Factors influencing implicit discourse relations in Czech

The explicitness and implicitness of discourse relations occur in a language in interplay with many other phenomena. In the present analysis, we concentrate on the correlation between the measure of discourse relation implicitness/explicitness, its semantics and frequency. In addition, we analyze some secondary signals that allow recognizing the meaning of the implicit discourse relations.

Research questions
The analysis was carried out on discourse relations from two semantic groups, temporal (synchrony, asynchrony) and contrastive relations (concession, confrontation). The annotation scenario is based on the modified PDTB approach.

The relation of confrontation is typically expressed with Czech equivalents of the discourse connectives whereas; within its arguments, different features are ascribed to entities which are expressed contrastively (A is big, whereas B is small).

Based on this language material, we tested the following hypotheses:

1. We expect that with relations with a lower frequency, the measure of explicitness is higher. These relations are not usual; therefore they need to be expressed explicitly with a discourse connective, in order to be understood.
2. We assume a higher measure of explicitness at relations with a marked, specific semantics than at relations with vague, unspecified semantics. Therefore we expect a higher explicitness with contrastive than temporal relations and differences between relations within these groups as well. For contrastive relations: compared to confrontation, concession contains a feature of negation, therefore it should be rather expressed than confrontation. As for temporal relations, asynchrony is more specific (and as such probably more often expressed explicitly) than synchrony which is closer to conjunction.
3. Furthermore, we analyze the implicit discourse relations in all the four groups (synchrony, asynchrony, concession, and confrontation) and look for explanation which other signals allow us to interpret the meaning of the relations.

Data
The data comes from the Enriched Discourse Annotation of Prague Discourse Treebank Subset (PDiT-EDA 1.0), containing annotation of explicit and implicit discourse relations, alternative lexicalizations, entity-based relations and other phenomena. The discourse annotation is combined with other types of annotations, such as syntax, syntactic semantics, information structure etc. The annotation was carried out on 2592 sentences (41877 tokens) coming from 100 written Czech journalistic texts. 15 text genres (as an interview, a weather forecast, a news report etc.) were chosen to cover the journalistic style representatively. The size of the corpus allows applying χ² test reliably.

Results

<table>
<thead>
<tr>
<th></th>
<th>Explicit</th>
<th>Implicit</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Intra-S</td>
<td>Inter-S</td>
<td>Total</td>
</tr>
<tr>
<td>TEMORAL</td>
<td>26</td>
<td>33</td>
<td>59</td>
</tr>
<tr>
<td>Synchrony</td>
<td>8</td>
<td>4</td>
<td>12</td>
</tr>
<tr>
<td>Asynchrony</td>
<td>18</td>
<td>29</td>
<td>47</td>
</tr>
<tr>
<td>CONTRAST</td>
<td>45</td>
<td>30</td>
<td>75</td>
</tr>
<tr>
<td>Confrontation</td>
<td>12</td>
<td>19</td>
<td>31</td>
</tr>
<tr>
<td>Concession</td>
<td>33</td>
<td>11</td>
<td>44</td>
</tr>
<tr>
<td>TOTAL</td>
<td>71</td>
<td>63</td>
<td>134</td>
</tr>
</tbody>
</table>

Table: Distribution of explicit and implicit discourse relations of synchrony, asynchrony, confrontation and concession in the PDiT-EDA 1.0 (number of occurrences)
**Implicitness and frequency of discourse relations**
The distribution of implicitness and explicitness of discourse relations significantly correlates with the frequency of their occurrence (confirmed with a $\chi^2$ test), the rarer relations are rather expressed explicitly. Thus, asynchrony and confrontation, having a higher occurrence in the corpus, are more often implicit (54%, 68%, respectively) in comparison with synchrony and concession 29%, 20%, respectively).

**Implicitness and semantics of discourse relations**
Within the analyzed data, contrastive discourse relations are not expressed explicitly significantly more often than temporal relations (56% and 44%, respectively). In this case, the idea that a strong semantic feature (contrastivity) is rather expressed was not confirmed.
Within the contrast relations, there is a significantly strong difference between confrontation and concession in the measure of explicitness, concession being very often expressed explicitly (concession 80%, confrontation 32%). We explain this as a result of a narrower semantics of concession, containing features of a condition as well as of a negation.
Although temporal relations differ in the same way, synchrony (71%) being more often expressed explicitly than asynchrony (46%), this difference is not significant due to the low occurrence of synchrony relations in the corpus as a whole (17 occurrences).
All the observations were confirmed with a $\chi^2$ test.

**What helps us to understand the semantics of implicit discourse relations?**
Certain regularities can be found in the structure of the analyzed implicit discourse relations. Temporal relations are signaled with the verbal aspect of the predicate in the first argument (imperfective for synchrony, cf. (1), perfective for asynchrony) and lexical temporal settings.

(1)
[Arg1: Vyjel se soupravou ze stanice na zelenou, rozjížděl se IMPERFECTIVE a nabíral IMPERFECTIVE rychlost.] **IMPLICIT SYNCHRONY** [Arg2: Náhle na kolejích ve tmě před sebou uviděl koncová světla jiného vlaku.]
[Arg1: He pulled his train out of the station on green, he was going and speeding up.] **IMPLICIT SYNCHRONY** [Arg2: Suddenly, he saw the tail lights of another train ahead on the tracks in the darkness.]

Contrastive relations can be deduced from stressed forms of pronomina in contrastive positions; for confrontation, specific course of the information structure is typical in which parts of the topic in both arguments are contrastive (see 2).

(2)
(Context: about exchange of goods between Russia and the Czech Republic)
[Arg1: Z ruské strany CONTRASTIVE TOPIC jde o suroviny, jako je plyn a ropa.] **IMPLICIT CONFRONTATION** [Arg2: z české CONTRASTIVE TOPIC o různé zboží.]
[Arg1: On the Russian side CONTRASTIVE TOPIC, there are raw materials, such as gas and oil.] **IMPLICIT CONFRONTATION** [Arg2: on the Czech side CONTRASTIVE TOPIC, there are various goods.]

**References**