

From: Jiri Hana (2007). Czech Clitics in Higher Order Grammar. PhD Thesis, The Ohio State University.

See http://ufal.mff.cuni.cz/~hana/bib.html#hana_diss

CHAPTER 3

BASICS OF CZECH WORD ORDER

3.1	Free word order	38
3.2	Elements with restricted word order	40
3.3	Information structure and Information Packaging	40
3.3.1	Theme – Rheme	42
3.3.2	Contrast	43
3.3.3	Theme Proper, Rheme Proper	44
3.3.4	Information packaging	45
3.3.5	Summary of the adopted Information Structure for Czech	48
3.4	Fronting	49
3.4.1	Short Constituent fronting – scrambling	49
3.4.2	Split fronting	50
3.4.3	Unbounded Dependencies	53
3.4.4	Multiple Fronted Expressions	54
3.4.5	Some restrictions on split fronting	58
3.4.6	Summary of §3.4	61

In this section, we describe and analyze the basic properties of Czech word order. First, we discuss word order in Czech in general. After that, we summarize the relation of word order and Information Structure. Then we briefly mention some elements with syntactically determined word order: prepositions and complementizers. Finally, we provide a slightly less elementary analysis of topicalization or fronting. In Chapter §5 these properties are analyzed within HOG.

A complete analysis of Czech word order phenomena is well beyond the scope of this thesis; below we present only the basic properties. We also leave out many other phenomena relevant to Czech word

order, notably the so-called wh-movement, comparatives and parentheticals. Clitics are discussed to a significantly greater depth in Chapter 4.

First a short note about examples. Appendix B discusses the presentation of data and their sources in more detail. I have tried to avoid constructing my own examples; instead I have used as many real utterances as possible – usually drawing them from various subcorpora of the Czech National Corpus (CNC) or the Prague Dependency Treebank (PDT). The Czech National Corpus includes two synchronic spoken corpora containing fiction, non-fiction and news (syn2000 abbreviated as syn0, syn2005/syn5), two spoken corpora (Oral2006, PMK), a corpus of private correspondence (KSK), news corpora (syn2006pub/syn6) and a few others. Any example that does not have a source listed is based on my own Czech native competence. Searching the corpora for evidence for a particular phenomenon is often far from trivial (c.f. e.g. Meurers 2005). While most of the corpora used are annotated with morphological and PDT also with syntactic information, the morphological annotation was mostly automatic and obviously is not perfect. The nature of current tagging technology means that errors are more common in less frequent constructions and especially in constructions involving discontinuities, both of concern in this thesis.

In the examples, information structure is marked in the English translation: Rheme is marked by the use of capitals and subscript R and contrast in theme by sans-serif and subscript C.

Finally, it is necessary to mention that there are two variants of Czech (see §A for more details): Official (Literary, Standard) Czech and Common (Colloquial) Czech. The two variants differ mainly in morphology and lexicon. One might argue that there are no native speakers of Official Czech. However, in the area of clitics, the grammatical differences are quite limited, and we discuss them where they arise. Simplifying somewhat, the spoken corpora can be seen as capturing Common Czech, and the written corpora, especially the news texts, as capturing Official Czech. The KSK corpus of private correspondence mixes features of both, sometimes even within the same sentence.

3.1 Free word order

Czech has exceptionally free word order in comparison with many other languages in general, and with English in particular. Unlike English, where word order is mostly fixed and is mainly used to express grammatical functions, word order in Czech is used to express Information Structure (see the next subsection).¹⁶ Thus for example, the four words in sentence (1) can be rearranged in all

¹⁶And probably also definiteness, as in Russian, another Slavic language (Brun 2000, 2001).

24 (=4!) possible ways. Each of the sentences has a different information structure, but all of them are grammatically correct.

- (1) Včera Petr viděl Marii.
yesterday Peter_N saw Mary_A
'Yesterday, Peter saw Mary.'

More precisely, Czech word order is very free with respect to the possibility of moving entire phrases; virtually any scrambling is possible. However, scrambling resulting in discontinuous phrases is much less common, although it is far more common than in English.¹⁷ It is mostly limited to discontinuities due to certain constructions (e.g., comparison), to clitics (see §4) and to sentences involving so-called *split fronting* (see §3.4). One of the first more systematic survey of discontinuous constructions in Czech can be found in (Uhlířová 1972). (Holan et al. 1998, 2000; Plátek et al. 2001) have suggested several measures expressing complexity of discontinuities and their reflection in the complexity of parsers.

Discontinuities in the Prague Dependency Treebank. Discontinuities in the surface syntax layer of the Prague Dependency Treebank (PDT; see §B) have been analyzed by (Hajičová et al. 2004; Zeman 2004). They report that about 23% of the 73,000 sentences contain some kind of nonprojectivity (roughly, discontinuity).¹⁸ However, many of the discontinuities are of a rather technical nature (many involve punctuation that is included as part of the syntactic structure) or are theory-dependent (e.g., they involve structures that could be analyzed as coordination of elliptical clauses, non-constituent coordination, gapping, etc., where only some of such analyses involve discontinuities). Finally note that PDT is a news corpus; the number and distribution of discontinuities in spoken and/or informal language are likely to be significantly different.

¹⁷According to (Holan et al. 2000), English allows a maximum of three discontinuity gaps in a phrase, while Czech does not impose any limit on the number of gaps. Of course, this is the competence point of view; the performance point of view is quite different – in a way parallel to, for example, relative-clause embedding which is also unlimited in competence but rather restricted in performance.

¹⁸Projectivity is defined on dependency trees. A dependency tree is a rooted ordered tree where the nodes are the words (tokens) of the sentence. In a dependency tree, the head word dominates its dependents (i.e., there is no distinction between a mother and its head daughter).

A dependency edge between a daughter d and mother m is projective iff all nodes that are between d and m in the word-order relation, are transitively dominated by m . A dependency tree is projective if all edges are projective, otherwise it is nonprojective. Various measures of degrees of non-projectivity have been explored, for example in (Havelka 2007).

3.2 Elements with restricted word order

While Information Structure (together with phrases embedding, see below) is the main factor determining word order in Czech, there are elements with fixed or highly restricted word order. In this section, we address prepositions and complementizers. Clitics, other set of elements with a restricted placement, are discussed in Chapter 4.

Prepositions. Prepositions immediately precede their NPs, as shown by *o* ‘for’ in (2a,c). There is no preposition stranding in Czech, as (2c) illustrates.

- (2) a. Požádali jsme je [o krátký rozhovor].
asked aux_{1pl} them_A for short interview
‘We asked them for a short interview.’ [syn6]
- b. O co jste je požádali?
for what aux_{1pl} them_A asked
‘For what did you ask them?’
- c. *Co jste je požádali o?
what aux_{1pl} them_A asked for
Intended: ‘What did you ask them for?’

Complementizers. Complementizers precede the clause as illustrated by *že* ‘that’ in (3):

- (3) Doufám, že [ses tam nenudila].
hope_{1sg} that aux_{2sg+refl} there not-bored
‘I hope you weren’t bored there.’ [ksk]

3.3 Information structure and Information Packaging

There is general agreement that different parts of an utterance make different informational contributions to the discourse. An utterance can be divided into two parts according to the informational contribution it makes. The new information communicated by the utterance is expressed by the part usually called *rheme* (e.g., in Firbas 1957; Steedman 2000a) or *focus* (e.g., in Sgall et al. 1986). On the other hand, the part usually called *theme* or *topic* connects *rheme* to the information already present in the common ground.¹⁹ Informally, one might say *rheme* is what the utterance says about

¹⁹Note that these terms are in some theories used differently. For example Steedman (2000a) uses *focus* to refer to contrast (both in theme and rheme). The term *topic* is sometimes used as synonymous to theme (e.g., Sgall et al.

the theme. Although there is some agreement about these basic properties of theme and rheme, anything beyond the intuitive characterization is controversial, including the exact nature of those items, their manifestation, existence of transitional items, etc. In the words of Enric Vallduví:

A number of proposals for the informational articulation of the sentence – sometimes incompatible – are found in the literature. The differences among them are significant [...]. What all the approaches have in common is the recognition that in the sentence there is some sort of informational split between a more informative part and a less informative part. Where that split is and what kind of split it is – a continuum or a dichotomy – is a matter of disagreement, but the split is nevertheless present. In our terms, it could be said that information is concentrated on a subpart of the sentence, while the remainder is licensed only as an anchoring vehicular frame for that informative part to guarantee an optimal entry into the hearer’s knowledge-store. (Vallduví 1993, p. 35).

The distinction was perhaps first suggested by Weil (1844).²⁰ Gabelentz (e.g., Gabelentz 1891) distinguished *psychological subject* (roughly theme) and *psychological predicate* (roughly rheme). In the Prague school, Information Structure has been studied extensively by Mathesius (1915, 1929, 1939), Firbas (1957, 1992, etc.; using the term Functional Sentence Perspective), Daneš (1974) and Sgall & Hajičová (Sgall et al. 1986, etc.; Topic-Focus Articulation). The Prague School’s main concern has been relation of the Information Structure to word order. The work by (Halliday 1967) is probably responsible for bringing the ideas to Generative Syntax (Jackendoff 1972; Selkirk 1984, and many others).

In this thesis, we treat Information Structure along the lines of Functional Generative Description (hence FGD; e.g., Sgall et al. 1986). The decision is primarily a pragmatic one; most of the empirical work on the Information Structure in Czech has been done in FGD or theories closely related. No other theory has been tested so extensively on Czech data. For example, in the Prague Dependency Treebank (see §B.1), about 50,000 have been manually annotated for Information Structure.²¹ The

1986), sometimes only as its contrastive part. Finally, *comment* is complimentary to *topic* in either of these meaning, so sometimes it is synonymous with rheme and sometimes refer to the part of the sentence that is not contrastive theme. See (Vallduví 1993, §3.1) for a comparison of terminology.

²⁰He calls *initial notion* or *point of departure* what we would call theme and *information being imparted* or *goal of discourse* what we would call rheme (Weil 1887 [1844], p. 30); he even suggests that what Latin expresses by word order, English expresses with emphasis (p.49 Note 7).

²¹To be precise, lexemes in the tectogrammatical layer of PDT are annotated for contextual boundness (see below). Information about theme and rheme can be derived from such annotation. On the portions that were processed by

theory differs from other theories in many important aspects; however at the level of detail needed here, it is largely compatible with many other treatments of Information Structure.

For some researchers, the terms *theme* and *rheme* refer to pragmatic or cognitive categories. While we do not dispute that such categories exist, we use the terms to refer to their syntactic counterparts (similarly as tense is related to time, aspect to *Aktionsart*, etc.).

3.3.1 Theme – Rheme

Following FGD, but also the general treatment of Information Structure in Czech syntax (e.g., Daneš et al. 1987), we partition the words in the tecto-structure of an utterance into *theme* and *rheme*.²² Theme and rheme are syntactic (tectogrammatical) categories that have cognitive/pragmatic counterparts and are expressed by various means, primarily by intonation and word order (see below). As examples, consider the four sentences from (Vallduví and Vilkuna 1998) in (4) (rhemes are marked by capitals).

- (4) a. What about pipes? In what condition are they?
The pipes are RUSTY._R
- b. What about pipes? What's wrong with them?
The pipes ARE RUSTY._R
- c. Why does the water from the tap come brown?
THE PIPES ARE RUSTY._R
- d. I have some rust remover. You have any rusty things?
THE PIPES_R are rusty.

Theme is the syntactic counterpart of being *given by the Question Under Discussion* (Roberts 1996), and rheme is the syntactic counterpart of *Information Focus* (Roberts 1998), which provides a (partial) answer to the question under discussion. In fact, FGD uses the so-called Question Test to identify focus (e.g., Sgall et al. 1986, §3.31). The difference is that in case of FGD, the questions are

several annotators, the agreement on tokens is about 76% (3 annotators, about 10,000 sentences) or about 68% (6 annotators, about 900 sentences) (Zikánová et al. 2007).

²²FGD usually uses the term *topic* for *theme* and *rheme* for *focus*. We chose theme and rheme because they seem to be less ambiguous across theories.

Also, including only words in the Information Structure is a simplification. FGD distinguishes topic/focus also for grammatical morphemes. For example, a past tense morpheme can belong to focus, while the verb itself belongs to topic, even though they are realized as a single word (at least in 3rd person).

just tests, while, in Roberts' theory, the questions under discussion are abstract entities modeling the discourse. Similarly FGD's themes and rhemes are very similar to themes and rhemes of Vallduví (e.g., Vallduví and Vilkkuna 1998) or Steedman (e.g., Steedman 2000*a*).

Although theme and rheme are related to old (familiar) and new information, they are still syntactic notions – they express how the speaker *decides* to modulate the information. Theme does not necessarily need to be old information. As Roberts (1996, p. 19) shows, theme can be used to communicate new information via presuppositions it triggers. On the other hand, rheme does not necessarily need to present new information – consider, for example, the dialog in (5) between a student and a professor. In an ideal situation, both know the answers, thus the rheme of the student's answer does not add to the common ground any information about Kepler discovering how planets work, but rather that the student knows the answer, is able to present it in an appropriate way, etc.

(5) Professor: What did Johannes Kepler discover while in Prague?

Student: He discovered TWO OF HIS PLANETARY MOTION LAWS_R.

3.3.2 Contrast

In addition to the theme-rheme distinction, it is common to distinguish between contrastive and noncontrastive elements. Consider the following dialog from (Jackendoff 1972):

(6) a. Well, what about FRED? What did HE eat?

b. FRED ate the BEANS.
 accent: fall-rise (B) fall (A)
 focus: independent dependent

(7) a. Well, what about the BEANS? Who ate THEM?

b. FRED ate the BEANS.
 accent: fall (A) fall-rise (B)
 focus: dependent independent

The fall accent (Jackendoff's *A-accent*) marks what Jackendoff calls *dependent focus*, and fall-rise accent (*B-accent*) marks *independent focus*. The difference is that (6b) cannot occur in the context of (7a) and (7b) cannot occur in the context of (6a). In Czech, the same distinction would be usually expressed by word order, with an optional fall-rise accent on the independent focus and fall accent on the dependent focus:

- (8) a. A co Fred? Co (ten) jedl?
 And what Fred? What (that) ate?
 ‘And about what Fred? What did he eat?’
- b. Fred jedl fazole.
 Fred ate beans
 ‘Fred ate the beans’
- (9) a. A co fazole? Ty jedl kdo?
 And what beans? Those ate who?
 ‘And what about the beans? Who ate them?’
- b. Fazole jedl Fred.
 beans ate Fred
 ‘Fred ate the beans’

For some researchers, the presence of contrasts implies the unit is rhematic. For example, according to Roberts (p.c.), both independent focus and dependent focus are rhematic. On the other hand, for some researchers contrast is orthogonal to the theme-rheme distinction, so parts of both theme and rheme can be contrasted. In such a view, Jackendoff’s independent focus is usually considered thematic and dependent focus rhematic. This is true, for example, for Steedman (1991, contrast is called *focus*), Valldví and Vilkuna (1998, *kontrast*), and probably also Kadmon (2001, *TOPIC-focus* = contrastive theme, *FOCUS-focus* = contrastive rheme).

FGD falls roughly into the latter group. Contrast is independent of the theme-rheme distinction, so there is a contrastive and noncontrastive theme (usually called contrastive and noncontrastive topic). The distinction between contrastive and noncontrastive rheme is not made for Czech. According to Sgall (p.c.), the reason is that while the distinction is cognitively relevant, it has no linguistic manifestation in Czech.

3.3.3 Theme Proper, Rheme Proper

According to FGD, in addition to the simple distinction of theme and rheme, there is a more fine-grained distinction of so called deep word order, a linear order expressing increasing communicative load (so-called *communicative dynamism*) of items in the utterance. Items in the theme come before items in the rheme in such ordering. Within the theme, the order of items reflects the items’ decreasing salience (see Hajičová and Vrbová 1982; Hajičová et al. 1990). The minimal item in such ordering, the most salient item, i.e., the most “thematic”-theme, is called *Theme Proper* (*Topic Proper*) and the most “rhematic”-rheme is called *Rheme Proper* (*Focus Proper*).

Theme proper and rheme proper usually correspond to individual clausal constituents, but there are exceptions. It is well known that they may correspond to a partial constituent, see for example (10) (the English translation corresponds to one of several possible interpretations, see §3.3.4.2.)

- (10) [Sportovec] je Pavel [dobrý].
 sportsman_{m.sg.nom} is Pavel good_{m.sg.nom}
 ‘Pavel is a GOOD_R sportsman_C.’ (As a sportsman, ...)

Although not discussed in the literature, in the light of multiple constituents involved in long-fronting (§3.4.3) and multiple constituents preceding clitics (§4.4.4), it seems reasonable to suggest that under certain circumstances a theme proper may consist of several constituents, or at least things that are traditionally regarded as multiple constituents.

In addition to the utterance level theme-rheme (topic-focus) dichotomy, the FGD theory of Information Structure distinguishes so-called contextually bound and contextually unbound elements (e.g., Sgall et al. 1986); they are primitive notions, but in a prototypical case, *context bound* corresponds to a contextually *given/familiar* and *context unbound* to a *new* expression. Neither of these notions is used in this thesis.

3.3.4 Information packaging

Different languages mark Information Structure in different ways. Distinct intonation and word order are the most common means in most languages, including Czech. In Czech, as a free word order language, the function of word order in expressing information structure is far more important than in languages like English.

3.3.4.1 Intonation

Until recently, relatively little attention was devoted to Czech prosody. Most of the statements about prosody are rather vague, with little or no grounding in exact phonetic experiments. The prosodic marking of rheme proper is usually called *intonation center* while contrastive theme is simply marked by *contrastive stress*, corresponding to Jackendoff’s B-accent.

According to Nino Peterek (p.c.), preliminary results suggest that contrastive theme is marked by a rising tune, but it is unclear whether it corresponds to something like L+H* or H*, or even L+H* L of the ToBI system developed for English (Silverman et al. 1992). Rheme has a falling tune; when positioned sentence finally, it is marked simply by L%. For discussion of various realizations of contrastive themes, see for example (Veselá et al. 2003).

3.3.4.2 Word order

Objective ordering Usually, sentences follow so-called *objective ordering* (Mathesius 1939, 1975). In that case, according to FGD:

1. The Intonation Center (the tune marking rheme proper) is at the end.
2. Thematic expressions precede rhematic expressions; contrastive theme tend to come before non-contrastive theme:

Theme Proper < other Theme < other Rheme < Rheme Proper

3. The order within the theme is constrained by salience, more salient items coming first (see, e.g., Hajičová and Vrbová 1982; Hajičová et al. 1990).
4. Rhematic expressions are usually ordered by a default word order, the so-called *systemic ordering* (Sgall et al. 1995).

The traditional and most straightforward way to interpret this is to see word order in Czech as the means of expressing theme and rheme. Thus Weil's statement that the "syntactic march is not the march of ideas" (Weil 1887 [1844], p. 21) is more true of English than of Czech.²³

There are many exceptions to this general pattern; see (Rosen 2001) for a summary. For example, word order in certain syntactic constructions is usually fixed regardless of IS (e.g., there is a strong preference for adjectives to precede their nouns); the finite verb occurs in the second position also more frequently than would be predicted by its IS function (this is probably an influence of German); as in many other languages, heaviness of constituents influences their placement; etc. Also, constituents with heterogenous IS (e.g., adjective belongs to rheme, noun belongs to theme) tend to stay continuous. However, the constituent might be split, especially if one part belongs to Theme Proper and the other to Rheme Proper. This is discussed in more detail in the following section.

²³One might say that in English, word order is relatively fixed and prosody is relatively free, while in Czech it is just the opposite. However, it is also possible to see the situation from a different perspective, along the lines suggested by (Roberts 1998, p. 146). In that view, word order in Czech would not express Information-Structure per se, but instead is only responsible for placing the rheme into the position where the Intonation Center can be realized. In our view, the problem with such a view is that (1) the IC can be under certain circumstances placed sentence non-finally (see below) and (2) the ordering within the theme by item salience would need to be considered a different phenomenon.

Subjective ordering In addition to the general objective ordering principle, there is a so-called *subjective ordering* (Mathesius 1939, 1975).²⁴ In this ordering, the Rheme Proper is placed at the beginning:

(11) Rheme Proper < Theme Proper < other Theme < other Rheme

Subjective ordering is usually used in excited speech; it is also quite common in newspapers, especially in titles (it probably adds some flavor of speed, urgency, etc.).

In addition to this simple case, there are also intermediate orders where a bigger portion or even the whole rheme is placed sentence initially. According to L. Uhlířová (p.c.) there is no systematic study on subjective ordering. We are therefore forced to leave this for future research and assume only the simplest possibility when only Rheme Proper occurs clause initially.

3.3.4.3 Analysis of Information Packaging

In the following sections and chapters we will the following reflection of Information Structure in word order.

Sentences having two parts:

1. The first part contains the theme proper (if there is any) in objective ordering and the rheme proper in subjective ordering. We will call such expression a *fronted* expression and analyze it in more detail in the following section.
2. Following the fronted expression is the rest of the sentence and it is ordered according to the increasing communicative dynamism:

Theme Proper < other Theme < other Rheme < Rheme Proper

Note that not all items must be present in this part of the sentence. A particular element may not present at all (only Rheme Proper is obligatory) or it could have been fronted.

²⁴In Weil (1887 [1844], pp. 43–47) the term *the pathetic order* refers to a similar phenomenon in Greek:

When the imagination is vividly impressed, or when the sensibilities of the soul are deeply stirred, the speakers enters into the matter of the discourse at the goal, and we do not become aware, till afterward, of the successive steps by which he could have entered had his mind been in a more tranquil state. (Weil 1887 [1844], p. 45)

This gives us the following two orders:

	fronted	rest of the sentence
objective:	Theme Proper	other Theme < other Rheme < Rheme Proper
subjective:	Rheme Proper	Theme Proper < other Theme < other Rheme

We assume that there are sentences without a fronted element. For example, the response in (12) is a rheme-only sentence in objective ordering. The first constituent is neither theme proper neither rheme proper, and we assume that it was not fronted.

(12) *Context: Proč máš takovou radost? – Why are you so happy?*

Martin odjel do Francie.
Martin went to France.
'Martin went to France.'

3.3.5 Summary of the adopted Information Structure for Czech

In the following, we assume the following basic view of Information Structure and Information Packaging in Czech. It is clear that more research is needed in this area.

1. Nature:

- (a) Every sentence is partitioned into theme and rheme. The rheme must be nonempty.
- (b) The most thematic/salient part of the theme is theme proper, the most rhematic part of the rheme is rheme proper.
- (c) Every item in the theme is either contrastive or noncontrastive.
- (d) Contrast is not linguistically distinguished for rheme (rheme proper might but need not express contrast).

2. Realization:

- (a) The word order reflects the IS of an utterance, either by objective ordering or subjective ordering. If there is a contrast in the theme, it tends to be on the theme proper.
- (b) The objective and subjective ordering differ in the nature of their initial (fronted, see next section) element: in the objective ordering it is the theme proper (if there is any), while in the subjective it is the rheme proper.

- (c) The rest of the sentence is ordered according to the following order:

Theme Proper < other Theme < other Rheme < Rheme Proper

- (d) Constituents with heterogeneous IS tend to stay continuous; however there are exceptions. For example, as discussed in the next section, even partial constituent can under certain circumstances undergo fronting.
- (e) Prosodically, the rheme proper is marked by the so-called Intonation Center. The contrastive theme is marked by a falling-rising tone, which is optional if the contrastive theme is sentence initial.
- (f) Some expressions (e.g., complementizers or clitics) are not ordered by IS

3.4 Fronting

In this section, we will explore the basic properties of a phenomenon usually called fronting or topicalization. We avoid the term topicalization, because this suggests the construction marks an expression as a topic (whether that means theme or only contrastive theme); which is true only in objective ordering. In subjective ordering, the fronted expression is rhematic.

In comparison with English or German, many aspects of Czech fronting are rather understudied. This applies mostly to so-called long fronting (where the expression occurs in a higher clause) and split fronting (where only part of a clausal constituent is fronted). Given the complexity and diversity of constraints on split and long fronting in other languages, it is unlikely that Czech would be significantly simpler in this area, yet these phenomena have been little discussed for Czech.

3.4.1 Short Constituent fronting – scrambling

As discussed in §3.3 above, theme proper (contrastive or not) and, in subjective ordering, rheme proper tend to occur sentence initially. For clausal constituents, this tendency is close to a strict rule. We analyze their presence in initial position, e.g., *housky* ‘rolls’ in (13) as simply a result of ordering the clausal constituents.

- (13) a. *Objective ordering:*

Context: Kdo koupil housky? – Who bought the rolls?

Housky koupil Martin.

rolls bought Martin.

‘MARTIN_R bought the rolls.’

b. *Subjective ordering:*

Context: Co koupil Martin? – What did Martin buy?

Housky koupil Martin.

rolls bought Martin.

‘Martin bought THE ROLLS_R.’

3.4.2 Split fronting

The situation when the theme or rheme proper correspond to only a part of a clausal constituent is more complex. We can distinguish two cases:

1. The whole constituent occurs in the position appropriate for the IS function of its head and the distinct IS of the subexpression is marked only by intonation. This possibility seems to be always available and it is not analyzed here.

(14) *Question: A co teda koupil makovýho a co kmínovýho? – And what did he buy with poppy-seeds and what with caraway?*

Martin koupil [makový HOUSKY_R] a kmínový ROHLÍKY_R.
Martin bought poppy-seed_{adj.pl.acc} rolls_{pl.acc} and caraway_{adj.pl.acc} bread-sticks_{pl.acc}

‘Martin bought poppy-seed ROLLS_R and caraway BREAD-STICKS_R.’

(As for poppy-seed things, Martin bought rolls and as for caraway things, he bought bread-sticks.)

2. The part of the constituent belonging to the theme proper or rheme proper is fronted, resulting in a discontinuity. This possibility is available only in certain circumstances, which are the topic of this section.

(15) *Question: A co teda koupil makovýho a co kmínovýho? – And what did he buy with poppy-seeds and what with caraway? (the same as in (14))*

[Makový] Martin koupil [– housky] a kmínový rohlíky.
poppy-seed_{adj.pl.acc} Martin bought rolls_{pl.acc} and caraway_{adj.pl.acc} bread-sticks_{pl.acc}

‘Martin bought poppy-seed ROLLS_R and caraway BREAD-STICKS_R.’

(As for poppy-seed things, Martin bought rolls and as for caraway things, he bought bread-sticks.)

The examples below show fronted partial expressions of various categories:²⁵

(16) Split NPs

a. AP from NP

[Makový] koupil [– housky].
 Poppy-seed_{adj.pl.acc} bought rolls_{pl.acc}

‘He bought poppy-seed rolls.’ (As for poppy-seed things, he bought rolls.)

b. N from NP

[Housky] koupil [makový –].
 rolls_{pl.acc} bought poppy-seed_{adj.pl.acc}

‘He bought poppy-seed rolls.’ (As for rolls, he bought poppy-seed ones)

c. PP from NP

[O syntaxi] jsem si půjčil [knihu –].
 about_{loc} syntax_{f.sg.loc} aux_{1sg} refl_D borrowed book

‘I have borrowed a book about syntax.’

[after De Kuthy 2002 (1)]

d. Possessive Adj from NP

[Dvořákovu] snesu [– operu], ale symfonii ani náhodou.
 Dvořák’s_{f.sg.acc} can-bear_{1sg} opera_{f.sg.acc}, but symphony not-even by-accident

‘I can bear Dvořák’s opera, but never his symphony.’

(17) Split predicative NPs

a. N from predicative NP

[Práce] to byla [galejnická –].
 job_{pl.acc} it was galley-like

‘It was a very hard job.’

[Uhlířová 1972 p. 174]

b. A from predicative NP

[Dobrý] je Pavel [– sportovec].
 good_{m.sg.nom} is Pavel sportsman_{m.sg.nom}

‘Pavel is a good sportsman.’ (As for good ...)

²⁵To support the orientation of a nonnative reader, the examples contain the symbol – in place where the fronted expression would be if it weren’t fronted (i.e., if it had the same IS function as the non-fronted part of the constituent). This is for expository reasons only; it is not meant to suggest that the analysis of the data should include the notion of a trace. Also, it shows only the phrase the fronted expression syntactically belongs to, not the exact position it would occur in if it weren’t fronted, which because of scrambling is not clear. The – is placed in an unmarked position.

(18) Split AP

a. Adj out of AP

[Hrdý] je [_ na své děti.]
proud is [_ on self children]

‘He is proud of his children.’

[after De Kuthy and Meurers 2001 (1c)]

b. PP out of AP

[Na své děti] je [hrdý _ .]
on self children is [proud _]

‘He is proud of his children.’

Verbal attribute. In the traditional Czech syntax, sentences involving (seemingly) split phrases are sometimes analyzed by means of a so-called complement²⁶ or verbal attribute. Informally, in this view, split NPs are analyzed as two sister phrases – an NP and a verbal attribute. The adjective agrees with the noun in the NP in the usual way. According to this analysis, the attribute relates semantically both to the verb and to the NP at the same time. Supposedly, it relates less to the NP than a normal adjective and less to the verb than a normal adjunct.

This view is roughly analogous to the reanalysis approach to similar phenomena in English or German (see De Kuthy and Meurers 2001 and references cited there). However, for Czech, this analysis has never been formally spelled out, especially its relation to semantics. Even informal analyses are rather limited (Svoboda 1969; Úličný 1969, 1970). There is little agreement in this area: some authors (Karlík et al. 1996) analyze all discontinuities with adjectives as verbal attributes, some (e.g., Daneš et al. 1987, p. 168) reject the notion completely, while others (Hajič et al. 1999; Uhlířová 1972) differ in the place of putting the boundary between the two cases. Unfortunately, the argument for or against never exceeds a few paragraphs.

Examples like (19), where the noun *hrušku* ‘pear_{f.sg.acc}’ might be replaced by a pronoun, suggest that analysis involving verbal attributes might be a better option than assuming discontinuous constituents. Because most analyses would assume that in (19b) *velkou* ‘big_{f.sg.acc}’ is not an attribute of the pronoun *ji* ‘her_{acc}’, it seems natural to assume that analogously, in (19b), it is not an attribute of the noun *hrušku* ‘pear_{f.sg.acc}’.²⁷

²⁶This term is not directly related to complements in phrase structure grammars. In this sense, a complement complements the verb in addition to its subject, objects and adjuncts. In addition to split fronting, complements are used to analyze control verbs and predicatives.

²⁷Although Jarmila Panevová (p.c.) suggests analyzing (19b) as replacement of the thematic noun *hruška* by a pronoun in the surface syntax layer of (a variant of) Functional Generative Description (Sgall et al. 1986).

- (19) a. Hrušku dal Martin Petrovi velkou.
 pear_{f.sg.acc} gave Martin Petr_{dat} big_{f.sg.acc}
 ‘Martin gave a BIG_R pear_C to Petr.’
- b. Martin ji dal Petrovi velkou.
 Martin her_{acc} gave Petr_{dat} big_{f.sg.acc}
 ‘Martin gave a BIG_R one to Petr.’

However, certain other cases suggest that an analysis involving discontinuous constituents is more plausible. For example, it seems more natural to analyze *o irským* ‘about Irish_{n.sg.loc}’ and *pivu* ‘beer_{n.sg.loc}’ in (20) as two parts of a split PP. Locative is strictly prepositional, thus analysis involving two continuous clausal constituents would require to treat the preposition-less *pivu* ‘beer_{n.sg.loc}’ as an exception.

- (20) *Context: Australský víno je dobrý. A co říkáš irským_C?*
 ‘Australian wine is good, and what do you think about Irish_C wine?’
- [O irským] jsem slyšel jen [– pivu].
 about Irish_{n.sg.loc} aux_{1sg} heard only beer_{n.sg.loc}
 ‘I have heard only about Irish_C beer.’

Note, however, that some speakers allow repeating the preposition, which would be an argument for a reanalysis view:

- (21) [O irským] jsem slyšel jen [o pivu].
 about Irish_{n.sg.loc} aux_{1sg} heard only about beer_{n.sg.loc}
 ‘I have heard only about Irish_C beer.’

Such constructions are however clearly impossible in my idiolect and my informants are split. For example, Jarmila Panevová (p.c.) judges them as better than those without the second preposition.

In the following, we assume the phrases are indeed discontinuous. The actual choice is not important for our purpose – we need *some* analysis of split-fronting so that we can analyze placement of clitics in the next chapter. Whether clitics follow the first part of a split constituent or a full reanalyzed constituent has the same consequences.

3.4.3 Unbounded Dependencies

As in English, the dependency between the fronted expression and its head (or the trace) can cross clausal boundaries. Unlike in the case of English (see e.g., Levine and Hukari 2006), this is a rather

understudied area of Czech, and we are not aware of any in-depth study of the phenomenon. Brief analyses of the phenomenon can be found in (Štícha 1996) and (Petkevič 1998).

- (22) a. [Makový] říkal Martin, že koupil [_ housky].
 Poppy-seed_{adj} said Martin that bought rolls.
 ‘Martin said he bought poppy-seed rolls.’
- b. [Makový] říkal Martin, že si myslí, že Petr koupil [_ housky].
 Poppy-seed_{adj} said Martin that refl_D thinks that Petr bought rolls.
 ‘Martin said he thinks that Petr bought poppy-seed rolls.’
- c. [Pivo] jsem přece hlásil, že podávají jenom [lahvové _].
 beer aux_{1sg} emph announced that serve_{spl} only bottled
 ‘I did announce that they serve beer only in bottles.’ [Rosen 1994 (37b)]

Such unbounded dependencies are also for non-split constituents:

- (23) a. [Housky] jsem si myslel, že říkal Petr, že koupil _ Martin.
 rolls aux_{1sg} refl_D thought that said Peter that bought Martin
 ‘The rolls, I thought Peter said Martin had bought.’
- b. [Toho kluka] si myslím, že jsem včera viděl _.
 That boy refl_D think that aux_{1sg} yesterday saw
 ‘That boy, I think I saw yesterday.’ [Petkevič (176)]
- c. [Zítřa] předpokládáme, že _ tlaková výše postoupí k jihu.
 tomorrow suppose_{1pt} that pressure height moves to south
 ‘Tomorrow, we suppose the pressure height will move to the south’[(Štícha 1996, p. 30) & Uhlířová]

3.4.4 Multiple Fronted Expressions

The theory of Information Structure in FGD implies that the theme and the rheme proper consist of a single (possibly partial) constituent.²⁸ However, examples of long fronting in (24) show that fronting of multiple constituents is possible. We are not aware of any analysis of multiple fronting in Czech, but Avgustinova and Oliva (1995) discuss a special case of this phenomenon: a clitic clusters preceded by multiple constituents. Generalizing and extending their data, we can conclude that multiple constituents can be fronted when all are contrasted, express a path (from – through – to), or are spatio-temporal stage adverbials.

²⁸As a dependency theory, FGD does not use the notion of constituents directly; here we mean a subtree of a node in a dependency tree.

(24) a. All contrasted:

[Petra do Francie] říkal Pavel, že si myslí, že Martin pošle hned.
Petr_A to France said Pavel that refl_D thinks that Martin will-send immediately
'Pavel said he thinks Martin would send Petr to France_C immediately.'

b. Path:

[Z Paříže na Remeš] si myslím, že říkal, že se stopuje blbě.
from Paris direction Reims refl_D think_{1sg} that said_{3sg} that hitch-hike badly
'I think he said that hitching from Paris in the direction of Reims does not go well.'

c. Period:

[Od pátku do neděle] očekáváme, že bude pršet.
from Friday till Sunday await_{1pl} that will rain.
'We expect that it will be raining from Friday till Sunday.'

d. Stage:

[Zítra ve vyšších polohách] očekáváme, že bude pršet.
tomorrow in higher altitudes await_{1pl} that will rain.
'We expect that it will be raining tomorrow in higher altitudes.'

3.4.4.1 Constituents?

The expressions participating in multiple fronting are traditionally analyzed as consisting of several constituents in Czech syntax. In fact, it is not clear how they could be analyzed differently, because Czech is traditionally analyzed in a dependency theory, which is radically endocentric (every constituent has a head) and lexicalist (there are no null heads).

The expressions, however, share some properties with single constituents. As discussed in §4.4.4, they can occur before the main clitic cluster, a place usually occupied by a single constituent. Another similarity is that they can be coordinated:

(25) a. Coordinated path, short fronting:

[[Z Varů] [do Chebu] a [z Paříže na Remeš]] se₁ mi₁ vždycky
from (Carls)bad to Cheb and from Paris in-direction Reims refl_A me_D always
stopovalo blbě.
hitchhiked badly

'I always had a hard time hitching from Carlsbad to Cheb and from Paris the direction of Reims.'

- b. Coordinated path, long fronting:

[[Z Varů] [do Chebu] a [z Paříže na Remeš]] říkal Martin, že
 from (Carls)bad to Cheb and from Paris in-direction Reims said Martin that
se₁ stopuje blbě.
 refl_A hitchhike badly

‘Martin said that it is hard to hitchhike from Carlsbad to Cheb and from Paris direction Reims.’

- c. Coordinated complements, short fronting:

[[Petra] [do Francie]] a [[Marii] [do Německa]] *bych* ještě poslal, ale Martina do
 Petr_A to France and Marie to Germany would_{1sg} still send but Martin_A to
 Maďarska ani náhodou.
 Hungary not-even by-accident

‘I could possibly send Petr_C to France_C and Marie_C to Germany_C, but never Martin_C to Hungary_C.’

- d. Coordinated complements, long fronting

[[Petra] [do Francie]] a [[Marii] [do Německa]] si myslím, že by šéf ještě
 Petr_A to France and Marie to Germany refl_D think_{1sg} that would₃ boss still
 poslal, ale ...
 sent, but ...

‘I think that the boss could possibly send Petr_C to France_C and Marie_C to Germany_C, but ...’

However, the expressions participating in multiple fronting also differ from constituents in many respects; for example it is hard to use a pronoun to refer to them.

3.4.4.2 “Internal” coordination

An interesting fact that we are not ready to provide analysis of is that not only the contrasted expressions can be coordinated as group with other contrasted expressions, but that the conjunction *a* ‘and’ can be inserted between them, as in (26). This adds a certain gradation of the contrast and is easier to accept when in a negative sentence or at least in a sentence contrasted with a negative one. For example (26a) suggests that Martin is a bad choice and together with Hungary it is even worse. Without the conjunction, the statement refers only to the whole combination (*Martin visiting Hungary*) as a bad choice, and the individual conjuncts might be possible, just not together (*Martin can go to Italy and Hungary can be visited by Eva*). A similar effect has the insertion of a pause instead of the conjunction.

- (26) a. [[Petra] a [do Francie]] *bych* ještě poslal, ale Martina a do Maďarska ani
 Petr_A and to France would_{1sg} still send but Martin_A a to Hungary not-even
 náhodou.
 by-accident
 Roughly: ‘I would send Petr_C to France_C, but never Martin_C to Hungary_C.’
- b. [[Všechny sny] a [najednou]] *se mu* určitě nesplní.
 All dreams and at-once refl_A him_D definitely not-fulfil.
 Roughly: ‘There is no way all his dreams will come true at the same time.’

The conjunction *a* or a prosodic boundary have similar consequences when inserted in a middle of a constituent. Consider (27a). The implication of the sentence is simply that I do not dare to babysit for the Nováks. However, when a pause is inserted in (27b) or the conjunction *a* in (27c), the implication is roughly along these lines: *I have a hard time with babysitting in general, and babysitting for Nováks is just something I do not dare at all.* These data suggest that even expressions that are traditionally analyzed as constituents with a single head can undergo multiple fronting.

- (27) a. [Hlídat₂ děti Novákům] *si*₁ teda netroufnu₁.
 watch_{inf} children Nováks_D refl_D so not-dare
 ‘I DO NOT DARE_R to babysit for the Nováks_C.’
- b. [Hlídat děti | Novákům] *si*₁ teda netroufnu₁.
 watch_{inf} children Nováks_D refl_D so not-dare
 ‘I DO NOT DARE_R to babysit_C for the Nováks_C.’
- c. [[Hlídat děti] a [Novákům]] *si*₁ teda netroufnu₁.
 watch_{inf} children and Nováks_D refl_D so not-dare
 ‘I DO NOT DARE_R to babysit_C for the Nováks_C.’

3.4.4.3 Constraints?

It is not clear whether any two (or more) expressions that can be fronted independently can be also fronted together. As we show in §4.4.4.2, the constraint suggested by (Avgustinova and Oliva 1995, pp. 36/37) in connection with clitics is too restrictive even for clitic placement. It is therefore, even more incorrect for fronting in general. In our opinion, the restrictions are more of a pragmatic than of a syntactic nature. Certain sentences with multiple frontings seem impossible simply because it is hard to imagine a context for them, especially if presented by themselves without sufficient context. We leave this issue for further study.

A similar phenomenon occurs in German, where the so-called *Vorfeld* has been argued to sometimes contain expressions that have been traditionally categorized as several constituents. Müller (2002,

2003, 2005) argues for analyzing them as a single constituent with an empty verbal head, which successfully constraints the Vorfeld to being interpreted as dependents of the same verbal head. However the meaning of such constructions is different in German than in Czech.

3.4.5 Some restrictions on split fronting

Czech is a free-constituent language, and therefore any clausal constituent can be fronted (with the exception of clitics; there are other constituents with restricted placement, such as determiners, but they are not clausal). However, as in other languages, there are limitations on split fronting. Below, we explore the more obvious ones.

3.4.5.1 Category Limitations.

Not every syntactic category can be fronted in a split fronting, and similarly not every category can be left behind. For example, while both relative clauses and prepositions can occur clause initially, they cannot be fronted as a result of the split fronting alone.

(28) Embedded Relative Clause

- a. Napsal jsem [knížku, která půjde dobře na odbyt].
wrote aux_{1sg} book which will-go well prep sale
'I wrote a book that will sell well.'
- b. * [Která půjde dobře na odbyt,] napsal jsem [knížku _].
which will-go well prep sale wrote aux_{1sg} book
- c. * [Která půjde dobře na odbyt,] si myslím, napsal jsem [knížku _].
which will-go well prep sale refl_D think wrote aux_{1sg} book

(29) Clausal Relative Clause

- a. [Která půjde dobře na odbyt,] jsem poznal hned.
which will-go well prep sale aux_{1sg} recognized right-away
'I recognized right away which one will sell well.'
- b. [Která půjde dobře na odbyt,] si myslím, že jsem poznal hned.
which will-go well prep sale refl_D think that aux_{1sg} recognized right-away
'I think I recognized right away which one will sell well.'

(30) Preposition (fronting NP from PP)

- a. * [Na] polož tu knihu _ stůl, ne pod (stůl).
on put the book table_A, not under table
Intended: 'Put the book on_C the table, not under_C it.'

- b. Polož tu knihu na_C stůl, ne pod_C stůl.
 put the book on table_A, not under table
 ‘Put the book on_C the table, not under_C the table.’

Fronting the demonstrative *ten* ‘the/this/that’ also does not seem possible.

- (31) a. Včera četl [ten básník] ze své knihy.
 yesterday read the poet from his book
 ‘Yesterday, the poet was reading from his book.’
- b. * [Ten] včera četl [– básník] ze své knihy.
 the yesterday read poet from his book
 ‘Yesterday, the poet was reading from his book.’
- c. [Ten básník] včera četl ze své knihy.
 the yesterday read poet from his book
 ‘Yesterday, the poet was reading from his book.’

3.4.5.2 Embedding Limitations.

Hajičová et al. (2004) claim that a contrastive expression has a strong tendency to stand in the initial position in the surface word order, no matter how deeply it is embedded in the underlying structure of the sentence. However, this does not seem to be correct. Generally only a clausal constituent can be split (in this respect, dependents of complex predicates and of prepositions act as clausal constituents). The existence of such limitations on embedding should not be really surprising; they exist in many other languages. See for example (De Kuthy 2002, p. 11) for constraints on split NPs in German.

In (32), only the whole complement of the verb or an dependent of that complement can be fronted (although stylistically this is not the best choice). Fronting of more embedded constituents as in (32d) is clearly out. It is also impossible to front the adjective *magisterských* ‘Master’s’, and or any other possible modifier of *diplomů* (e.g., *všech univerzit* ‘of all universities’.)

- (32) a. Vláda předepisuje [velikost písmen [na deskách [magisterských diplomů.]]]
 government_N regulates size letters_G on covers Master’s_G diplomas_G
 ‘The government regulates the character size on the covers of Master’s diplomas.’
- b. [Velikost písmen [na deskách [magisterských diplomů]]] vláda předepisuje.
 size letters_G on covers Master’s_G diplomas_G government_N regulates
 ‘The government regulates the character size on the covers of Master’s diplomas.’
- c. ? [Na deskách [magisterských diplomů]] vláda předepisuje [velikost písmen –]
 on covers Master’s_G diplomas_G government_N regulates size letters_G
 ‘The government regulates the character size on the covers of Master’s diplomas.’

- d. * [Magisterských diplomů] vláda předepisuje [velikost písmen [na deskách _]]
 Master's_G diplomas_G government_N regulates size letters_G on covers

Similarly, the PP in (33) is “too embedded” to be fronted. The intended meaning can be expressed by fronting the whole PP and using the intonation to put contrast on the adjective *kategoriální* ‘categorical’.

- (33) a. * [O kategoriální] jsem si půjčil [knihu [_ gramatic]].
 about_{loc} categorial_{f.sg.loc} aux_{1sg} refl_D borrowed book grammar_{f.sg.loc}
 Intended: ‘I have borrowed a book about categorial_C grammar.’
- b. * [Kategoriální] jsem si půjčil [knihu [o _ gramatic]].
 categorial_{f.sg.loc} aux_{1sg} refl_D borrowed book about_{loc} syntax_{f.sg.loc}
 Intended: ‘I have borrowed a book about categorial_C grammar.’
- c. [O kategoriální_C gramatic] jsem si půjčil [knihu _] (o závislostní
 about_{loc} categorial_{f.sg.loc} grammar_{f.sg.loc} aux_{1sg} refl_D wrote book
 článek).

‘I have borrowed a book about categorial_C grammar (about dependency grammar, I borrowed an article).’

3.4.5.3 Prepositions.

Rosen (2001, p. 195) shows that in a split PP, the preposition and attribute must be fronted together, as in (34). This is similar to the situation in Polish (Kupšć 2000, §2.4.2) and Serbo-Croatian (e.g., Penn 1999a, p. 179).²⁹

- (34) AP from PP
- a. [O [jak dotovanou] se jedná [_ soutěž]].
 about how financed_{f.sg.loc} refl_A is talked competition_{f.sg.loc}
 ‘How financed a competition is it?’ [Rosen 2001 (150b)]

²⁹Penn discusses split PP in connection with so-called 2W placement of clitics. In such placement, the clitics follow the first prosodic word of a sentence and can thus split the initial constituent. It is claimed (Halpern 1998, p. 111) that at least some 2W placement cannot be explained by independently split constituents, e.g., due to fronting. In Czech, clitics can follow a partial constituent only in cases when the constituent is split for other reasons.

Penn’s concern is thus opposite to ours. In his analysis, it is natural to ask why anything else is required to stand initially with the preposition. In our case, it is natural to ask why the preposition is required to stand initially when anything else is fronted.

- b. *Context: Australský víno je dobrý. A co říkáš irskému_C?*
 ‘Australian wine is good, and what do you think about Irish_C wine?’
 [O irským] jsem slyšel jen [– pivu].
 about Irish_{n.sg.loc} aux_{1sg} heard only beer_{n.sg.loc}
 ‘I have heard only about Irish_C beer.’
- c. * [Irským] jsem slyšel jen [o – pivu].
 about Irish_{n.sg.loc} aux_{1sg} heard only beer_{n.sg.loc}
 ‘I have heard only about Irish_C beer.’

However, the situation applies to any PP-split; the preposition must precede even a fronted noun. (Recall that – denotes the unmarked position of the fronted expressions, not traces.)

(35) P+N from PP

- a. [O pivu] jsem slyšel jen [– irským –].
 about beer_{n.sg.loc} aux_{1sg} heard only Irish_{n.sg.loc}
 ‘I have heard only about Irish beer_C.’
- b. * [Pivu] jsem slyšel jen [o irským –].
 beer_{n.sg.loc} aux_{1sg} heard only about Irish_{n.sg.loc}

It does not seem that the constraints behind the examples above could be prosodic. While certain Czech prepositions are indeed proclitics, the situation applies even to multisyllabic non-clitic preposition like *kolem* ‘around’. On the other hand, it is possible that the constraint is a generalization of an (originally?) prosodic constraint.

3.4.6 Summary of §3.4

In a simple case, theme proper and rheme proper correspond to clausal constituents; in objective ordering theme proper and in subjective ordering rheme proper are fronted – i.e., they occur clause initially, and can climb to higher clauses.

Split fronting, i.e., fronting of expressions that are not clausal constituents is also possible. When a theme/rheme proper does not correspond to a clausal constituent, the expression can be topicalized if the minimal constituent containing it is a clausal constituent. In this respect NPs of clausal PPs and dependents of complex predicates act as clausal constituents. The topicalized expression may, but need not, include a head of the clausal constituent it is part of. There are certain additional syntactic restrictions, for example, prepositions and non-clausal relative clauses cannot be topicalized.

The topicalized expression may consist of several expressions if they are all contrasted, if they are so-called stage adverbials, or if they express path or period.