}}$ ***(Charles University in PRAGUE***$_{\text{focus\_2}}$***)***

(79) *Agnes, Barbara and Carol*$_{\text{topic\_1}}$ *– they all*$_{\text{topic\_2}}$ ***are MODELS***$_{\text{focus\_2}}$ **–** ***are new faces of our COSMETICS***$_{\text{focus\_1}}$.

(80) *Empress Michiko*$_{\text{topic\_1}}$ ***celebrated her 80th BIRTHDAY***$_{\text{focus\_1}}$ *(we*$_{\text{topic\_2}}$ ***CONGRATULATE***$_{\text{focus\_2}}$*).*

(81) *Regarding your complaint, we*$_{\text{topic}}$ ***will do our best to solve the PROBLEM***$_{\text{focus}}$.
(= *Since we regard your complaint...*)

(82) *Seen from above, we*$_{\text{topic}}$ ***would look like a chain of little worker ANTS***$_{\text{focus}}$.
(= *If we look from above...*)

8.6.1 Local Topic–Focus Articulation

As mentioned above, dependent clauses themselves are as a whole part of the global topic–focus articulation. However, deeper in sentence structure (in the dependent clauses), we can look for the **local topic–focus articulation** – see the following examples:

(83) *The Wall Street Journal has selected 100 legacies from World War I that*$_{\textbf{topic\_local}}$ ***continue to shape our lives today***$_{\textbf{focus\_local}}$

(84) *Two separate attacks on men from India's northeast cannot be prosecuted as racist crimes because India*$_{\textbf{topic\_local}}$ ***doesn't have a law to protect northeasterners against racist violence***$_{\textbf{focus\_local}}$

(85) *I'd like to go home now (if you*$_{\textbf{topic\_local}}$ ***don't mind***$_{\textbf{focus\_local}}$*).*

(86) *Regarding*$_{\textbf{topic\_local}}$ ***your complaint***$_{\textbf{focus\_local}}$*, we will do our best to solve the problem.*

(87) *Seen*$_{\textbf{topic\_local}}$ ***from above***$_{\textbf{focus\_local}}$*, we would look like a chain of little worker ants.*

For example, the item *that* from Example (83) is contextually bound and is part of the focus of the governing sentence (in terms of the global topic–focus articulation). From the perspective of the local topic–focus articulation, it is also contextually bound and is a part of the topic of the dependent sentence (dependent relative clause). Generally, **each node has exactly one and invariable value of contextual boundness** and, at the same time, it may be part of topic or focus regarding the perspective of local vs. global topic–focus articulation.

8.7 Annotation of the First Sentence in Text

The first sentence of a text exhibits (to a certain extent) several specifics. It opens a new communication unit and new communication theme. It may be thus supposed that all of its items could be contextually non-bound.

However, this is not true in all cases. For example, a text may be contextually connected with the set of other texts, it may inform about something that readers or hearers have activated in their consciousness (i.e. about something they are aware of, that is close to them etc.).

Typical examples of texts where the first sentences contain items that are contextually bound are newspaper articles bringing (after some time) some new information about one theme, see the following examples of newspaper headings concerning the Ukrainian-Russian crisis:

(88) 16. 7. 2014: *Obama Announces Sanctions*$_{.f}$ *On RUSSIA*$_{.f}$

(89) 23. 7. 2014: *EU imposes new sanctions*$_{.f}$ *on Russian*$_{.f}$ *OFFICIALS*

(90) 2. 8. 2014: *Beyond SANCTIONS*$_{.f}$*: What's the West's Strategy on RUSSIA*$_{.f}$*?*

(91) 11. 9. 2014: *The West's Sanctions*$_{.f}$ *on RUSSIA*$_{.f}$*: Would it Lead to Cold War 2.0?*

(92) 15. 9. 2014: *HOW far do EU-US sanctions*$_{.t}$ *on Russia*$_{.t}$ *go?*
(= *I have a question concerning the sanctions on Russia that we have already written about: How far do they go?*)

(93) 26. 9. 2014: *Will Sanctions*$_{.t}$ *Push Russia*$_{.f}$ *Into RECESSION?*
(= *We all know that there are Sanctions on Russia and the question is whether they cause a Russian recession.*)

All the headings of articles open new texts and, at the same time, they all contain the items *sanctions* and *Russia* (*Russian*). When the articles were written, the theme of "sanctions on Russia" was very actual and it appeared in media very often. The readers (or viewers and hearers) as well as the authors of texts considered this theme to be known. Therefore, the authors used these expressions as contextually bound. However, we cannot say that each occurrence of the words *sanctions* and *Russia* was (at that time) contextually bound – it always depends on the particular use in the concrete text and concrete communication situation. With the increasing time when the theme is not discussed so often any more, also its activation in the consciousness of potential readers or hearers (salience) obviously descends.

The first sentence of a text may thus contain not only the items that are contextually non-bound, but also contextually bound.

Many written texts (especially newspaper articles) begin with a heading. The heading (as the first sentence of the text) is unique in many ways – it captures the theme of the whole text, see Example (94) and sometimes, it contains the whole main idea (the headings of newspaper articles should attract the reader to read the whole article, but at the same time, the authors take into account that very often he or she reads only the heading; therefore, the heading often includes the main message of the article), see Example (95).

(94) ***Kung PAO Chicken Recipe*****focus**

(95) ***Hungary Drops Internet Tax PLAN*****focus**

The heading (in the case of newspaper texts) may be accompanied by a subheading or it is followed directly by the first sentence of the first paragraph. The subheading or the first sentence of the first paragraph very often just repeat (or only slightly expand) the information that the reader has learned from the heading, see Examples (96) and (97).

(96) ***OBAMA*** **focus_1*****: 'My Credit Card Was REJECTED'*****focus_2**
*President Barack Obama said*_**topic_2** ***his credit card was DECLINED***_**focus_2**

(97) ***McDonald's Japan Bans SMOKING in Its Restaurants***_**focus_1**
*The burger chain has prohibited smoking*_**topic_2** ***in ALL of its outlets in the country***_**focus_2**

The subheadings of Examples (96) and (97) almost literally repeat the information that the reader has already got from the heading: the president's credit card was rejected; McDonald's in Japan prohibited smoking in its restaurants. The subheadings give basically the same message as the headings – often they use contextual synonyms: *reject – decline*, *ban – prohibit*, *restaurant – outlet*, *Japan – country*, but also literal repetitions: *credit card – credit card*, *smoking – smoking*, or sometimes explicit expressions of what was only implicit in the heading: *Obama – President Barack Obama*, *: – say*. In both cases, the author even uses the same communication perspective in the heading and subheading. The core of communication (focus proper) is the same. The subheading thus does not tell the reader anything new. Yet the sentences of the subheadings are not marked as sentence topic as a whole (every sentence must have a focus), see Example (98), and all nodes as contextually bound, but they certainly have their own focus (even though all the sentence items have been already used in the preceding context). However, the focus extent is not entirely clear.

Therefore, the focus does not have to always bring completely new or surprising information that the reader did not know before. It depends on whether **certain information is presented as "non-derivable" by the author**.

(98) ***McDonald's Japan Bans SMOKING in Its Restaurants***$_{\textbf{focus\_1}}$

** The burger chain has prohibited smoking in all of its outlets in the country*$_{\textbf{topic\_2}}$

We assume that repetition of what was said is a part of the author's communication strategy. He or she probably wants to attract the reader's attention and to achieve that the reader remembers the main information. The repeated information thus may have its function in the text (at least from the author's perspective) and therefore, it must have its focus.

The question is how much functional the repetition of nearly identical information in texts is for the reader. However, newspaper articles use a similar style of writing rather frequently.

In case that the repetition in the subheading is not functional and "specific" as in examples above, we take it in the same way as the repetition within a text, see Example (99) – the all-focus sentence in the heading (it is the so-called topicless sentence) becomes topic in the subheading.

(99) ***Restaurant Revival in PRAGUE***$_{\textbf{focus\_1}}$

Prague's restaurant revival$_{\textbf{topic\_2}}$ ***is raising the profile of the local CUISINE***$_{\textbf{focus\_2}}$

9 Recognizing Boundary between Contextually Bound and Non-Bound Nodes in English

The central question regarding the annotation of topic–focus articulation is how the annotator (i.e. the listener or reader) can recognize where the speaker or writer put the boundary between contextually bound and non-bound nodes (at the level of items directly dependent on the governing verb). Each language has its special means to signal the contextual boundness of the individual sentence items. In this section, we present the means for English as an analytic language.

Some sentence items are “predestined” to act rather as contextually bound or non-bound mainly due to their lexical or grammatical properties. In other words, communication need to express some sentence item as bound or non-bound can be fulfilled by using certain formal means which are appropriate for the given function. Here is a commented overview of these “formal means”. It can serve as a guide for the annotation.

9.1 Surface Word Order

Surface word order in English is relatively fixed by the grammatical factor. To a limited extent, however, it may also be an indicator of contextual boundness. In the English unmarked indicative sentence, some items in the initial position are usually contextually bound (and at the same time, they usually have a lower degree of communicative dynamism), while the items in the final position tend to be contextually non-bound and more dynamic (however, not absolutely in all cases). The surface word order can indicate, for example, the contextual boundness of some causal verb modifications, see Examples (100) and (101).

(100) *Because of Ebola disease fears*$_{\textbf{topic part}}$, *North Korea restricted TOURISM.*

(101) *North Korea restricted tourism because of EBOLA disease fears*$_{\textbf{focus part}}$.

In the first case, the causal verb modification (*because of Ebola disease fears*) is at the beginning of a sentence and it is contextually bound. The main aim of the author is to say that North Korea restricted tourism. In the latter case, it stands at the end of the sentence and it is contextually non-bound (in this case, it is the focus proper). The main objective is to say the reason for the restriction of tourism in North Korea (*because of Ebola disease fears*).

The similar cases are also the following examples.

(102) [*One of the biggest full moons of the year – a so-called "supermoon" – will light up the night sky on Saturday (July 12). Supermoons occur when the moon reaches the closest point to Earth, called "perigee," during its month-long orbit.*] *On July 12*$_{\textbf{topic part}}$, *the moon will reach its full phase at 7:25 a.m. EDT (11:25 GMT)* $_{\textbf{focus part}}$.

The temporal modification at the beginning of the last sentence in the text (*on July 12*) represents a "temporal scene or perspective" for further action – it stands at the beginning of a sentence and it is contextually bound (in the previous text, it has also been mentioned – as the focus proper of the first sentence in the text). The last sentence of the text thus outlines what will happen on July 12. On the contrary, the temporal modification at the end of the sentence (*at 7:25 a.m. EDT (11:25 GMT)*) expresses the focus proper, i.e. the most important

information of given statement and is therefore contextually non-bound. The aim of the last sentence is to say when exactly during July 12 the moon comes into its climactic phase.

In Example (103), the verb modification of purpose (*to reach this*) clearly refers to the previous context (it is contextually bound) and at the same time, it stands at the beginning of a sentence. The author's intention was primarily to say what we will do to achieve the mentioned goal, not for what purpose we will do it.

(103) [*The final aim of INtheMC is to get more participation in both internationalization programmes and student mobility in Europe.*] *To reach this*$_{\textbf{topic part}}$ *we will target all stakeholders: teachers, management and STUDENTS.*

However, the own position of certain sentence item in the surface word order cannot be the main and the only guide for determining the contextual boundness. In the following example, the sentence participant *to reach this figure* is contextually bound but in the surface word order, it stands at the very end of the sentence.

(104) [*It took Facebook just 3 years and 8 months to reach 50 million users.*] *In comparison, it took 38 years for radio's audience to reach this figure*$_{\textbf{topic part}}$.

In this case, the intent of the author was to say how long it took before radio reached the same number of users as Facebook (38 years). The aim of the statement was not to say what took radio 38 years – the expression "to reach this figure" is not contextually non-bound (even though it stands in the rightmost position in the surface word order – in the deep word order, it stands on the left).

The position of a sentence item in the surface word order in English sentences is therefore only one of the criteria which can be taken into account in determining its contextual boundness. However, it is not a crucial criterion.

The position of subject in English is very stable – it usually stands at the very beginning of the sentence in the surface word order (even if it is focus proper):

(105) *A CAR*$_{\textbf{focus part}}$ *stopped in the street.*

In this case, the grammatical factor affects the surface word order in English very strongly – it is stronger than the factor of contextual boundness and communicative dynamism. A clear contextual non-boundness of the subject *a car* (in this case, the subject is the focus proper of the given sentence) and very high (in the given sentence even the highest) degree of communicative dynamism together "attract" the subject to stand at the end of the sentence (i.e. to the right). On the other hand, grammatical factor (following the fixed order of SVO for the individual items of the English sentence: subject – governing verb – object) "pulls" the subject to the initial position – at the very beginning of the sentence (i.e. to the left). In this case, the grammatical factor "wins" in English: the contextually non-bound subject with the highest degree of communicative dynamism stands to the left in the surface word order.

However, there are languages (e.g. inflected Czech) where the contextual boundness along with communicative dynamism prevail over the grammatical factor in the same situation – i.e. the subject in the surface word order stands in the rightmost position, at the very end of the sentence, see the following example (it is a translation of the English Example (105) to Czech).

(106) *Na silnici zastavilo* *AUTO*$_{\textbf{focus part}}$.

However, even in English, there is a clear tendency (at least in some cases) of the contextually non-bound item with a high degree of communicative dynamism to stand to the right of the governing verb in the surface word order. For this purpose, English has several means how to move the communicatively most important item more to the right – e.g. by using passive, cleft and pseudo-cleft sentences or existential constructions *there is, there exists* etc. (see the following examples).

(107) *It was* *ME*$_{\textbf{focus part}}$ *who left the relationship because I realized how destructive it was.*

(108) *What I gave him were just little* *fragments*$_{\textbf{focus part}}$ *of MEMORIES.*

(109) *We have no rational evidence that there exits another* *WORLD*$_{\textbf{focus part}}$.

(110) *All you need is* *LOVE*$_{\textbf{focus part}}$.

9.1.1 Structures with Direct Speech

English has also other structures that violate the strong tendency to SVO word order – structures with direct and indirect speech which often appear in newspaper texts. They often have OVS word order, see Example (111), but also OSV word order, see Example (112). SVO word order is also possible, see Example (113).

(111) *The young lady was stabbed in a WILD way, said police report*$_{\textbf{topic part}}$.

(112) *Kaci Hickox will be privately transported back to MAINE, Gov. Chris Christie said*$_{\textbf{topic part}}$.

(113) *Gov. Chris Christie said*$_{\textbf{topic part}}$ *that a nurse being held at a Newark hospital would be allowed to leave Monday and would be privately transported back to MAINE.*

Examples (111) and (112) demonstrate that to a certain extent, even English surface word order may be influenced by topic–focus articulation (despite the strong grammatical factor). The focus of the statement (i.e. what the police report said) occurs in the first part of the sentence and the less dynamic items (the governing verb and subject) are in the following part. If the sentence in the surface appears with the order of focus – topic, it is considered to be marked and it is called the **subjective word order**. It often appears in spoken language or in the information media where it serves as a mean how to attract the attention of readers and viewers.

The surface order of topic – focus (see Example (113) – the focus items stand in the final position of the sentence) is considered unmarked (or basic) and it is viewed as the **objective word order**.

During the annotation of communicative dynamism, the mutual sequence of nodes in the trees is arranged in the way that the contextually bound items stand before the governing verb of the main clause and contextually non-bound items behind it. **The deep order of nodes** thus **does not copy** the potential **subjective word order** that appears in the surface structure of the sentence, but it is arranged as **the objective word order**, see the dependency tree 3B in Section 13.2 *Text of the Article B.* Therefore, the order of subject, verb and object in Examples (111) and (112) will be modified in a dependency tree to the order of subject –

verb – object (topic – focus). Verbs introducing direct and indirect speech (the most frequent are verba dicendi) are considered contextually bound in most cases.

9.2 Intonation

One of the most important criteria for determining the contextual boundness is intonation. The intonation centre always lies in the focus proper (i.e. in the communicatively most important item). The focus proper is always contextually non-bound. During the annotation of spoken texts, the annotator hears where the author (speaker) puts the intonation centre of the sentence. However, during the annotation of written texts, it is necessary for the annotator to read first the given sentence (or text) aloud and to watch where he or she puts the intonation centre. We assume that the annotator is able to understand and interpret the annotated text properly and therefore to put the intonation centre in the same place as the author of the text.

As it is evident from the Example (105), the intonation centre in an English declarative sentence does not have to be situated always at its end, as it is e.g. in the inflected Czech in the sentence with the objective word order.

Intonation centre may be in English (but also in inflected languages like Czech) placed anywhere (in terms of surface word order) – in the first, last or any other place in the sentence.

Intonation is generally a stronger factor reflecting contextual boundness and communicative dynamism than the surface word order.

During the annotation of topic–focus articulation in texts (whether written or spoken and in any language), it is always necessary to take intonation into account. Different meanings of sentences with the same surface word order can often be distinguished thanks to intonation, see the following example:

(114) *OBAMA*$_{\textbf{focus part}}$ *is the favorite.* / *Obama is the FAVORITE*$_{\textbf{focus part}}$*.* / *Obama IS*$_{\textbf{focus part}}$ *the favorite.*

In the first case, we say that it is Obama who is the favorite. In the second case, we say about Obama that he is the favorite. Finally, we say that Obama is the favorite and not that he is not. The intonation helps the hearer to understand which of the three possible meanings is the right one that was intended by the author.

9.3 Type of Determination

English has a definite, indefinite and zero article. The kind of determination used by the author is closely linked to the topic–focus articulation. The indefinite article is usually accompanied by an entity that the speaker represents as new and communicatively most important. It can thus be assumed that the indefinite article accompanies contextually non-bound entities, cf. the following example which is a heading of the article from the *Wall Street Journal*. The expression *a festive dining room* is contextually non-bound – contextual non-boundness is signaled by an indefinite article *a* in this case.

(115) *Designing a Festive DINING Room* **focus part**

Contextual non-boundness of individual nodes can be determined also by the zero article. It may occur, for example, with nouns in plural that would have an indefinite article in singular – in these cases, it indicates contextual non-boundness:

(118) *Chad Oppenheim's Minimalist Miami Home – Kids* **focus part** *WELCOME*

The expression *kids* is accompanied by the zero article which in this case indicates that the item is contextually non-bound (the example is again a heading of the article from the *Wall Street Journal*).

On the other hand, the definite article is inherently predetermined rather for contextually bound nodes, cf. the expression *the piece* in the following example which belongs to the sentence topic.

(116) [*The "Esse Sofa" by Italian company Edra is covered in synthetic ostrich skin, perfect for a room with juice boxes and sticky fingers.*] *The piece* **topic part** *came from the couple's previous HOME and is a good fit with the room's neutral palette.*

The expression *the piece* is clearly synonymous with *the "Esse Sofa" by Italian company Edra* and is contextually bound. Its contextual boundness is indicated by the definite article *the* in this case.

However, it does not apply to all cases. The use of the definite article is not a completely reliable indicator of contextual boundness because in many cases, it also accompanies nodes that are contextually non-bound (cf. the expression *the best seat* and *in the house* in Example (117) which is again a heading of the article from the *Wall Street Journal* – the items are in the sentence focus, i.e. they are contextually non-bound, "even though" they are accompanied by the definite article *the*).

In some cases, the use of the definite article is thus determined by the grammatical factor (cf. examples like *the roof of our house*) rather than by contextual boundness.

(117) *Window Seats: The Best Seat*$_{\textbf{focus part}}$ *in the HOUSE*$_{\textbf{focus part}}$

9.4 Rhematizers

The indicators of contextual (non)boundness of certain nodes are the so-called rhematizers. Rhematizers usually occur with items belonging to the sentence focus, cf. the following example:

(120) [*The Careful Renovation of an Architecturally Important House*]
In Napa Valley, business partners tackle a derelict house with moldy wood, a bedroom that can only$_{\textbf{focus part}}$ *be accessed via the SHOWER*$_{\textbf{focus part}}$.

In this case, the rhematizer *only* indicates by its scope where we may find the focus proper of the subordinate relative clause: *via the shower*. Although the scope of the rhematizer concerns only one expression, it is placed before the verb in the surface word order, which is typical for English.

In other languages (as, for example, Czech), the scope of rhematizers is connected more to the surface word order. In Czech, the rhematizer would stand right before the item to which it is related, see Example (121) (a Czech translation of the Example (120)).

(121) [*Pečlivá rekonstrukce architektonicky významného domu*]
V Napa Valley řeší obchodní partneři opuštěný dům s plesnivým dřevem, ložnici, která je přístupná pouze$_{\textbf{focus part}}$ *přes SPRCHY*$_{\textbf{focus part}}$.

In English, it is not possible to rely on that the scope of rhematizer includes all the items that are placed behind it in the surface word order. The annotator must interpret its true scope from the context. Other rhematizers in English (besides the mentioned *only*) are, e.g., the following expressions: *not* (i.e. sentential and constituent negation), *even*, *particularly*, *in particular*, *especially*, *also*, *for example*, *principally*, *exclusively*, *just*, *at least*, *maximally*, *extremely*, *already*, *almost*, *extra*, *largely*, *mostly*, *mainly*, *contrariwise* and many others.

However, in some cases, even the rhematizer may be connected to sentence items that are contextually bound.

(122) [*After a full week of market-moving data, economy-watchers face a very light slate of reports scheduled next week. That will enable Federal Reserve events to hog the spotlight.*] *But* *even*$_{\textbf{topic part}}$ *in light week*$_{\textbf{topic part}}$*, here are FIVE items not to miss.*

In this case, the rhematizer *even* is connected with the expression *in light week* that is contrastive contextually bound.

9.5 Pronominalization

Expressions that are represented or accompanied by a pronoun (especially personal and demonstrative) are used often as identifiable by the speaker, and thus as contextually bound.

(123) [*Robin Williams, Actor, Remembered By Friends and Fans*]
WHAT friends of the late actor Robin Williams said about him$_{\textbf{topic part}}$ *on Twitter.*

Example (123) contains the prepositional phrase *about him* that is coreferential with the expression *the late actor Robin Williams*. In this example, it is contextually bound and belongs to sentence topic.

However, we cannot say that any pronominal expression is contextually bound in any case – see Example (124).

(124) *Hilarious paramedic guy Harry claims Harry Styles was named after HIM*$_{\textbf{focus part}}$*, but Anne Twist denies it.*

In this example, the expression *after him* is the focus proper of the dependent clause and is thus contextually non-bound (even if the reader is able to interpret from the context to whom the pronoun *him* refers).

In this particular case, however, we cannot exclude the double reading of the sentence, i.e. its ambiguity, cf. Examples (125) and (126).

(125) *Hilarious paramedic guy Harry claims Harry Styles was named* <u>*after HIM*</u>**focus part**
(i.e. *Regarding Harry Styles, it was named after HIM.*)

(126) *Hilarious paramedic guy Harry claims HARRY Styles was named* <u>*after him*</u>**topic part**
(i.e. *According to him, HARRY Styles was named after him.*)

If the intonation centre lies in the node *Harry*, the expression *after him* is contextually bound and *Harry Style* is the focus proper of the subordinate clause ("*even Harry Styles* was named after him"). If the intonation centre lies in the node *him*, this expression will be contextually non-bound and it will be the focus proper of the subordinate clause ("Harry Styles was named *after him*").

From the example, it is obvious that **in some cases, there is no single "correct" way how to understand topic–focus articulation in the given sentence**, but that it always (to a certain extent) depends on the perception of a particular reader (annotator).

Items expressed by indefinite pronouns and pronominal adverbs or numerals (such as *some*, *someone*, *something*, *sometime*, *somewhere*, *somehow*, *once*) express inherently something unknown, vague or yet unmentioned. Therefore, they often act as contextually non-bound (however, not absolutely in all cases), see Example (127).

(127) *James BROWN: Lost* *Someone*$_{\textbf{focus part}}$ *LYRICS*

I lost *someone*$_{\textbf{focus part}}$*, my LOVE*
Someone$_{\textbf{topic part}}$ *who's GREATER*
Than the STARS above
Someone$_{\textbf{topic part}}$ *who I NEED*
Someone$_{\textbf{topic part}}$ *who don't*
Let my heart BLEED

Someone$_{\textbf{topic part}}$ *that's the ONE*
That's the *SOMEONE*$_{\textbf{focus part}}$
That's the *someone*$_{\textbf{focus part}}$ *that I LOST*

The instances of the word *someone* marked as "topic part" are contextually bound. The aim is to say the properties of *someone* that I lost. On the contrary, the other instances of the word *someone* are contextually non-bound. In these cases, the aim of the statement is to say that this is "that someone" that I lost.

When deciding on contextual boundness, it is always necessary to take into account the context and the likely author's intent. At the same time, it is not excluded that some sentences admit different readings and are thus ambiguous (at least in their written form) – as mentioned above. The last verse of the poem could be probably interpreted either as "this is that SOMEONE that I lost" (intonation centre would be on the word *someone*), or "someone that I lost is THAT" (intonation centre would be on the word *that*). In the first case, the expression *someone* would be contextually non-bound (*That's the SOMEONE...*), in the second one, contextually bound (*THAT's the someone...*).

9.6 Deictic Expressions

Not only personal and demonstrative pronouns but also deictic expressions can be generally assumed to be most likely contextually bound (they refer to something or substitute something). Deictic expressions refer mainly to persons, places and time, see the following examples.

(128) *"Our new mission* <u>*here*</u>**topic part** *is not simply to OBSERVE; it's also to HEAL," Bill Masters says.*

The deictic word *here* is integrated into the broader situational context and therefore, it was annotated as contextually bound (*here* or the place to which it refers is probably identifiable for the recipient as well as author of the text).

However, there are also cases where the deictic expression is contextually non-bound – see Example (129).

(129) [*Over the past decade or so, Google's rise as the dominant figure in online advertising has given the company an air of being an unstoppable force, almost untouchable.*] *But that is* <u>*HERE*</u>**focus part** *in the U.S.* [*It is a very different story in Europe.*]

In Example (129), the deictic expression *here* is a part of focus and thus it is contextually non-bound (although the reader knows to which place it refers in this case – to the USA). The author of the text used *here* in the same way how contextually non-bound expressions are usually used. In this case, the item *here* brings an unrecoverable (core) information of the statement ("but this only applies HERE") and is therefore contextually non-bound. However, similar cases where deictic expressions are contextually non-bound are in minority.

Other examples of deictic expressions are *there*, *this*, *that*, *their*, *own*, *myself*, *now*, *then*, *soon*, *today*, *yesterday*, *year* etc.

9.7 Coreference and Associative Anaphora

Expressions that are in a coreferential or associative anaphoric relation with items mentioned in the previous (con)text tend to be contextually bound, see Example (130).

(130) [*Thomas F. McLarty: How Obama Can Salvage His Last Two Years*]
The president$_{\textbf{topic part}}$ *faces complex CRISES, including Ebola and Islamic State.*

The example demonstrates that the expression *the president* refers to the previously mentioned item *Obama*. In this case, the expression *the president* is contextually bound.

In authentic texts, we may deal with grammatical coreference, see Example (131), text coreference, see Example (132), or associative (bridging) anaphora, see Example (133). More details to coreference and anaphora in Prague treebanks are in Nedoluzhko (2011).

(131) *Actors who*$_{\textbf{topic part}}$ *are cast in New York's elaborate Halloween haunted-house shows are allowed to hone serious theatrical SKILLS.*

(132) *More companies are tailoring their*$_{\textbf{topic part}}$ *computer purchases to individual workers' NEEDS.*

(133) *As autumn days shorten, trees change their color. The leaves*$_{\textbf{topic part}}$ *FALL OFF.*

Expressions in a coreferential or anaphoric relation with items from the previous context, i.e. *who*, *their* and *leaves*, are contextually bound in these examples.

However, the fact that a participant is a part of a coreferential or anaphoric chain cannot be considered contextually bound automatically – see Examples (134) and (135).

(134) [*School Should Start Later So Teens Can Sleep, Urge Doctors*]
The American Academy of Pediatrics supports pushing back start times for OLDER kids$_{\textbf{focus part}}$.

(135) *The internet is angry with President Obama over saluting Marines with the same* $hand_{\textbf{focus part}}$ *he was using to hold his COFFEE cup.*

The expressions *older kids* and *hand* obviously have a relation to the previously mentioned entities (*older kids* is a synonym for the expression *teens*; the expressions *hand* and *President Obama* are in an associative anaphoric relation of part and whole) even though they are used as contextually non-bound in the examples above. The aim of the statement in Example (134) is that "the school start should move for older kids". In Example (135), the expression *hand* is even a part of focus proper – the core of the displeasure was that "President Obama saluted *with the same hand* he was using to hold his coffee cup".

9.8 Implicitness

On the level of deep syntax (on the tectogrammatical layer), also items (nodes) are reconstructed that are not expressed in the surface (e.g. unexpressed subjects in Czech, unexpressed participants or obligatory free modifications from the valency frame of verbs in Czech and English etc.). The fact that the speaker does not feel the need to explicitly express certain information means that this information is considered most likely as given and thus contextually bound, see Example (136).

(136) *When a review established that some four hundred persons had already emigrated, local authorities recommended a change of STRATEGY.*

The verb *to recommend* from the example has the following participants in its valency frame: Actor ("who recommends"), Addressee ("to whom it is recommended") and Patient ("what is recommended"). In the example, the author of the sentence explicitly expresses only the Actor (*authorities*) and Patient (*change*). The Addressee is not expressed in the surface structure (but it is present in its deep structure). However, even these unexpressed items (present only in the deep structure) are provided by the value of contextual boundness be it t, c or f.

In this case, the addressee is annotated as contextually bound – we can imagine e.g. the expression *to the public* in its place.

The majority of reconstructed nodes on the tectogrammatical layer may be marked as contextually bound. However, there are some exceptions – see Example (137).

(137) *The Difference between Your <u>First and Second CHILD</u>*

There is an ellipsis in the sentence: *first child and second child* (it is not the same child, but two children). The omitted expression *child* appears first in the sentence (in its deep structure). Therefore, it is viewed as contextually non-bound. Its second occurrence (which appears in the surface structure of the sentence) may be seen as a repetition of the first one and therefore it is contextually bound: *first* [*child* (= contextually non-bound item)] *and second child* (= contextually bound item).

Therefore, the fact that a sentence item is not expressed in the surface structure of the sentence is not a sufficient reason to automatically understand this item as contextually bound without further investigation.

9.9 Coordination and Apposition

We can assume that members of coordination (in the semantic relation of conjunction) and apposition have the same value of contextual boundness, see Examples (138) and (139).

(138) *I like Mr. <u>Bean</u>*$_{\textbf{focus part}}$ *and his <u>TEDDY</u>*$_{\textbf{focus part}}$.

Expressions *Bean* and *Teddy* are members of coordination and in this case, they are both contextually non-bound (both are part of focus which expresses what or whom I like).

(139) *Forget any potential catfight between Brad Pitt's former wife, Jennifer Aniston, and his <u>current partner</u>*$_{\textbf{focus part}}$, *<u>Angelina JOLIE</u>*$_{\textbf{focus part}}$, *the actor says in a new interview.*

The expression *current partner, Angelina Jolie* is an example of apposition. The nodes *partner* and *Jolie* have the same value of contextual boundness – they are contextually non-bound.

There are a few cases where members of coordination or apposition have a different value of contextual boundness – practically it is only the case where one of the members literally repeats from the previous context (in the deep structure) as in Example (137) items *child* and *child*.

9.10 Phenomena Appearing on the Scene

Phenomena which the author newly introduces on the scene are contextually non-bound, as the things or situations that are just appearing (or that already exist or are disappearing) are informationally very important. They are very often the focus proper, see Example (140) which is the name of the well-known song of the *Beatles*.

(140) *Here Comes the* <u>*SUN*</u>**focus part**

The phenomenon introduced on the scene (*the Sun*) is the item carrying the highest degree of communicative dynamism. It is contextually non-bound and it is the element carrying the intonation centre – in this case, it is the focus proper.

Other examples of the phenomena appearing on the scene are:

(141) *Once upon a time, there was a* <u>*KING*</u>**focus part** *who ruled a great and glorious nation.*

(142) *There is no* <u>*EXCEPTION*</u>**focus part**.

The phenomenon introduced on the scene is usually the subject of the sentence. At the same time, in English, there is a strong tendency for a subject to stand immediately before the governing verb in the surface word order. Therefore, the subject is often grammatically represented by the so-called anticipatory *it* (see Section *8.1 A Layer of Deep Syntax – Tectogrammatical Layer*), cf. the following example:

(143) *It's nice* <u>*to MEET*</u>**focus part** *you.*

In languages where the grammatical factor does not affect the word order so strongly as in English (e.g. in Czech), a formal subject (like *there, it*) does not occur, see the following examples which are Czech translations of the sentences mentioned above.

(144) *Zde přichází* <u>*SLUNCE*</u>**focus part**.

(145) *Byl jednou jeden* <u>*KRÁL*</u>**focus part**, *který vládl velkému a slavnému národu.*

(146) *Neexistuje žádná* *VÝJIMKA*$_{\textbf{focus part}}$.

(147) *Je hezké* *SETKAT SE*$_{\textbf{focus part}}$ *s vámi.*

However, the consequence is that some sentences that would be divided into topic and focus in English (because semantically empty expressions like *there* and *it* at the beginning of a sentence exhibit similar features like contextually bound nodes) would be sentences without topic in Czech (unless there is another contextually bound item directly dependent on the governing verb). However, the communicative value of these sentences is in both Czech and English the same. The Czech and English version of the same sentence certainly does not have a different topic–focus articulation.

We perceive the "division" of subject in an English sentence only as a surface expression of a single item from the deep structure (caused by a grammatical factor affecting the surface word order of English sentences). In the deep structure, it is just one single item. Therefore, we consider the anticipatory *it* and existential *there* as a part of focus, not topic, see Examples (148) and (149).

(148) ***There is no EXCEPTION***$_{\textbf{focus}}$

(149) ***It's nice to meet YOU***$_{\textbf{focus}}$

9.11 Semantic Type of a Sentence Item (Node)

Regarding topic–focus articulation, some types of nodes in the sentence clearly incline as a whole to behave in the same way. For example, words expressing the **speaker's attitude** towards the content of the utterance (or of its part) are **contextually bound** in most cases (they are connected with the speaker that is involved in the communication situation and its context). In the Prague dependency treebanks, these nodes are marked as "functor = ATT", see Example (150).

(150) *Manish PANDEY:* *Undoubtedly*$_{\textbf{topic part}}$ *today is one of the best days of my CAREER*

However, if such an expression forms an independent clause, it is contextually non-bound, see Example (151).

(151) *Will you still love me tomorrow? – UNDOUBTEDLY*$_{\textbf{focus part}}$.

Contextually bound (basically in all cases) are also **discourse connectives connecting the given statement with the previous context**, see Example (152). In the Prague dependency treebanks, these nodes are marked as "functor = PREC".

(152) *I will arrive in Geneve at 23:00 hours. So*$_{\textbf{topic part}}$ *I guess I can´t get the shuttle service from the HOTEL.*

Contextually bound are very often also the so-called **false subjects**, i.e. expressions in the position of subject that are, however, not captured within the valency frame of the given verb. In the Prague dependency treebanks, these nodes are labeled as "functor = INF", see Example (153).

(153) *It*$_{\textbf{topic part}}$ *is RAINING.*

On the contrary, **contextually non-bound** items are very often **expressions modifying or specifying nouns**. They often bring some additional or new information about the noun they modify. They tend to be contextually non-bound even if their governing noun is contextually bound, see Example (154). In the Prague dependency treebanks, these nodes are marked as "functor = RSTR, ID".

(154) *I am Bond. JAMES*$_{\textbf{focus part}}$ *Bond.*

In Example (154), the attribute *James* is contextually non-bound and, at the same time, it is the focus proper of the statement (it is the only one focus item).

Obviously, we may find cases where expressions functioning as attributes are contextually bound, see Example (155).

(155) *I have not mentioned the CD-ROM sources to allow the reader to read the article in a smooth flow and to "go into" the teachings without the jarring effect of the* *mentioned*$_{\textbf{focus part}}$ *sources and REFERENCES.*

In Example (155), the attribute *mentioned* is contextually bound (the verb *to mention* already appears in the previous context). However, it is a part of focus because its highest governing node (*sources*) directly dependent on the governing verb is contextually non-bound.

Contextually non-bound are in most cases also roots of an independent nominative clause which is not a parenthesis (it is the basic item of **sentence equivalents**). In the Prague dependency treebanks, these nodes are labeled as "functor = DENOM", see Example (156).

(156) *Page*$_{\textbf{focus part}}$ *56*

Contextually non-bound are usually also items expressing modality – they express especially **necessity**, **possibility** or **probability**. In the Prague dependency treebanks, these nodes are marked as "functor = MOD", see Example (157).

(157) *10 LIES You Were* *Probably*$_{\textbf{focus part}}$ *Taught In School.*

Contextually non-bound are basically in all cases interjections or particles representing roots of the so-called independent interjectional clause. In the Prague dependency treebanks, these nodes are labeled as "functor = PARTL", see Examples (158) and (159).

(158) *YES*$_{\textbf{focus part}}$*, that's exactly what I MEANT.*

(159) *OOPS*$_{\textbf{focus part}}$*, something went WRONG.*

Contextually non-bound are very often also the so-called **rhematizers** that are usually expressed by particles and adverbs or by items expressing negation or affirmation. In the Prague dependency treebanks, these nodes are marked as "functor = RHEM", see Section 9.4 *Rhematizers*. They are contextually non-bound if they rhematize the expressions in focus, see Example (160).

(160) *Only*$_{\textbf{focus part}}$ *LOVERS Left Alive*

However, if the rhematizer belongs to contextually bound expressions, also the rhematizer is contextually bound, see Example (161).

(161) *Both of them were hungry. However, only*$_{\textbf{topic part}}$ *she wanted to cook MEAL.*

10 Topic–Focus Articulation from the Perspective of Translators

Functional Generative Description considers topic–focus articulation as a phenomenon associated with the deep structure of sentences because the change of topic–focus articulation entails also a change in the meaning of the sentence, cf. Examples (162) to (165).

(162) *It is not true that* <u>*JOHN*</u>**focus part** *stole my heart. (It was JACK.)*

(163) *It is not true that John* <u>*STOLE*</u>**focus part** *my heart. (He BORROWED it only.)*

(164) *It is not true that John stole* <u>*MY*</u>**focus part** *heart. (He stole HER heart, unfortunately.)*

(165) *It is not true that John stole my* <u>*HEART*</u>**focus part**. *(He stole my HANDBAG.)*

All the given sentences in English are expressed in the same way (in the written form, i.e. they have the same surface word order). Yet each of them has a different meaning. The difference in meaning is expressed by the individual languages differently – e.g. by the intonation (in the spoken form of the sentence) or word order (word order as a means of expressing topic–focus articulation is used mainly by languages with the so-called free word order like Czech). English uses for the distinction of sentence meaning with different topic–focus articulation mainly intonation, word order only to a limited extend. On the other hand, Czech uses both very often – in Czech, the focus proper has the intonation centre and it also tends to occupy the last position in the sentence (in unmarked declarative sentences with the so-called objective word order), see the following examples (translations of the English examples above).

(166) *Není pravda, že moje srdce ukradl* <u>*JOHN*</u>**focus part**.

(167) *Není pravda, že John moje srdce* <u>*UKRADL*</u>**focus part**.

(168) *Není pravda, že John ukradl* <u>*MOJE*</u>**focus part** *srdce.*

(169) *Není pravda, že John ukradl moje* <u>*SRDCE*</u>**focus part**.

However, if the focus proper is the attribute modifying a noun, as in Example (168), the grammatical factor outweighs the topic–focus articulation even in Czech (in most cases) and the attribute remains in its typical position, i.e. before the noun it modifies.

The examples demonstrate that even if the sentences have different word-order realization in English and Czech, the meaning remains the same in both languages. For example, it is not possible to translate Example (162) in English as Example (169) in Czech, but only as Example (166).

Therefore, when translating from one language to another, it is necessary to transfer not only the corresponding lexical meanings, but also **the relationships between these meanings**, i.e. the communicative importance of each translated semantic element and information structure of the statement. In other words, whether a particular sentence (as a part of broader context) is expressed (in its surface realization) in any language, its deep (semantic) information structure is always the same. **It is thus not possible for the same sentence in the same context to have different topic–focus articulation in English, German or Russian.** The communication intention must remain the same regardless of the means by which it is expressed in the particular language, see the following Examples (170) to (173) that express the answer to the question *What did you do yesterday?* in English, Czech, German and Russian.

(170) English

Surface structure: *I played football yesterday.*

Deep structure: $I_{.t}$ $yesterday_{.t}$ $\boldsymbol{played}_{.f}$ $\boldsymbol{FOOTBALL}_{.f}$

(171) Czech

Surface structure: *Včera jsem hrál fotbal.*

Deep structure: [$Já_{.t}$] $včera_{.t}$ $\boldsymbol{\{jsem\ hrál\}}_{.f}$ $\boldsymbol{FOTBAL}_{.f}$

(172) German

Surface structure: *Gestern habe ich Fußball gespielt.*

Deep structure: $Ich_{.t}$ $gestern_{.t}$ $\boldsymbol{\{habe\ gespielt\}}_{.f}$ $\boldsymbol{FUßBALL}_{.f}$

(173) Russian

Surface structure: *Вчера я играл в футбол.*

Deep structure: $Я_{.t}$ $вчера_{.t}$ $\boldsymbol{играл}_{.f}$ $\boldsymbol{\{в\ ФУТБОЛ\}}_{.f}$

As seen from the examples, the annotation of topic–focus articulation is the same in all cases. The contextually bound items *I* and *yesterday* belong to topic and the contextually non-bound items *play* and *football* belong to focus. Also the communicative dynamism of the individual items is the same in all languages (in the given context).

However, in the surface word order, e.g. the node *yesterday* stands at the very end of the sentence in English, i.e. in the place where Czech sentence usually puts the focus proper. The focus proper *football* (i.e. the node with the highest communicative dynamism) stands rightmost in the surface structure in Czech and in Russian, but not in German and English. On the contrary, the node with the lowest degree of communicative dynamism *I* stands most to the left in the surface word order only in English. In German, it is behind the finite verb, in Russian before a finite verb and, at the same time, behind the expression *yesterday* and in Czech, *I* is not expressed in the surface structure at all.

The extent of the individual nodes is also different. The predicate *to play* is expressed by a simple word in the surface structure in Russian and English but by an analytical predicate in Czech and German (i.e. by a lexical verb plus auxiliary verbs). In all cases, however, it is just one node in the deep structure.

Not only the annotation of contextual boundness of the individual nodes is the same, but also the communicative dynamism. It is indicated in Examples (170) to (173) by the deep word order (in the deep structure). From this, we may see that nodes in topic have a lower degree of communicative dynamism than nodes in focus. The lowest degree of communicative dynamism carries the subject *I* and the highest the focus proper *football*.

The fact that the surface expression of a sentence in different languages has the same topic–focus articulation in the deep structure (in terms of contextual boundness as well as communicative dynamism) is important for the annotation of multilingual parallel corpora such as the *Prague Czech-English Dependency Treebank*. It contains the original English texts from the *Wall Street Journal* and their Czech translations. The annotation of topic–focus articulation in Czech and English part thus must be uniform.

On the other hand, it is clear that different languages can express the same extralinguistic reality in different ways. In the consequence, the deep dependency trees are not absolutely the same for different languages in all cases – e.g. they may have a different number of nodes although they represent the same sentence. For example, German very often uses compounds even in cases where another language does not use them etc. However, a compound is considered a single lexical unit as well as a single dependency node. In this

way, a certain extralinguistic reality in one language may have a tree with more nodes (that require their own annotation of contextual boundness and communicative dynamism) than in another language, see Examples (174) to (177).

Czech

(174) *Čajová lžička je vhodná pro míchání čaje nebo KÁVY.*

Deep structure: {*lžička*}$_{.f}$ {*čajová*}$_{.f}$

Russian

(175) *Чайная ложка подходит для смешивания чая или КОФЕ.*

Deep structure: {*ложка*}$_{.f}$ {*чайная*}$_{.f}$

English

(176) a *teaspoon is suitable for stirring of tea or COFFEE.*

Deep structure: {*a teaspoon*}$_{.f}$

German

(177) *Ein Teelöffel ist geeignet zum Rühren Tee oder KAFFEE.*

Deep structure: {*ein Teelöffel*}$_{.f}$

While German and English have a single lexical unit for *a teaspoon* (represented by a single node in the deep structure), Czech and Russian use a connection of an attribute and a noun, i.e. two lexical units (represented by two different nodes in the deep structure). In English and German, the node representing a compound is thus assigned a single value of contextual boundness and the communicative dynamism is not marked within this node.

On the other hand, in Czech and in Russian, both nodes get their own value of contextual boundness and their mutual order will be annotated in accordance with the degree of communicative dynamism – the attribute which carries a higher degree of communicative dynamism will be placed to the right from its governing (and less dynamic) noun. However, it is not admissible for the nodes *teaspoon* and *Teelöffel* to get a different value of contextual boundness than nodes *lžička a ложка* which are the governing nodes for phrases *čajová lžička* and *чайная ложка*.

11 Annotation of Topic–Focus Articulation in Dependency Treebanks – Some Guidelines Principles

1. We carefully watch the preceding context.

2. In the annotated sentence, we firstly find its **focus proper** (i.e. the most important part due to which the author realized the sentence) according to the following points:

a) **importance of the information** that is carried by this part of sentence (the focus proper is such part that is considered the most important for the reader);

b) location of the **intonation centre**: we read the sentence aloud (in the context of the previous statement) and we recognize where we put the intonation centre;

c) the part of the sentence where both a) and b) meet is the focus proper and we assign it the value "f". We move the node of the focus proper (with its whole subtree) to the rightmost position in the dependency tree.

Note: Within a nominal group, the intonation centre may be moved from the focus proper to some of the noun modifications. The focus proper may thus be also a node hierarchically higher than that one carrying the intonation centre.

Note: The focus proper does not have to be identical with the grammatical determination of sentence elements (i.e. e.g. not all nodes syntactically dependent on the focus proper must necessarily have the value "f"), see Example (178):

(178) [*How much dollars do you spent on your mobile phone?*] *About 40*$_{\mathbf{f}}$ *dollars*$_{\mathbf{t}}$.

Note: The focus proper in English sentences does not have to be necessarily placed at the final position in the surface word order.

Note: We translate the sentence into the language that reflects topic–focus articulation in the surface word order – e.g. in Czech, the focus proper is placed at the very end of the sentence in most cases (i.e. in the sentence with an unmarked, objective word order).

3. We carry out the **question test**. We ask a question so that the annotated sentence is an adequate answer to this question. The question must contain all items known from the broad context of the given communication situation (they do not have to be explicitly mentioned in the previous text), see Example (179).

(179) [*Yesterday I upgraded my Linux distribution. Although I like the improvements, a problem that was already bad seems to have gotten worse: Size limits for plasmoids are too big!*] *Because of that I had to remove it after upgrading.*

Question test:
An example of a question representing the communication situation outlined in the square brackets:
***WHAT did** I **do** with it because of the problem after upgrading?*
Answer: *Because of that I **had to REMOVE** it after upgrading.*

The items occurring both in the question and answer (directly dependent on the governing verb) are contextually bound (part of topic) and the items (directly dependent on the governing verb) after which the question asks are contextually non-bound (they are part of focus):

Because of that $I_{\textbf{topic\_1}}$ ***had to REMOVE***$_{\textbf{focus}}$ *it after upgrading*$_{\textbf{topic\_1}}$.

Note: If we cannot ask any similar question (the annotated sentence is too far from the available context as in Example (180)), we ask the question ***What happened?*** Then it is the so-called topicless sentence, i.e. a sentence without topic (not every sentence must have a topic, but it always have a focus). Sentences without topic tend to be the first sentences of a text or its heading. However, it is not true that all first sentences and all headings are without topic. They may include segments involved into the wide situational and experiential context.

(180) [*London –*] ***New media have become far more than just toys for POLITICIANS***$_{\textbf{focus}}$

4. To check our judgements, we carry out also the **test with negation**. Its results should be consistent with the results of the question test.

*Because of that I had not to remove it after upgrading *but because of lack of time.*
*Because of that I had not to remove it after upgrading *but my brother.*
*Because of that I had not to remove it after upgrading *but a car.*
*Because of that I had not to remove it after upgrading *but after lunch.*

Because of that I had not to remove it after upgrading ***but I had to reinstall it****.*

Text segments that are in the scope of negation are in focus whereas those that are not in the scope of negation are in topic.

Because of that $I_{\mathbf{topic\_1}}$ ***had to REMOVE***$_{\mathbf{focus}}$ *it after upgrading*$_{\mathbf{topic\_1}}$**.**

5. Text segments that are "asked for" in the question test and which may be in the scope of negation are assigned the value "f" (on the level of nodes directly dependent on the governing verb; deeply embedded nodes in the tree can have, of course, a different value of contextual boundness).

6. We have distinguished the items in the focus (contextually non-bound) and in the topic (contextually bound) – on the level of nodes directly dependent on the governing verb. Concerning items in the topic, it is necessary to decide whether they are **contrastive** (then they are assigned the value "c") or non-contrastive (then they receive the value "t").

Note: We read the sentence aloud. We observe whether we may put a contrastive stress on any of the items (it has a rising melody). If so, it gets the value "c". If not, all contextually bound nodes are marked as "t".

7. We assign the value of contextual boundness to nodes that are **deeply embedded** in the sentence structure (they modify the nodes directly dependent on the governing verb of the main clause). Each of them has its own value of contextual boundness regardless of whether it is dependent on the item in the focus or in the topic.

8. We carry out the annotation of **communicative dynamism** – in each level of the dependency tree separately. We move the nodes in the topic directly depending on the governing verb before the governing verb (i.e. to the left) and nodes in the focus behind it (i.e. to the right). According to the rules for ordering nodes, we regulate the order of nodes in all the levels of the dependency tree.

12 Summary

In our manual, we have briefly introduced the theory of topic–focus articulation from the perspective of the Functional Generative Description. The fundamental features which the Functional Generative Description works with are contextual boundness and communicative dynamism. These two phenomena also serve as grounds for delimitation of topic and focus.

The theory of the Functional Generative Description also served as a basis for the annotation of topic–focus articulation in the Prague Czech-English Dependency Treebank. Topic–focus articulation is annotated on the tectogrammatical (deep syntactico-semantic) layer of language in two steps – as contextual boundness and communicative dynamism in dependency trees. The division of sentences into topic and focus is not explicitly marked but it is clearly deducible from the annotation of contextual boundness and communicative dynamism.

The annotation of topic–focus articulation in the Prague treebanks belongs to the phenomena with quite high inter-annotator agreement – despite the fact that the annotation of authentic texts depends to some extent on the annotator's interpretation (see above mentioned ambiguous sentences). The inter-annotator agreement in assigning the value to individual nodes in the annotation of topic–focus articulation in PDT was 82% (see Mírovský, 2015).

There are still a few open questions. One of them is a further study of contrastive contextually bound nodes. During annotations of written texts, it turned out that the annotators are not sure in some cases whether the given node can or cannot bring the facultative contrastive stress. Yet, the possible occurrence of the contrastive stress is crucial in order to decide whether the sentence element is contrastive or non-contrastive contextually bound. In such cases, the annotators have to rely on their language consciousness and experience to some extent.

13 Sample Annotation of Topic–Focus Articulation in Dependency Trees

For a practical demonstration of annotation of topic–focus articulation, we have chosen two separate texts (here marked as A and B) from the *Prague Czech-English Dependency Treebank* (PCEDT). PCEDT contains texts from the *Wall Street Journal* collected in the corpus Penn Treebank. The texts are from the newspaper articles.

Topic–focus articulation is annotated in the dependency trees in two steps – it concerns two individual phenomena:

a) **contextual boundness** (its values are "c" for nodes that are contrastive contextually bound and are marked in green, "t" for nodes that are non-contrastive contextually bound and are marked in white and "f" for nodes that are contextually non-bound and are marked in yellow);

b) **communicative dynamism** (it is annotated through the mutual sequence of nodes in each individual level of a dependency tree).

For illustration, we mark also the division of a sentence into (global) topic (not in bold) and (global) focus (in bold) (local topic–focus articulation, e.g. in dependent clauses, is not explicitly marked but it is well visible from the dependency trees). The topic consists of all contextually bound nodes directly dependent on the governing verb (in the independent clause) and all nodes depending on them. The focus consists of all contextually non-bound nodes directly dependent on the governing verb (in the independent clause) and all nodes depending on them.[19]

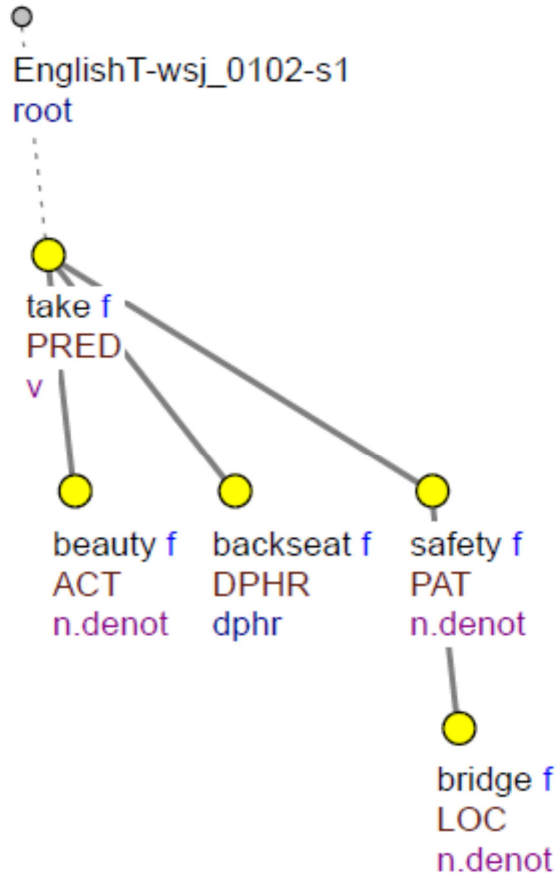
[19] The arrows seen in the figures capture the relations of coreference and associative anaphora.

13.1 Text of the Article A

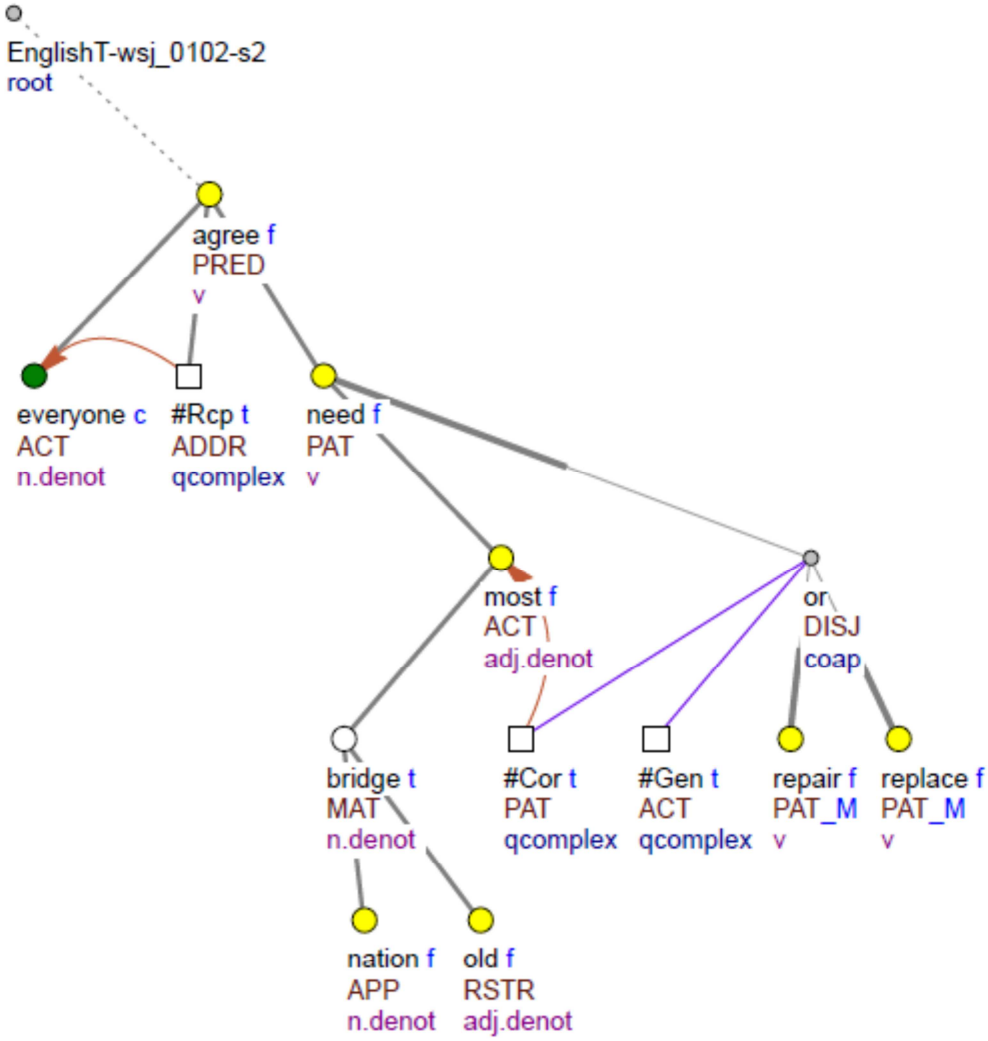
1) ***Beauty Takes Backseat To Safety on Bridges***

2) *Everyone agrees that most of the nation's old bridges need to be repaired or replaced.*
3) *But there's disagreement over how to do it.*
4) *Highway officials insist the ornamental railings on older bridges aren't strong enough to prevent vehicles from crashing through.*
5) *But other people don't want to lose the bridges' beautiful, sometimes historic, features.*
6) *"The primary purpose of a railing is to contain a vehicle and not to provide a scenic view," says Jack White, a planner with the Indiana Highway Department.*
7) *He and others prefer to install railings such as the "type F safety shape," a four-foot-high concrete slab with no openings.*
8) *In Richmond, Ind., the type F railing is being used to replace arched openings on the G Street Bridge.*
9) *Garret Boone, who teaches art at Earlham College, calls the new structure "just an ugly bridge" and one that blocks the view of a new park below.*
10) *In Hartford, Conn., the Charter Oak Bridge will soon be replaced, the cast-iron medallions from its railings relegated to a park.*
11) *Compromises are possible.*
12) *Citizens in Peninsula, Ohio, upset over changes to a bridge, negotiated a deal: The bottom half of the railing will be type F, while the top half will have the old bridge's floral pattern.*
13) *Similarly, highway engineers agreed to keep the old railings on the Key Bridge in Washington, D.C., as long as they could install a crash barrier between the sidewalk and the road.*

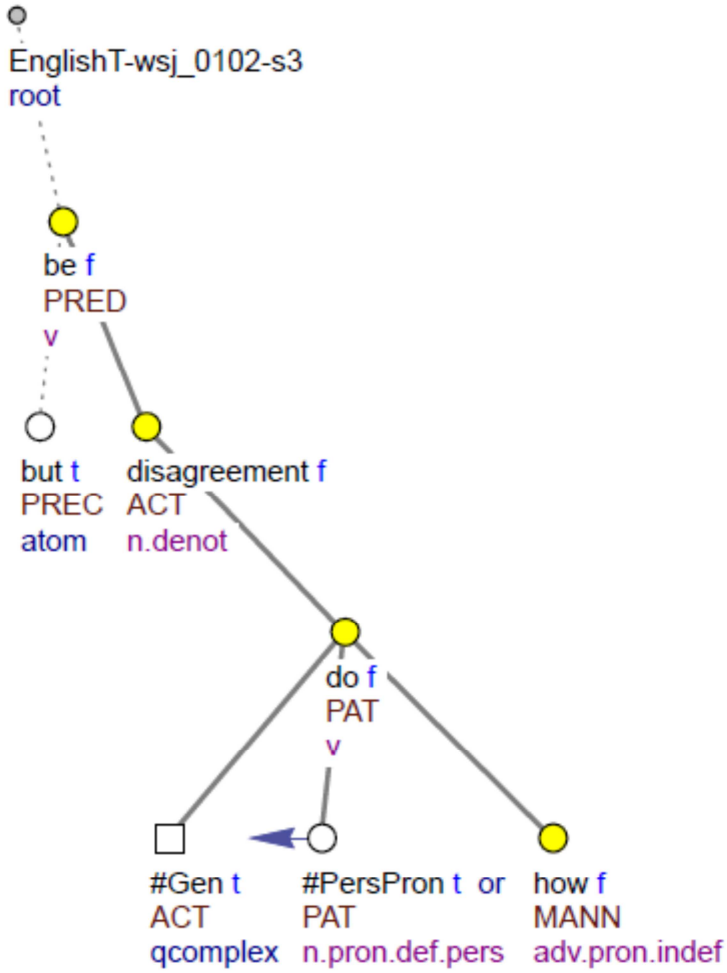
1A) ***Beauty Takes Backseat To Safety on Bridges***$_{\textbf{focus}}$



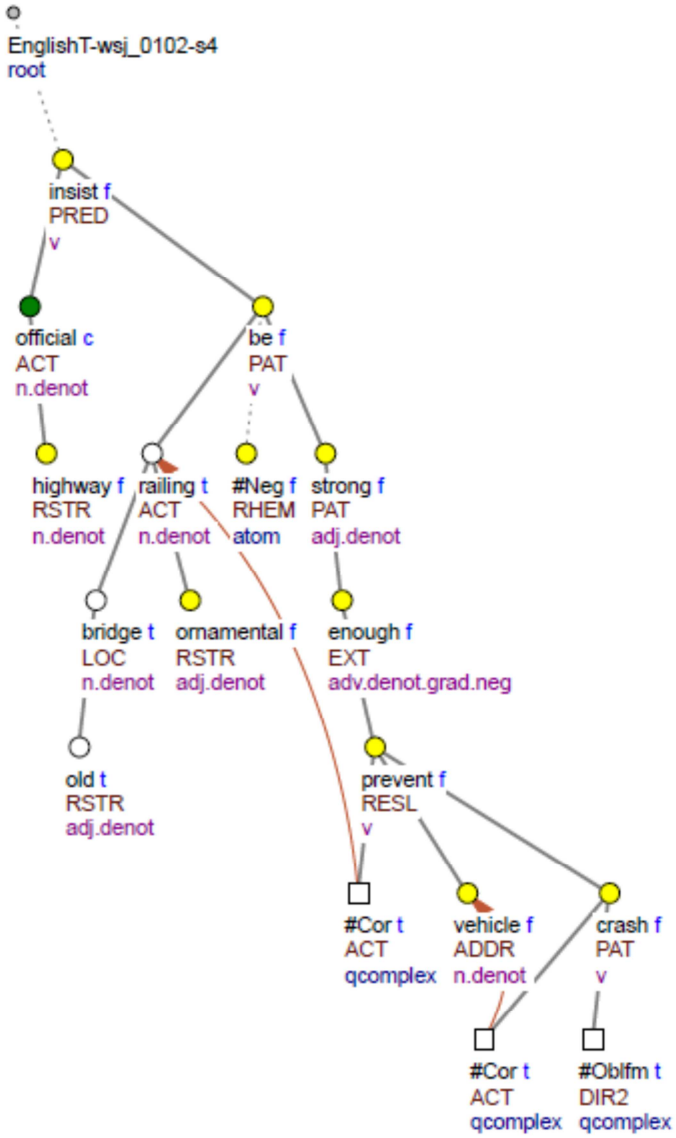
2A) *Everyone*$_{\textbf{topic}}$ ***agrees that most of the nation's old bridges need to be repaired or replaced***$_{\textbf{focus}}$.



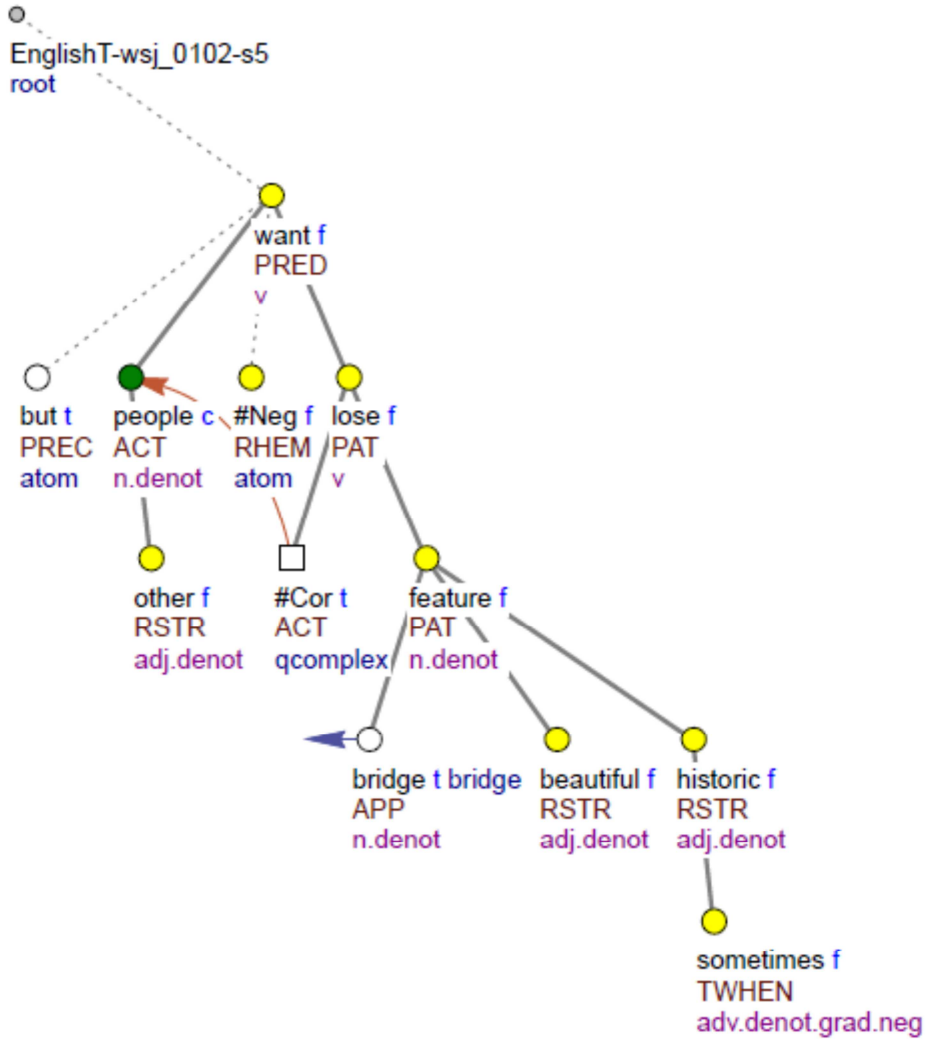
3A) *But*$_{\textbf{topic}}$ ***there's disagreement over how to do it***$_{\textbf{focus}}$.



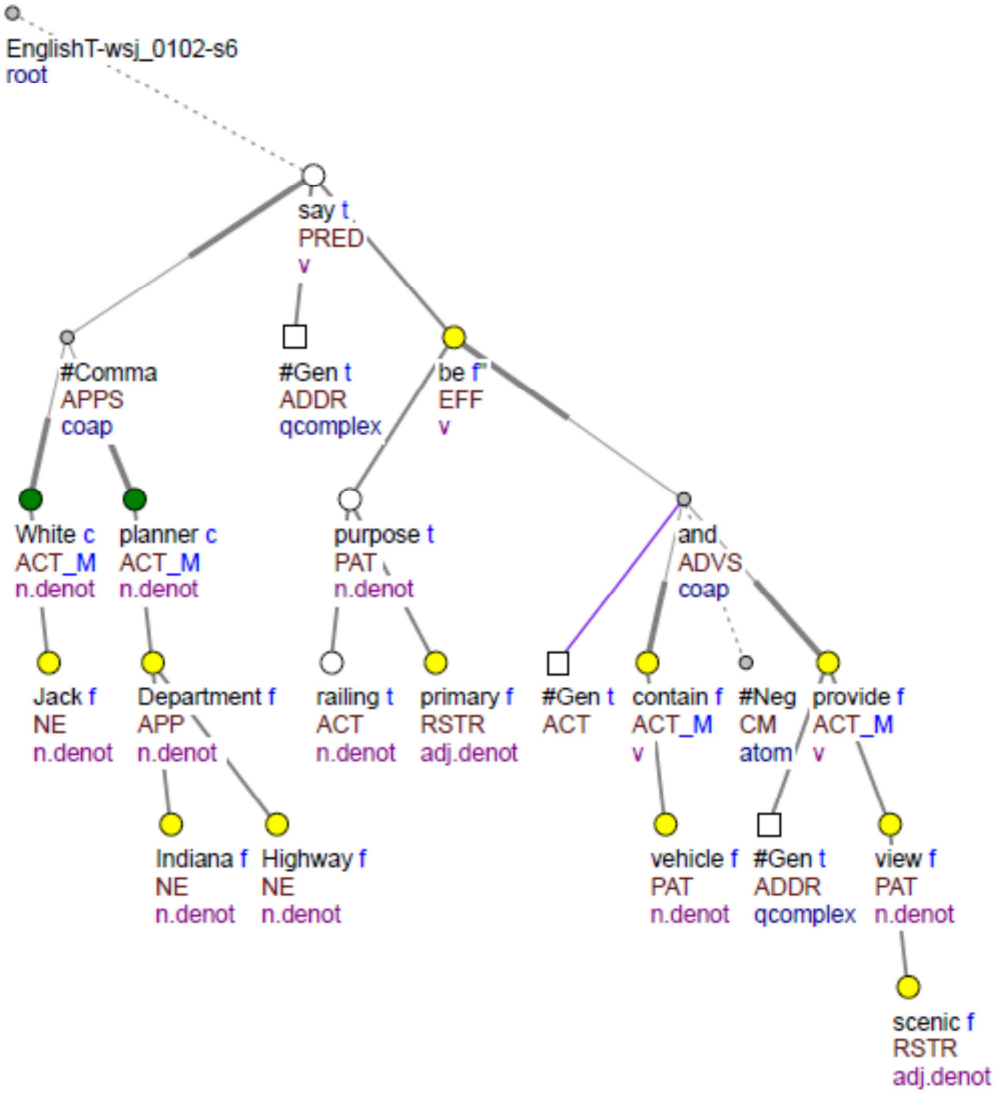
4A) *Highway officials*$_{\textbf{topic}}$ ***insist the ornamental railings on older bridges aren't strong enough to prevent vehicles from crashing through***$_{\textbf{focus}}$.



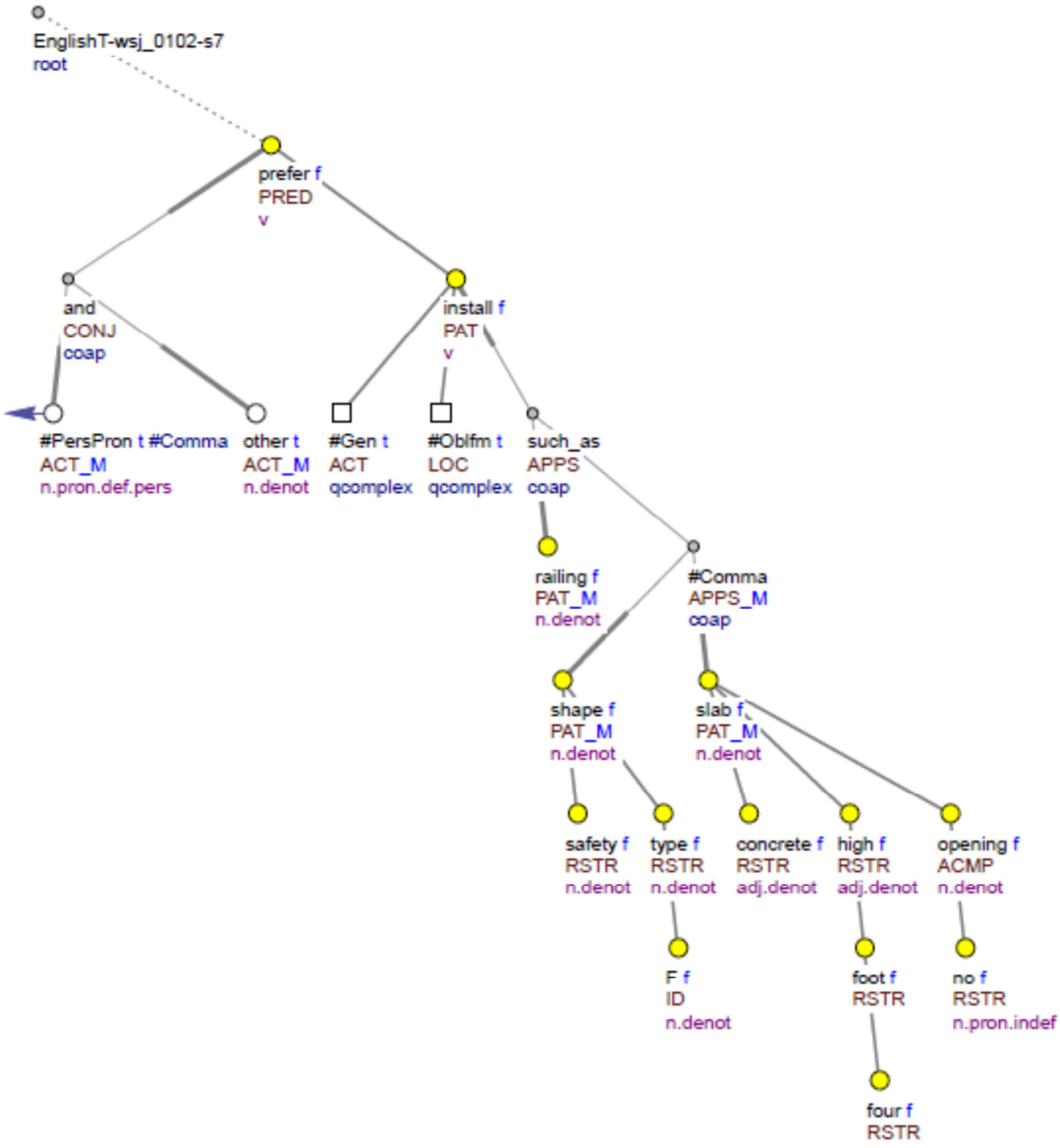
5A) *But other people*$_{\textbf{topic}}$ ***don't want to lose the bridges' beautiful, sometimes historic, features***$_{\textbf{focus}}$.



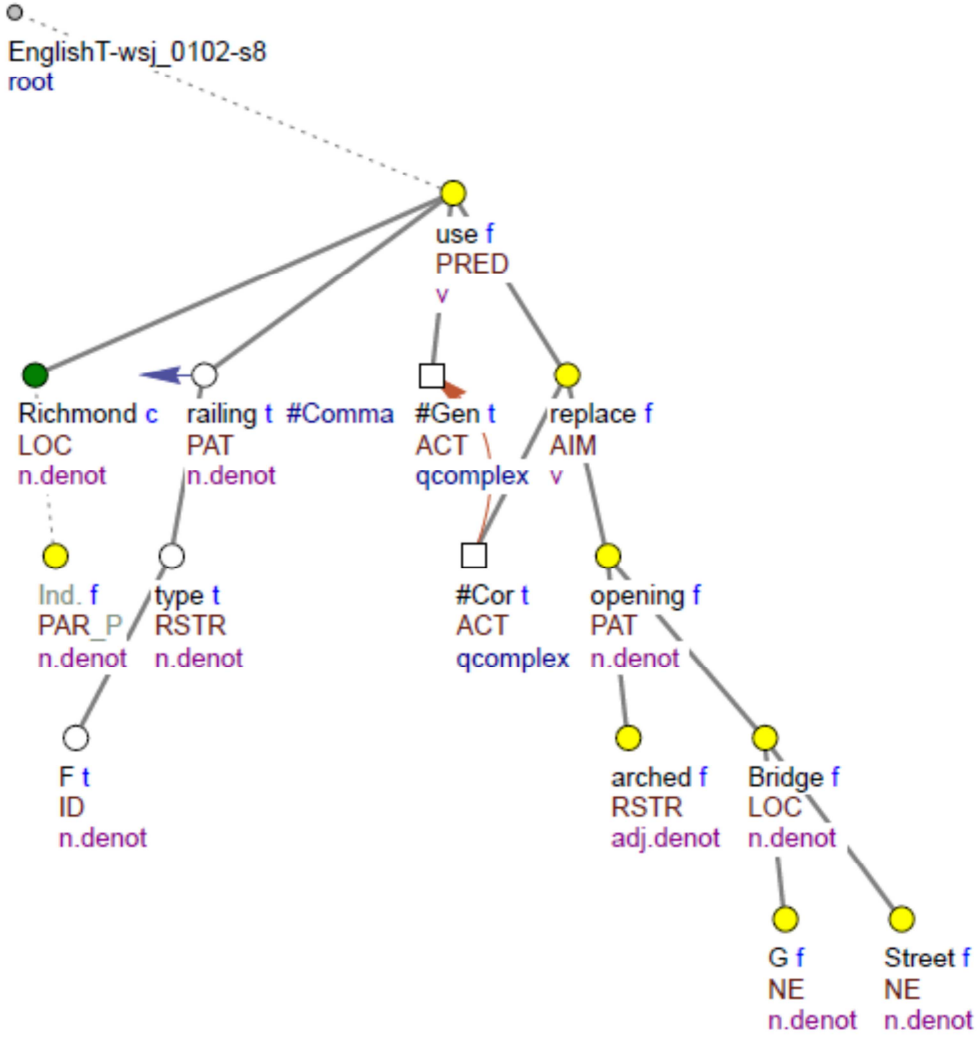
6A) ***"The primary purpose of a railing is to contain a vehicle and not to provide a scenic view,"***$_{\text{focus}}$ *says Jack White, a planner with the Indiana Highway Department*$_{\text{topic}}$.



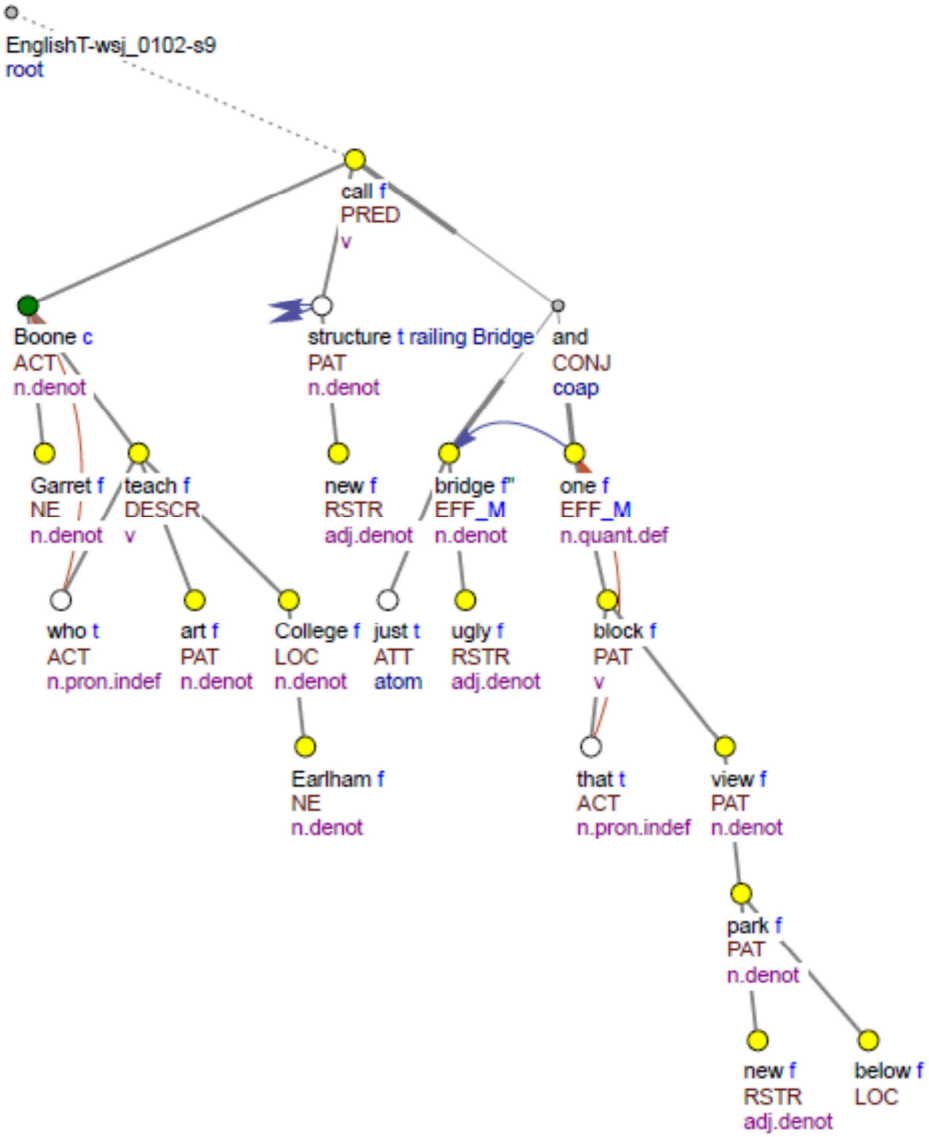
7A) *He and others*$_{\textbf{topic}}$ ***prefer to install railings such as the "type F safety shape," a four-foot-high concrete slab with no openings***$_{\textbf{focus}}$.



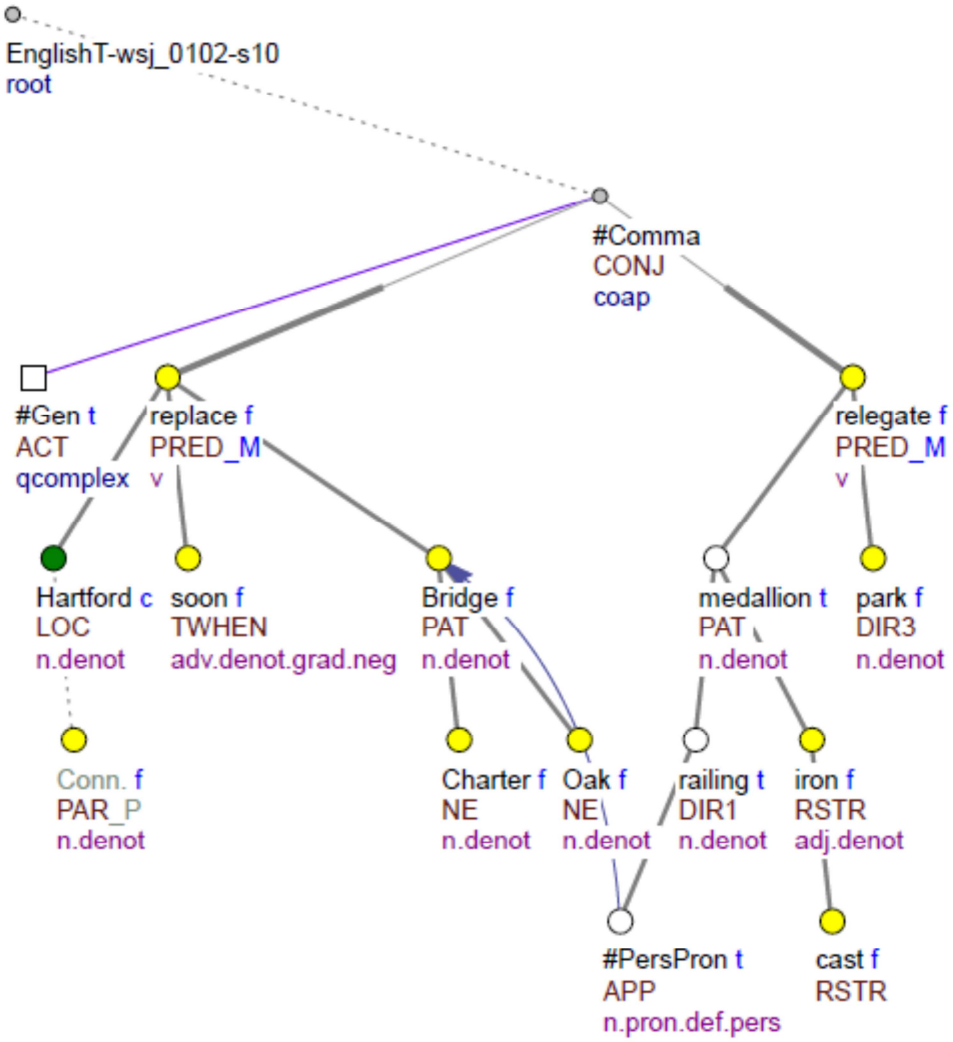
8A) *In Richmond, Ind., the type F railing*$_{\textbf{topic}}$ ***is being used to replace arched openings on the G Street Bridge***$_{\textbf{focus}}$**.**



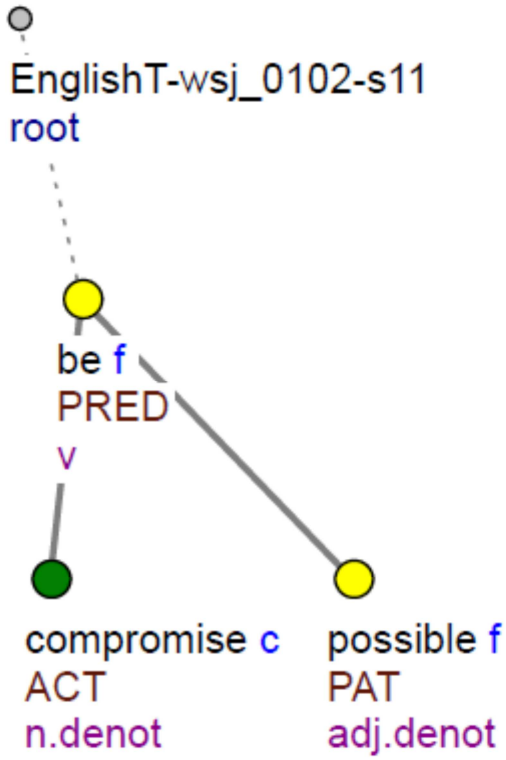
9A) *Garret Boone, who teaches art at Earlham College,*$_{\text{topic\_1}}$ ***calls***$_{\text{focus\_1}}$ *the new structure*$_{\text{topic\_1}}$ ***"just an ugly bridge" and one that blocks the view of a new park below***$_{\text{focus\_1}}$**.**



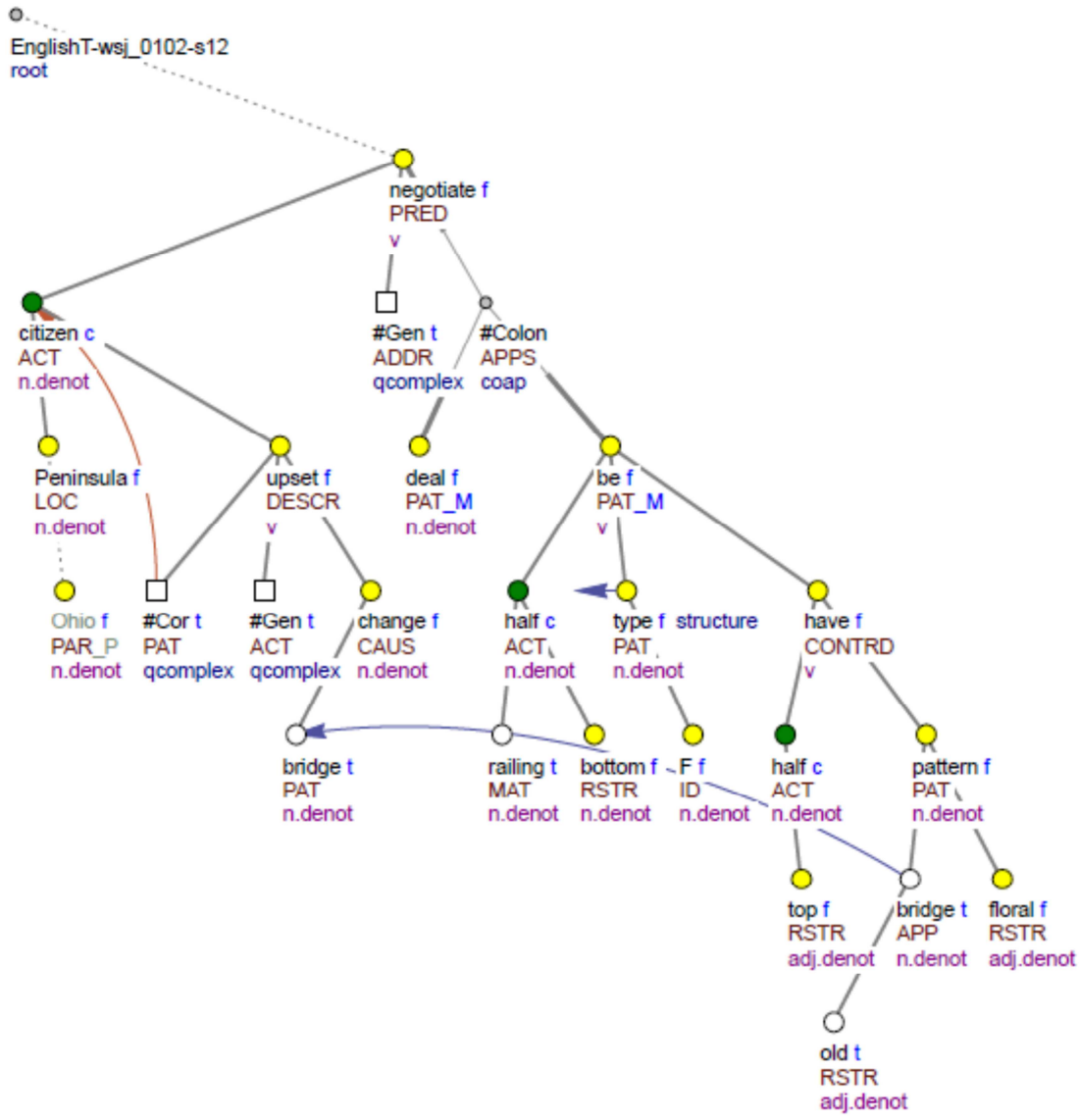
10A) *In Hartford, Conn.,*$_{\text{topic\_1}}$ ***the Charter Oak Bridge will soon be replaced***$_{\text{focus\_1}}$***,***
the cast-iron medallions from its railings$_{\text{topic\_2}}$ ***relegated to a park***$_{\text{focus\_2}}$***.***



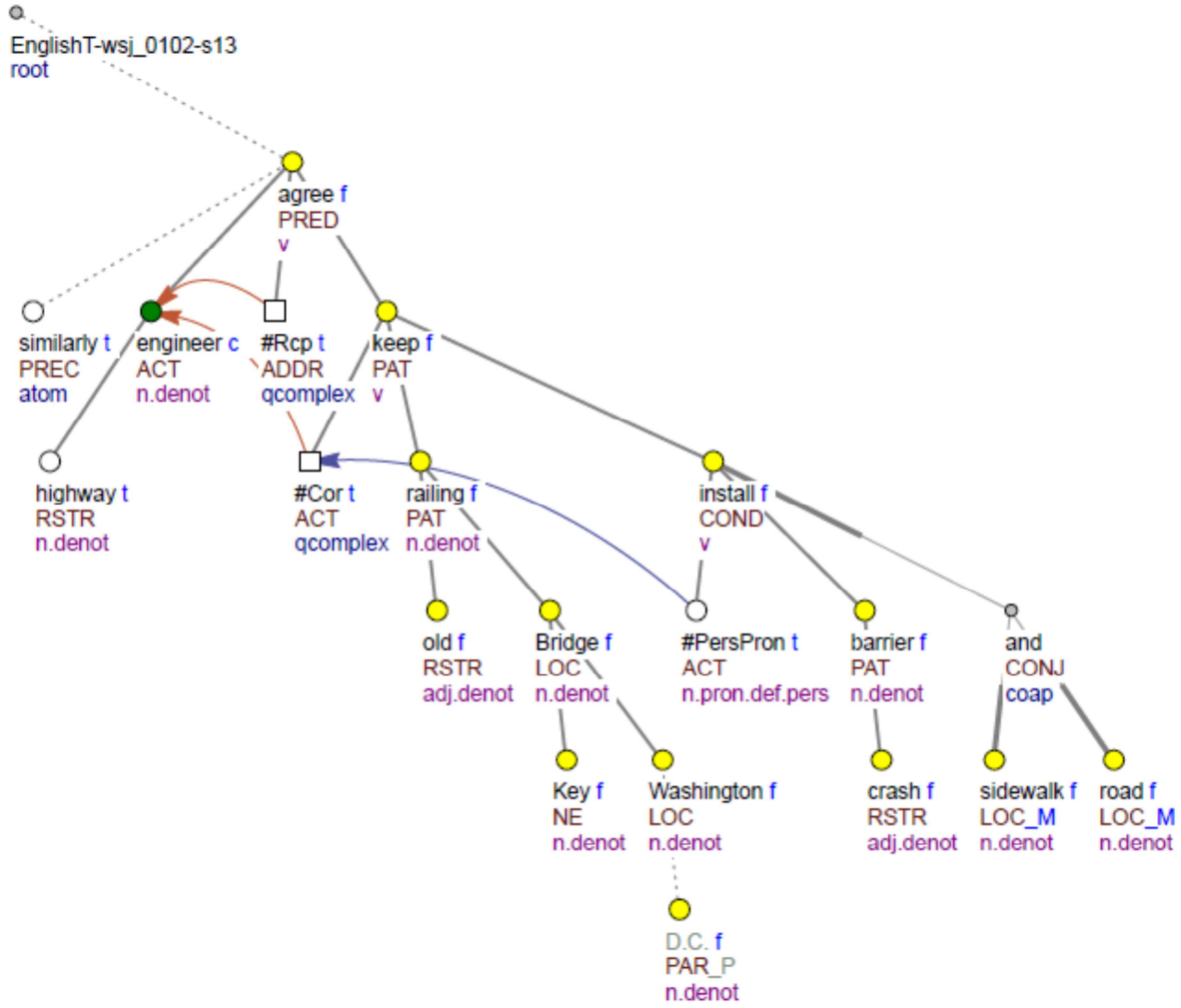
11A) *Compromises*$_{\textbf{topic}}$ ***are possible***$_{\textbf{focus}}$.



12A) *Citizens in Peninsula, Ohio, upset over changes to a bridge,*$_{\textbf{topic}}$ ***negotiated a deal: The bottom half of the railing will be type F, while the top half will have the old bridge's floral pattern***$_{\textbf{focus}}$.



13A) *Similarly, highway engineers*$_{\textbf{topic}}$ ***agreed to keep the old railings on the Key Bridge in Washington, D.C., as long as they could install a crash barrier between the sidewalk and the road***$_{\textbf{focus}}$.

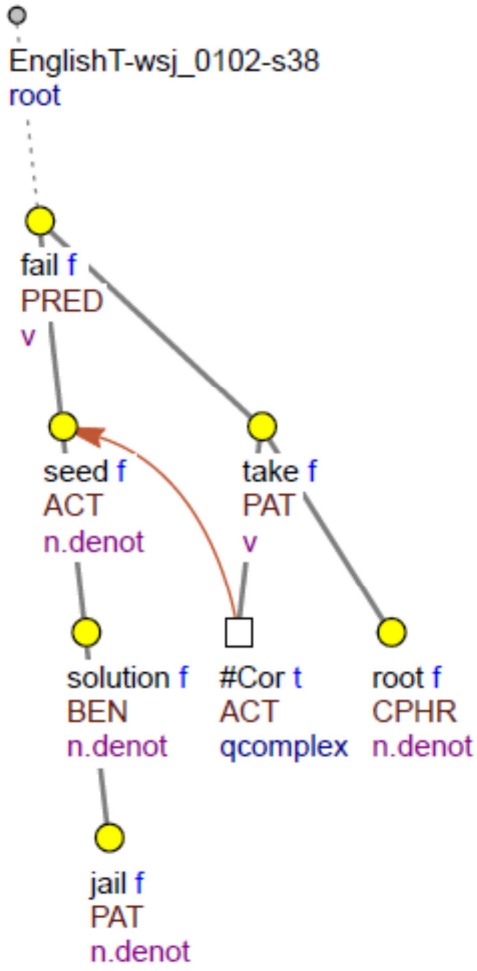


13.2 Text of the Article B

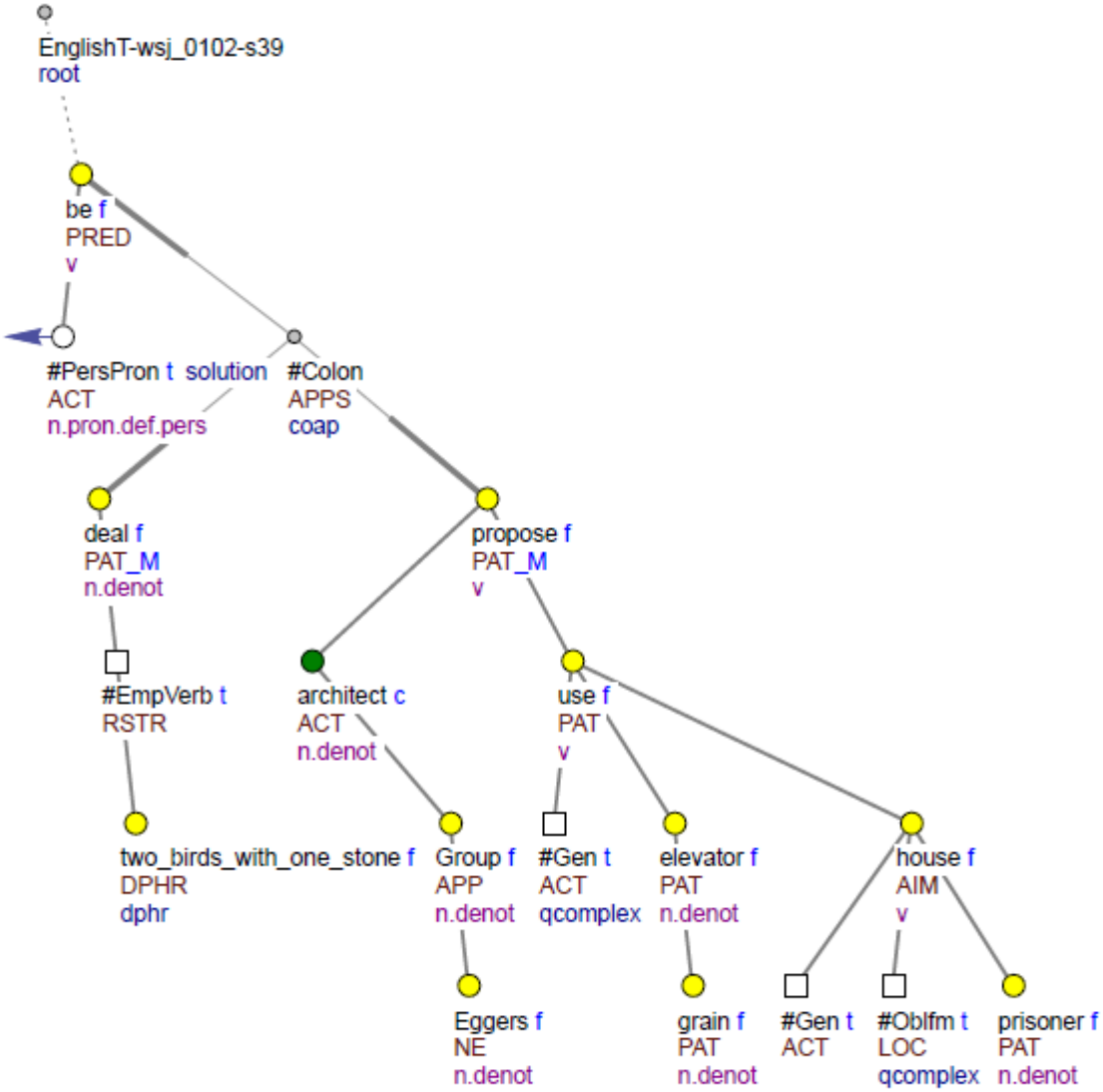
1) ***Seed for Jail Solution Fails to Take Root***

2) *It's a two birds with one stone deal: Eggers Group architects propose using grain elevators to house prisoners.*
3) *It would ease jail overcrowding while preserving historic structures, the company says.*
4) *But New York state, which is seeking solutions to its prison cell shortage, says "no."*
5) *Grain elevators built in the 1920s and '30s have six-inch concrete walls and a tubular shape that would easily contain semicircular cells with a control point in the middle, the New York firm says.*
6) *Many are far enough from residential areas to pass public muster, yet close enough to permit family visits.*
7) *Besides, Eggers says, grain elevators are worth preserving for aesthetic reasons – one famed architect compared them to the pyramids of Egypt.*
8) a *number of cities – including Minneapolis, Philadelphia and Houston – have vacant grain elevators, Eggers says.*
9) a *medium-sized one in Brooklyn, it says, could be altered to house up to 1,000 inmates at a lower cost than building a new prison in upstate New York.*
10) a *spokesman for the state, however, calls the idea "not effective or cost efficient."*

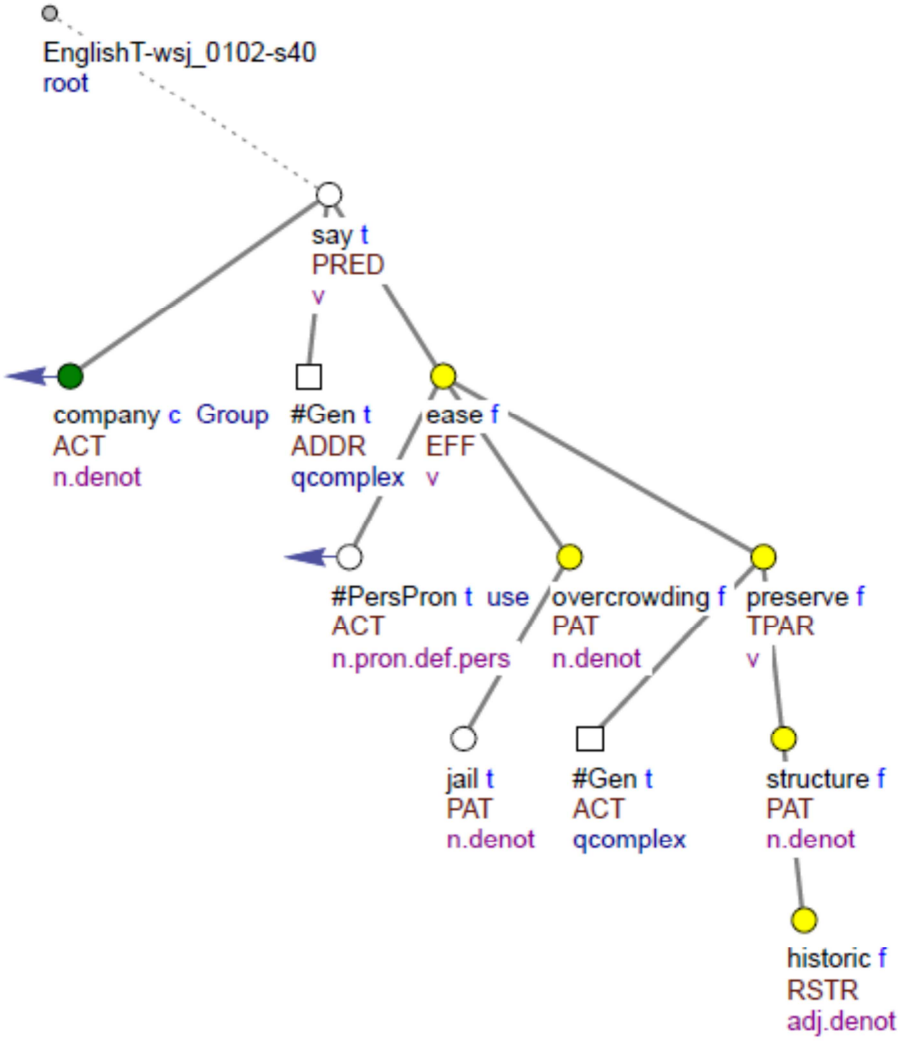
1B) ***Seed for Jail Solution Fails to Take Root*** $_{\textbf{focus}}$



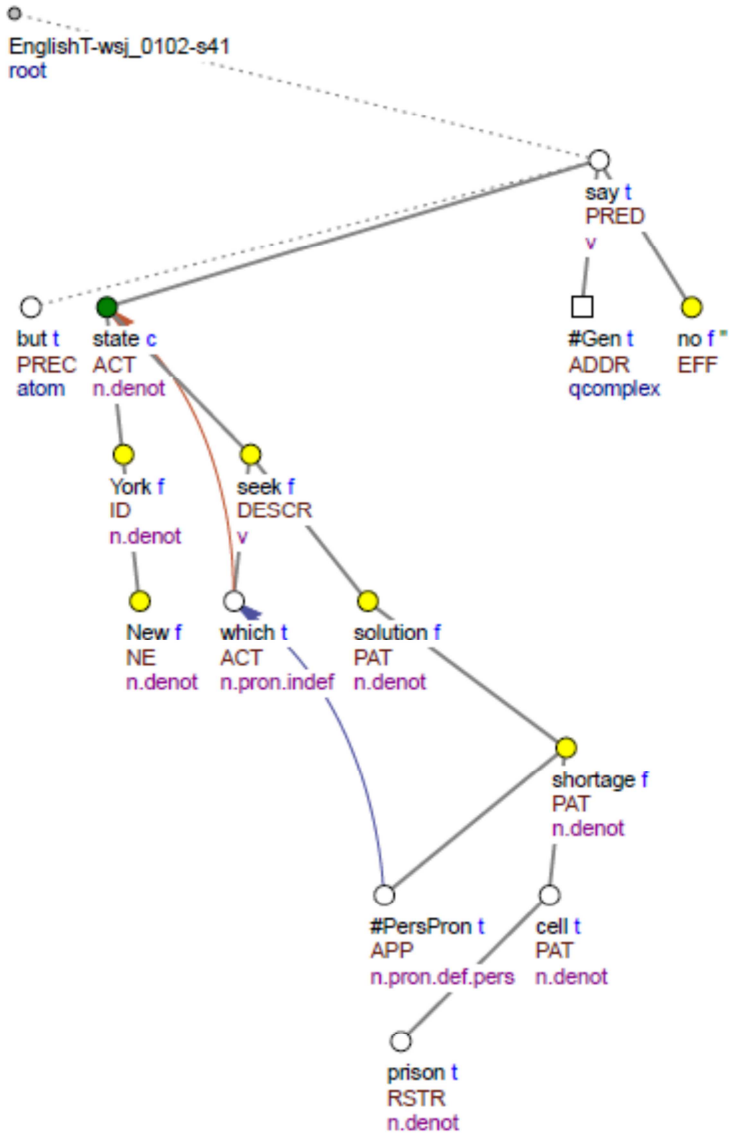
2B) *It*$_{\text{topic\_1}}$***'s a two birds with one stone deal***$_{\text{focus\_1}}$*: Eggers Group architects*$_{\text{topic\_2}}$ ***propose using grain elevators to house prisoners***$_{\text{focus\_2}}$**.**



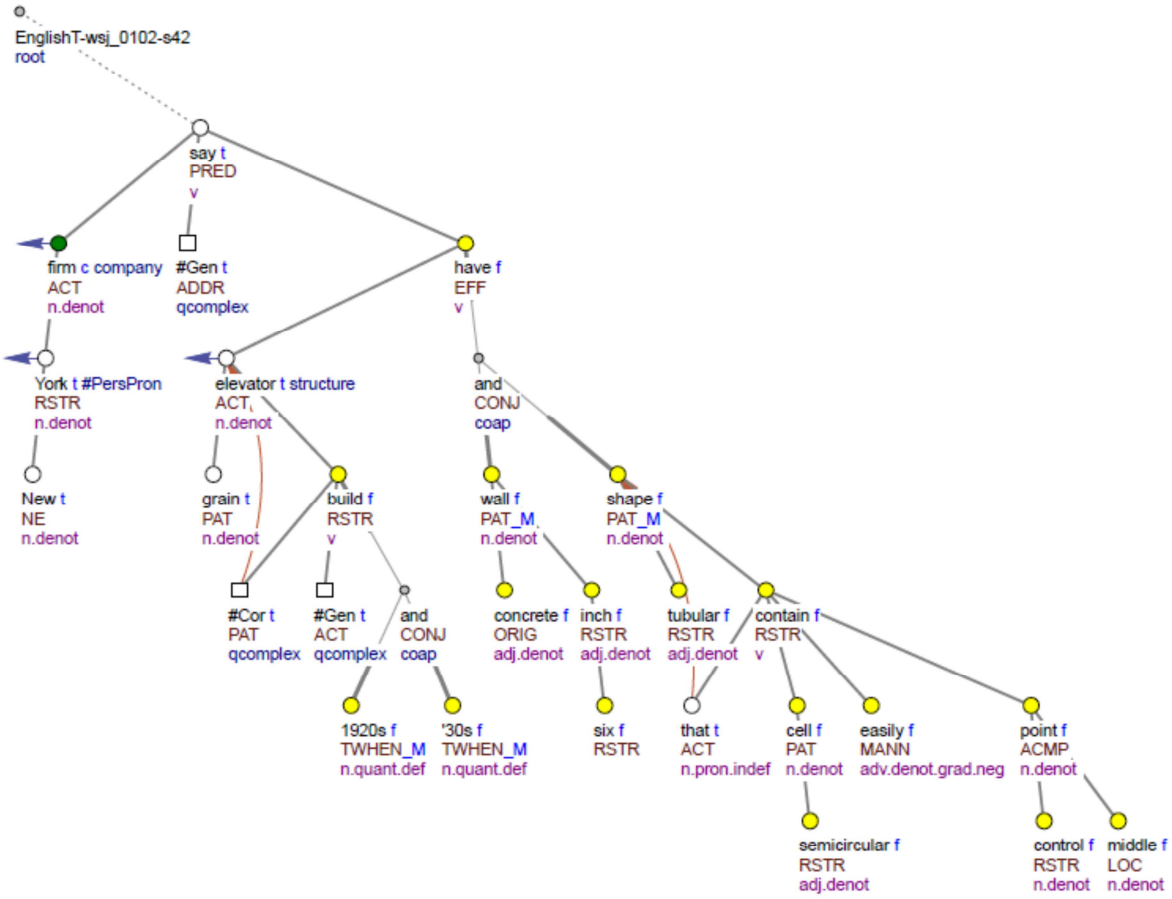
3B) ***It would ease jail overcrowding while preserving historic structures,***$_{\textbf{focus}}$ *the company says*$_{\textbf{topic}}$.



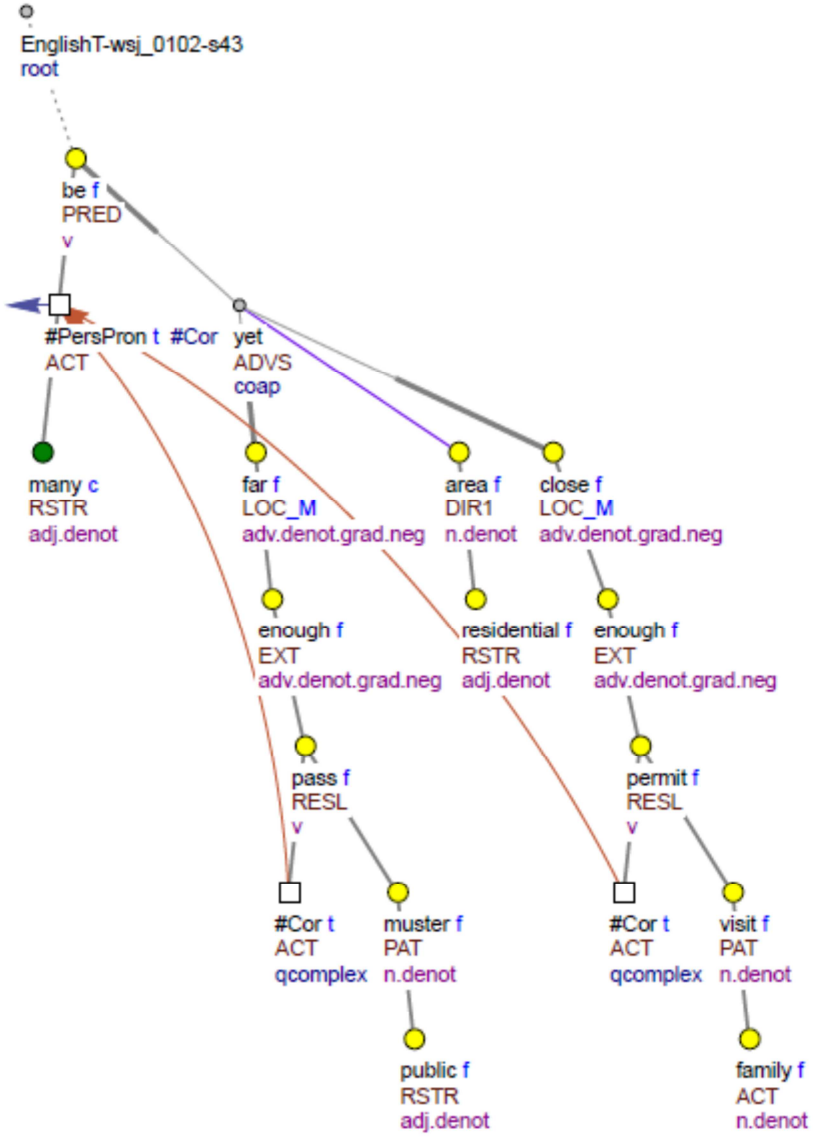
4B) *But New York state, which is seeking solutions to its prison cell shortage, says*$_{\textbf{topic}}$ ***"no."***$_{\textbf{focus}}$



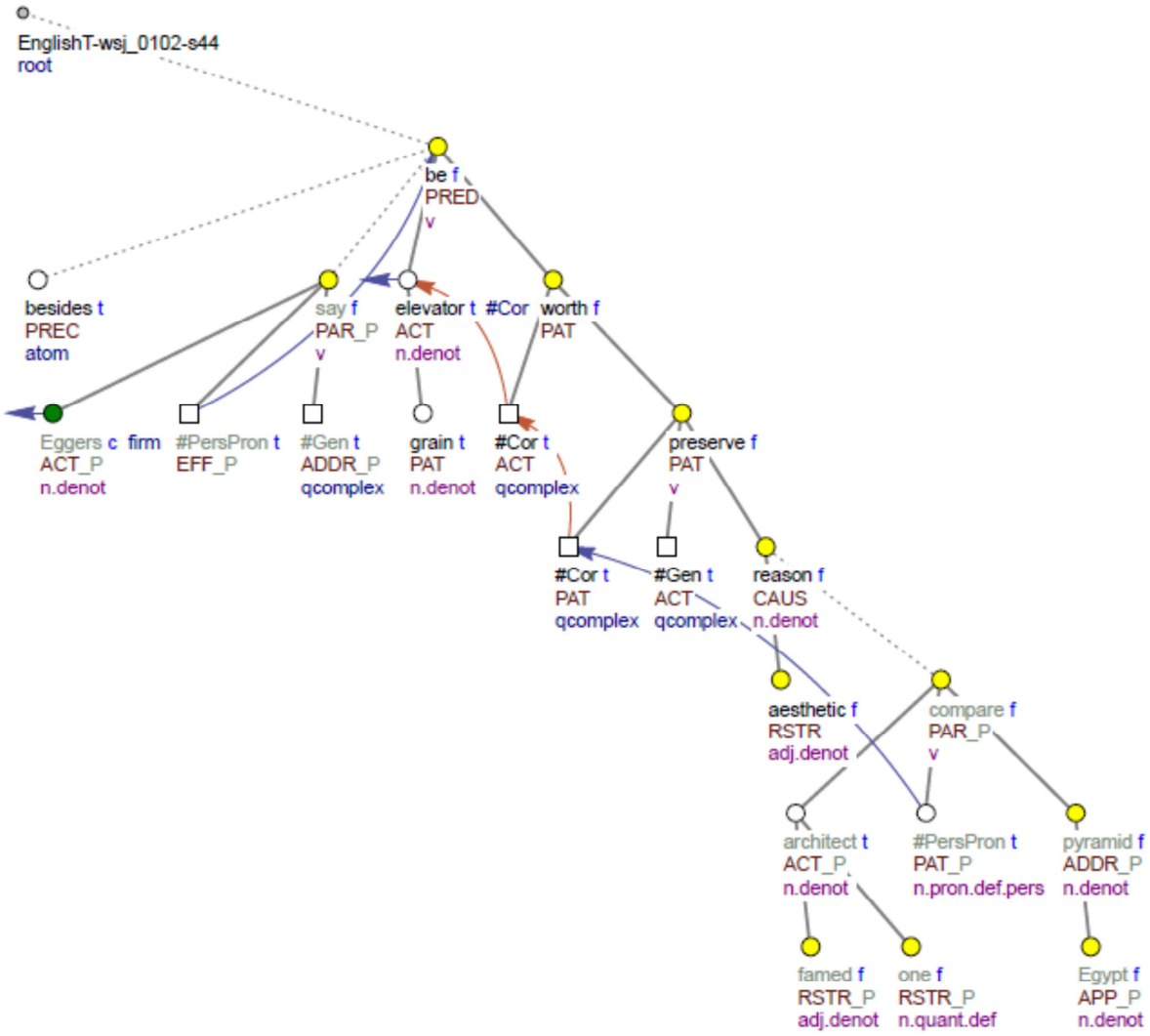
5B) ***Grain elevators built in the 1920s and '30s have six-inch concrete walls and a tubular shape that would easily contain semicircular cells with a control point in the middle,*** **focus** *the New York firm says* **topic**.



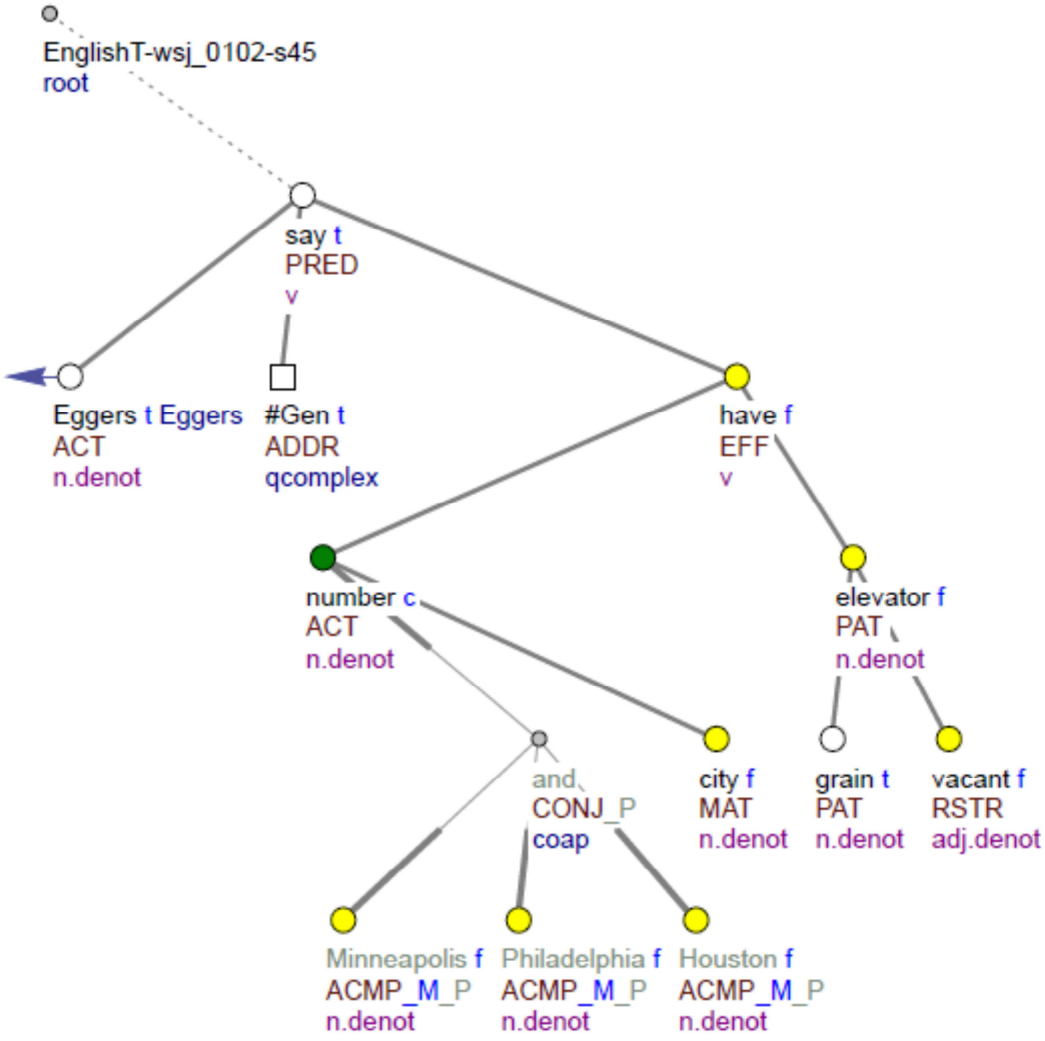
6B) *Many*$_{\textbf{topic}}$ ***are far enough from residential areas to pass public muster, yet close enough to permit family visits***$_{\textbf{focus}}$**.**



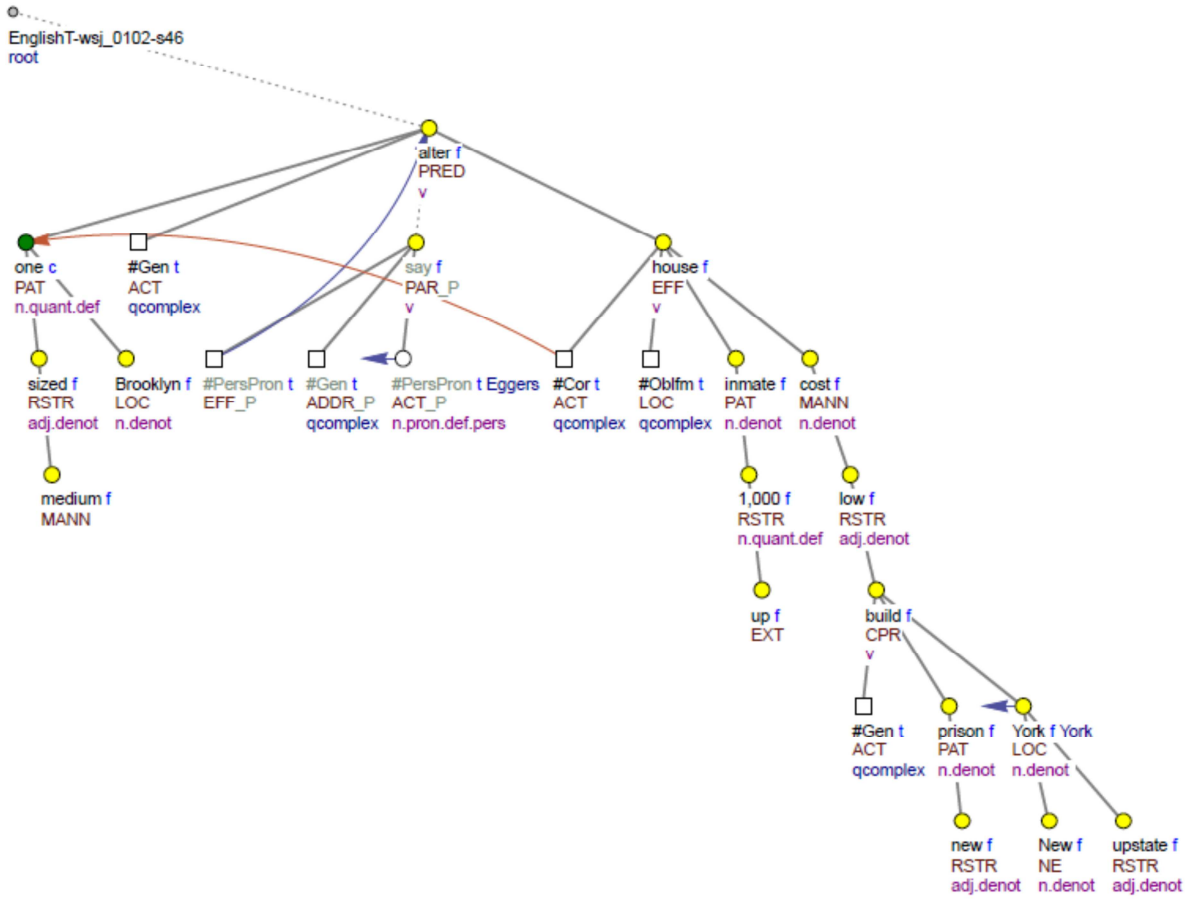
7B) *Besides*$_{\text{topic\_1}}$, *Eggers*$_{\text{topic\_2}}$ ***says***$_{\text{focus\_2}}$, *grain elevators*$_{\text{topic\_1}}$ ***are worth preserving for aesthetic reasons***$_{\text{focus\_1}}$ – *one famed architect*$_{\text{topic\_3}}$ ***compared them to the pyramids of Egypt***$_{\text{focus\_3}}$.



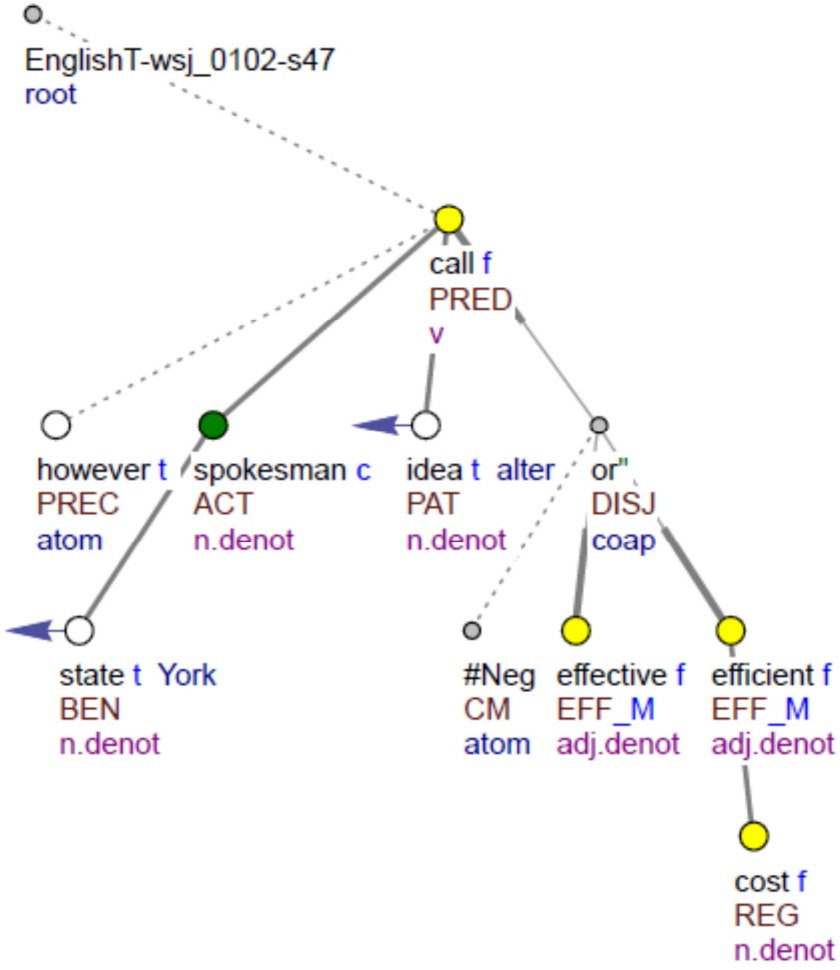
8B) a ***number of cities – including Minneapolis, Philadelphia and Houston – have vacant grain elevators,***$_{\text{focus}}$ *Eggers says*$_{\text{topic}}$.



9B) a *medium-sized one in Brooklyn*$_{\textbf{topic\_1}}$, *it*$_{\textbf{topic\_2}}$ ***says***$_{\textbf{focus\_2}}$***, could be altered to house up to 1,000 inmates at a lower cost than building a new prison in upstate New York***$_{\textbf{focus\_1}}$**.**



10B) a *spokesman for the state, however,*$_{\text{topic\_1}}$ ***calls***$_{\text{focus\_1}}$ *the idea*$_{\text{topic\_1}}$ ***"not effective or cost efficient."***$_{\text{focus\_1}}$



References and Sources

Baumann, Stefan, Caren Brinckmann, Silvia Hansen-Schirra, Geert-Jan M. Kruijff, Ivana Kruijff-Korbayová, Stella Neumann, Erich Steiner, Elke Teich and Hans Uszkoreit. 2004. The MULI Project: Annotation and Analysis of Information Structure in German and English. In Maria Teresa Lino, Maria Francisca Xavier, Fátima Ferreira, Rute Costa, Raquel Silva, Carla Pereira, Filipa Carvalho, Milene Lopes, Mónica Catarino and Sérgio Barros (eds.), *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)*. Lisbon: European Language Resources Association, 1489–1492.

Bejček, Eduard, Eva Hajičová, Jan Hajič, Pavlína Jínová, Václava Kettnerová, Veronika Kolářová, Marie Mikulová, Jiří Mírovský, Anna Nedoluzhko, Jarmila Panevová, Lucie Poláková, Magda Ševčíková, Jan Štěpánek and Šárka Zikánová. 2013. *Prague Dependency Treebank 3.0*. Data/software. Prague: Charles University in Prague, Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics. [cit. 2015_09_10]

Bohnet, Bernd, Alicia Burga and Leo Wanner. 2013. Towards the Annotation of Penn TreeBank with Information Structure. In Ruslan Mitkov and Jong C. Park (eds.), *Proceedings of the Sixth International Joint Conference on Natural Language Processing*. Nagoya: Asian Federation of Natural Language Processing, 1250–1256.

Calhoun, Sasha, Malvina Nissim, Mark Steedman and Jason Brenier. 2005. A Framework for Annotating Information Structure in Discourse. In Adam Meyers (ed.), *Proceedings of the Workshop on Frontiers in Corpus Annotations II: Pie in the Sky*. Ann Arbor: Association for Computational Linguistics, 45–52.

Chomsky, Noam. 1969. *Deep Structure, Surface Structure, and Semantic Interpretation*. Bloomington: Indiana University Linguistics Club.

Cinková, Silvie, Jan Hajič, Marie Mikulová, Lucie Mladová, Anja Nedolužko, Petr Pajas, Jarmila Panevová, Jiří Semecký, Jana Šindlerová, Josef Toman, Zdeňka Urešová and Zdeněk Žabokrtský. 2006. *Annotation of English on the tectogrammatical level. Reference book.* Prague: Charles University in Prague, Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics. [cit. 2015_09_10]

Daneš, František. 1949. Intonace otázky. Kapitola z větší práce [Intonation of a Question. Chapter from Larger Work]. *Naše řeč*, 33, 62–68.

Daneš, František. 1974. *Papers on Functional Sentence Perspective*. Prague: Academia.

Dipper, Stefanie, Michael Götze and Stavros Skopeteas. 2007. *Information Structure in Cross-Linguistic Corpora: Annotation Guidelines for Phonology, Morphology, Syntax, Semantics and Information Structure*. Potsdam: Universitätsverlag Potsdam.

Firbas, Jan. 1964. On Defining the Theme in Functional Sentence Perspective. *Travaux Linguistique de Prague*, 1(1), 267–280.

Firbas, Jan. 1971. On the Concept of Communicative Dynamism in the Theory of Functional Sentence Perspective. *Brno Studies in English*, 7(19), 12–47.

Firbas, Jan. 1976. A study in the functional sentence perspective of the English and the Slavonic interrogative sentences. *Studies in English*, 12, 9–56.

Firbas, Jan. 1992. *Functional Sentence Perspective in Written and Spoken Communication*. Cambridge, UK: Cambridge University Press.

von der Gabelentz, Georg. 1868. Ideen zu einer vergleichenden Syntax – Wort- und Satzstellung. *Zeitschrift für Völkerpsychologie und Sprachwissenschaft*, 6(1), 376–384.

Hajič, Jan, Eva Hajičová, Jarmila Panevová, Petr Sgall, Ondřej Bojar, Silvie Cinková, Eva Fučíková, Marie Mikulová, Petr Pajas, Jan Popelka, Jiří Semecký, Jana Šindlerová, Jan Štěpánek, Josef Toman, Zdeňka Urešová and Zdeněk Žabokrtský. 2012. Announcing Prague Czech-English Dependency Treebank 2.0. In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012)*, Istanbul: European Language Resources Association, 3153–3160. [cit. 2015_09_10]

Hajičová, Eva. 1972. Some Remarks on Presuppositions. *The Prague Bulletin of Mathematical Linguistics*, 17, 11–23.

Hajičová, Eva. 1973. Negation and Topic vs. Comment. *Philologica Pragensia*, 16(1), 81–93.

Hajičová, Eva. 1975. *Negace a presupozice ve významové stavbě věty* [Negation and Presupposition in Semantic Structure of Sentence]. Prague: Academia.

Hajičová, Eva. 1984. On presupposition and allegation. In *Contributions to Functional Syntax, Semantics and Language Comprehension*. Amsterdam: John Benjamins, 99–122.

Hajičová, Eva. 1995. Postavení rematizátorů v aktuálním členění věty [Position of Rhematizers in the Topic–Focus Articulation]. *Slovo a slovesnost*, 56(4), 241–251.

Hajičová, Eva. 2012. Topic–Focus Revisited (Through the Eyes of the Prague Dependency Treebank). In Jurij D. Apresjan (ed.), *Smysly, teksty i drugie zachvatyvajuščie sjužety. Sbornik statej v čest 80-letija Igorja Aleksandroviča Melčuka*. Moscow: Jazyky slavjanskoj kultury, 218–232.

Hajičová, Eva, Jiří Havelka and Kateřina Veselá. 2005. Corpus Evidence of Contextual Boundness and Focus. In Pernilla Danielsson (ed.), *Proceedings of the Corpus Linguistics Conference Series*. Birmingham: University of Birmingham, 1–9.

Hajičová, Eva, Barbara Partee and Petr Sgall. 1998. *Topic–Focus Articulation, Tripartite Structures, and Semantic Content*. Dordrecht: Kluwer Academic Publishers.

Lambrecht, Knud. 1996. *Information Structure and Sentence Form: Topic, Focus, and the Mental Representations of Discourse Referents*. Cambridge, UK: Cambridge University Press.

Mathesius, Vilém. 1907. Studie k dějinám anglického slovosledu [A Study on History of English Word Order]. *Věstník České akademie*, 16(1), 261–265.

Mathesius, Vilém. 1929. Zur Satzperspektive im modernen Englisch. *Archiv für das Studium der neueren Sprachen und Literaturen*, 155(29), 202–210.

Mathesius, Vilém. 1939. O tak zvaném aktuálním členění větném [On the So-Called Topic–Focus Articulation]. *Slovo a slovesnost*, 5(4), 171–174.

Mathesius, Vilém. 1941. Základní funkce pořádku slov v češtině [The Main Functions of the Word Order in Czech]. *Slovo a slovesnost*, 7, 169–180.

Mel'čuk, Igor A. 1981. Meaning–Text Models: A Recent Trend in Soviet Linguistics. *Annual Review of Anthropology*, 10(1), 27–62.

Mikulová, Marie, Alevtina Bémová, Jan Hajič, Eva Hajičová, Jiří Havelka, Veronika Kolářová, Lucie Kučová, Markéta Lopatková, Petr Pajas, Jarmila Panevová, Magda Razímová, Petr Sgall, Jan Štěpánek, Zdeňka

Urešová, Kateřina Veselá and Zdeněk Žabokrtský. 2005. *Anotace na tektogramatické rovině Pražského závislostního korpusu. Anotátorská příručka* [Annotation on the Tectogrammatical Layer in the Prague Dependency Treebank. Annotator´s Manual]. Technical report TR-2005-28. Prague: Charles University in Prague, Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics. [cit. 2015_08_30]. Available from <https://ufal.mff.cuni.cz/pdt2.0/doc/manuals/cz/t-layer/pdf/t-man-cz.pdf>.

Mírovský, Jiří. 2015. Inter-Annotator Agreement. In Zikánová Šárka, Hajičová Eva, Hladká Barbora, Jínová Pavlína, Mírovský Jiří, Nedoluzhko Anna, Poláková Lucie, Rysová Kateřina, Rysová Magdaléna, Václ Jan, *Discourse and Coherence. From the Sentence Structure to Relations in Text.* Prague: Charles University in Prague, Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics, 89–96.

Mírovský, Jiří, Kateřina Rysová, Magdaléna Rysová and Eva Hajičová. 2013. (Pre-)Annotation of Topic–Focus Articulation in Prague Czech-English Dependency Treebank. In *Proceedings of the Sixth International Joint Conference on Natural Language Processing*. Nagoya: Asian Federation of Natural Language Processing, 55–63.

Nedoluzhko, Anna. 2011. *Rozšířená textová koreference a asociační anafora: koncepce anotace českých dat v Pražském závislostním korpusu* [Extended Nominal Coreference and Bridging Anaphora (An Approach to Annotation of Czech Data in Prague Dependency Treebank)]. Prague: Charles University in Prague, Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics.

Paggio, Patrizia. 2006. Annotating Information Structure in a Corpus of Spoken Danish. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk and Stelios Piperidis (eds.), *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*. Genoa: European Language Resources Association, 1606–1609.

Paul, Hermann. 1886. *Prinzipien der Sprachgeschichte*. Halle: Max Niemeyer.

Rysová, Kateřina. 2011. The Word Order of Inner Participants in Czech, Considering the Systemic Ordering of Actor and Patient. In Kim Gerdes, Eva Hajičová and Leo Wanner (eds.), *Depling 2011: Proceedings, International Conference on Dependency Linguistics*. Barcelona: Universitat Pompeu Fabra, 183–192.

Rysová, Kateřina. 2014. *O slovosledu z komunikačního pohledu* [On Word Order from the Communicative Point of View]. Prague: Charles University in Prague, Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics.

Rysová, Kateřina and Jiří Mírovský. 2014. Valency and Word Order in Czech – A Corpus Probe. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard and Joseph Mariani (eds.), *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*. Reykjavik: European Language Resources Association, 975–980.

Rysová, Kateřina, Jiří Mírovský and Eva Hajičová. 2015. On an Apparent Freedom of Czech Word Order. A Case Study. In *14th International Workshop on Treebanks and Linguistic Theories (TLT 2015)*. Warszawa: IPIPAN, 93–105.

Sgall, Petr. 1967. *Generativní popis jazyka a česká deklinace* [Generative Description of Language and Czech Declension]. Prague: Academia.

Sgall, Petr. 1975. On the Nature of Topic and Focus. In Hakan Ringbom (ed.), *Style and Text (Studies Presented to Nils Erik Enkvist)*. Stockholm: Scriptor, 409–15.

Sgall, Petr. 1979. Towards a Definition of Focus and Topic. *Prague Bulletin of Mathematical Linguistics*, 31, 3–25.

Sgall, Petr, Eva Hajičová and Eva Benešová. 1973. *Topic, Focus and Generative Semantics*. Kronberg/Taunus: Scriptor.

Sgall, Petr, Eva Hajičová and Eva Buráňová. 1980. *Aktuální členění věty v češtině* [Topic–Focus Articulation in Czech]. Prague: Academia.

Sgall, Petr, Eva Hajičová and Jarmila Panevová. 1986. *The Meaning of the Sentence in Its Semantic and Pragmatic Aspects*. Dordrecht: Reidel Publishing Company.

Stede, Manfred and Arne Neumann. 2014. Potsdam Commentary Corpus 2.0: Annotation for Discourse Research. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk and Stelios Piperidis (eds.), *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*. Reykjavik: European Language Resources Association, 925–929.

Steedman, Mark. 1991. Structure and Intonation. *Language*, 67(2), 260–296.

Škvorecký, Josef. 1986. *Dvorak in Love. A light-hearted dream*. Transl. from the Czech original *Scherzo capriccioso* by Paul Wilson. Toronto: Lester & Orpen Dennys.

Škvorecký, Josef. 1991. *Scherzo capriccioso: Veselý sen o Dvořákovi*. Prague: Odeon.

Štěpánková, Barbora. 2014. *Aktualizátory ve výstavbě textu, zejména z pohledu aktuálního členění* [Foculizing Particles in Text Structuring, especially from Topic–Focus Articulation Perspective]. Prague: Charles University in Prague, Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics.

Veselá, Kateřina, Nino Peterek and Eva Hajičová. 2003. Some Observations on Contrastive Topic in Czech Spontaneous Speech. *Prague Bulletin of Mathematical Linguistics*, 79–80, 5–22.

Wegener, Philipp. 1885. *Untersuchungen über die Grundfragen des Sprachlebens*. Amsterdam: Benjamins.

Weil, Henri. 1844. *Question de grammaire générale: de l'ordre des mots dans les langues anciennes comparées aux langues modernes (thèse française)*. Paris: Joubert. Transl. by Charles W. Super as Weil, Henri. 1887. *The Order of Words in the Ancient Languages Compared with That of the Modern Languages*. Boston: Ginn.

Weil, Henri. 1978. *The Order of Words in the Ancient Languages Compared with That of the Modern Languages*. Amsterdam/Philadelphia: Benjamins. Aldo Scaglione (ed.), edited and reprinted version of Weil (1887), see Weil (1844).

Zikánová, Šárka. 2006. What do the Data in Prague Dependency Treebank Say about Systemic Ordering in Czech? *The Prague Bulletin of Mathematical Linguistics*, 86, 39–46

Zikánová, Šárka, Miroslav Týnovský and Jiří Havelka. 2007. Identification of Topic and Focus in Czech: Evaluation of Manual Parallel Annotations. *The Prague Bulletin of Mathematical Linguistics*, 87, 61–70.

Appendix: Abbreviations of Functors in the Prague Dependency Treebanks

(see Mikulová et al., 2005)

ACMP	adjunct expressing accompaniment (in the broad sense of the word)	*tatínek s maminkou*.ACMP *(=Father with Mother)*
ACT	argument – Actor	*Otec*.ACT *pracuje. (=Father is working)*
ADDR	argument – Addressee	*Poslal dárek příteli*.ADDR *(=He sent a present to a friend)*
ADVS	paratactic structure root node – adversative relation	*Viděl, ale*.ADVS *neslyšel. (=He saw (it) but he didn't hear a thing)*
AIM	adjunct expressing purpose	*Cvičí, aby zhubla*.AIM *(=She does exercises in order to lose weight)*
APP	adnominal adjunct expressing appurtenance	*můj*.APP *hrad (=my castle)*
APPS	the root node of an appositional structure	*substantivum, neboli*.APPS *podstatné jméno (=substantive, or noun)*
ATT	atomic expression expressing the speaker's attitude	*Je to samozřejmě*.ATT *pravda. (=Of course, it is true)*
AUTH	adnominal adjunct referring to the author (of sth)	*Nezvalovy*.AUTH *verše (=Nezval's poems)*
BEN	adjunct expressing that sth is happening for the benefit (or disadvantage) of sb/sth	*Pracuje pro firmu*.BEN *(=He is working for the company)*
CAUS	adjunct expressing the cause (of sth)	*Z důvodu nemoci*.CAUS *zavřeno. (=It is closed because of illness)*
CNCS	adjunct expressing concession	*Navzdory studijním úspěchům*.CNCS *se v praxi neuplatnil. (=Despite he was successful as a student, he wasn't equally successful in practice)*
CM	conjunction modifier	*otec a také*.CM *syn (=lit. Father and as well as his son)*
COMPL	adjunct – predicative complement	*Vrátila se unavená*.COMPL *(=She returned tired)*
COND	adjunct expressing a condition (for sth else to happen)	*Když spí*.COND, *nezlobí. (=If he sleeps, he is good)*
CONFR	paratactic structure root node – confrontation	*Pavel se zlepšuje, kdežto*.CONFR *Jan dostává čtyřky. (=Pavel is getting better while Jan is getting fours/bad marks)*
CONJ	paratactic structure root node – simple coordination/conjunction	*Pavel a*.CONJ *Jan (=Pavel and Jan)*

CONTRA	paratactic structure root node – two entities are in conflict (in a match, fight etc.)	*otec versus*.CONTRA *syn (=Father vs. son)*
CONTRD	adjunct expressing confrontation	*Zatímco mzdy klesají*.CONTRD *, ceny se zvyšují. (=While wages are going down, prices are going up)*
CPHR	the nominal part of a complex predicate	*mít plán*.CPHR *(=to have a plan)*
CPR	adjunct expressing comparison	*víc než tisíc*.CPR *korun (=more than one thousand crowns)*
CRIT	adjunct expressing a criterion/measure/standard	*Seřaď slova podle abecedy*.CRIT *(=Put the words in the alphabetical order, lit. organize words according to alphabet)*
CSQ	paratactic structure root node – consequential relation	*Pracoval nezodpovědně, a*.CSQ *proto dostal výpověď. (=He wasn't responsible (in his work), and therefore, he was fired)*
DENOM	effective root node of an independent nominal clause (which is not parenthetical)	*Základní škola*.DENOM *(=Primary school)*
DIFF	adjunct expressing a difference (between two entities, states etc.)	*Je vyšší o dva centimetry*.DIFF *(=He is two centimeters taller)*
DIR1	directional adjunct – answering the question "odkud (=where from?)"	*Přijel z Prahy*.DIR1 *(=He came from Prague)*
DIR2	directional adjunct – answering the question "kudy (=which way?)"	*Jdou lesem*.DIR2 *(=They are walking through the woods)*
DIR3	directional adjunct – answering the question "kam (=where to?)"	*Přišel domů*.DIR3 *(=He came home)*
DISJ	paratactic structure root node – disjunctive relation	*Pojedu já, nebo*.DISJ *ty. (=Either I will go, or you)*
DPHR	the dependent part of an idiomatic expression	*křížem krážem*.DPHR *(=crisscross)*
EFF	argument – Effect	*Jmenovali ho předsedou*.EFF *(=They appointed him as a chairman)*
EXT	adjunct expressing extent	*V nádobě je přesně*.EXT *litr vody. (=The pot contains exactly one liter of water)*
FPHR	part of a foreign-language expression	*hotovostní*.FPHR *tok*.FPHR
GRAD	paratactic structure root node – gradation	*Běžel, ba*.GRAD *utíkal. (=He not only ran, but he ran helter-skelter)*
HER	adjunct expressing inheritance	*šátek po matce*.HER *(=lit. scarf after Mother, i.e. inherited)*
ID	the nominative of identity and	*hrad Karlštejn*.ID; *trest smrti*.ID *(= the castle*

	explicative genitive	*Karlštejn, death penalty; lit. penalty death*$_{.genitive}$*)*
INTF	atomic expression referring to the "false (expletive) subject"	*Ono*.INTF *prší. (=It is raining)*
INTT	adjunct expressing intention	*Šel nakoupit*.INTT *(=He went shopping)*
LOC	locative adjunct – answering the question "kde (=where?)"	*Pracuje v Praze*.LOC *(=She works in Praha)*
MANN	adjunct expressing the manner (of doing sth)	*Mluví hlasitě*.MANN *(=He is talking loudly)*
MAT	adnominal argument referring to the content of a container	*sklenice vody*.MAT *(=a glass of water)*
MEANS	adjunct expressing a means (of doing sth)	*Píše perem*.MEANS *(=She is writing with a pen)*
MOD	atomic expression with a modal meaning	*Pracuje asi*.MOD *na půl úvazku. (=She works probably part-time)*
OPER	paratactic structure root node referring to a mathematical operation or interval	*pět až*.OPER *deset hodin (=from five to ten hours)*
ORIG	argument – Origo	*Vyrábí nábytek ze dřeva*.ORIG *(=He makes furniture out of wood)*
PAR	effective root node of a parenthetic (verbal or nominal) clause	*Přijedu 13. prosince (pátek*.PAR *). (=I am coming on December 13th (Friday))*
PARTL	effective root node of an independent interjectional clause	*Hurá*.PARTL *, vyhráli jsme! (=Hurray, we won!)*
PAT	argument – Patient	*Vaří oběd*.PAT *(=He is cooking lunch)*
PREC	atomic expression referring to the preceding context	*A*.PREC *pak odešel. (=And then he left)*
PRED	effective root node of an independent verbal clause (which is not parenthetical)	*Pavel dal*.PRED *kytku Martině. (=Pavel gave a flower to Martina)*
REAS	paratactic structure root node – causal relation	*Dostal výpověď, neboť*.REAS *pracoval nezodpovědně. (=He was fired, since he wasn't responsible)*
REG	adjunct expressing with regard to what sth is asserted	*Vzhledem k počasí*.REG *nelze nic plánovat. (=Considering the weather, it's not possible to plan anything)*
RESL	adjunct expressing the result/effect of something	*Mluví tak potichu, že mu nerozumíme*.RESL *(=He is speaking so softly that we can't understand what he's saying)*
RESTR	adjunct expressing an exception / restriction	*Kromě tebe*.RESTR *tam byli všichni. (=Except for you, everybody was there)*
RHEM	atomic expression – rhematizer	*Jen*.RHEM *Karel odešel. (=Only Karel left)*

RSTR	adnominal adjunct modifying its governing noun	*velký*.RSTR *dům (=a big house)*
SUBS	adjunct expressing that sb/sth substitutes for sb/sth else	*Za otce*.SUBS *jednal strýc. (Instead of Father, our uncle took action)*
TFHL	temporal adjunct – answering the question "na jak dlouho? (=for how long?)"	*Přijel na měsíc*.TFHL *(=He came for a month)*
TFRWH	temporal adjunct – answering the question "ze kdy? (=from when?)"	*Přeložil jednání ze soboty*.TFRWH *na dnešek. (=He shifted the negotiations from Saturday to today)*
THL	temporal adjunct – answering the questions "jak dlouho? (=how long?)" and "za jak dlouho? (=after how long?)"	*Stihnul to za týden*.THL *(=He managed to do it in a week)*
THO	temporal adjunct – answering the questions "jak často? (=how often?)" and "kolikrát? (=how many times?)"	*Pracuju na tom každý den*.THO *(=I work on that every day)*
TOWH	temporal adjunct – answering the question "na kdy? (=to when?)"	*Přeložil jednání ze soboty na dnešek*.TOWH *(=He moved the negotiations from Saturday to today)*
TPAR	temporal adjunct – answering the questions "současně s čím? (=in parallel/simultaneously with what?)" and "během jaké doby? (=during what time?)"	*Během naší dovolené*.TWHEN *ani jednou nepršelo. (=During our holiday it didn't rain once)*
TSIN	temporal adjunct – answering the question "od kdy? (=since when?)"	*Budu pracovat od zítra*.TSIN *(=I will be working from tomorrow)*
TTILL	temporal adjunct – answering the question "do kdy? (=until when?)"	*Udělám to do pátku*.TTILL *(=I will do it before Friday)*
TWHEN	temporal adjunct – answering the question "kdy? (=when?)"	*Přijdu zítra*.TWHEN *(=I'll come tomorrow)*
VOCAT	effective root node of an independent vocative clause	*Hanko*.VOCAT *, podej mi to. (=Hanka, give it to me)*

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

ÚFAL TR-2013-53 Marie Mikulová, Eduard Bejček, Jiří Mírovský, Anna Nedoluzhko, Jarmila Panevová, Lucie Poláková, Pavel Straňák, Magda Ševčíková, Zdeněk Žabokrtský, *Úpravy a doplňky Pražského závislostního korpusu (Od PDT 2.0 k PDT 3.0)*

ÚFAL TR-2013-54 Marie Mikulová, Eduard Bejček, Jiří Mírovský, Anna Nedoluzhko, Jarmila Panevová, Lucie Poláková, Pavel Straňák, Magda Ševčíková, Zdeněk Žabokrtský, *From PDT 2.0 to PDT 3.0 (Modifications and Complements)*

ÚFAL TR-2014-55 Rudolf Rosa, *Depfix Manual*

ÚFAL TR-2014-56 Veronika Kolářová, *Valence vybraných typů deverbativních substantiv ve valenčním slovníku PDT-Vallex*

ÚFAL TR-2014-57 Anna Nedoluzhko, Eva Fučíková, Jiří Mírovský, Jiří Pergler, Lenka Šíková, *Annotation of coreference in Prague Czech-English Dependency Treebank*

ÚFAL TR-2015-58 Zdeňka Urešová, Eva Fučíková, Jana Šindlerová, *CzEngVallex: Mapping Valency between Languages*

ÚFAL TR-2015-59 Kateřina Rysová, Magdaléna Rysová, Eva Hajičová, *Topic–Focus Articulation in English Texts on the Basis of Functional Generative Description*