Abstract

This paper is focused on description of hypotactic constructions (constructions with subordinating conjunctions) with “elaborative” meanings. In dependency-based linguistic literature, they are referred to as hypotactic coordinations or also (a subset of) false dependent clauses. The analysis makes use of syntax- and discourse-annotated corpora of Czech and English and thus offers an empirically grounded contrastive study of the phenomena.

1 Motivation and Background

One of the basic means of expressing syntactic dependency are subordinating conjunctions (henceforth subordinators). They also signal the semantic type of the dependency relation, i.e. the semantic relation holding between the dependent and the governing clause. Some of them have several semantic interpretations. In this paper, we describe those uses of subordinators that operate between two syntactically dependent but semantically independent contents. In other words, the clause they introduce is formally dependent, but semantically it expresses an elaborative (coordinating, restating, etc.) meaning. For the purposes of this paper, we call them hypotactic coordinations (see Panevová 2012).1

The analysis is anchored in the theoretical framework of Prague School of structuralism and its extension – functional generative description (FGD, Sgall et al. 1986). It was carried out on 50 thousand Czech sentences from the Prague Discourse Treebank, and on a similar amount of English data from the Wall Street Journal – Penn Discourse Treebank.

In linguistic theories of dependency, there are several ways of understanding the relation between formal and semantic principles of a sentence composition. Czech linguistic tradition usually distinguishes hypotaxis and parataxis as two basic formal principles of combining clauses to create a compound sentence. In majority, the linguistic community agrees that hypotaxis corresponds mostly to the semantic relation of determination (one clause semantically complements or enriches the other, building together one content), and parataxis corresponds to the semantic principle of coordination (connecting two semantically autonomous contents – the second clause adds some new information to the first clause) (Hrbáček 2000). There are, however, discrepancies between these forms and their functions.

Such a phenomenon (correspondence between hypotaxis–determination, parataxis–coordination and also their discrepancies) was described earlier in structuralist works (Karcevskij 1929) and later in FGD for morphological and also syntactical level of linguistic analysis as asymmetric dualism of forms and functions (Panevová 1980, 2012). The phenomenon is depicted in Figure 1. The solid arrows symbolize the most common relations between the form and the meaning, the dashed arrows stand for other relations, i.e. coordination realized in the hypotactic form and determination realized in the paratactic form.

![Figure 1: A schema of the asymmetric dualism between formal and semantic relations (Panevová 2012)](image-url)

In the annotation of discourse structure, the account of semantic types2 of relations between discourse units deliberately disregards the notion

---

1 Primarily, the term hypotactic coordination was used on the level of a simple clause description, for constructions such as “mum with dad”.

2 e.g. temporality, causality, contrast
of syntactic parataxis/hypotaxis, in order to liberate the perception of discourse structure from the sentential syntax.

In this paper, we look back from the discourse structure to the sentential syntax. From this point of view, for one discourse-semantic (or cognitive) type (e.g. causality), there are several language means (forms) of expression (e.g. hypotactic and paratactic constructions on the inter-clausal level).

On the general level, we are interested in the question how discourse semantics is realized formally in the sentence, our specific question is to what degree the correspondence hypotaxis–determination and parataxis–coordination on the sentence level analysis holds also for the discourse level analysis. In other words, we want to see if e.g. causality, a basic semantic concept of connecting propositions in discourse, is a matter of hypotactic constructions or if it is rather a matter of parataxis. Jinová et al. (2011) offered an overview of intra- and inter-sentential distributions of discourse relations in the Prague Discourse Treebank. Here, we are interested in hypotactic/paratactic distributions of discourse relations\(^3\) in order to either confirm or disprove that tendencies holding for the principles of sentence composition hold also for discourse composition.

In the study presented in this paper we focus on one part of the problem stated above – subordinators with elaborative meaning. The annotation of explicit\(^4\) discourse connectives (with subordinators being a subset of them) and their discourse functions revealed some discrepancies in the perception of the sentence and discourse meanings. Subordinating connectives in constructions which we call hypotactic coordinations became one of the most visible differences between the sentence and discourse analysis in the Prague approach.

Only thanks to the more cognitive-based\(^5\) discourse annotation against the background of the dependency-based syntactical tagging of the same data, we were first able to identify these constructions and study them empirically. As constructions with subordinating conjunctions,

---

\(^3\) those realized within a compound sentence

\(^4\) present on the surface

\(^5\) or content-based, according to Panevová (2010)
they are tagged – accordingly to their form on the level of language meaning – as constructions with dependent clauses\(^6\). As discourse connectives, these subordinators are tagged in accordance with the elaborative meaning they express (the level of cognitive content), see Example (1) and Figure 2. The tree diagram shows the syntactic dependency of the clause introduced by *když* (*when*), and at the same time, the discourse tag “spec” for *specification*, which is a typical elaborative category (the prototypical temporal reading is considered inappropriate or marginal in this case).

(1) Další zajímavý výzkum provedla agentura NEOBLBA, *když* zkoumala souvislost mezi barvou obrouček u brýlí a politickým přesvědčením.

Another interesting research was conducted by the NEOBLBA agency *when* it examined a connection between the color of glasses rims and political beliefs.

Apart from the analysis of Czech subordinators, we were also interested in another theoretical issue, where the empirical data of the kind we had at our disposal could lead to other findings: Do other languages demonstrate the same or similar examples of the described asymmetric dualism? We were able to look into this issue on discourse- and syntax-annotated English data of the Penn Discourse Treebank.

The paper is structured as follows: in Section 2, the two corpora used for the analysis (Prague Discourse Treebank 1.0 and Penn Discourse Treebank 2.0) are briefly introduced. Section 3 presents the distribution of types of intra-sentential discourse relations (in total and in hypotactic constructions) in the Prague Discourse Treebank. Sections 4 and 5 are devoted to the analyses of Czech and English subordinators in *hypotactic coordinations*, respectively, and we summarize our findings in the concluding Section 6.

2 Resources used (PDT and PDTB)

Prague Discourse Treebank 1.0\(^7\) (PDT, Poláková et al. 2012) is an annotation extension of the Prague Dependency Treebank 2.5\(^8\) (PDT, Bejček et al. 2012). PDT consists of approx. 50 thousand sentences of Czech newspaper texts manually annotated with discourse relations anchored by explicit (i.e. surface present) connectives. The annotation was carried out directly on the dependency trees (of the tectogrammatical (or syntactico-semantic) layer of PDT, see Sgall et al. 1986).

Penn Discourse Treebank 2.0\(^9\) (PDTB, Prasad et al. 2008) is a manually annotated treebank of English texts from the Wall Street Journal, its size is comparable to the PDT (again, approx. 50 thousand sentences). The annotation comprises both explicit and implicit discourse relations. In comparison with the Prague approach, the annotation was carried out on raw texts and only then mapped onto the syntactic trees.

Let us emphasize that all numbers and examples from PDT that we present in this paper have been measured on and taken from the *training* and *development test* parts of the data (9/10 of the treebank, approx. 44 thousand sentences). The *evaluation test* part of the data thus remains unobserved.

3 Discourse relations and hypotactic structures in PDT

In order to examine how discourse level analysis is related to the principles of sentence composition, all realizations of discourse relations within one (compound) sentence were measured over the data of PDT. Then, as our interest here lies in subordinators, the percentage of subordinate structures among all intra-sentential realizations was measured.

Distribution of individual types of discourse relations\(^10\) for subordinate structures is given in Table 1. It displays the total number of intra-sentential realizations for each semantic type of discourse relation and the percentage of subordinate structures for each type of relation. The remaining fraction consists of predominantly paratactic forms and a small number of parenthetical and other marginal structures. On the basis of these data, the following observations can be made.

First, all discourse intra-sentential relations whose syntactic parallels are treated by in Czech linguistic tradition as cases of *determination* (or content-dependency) – i.e. *purpose, condition – result of the condition, synchrony, conces-
sion, precedence – succession and reason – result (Hrbáček 2000, Daneš et al. 1987) are realized in PDiT as subordinate structures in 50% of cases or more. The least distinctive is this result for the relation of reason – result, the most distinctive for the relation of purpose (purpose was only realized in the hypotactic structure).

<table>
<thead>
<tr>
<th>type of discourse relation</th>
<th>number of occurrences within one sentence in PDiT(^\text{11})</th>
<th>hypotactic structures (in %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>purpose</td>
<td>372</td>
<td>100</td>
</tr>
<tr>
<td>condition – result of the condition</td>
<td>1,171</td>
<td>99</td>
</tr>
<tr>
<td>synchrony</td>
<td>140</td>
<td>84</td>
</tr>
<tr>
<td>concession</td>
<td>561</td>
<td>82</td>
</tr>
<tr>
<td>precedence – succession</td>
<td>495</td>
<td>68</td>
</tr>
<tr>
<td>confrontation</td>
<td>312</td>
<td>55</td>
</tr>
<tr>
<td>reason – result</td>
<td>1,428</td>
<td>51</td>
</tr>
<tr>
<td>explication</td>
<td>89</td>
<td>26</td>
</tr>
<tr>
<td>restrictive opposition</td>
<td>87</td>
<td>15</td>
</tr>
<tr>
<td>specification</td>
<td>453</td>
<td>13</td>
</tr>
<tr>
<td>exemplification</td>
<td>22</td>
<td>5</td>
</tr>
<tr>
<td>correction</td>
<td>300</td>
<td>4</td>
</tr>
<tr>
<td>opposition</td>
<td>1,235</td>
<td>4</td>
</tr>
<tr>
<td>conjunction</td>
<td>5,389</td>
<td>1</td>
</tr>
<tr>
<td>gradation</td>
<td>196</td>
<td>0.5</td>
</tr>
<tr>
<td>conjunctive alternative</td>
<td>62</td>
<td>0</td>
</tr>
<tr>
<td>disjunctive alternative</td>
<td>234</td>
<td>0</td>
</tr>
<tr>
<td>equivalence</td>
<td>38</td>
<td>0</td>
</tr>
<tr>
<td>generalization</td>
<td>8</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 1: Intra-sentential discourse relations in PDiT

Second, with the exception of confrontation (see below in this section), all relations whose syntactic parallels are treated as cases of coordination (or content parallelism) are realized as hypotactic structures much less often than the first group. For four types of relations (disjunctive and conjunctive alternative, equivalence and generalization), no hypotactic realization was found in PDiT.

These findings corroborate the hypothesis about a symmetrical relation between hypotaxis and determination on one side, and parataxis and coordination on the other. Of course, with the exception of purpose, all discourse types whose syntactic parallel is treated as determination have also paratactic realizations documented in PDiT (for reason – result, they represent almost a half of the occurrences) and the majority of discourse types whose syntactic parallel is treated as content parallelism (coordination) was documented also as a hypotactic form. These hypotactic forms are, however, in sum much less frequent than paratactic forms of relations in the first group, and thus they represent a linguistically interesting phenomenon that has not been described yet on the basis of a larger corpus material. Therefore, in Section 4, we introduce a detailed analysis of types of these structures according to their formal characteristics. We call them hypotactic coordinations further on.

Before we proceed further, two types of the PDiT discourse relations, namely confrontation and explication, require a special comment. The syntactic parallel of confrontation is treated as a type of semantic coordination (Daneš et al. 1987, p. 462) – two pieces of content are put side by side and compared (see Example (8)). On the other hand, comprehensive description of Czech syntax distinguishes paratactic and hypotactic means of its realization (ibid.) and thus reflect its special status among coordinations. Our data confirms this status – the relation of confrontation is in 55% of cases realized in hypotactic structures.

The second PDiT relation that deserves a special comment is explication – it is not a basic relation in grammatical descriptions of the Czech syntax, it was newly introduced for the discourse level analysis of PDiT. From the semantic point of view, it has a mixed nature between determination (an explanation of the content of one text unit is given in the second text unit) and content parallelism (these contents are somehow similar).\(^\text{12}\) Because of this mixed nature, we ex-

\(^{11}\) Please note again that all numbers related to PDiT refer to the training and development test parts of the data (9/10 of the treebank, 43,955 sentences).

\(^{12}\) Cf. for example the context (A), where the dependent clause expresses an explanation of the fact of a late interest by saying what “late” means.

\(\text{(A)}\) O studium svého syna jste se začal zajímat pozdě, protože oficiální termín přihlášek na střední školy a učňovská zařízení vypršel s koncem února.
clude explication from further analysis. Typical connectives for this relation in Czech are paratactic. Hypotactic realizations of this relation employ the same connectives as reason–result (26% of intra-sentential realizations in PDiT).

4 Subordinators with elaborative meanings in PDiT

According to their formal structure, we can distinguish four main types of dependent clauses expressing elaborative meanings.

4.1 Clauses with a specific unambiguous structure

First, there are certain hypotactic formal means in Czech that only express one particular coordination relation and no others. We call them specific structures. These hypotactic structures are not very frequent in our data and they were only documented for correction (11 occurrences, connective místo (toho, aby) a conjunction (2 occurrences, connective kromě toho, že (besides, lit. besides that, that)). These structures are exemplified in (2) and (3).

(2) Kromě toho, že je kompatibilní s MS-DOS, podporuje řadu programů pro postižené osoby.

Besides being compatible with MS-DOS, it supports a variety of programs for disabled people.

(Lit: Besides that, that it is compatible with MS-DOS...)

(3) Místo aby clo od poslanců vymáhali, říkali jim "jen jedete, jen jedete".

Instead of exacting the customs from the members of parliament, they told them “just go, just go”.

(Lit: Instead that the customs from the members-of-parliament they-exacted, they-told them “just go, just go”.)

These examples suggest that the specific status of these hypotactic coordinations is connected with the form of the subordinators – they are not regular conjunctions, they are composed of several elements: a preposition, (optionaly) of a relative pronoun, and a conjunction.

4.2 Relative clauses

Second, some relative clauses are known to have other functions than only to determine the noun phrase. Rather, they provide additional information which can be expressed easily in a separate sentence (often they also express temporal succession of events). There is a possibility to consider these cases relevant for discourse analysis in our sense. In the Czech description of syntax, they are mostly called false relative clauses (Daneš et al. 1987, p. 533)13, in English they are viewed as non-restrictive relative clauses (Quirk et al. 1992). As far as we know, however, there are no clear criteria for distinguishing semantically autonomous contents from determined contents in relative clauses, or, in our view, discourse-relevant cases from the other ones. Often it is impossible to say whether the relative clause only determines the noun phrase or continues the discourse. For the PDiT annotation, it was decided that only those cases are marked where there is (apart from the relative pronoun/adverb) an explicit connective present in the relative clause.

In PDiT data, we were able to document 45 cases of opposition, 24 cases of conjunction, 6 cases of restrictive opposition, 2 cases of confrontation and 1 case of correction expressed between a relative clause and its governing clause. Examples of such a realization of opposition and conjunction are given in (4) and (5).

(4) Chtěli jsme hrát nátlakový fotbal, který však ztrskotal na kvalitní obraně Benešova.

Lit: We wanted to play an aggressive football, which however failed on a high-quality defence of Benešov.

(5) Kuvajťan byl rychlou záchrannou službou převezlen do pražské Thomayerovy nemocnice, kde byl také operován.

The Kuwaiti was transported by the ambulance to the Prague Thomayer hospital, where he also underwent a surgery.

---

13 or improper relative clauses, in Czech nepravé véty vedlejší
4.3 Clauses formally equal to regular hypotactic structures

The third group of hypotactic coordinations is represented by structures formally indistinguishable from regular dependency structures. In Czech linguistic tradition, these types of dependent clauses are also often called “false” (or “improper”), as they formally signal dependency but semantically express an elaborative relation between two independent propositions. Table 2 lists all types and number of occurrences of these structures that we were able to document in the PDiT data.

Example (1) from the introductory section shows such a case of specification, which is formally expressed as a construction with a dependent temporal clause and the subordinator když (when). Example (6) below shows the same situation for confrontation, which is formally expressed as a construction with a dependent conditional clause and the subordinator jestliže (if).

(6) Jestliže v roce 1993 jich bylo 8650, což je vytížení kapacity lázní asi na 65 až 70 procent, tak v letošním roce by jich mělo být již 9745.

If there were 8,650 of them in 1993, which represents the capacity utilization of the spa to about 65 to 70 percent, then this year there should be already 9,745 of them.

Despite this rare use of subordinators in hypotactic coordinations, there is one subordinator in Czech, namely zatímco (while), which is used in hypotactic coordinations regularly and frequently. Zatímco in Czech either expresses temporal synchronicity (7) or confrontation (8) and both these uses are perceived as regular, in the sense not “false” or “improper”. The confrontational use, however, is treated as a semantic coordination, not determination, see Section 3. Therefore we claim that the connective zatímco in the confrontational use is the only regular form of expressing the asymmetric dualism in Czech on the syntax-discourse level of analysis. For comparison, it is added to Table 3 (the last row).

There are 160 occurrences of confrontation with the connective zatímco (while) documented in the PDiT data.

(7) [...] zatímco Sára ještě spí, zapřáhne osla.
while Sarah is still sleeping, he hitches up the donkey.

President Václav Havel enjoys the confidence of 75 percent of citizens, while the confidence in the Prime Minister Václav Klaus declined to 54 percent.

4.4 Connective s tím, že (along with)

One subordinator in Czech – s tím, že (roughly along with or saying also that, lit. with that, that) – is semantically vague and can serve as a connective for many relations (in PDiT data, there are eight different types of relation expressed by this connective). The type of the relation is inferable only from the context. From the point of view of hypotactic coordination in PDiT, it serves as a connective of conjunction in 14 cases and as a connective of specification in 3 cases. Examples of these contexts are given in (9) and (10), respectively.

(9) Doplňený návrh by měl obsahovat dvě varianty řešení s tím, že se k němu správní rada Českých drah znovu sejde 3. března.

The completed proposal should contain two variants of the solution and (lit. with that that) the Board of Czech Railways will reconvene to address it again on the third of March.

(10) K oběma vraždám se přiznal s tím, že chtěl získat skromný majetek důchodců a drobné peněžní částky.

He confessed to both murders saying that he wanted to get the modest possessions of the retirees and small amounts of money.

5 Subordinators with elaborative meanings in PDTB

Thanks to Penn Discourse Treebank 2.0, we have at our disposal English subordinators annotated for their discourse semantics (Prasad et al. 2008). Having left their prevalent uses out of this analysis, we were able to draw (at least partial) parallels between their “non-standard” uses in Czech and in English.

We translated into English the Czech subordinators that took part in hypotactic coordinations (e.g. když = when; jestliže = if) and searched the PDTB for similar patterns. Even though such constructions may be language-specific, and, for English, they are scarcely documented in linguistic handbooks14, some correspondence between Czech and English in our data is evident, compare Examples (11)–(13).

Subordinator: if, PDTB tag: Comparison:Contrast

(11) If Mr. Wilbur's translation is a finely ground lens through which we see the pettiness and corruption of 17th-century Paris, Mr. Falls's production is a mirror in which we see ourselves.

Subordinator: if, PDTB tag: Comparison:Contrast:Juxtaposition

(12) If the political establishment is reluctant to forgive sexual misadventures, the private sector sometimes will.


(13) In the same sentence he contradicts himself when he reports that the government still retains 40% of the total equity of the airline.

In these examples, the predominantly conditional if (11), (12) and the predominantly temporal when (13) express elaborative meanings – the same ones that we were able to document for Czech in Section 3.3, i.e. in Example (6) confrontation, in Example (1) specification.

A similar correspondence was documented for “false” relative clauses in English. Example (14) from PDTB shows a relative clause introduced by which that also contains a contrastive connective nonetheless. This co-occurrence in our view clearly signals the presence of a semantically autonomous content in the dependent clause (e.g. a coordination of the two contents rather than a determination) and so it corresponds to the Czech sentence in (4).

14 with the exception of some non-restrictive relative clauses with a coordinative meaning (Quirk et al. 1992, p. 648), or “false” infinitives of purpose, such as: I awoke to find the room flooded by sunshine. (Dušková 1992, p. 562)
Similarly as in Czech, the only subordinators that regularly signal a coordinative meaning are while and whereas: in terms of PDiT relations, they express confrontation. Of course, Czech zatímco and English while cannot be mapped 1:1 (e.g. the English while regularly expresses also causality) but for both languages they represent the most frequent subordinator with a coordinative meaning.

Having found evidence for parallels in use of subordinators in hypotactic coordinations in English and Czech, we are aware of the fact that the direction of analysis from Czech to English may have not revealed all such relevant structures in English. The existence of other types of hypotactic coordinations cannot be excluded and it is a possible topic for further linguistic research.

However, in spite of the assumed language-specificity in the repertoire of connective means and their use, our findings on a relatively small amount of data and a restricted language domain (financial journal) suggest that when subordinators deviate from their usual functions, they tend to do it in the similar way in Czech and English.

6 Conclusion

In this paper, we surveyed structures where subordinators convey coordinative meaning (hypotactic coordinations). These structures represent an irregular relation between formal and semantic principles of sentence composition, since coordinative meanings are prototypically realized in paratactic structures. On the basis of PDiT and PDTB, we described this phenomenon for Czech and we have drawn some comparisons of their use in English.

As the first step, the distribution of discourse relations as hypotactic versus paratactic structures in PDiT was measured to see to what extent the hypothesis of correspondence determination – hypotaxis, coordination – paratsix is also applicable for discourse semantics. We found that with the exception of confrontation, whose syntactic (and sentence-level semantic) counterpart is treated as coordination and which appears in our data quite regularly both as paratactic and hypotactic structures, relations whose syntactic counterpart is treated as coordination are realized as hypotactic structures rather rarely.

Further, we analyzed four types of hypotactic coordinations in Czech according to the characteristics of their respective subordinators. Some subordinators (e.g. kromě toho, že (lit. besides that that)) are specific only for coordination, a majority of them is used regularly for other relations than coordination (e.g. jestliže (if) is a regular subordinator for condition, some uses of it however express confrontation) etc.

Finally, subordinators whose “non-standard” meaning was documented for Czech (e.g. jestliže (if)) were translated and looked for also in the English data. Despite of the assumed language-specificity in connective functions, we were able to document English examples corresponding to the Czech structures.

Our findings are of course limited by the size and type of the language resources available for such a comparative study. Nevertheless, it should be highlighted that only the existence of such manually and specifically annotated corpora that gather linguistic information from different levels of language description makes it possible for the first time to carry out such a linguistic analysis.

Acknowledgments

We gratefully acknowledge support from the Grant Agency of the Czech Republic (grants P406/12/0658 and P406/2010/0875) and from the Ministry of Education, Youth and Sports in the Czech Republic (the LINDAT-Clarin project LM2010013). This research was also supported by SVV project number 267 314.

References


