

The meaning of the conditional mood within the tectogrammatical annotation of Prague Dependency Treebank 2.0

Magda Ševčíková

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

Abstract. The conditional form is one of the moods of Czech verbs, and it renders several meanings in contemporary Czech texts (Sect. 2). The present paper focuses on the primary function of this mood, which is to express hypothetical events (Sect. 3). In Section 4, we briefly mention how modality has been treated up to now in PDT 2.0 and some other treebanks and finally in Section 5 we propose a new way how the primary meaning of the conditional mood should be captured in the annotation scheme of the tectogrammatical layer of PDT 2.0.

1 Introduction

Verbal mood as a universal morphological category occurs in many languages of the world, however, it is structured differently in different languages (see, e.g., [1], [10]). In Czech, three moods are traditionally distinguished: indicative, imperative, and conditional. In this paper, attention is focused on the Czech conditional mood, in particular on its primary function. As the corpus data document, the conditional is primarily used to refer to events which may be generally characterized as hypothetical; more specifically, to events which are contingent on realization of other events, to events which cannot happen any more etc. We want to demonstrate that in contemporary Czech the conditional as a means expressing hypotheticality of an event is a semantically relevant means which contributes to the meaning of the whole sentence. Thus, it has to be included in the representation of the meaning of the sentence.

In Prague Dependency Treebank 2.0 (PDT 2.0 in the sequel), the linguistic meaning of the sentence is represented as a tectogrammatical tree, which consists of nodes and edges with a set of attributes (see [6]). However, neither in PDT 2.0 nor within Functional Generative Description (FGD in the sequel; c.f. [14]), on the basis of which the annotation scenario of PDT 2.0 was built, any considerable attention has been paid to the semantics of the morphological category of mood.

After an overview of the functions of the conditional mood (Section 2), the semantic relevance of the conditional in its primary function is documented in Section 3 of the present paper. In Section 4, the treatment of the conditional and other modal means in PDT 2.0 (and in FGD) as well as in some other treebanks

is described. A new formal means (called *grammateme* in PDT 2.0 and FGD) for capturing the primary function of the conditional within the tectogrammatical annotation of PDT 2.0 is proposed in Section 5. The paper concludes with some final remarks in Section 6.

2 Functions of the conditional mood

2.1 The primary function of the conditional

The morphological category of verbal mood is acknowledged to be one of the means by which modality is expressed in Czech. Modal meanings are further expressed by many other means such as modal verbs, modal particles, prosody etc. The primary function of the conditional is described under several terms in Czech linguistic literature. In this paper, we use the term *hypotheticality* to refer to the semantics of this mood. In its primary function, the conditional is opposed to the indicative by means of which events are simply asserted, presented as given. In Grammar of Czech (see [8], [2]), the semantic opposition of hypotheticality vs. assertion is called *factual modality*. The conditional mood is considered to be the marked member in this opposition.

Two types of hypothetical events are further distinguished: events which could happen (potential events) vs. events which cannot happen (irreal events).¹ Potential events are expressed by the so called present conditional as one of two subcategories of the conditional mood. The second subcategory of the conditional, the so called past conditional, expresses irreal events unambiguously. However, the present conditional is frequently (or even predominantly) used instead of the past conditional in Czech in the last decades. Such a substitution has been regarded as acceptable if the irrealty is signaled by an adverb etc. (e.g. in [5]). Nevertheless, in the analyzed corpus data, the present conditional was used as an expression of irreal events also in cases in which the irrealty could be resolved only on the basis of a very broad context or of knowledge of situation. Such an ambiguity causes substantial problems with annotation.

2.2 Secondary functions of the conditional

Besides the primary function, the conditional fulfills also other (secondary) functions in Czech. For example, in the sentences *Doporučil bych vám tu smlouvu podepsat* ‘I would recommend you to sign the contract’ or *Uvedl byste konkrétní příklad?* ‘Would you give a concrete example?’, the speaker uses the conditional to express his recommendation or appeal in a more polite way than using the indicative.

The conditional can further express the speaker’s will to do something (c.f. *Kouřil bych* ‘I would smoke’, which actually means “I want to smoke”) or his conviction whether the content will be realized or not (e.g., *Šel bych* ‘I would go’,

¹ The term ‘irreal’ is used according to [17], i.e., it refers to events which are impossible rather than to unreal events.

which can be used instead of “I will probably go” in an appropriate context). This paper concerns with the primary function of the conditional; contexts in which it is used in its secondary functions are not analyzed here.

3 The conditional mood as a means for expressing hypothetical events

3.1 Corpus material analyzed

The primary function of the conditional mood was studied on the basis of language data from two corpora: from PDT 2.0 (<http://ufal.mff.cuni.cz/pdt2.0>; see [6]) and from SYN2005 corpus (<http://ucnk.ff.cuni.cz>).

PDT 2.0 is a collection of contemporary Czech newspaper texts to which a morphological annotation and annotation at two syntactic layers was assigned, at the so called analytical layer (layer reflecting the surface shape of the sentence) and at the tectogrammatical layer (layer of the linguistic meaning of the sentence). Annotation of all three types is available for more than 49 thousand sentences (i.e., more than 830 thousand tokens); the train data set (about 80 % of the full data) was used to search for sentences to be examined.

SYN2005 is a representative corpus of contemporary written Czech which contains 100 million tokens. In comparison with PDT 2.0, only morphological annotation was assigned to the data of SYN2005.

3.2 Semantic relevance of the conditional

In order to show the semantic relevance of the conditional mood in its primary function, we proposed a substitution test. In the test, the conditional, which is supposed to be the marked member of the opposition of the factual modality, was substituted for the indicative as for the unmarked member of this opposition. The substitution test was performed in several types of sentences from PDT 2.0 and SYN2005 corpus in the following way: in a concrete sentence which involved a present or past conditional verb form, the conditional was replaced by the indicative. Then, we tested whether the resulting sentence with the indicative can be used as an expression of the hypothetical event in question and whether it is acceptable with regard to the immediate as well as broader context, to our knowledge of situation etc. (c.f. [16] for more details on the substitution test).²

The acceptability of the indicative in the examined sentences seems to be related to the type of hypotheticality expressed by the sentence, i.e., whether a potential or an irreal event was expressed. Irreal events can be usually expressed only by the past or present conditional. For instance, in sentences (1) and (2), which both render irreal events, the present and past conditional verb forms are used, respectively; the conditional cannot be substituted for indicative verb forms here due to the given contexts, cf. in the sentence (1') the indicative is

² In [7], the semantics of the past conditional was tested in a similar way.

unacceptable due to the fact that Joseph Roth is not alive any more, in (2') since they (the photos) were not reminded.

On the contrary, when expressing potentiality, the present conditional can be often replaced by the indicative; see examples (3) and (3'), (4) and (4'). This substitutability is connected, in our opinion, with the semantic closeness of the potentiality of the conditional and the future meaning of the indicative (cf. [17]).

(1) *Svatý pijan Joseph Roth **by** dnes **oslavil** rovnou stovku.* 'The saint drunkard Joseph Roth **would celebrate** his 100th birthday today.' (PDT 2.0)

(1') #*Svatý pijan Joseph Roth dnes **oslaví** rovnou stovku.* '#The saint drunkard Joseph Roth **celebrates** his 100th birthday today.'³

(2) ... *připomínat si je **by bylo bývalo** bolestivé.* '... to remind oneself of them **would have been** painful.' (SYN2005)

(2') #... *připomínat si je **bylo** bolestivé.* '#... to remind oneself of them **was** painful.'

(3) *Uhrát tu remízu **by bylo** úspěchem.* 'To draw the game **would be** a success.' (PDT 2.0)

(3') *Uhrát tu remízu **bude** úspěchem.* 'To draw the game **will be** a success.'

(4) *Podle A. Röschové (ODS) **by** pro Kozlův návrh **byla** asi polovina klubu ODS, ale osobně **by byla** spíš proti.* 'According to A. Röschová (ODS), approximately a half of the club of ODS **would vote** for Kozel's proposal but she personally **would vote** against it.' (PDT 2.0).

(4') *Podle A. Röschové (ODS) **bude** pro Kozlův návrh asi polovina klubu ODS, ale osobně **bude** spíš proti.* 'According to A. Röschová (ODS), approximately a half of the club of ODS **will vote** for Kozel's proposal but she personally **will vote** against it.'

3.3 Conditional vs. indicative in selected contexts

However, in the corpus data, several sentences expressing potential as well as irreal events occurred which violated our hypothesis of the relation of the hypotheticality type and the acceptability of the indicative. On the one hand, the indicative could be used instead of the conditional in some irreal contexts. On the other hand, the indicative was not acceptable in several sentences which expressed potential events. These two cases could be characterized as follows:

- An irreal event could be expressed both by the (present or past) conditional and the indicative, for instance, in governing clauses of the examined con-

³ The hash mark # indicates that the given sentence (with the indicative) in comparison to the original sentence (with the conditional) is unacceptable since it does not express a hypothetical event without regard to the fact that it is mostly a grammatical Czech sentence.

ditional complex sentences. In our opinion, the indicative comes into consideration in these sentences since the irreality is signaled explicitly by another means besides the verbal mood, e.g., by the conditional clause; see the sentences (5) and (5').

- Concerning the indicative in potential contexts, the loss of the hypotheticality which is connected with the examined substitution plays an indispensable role when judging the acceptability of the indicative instead of the conditional. The conditional mood cannot be replaced by the indicative in such sentences which, according to our interpretation, express potential events whose realization the speaker does not accept (cf. the example (6) vs. the sentence (6'), which cannot be accepted in the given context).

(5) ... *kdyby dnes přišel Thomas Alva Edison do české banky se žádostí o úvěr na výrobu jakýchsi "žárovek, které změní svět", **skončil by** s nepořízenou ...* '... if Thomas Alva Edison would come to a Czech bank with an appeal for a credit for production of a kind of "light bulbs which change the world" in these days, he **would go away** empty-handed ...' (PDT 2.0)

(5') ... *kdyby dnes přišel Thomas Alva Edison do české banky ..., **skončí** s nepořízenou ...* '... if Thomas Alva Edison would come to a Czech bank ... in these days, he **goes away** empty-handed ...'

(6) *V každém případě **by** v Polsku **vypukla** vážná politická krize.* 'In all cases a serious political crisis **would break** out in Poland.' (PDT 2.0)

(6') # *V každém případě v Polsku **vypukne** vážná politická krize.* '#In all cases a serious political crisis **breaks out** in Poland.'

To resume the results of the substitution test, the conditional cannot be replaced by the indicative in most sentences expressing irreal events and in such sentences rendering potential events whose realization is not expected by the speaker. These cases are considered to be an evidence that the conditional is an irreplaceable means in contemporary Czech texts and has to be included in the representation of the meaning of the sentence. On the contrary, the conditional can be substituted for the indicative in many sentences expressing potential events but also in some sentences with irreal meaning. The relatively free substitutability of the two moods in potential contexts is possible due to the close relation between the meanings of potentiality and future. In sentences which express irreal events, the conditional can be substituted for the indicative only if the irreality is clearly expressed by some other means; these cases do not deny the semantic relevance of the conditional.

4 Modality in PDT 2.0 and some other treebanks

4.1 Description of modality in PDT 2.0 and FGD

In theoretical works based on FGD, neither verbal mood nor other modal means have been studied yet in more detail in spite of the importance of these means

in constituting sentence semantics. In PDT 2.0, a special grammateme *verbmod* was defined. However, since the values of this grammateme directly correspond to the morphological moods occurring in the sentence, the grammateme is not to be considered as a semantic counterpart of the morphological category of mood but rather as a provisional, “technical” solution which requires further investigation.

Within the broad area of modality, main attention has been paid to modal verbs in FGD (see esp. [12]). On the basis of syntactic and semantic criteria, a group of so called proper modal verbs were identified; e.g. *muset* ‘must / have to’, *mocht* ‘can’. Meanings of these verbs are considered as modal features “added” to the meaning of an autosemantic verb. In the tectogrammatical tree, meanings of proper modal verbs are represented by a grammateme belonging to the node representing the autosemantic verb in question (i.e., similarly to meanings expressed by the morphological category of tense etc.). Within PDT 2.0, the modality expressed by modal verbs was captured in the same way as in FGD.

4.2 Modality in some other treebanks

Also other treebanks, if paying any attention to modality at all, confined to modal verbs whereas semantics of verbal moods is omitted (the mood is usually reflected just within the morphological annotation). Semantics of modal verbs is described, e.g., in Proposition Bank, which is a corpus annotated with semantic roles, or in the annotated treebank of Bulgarian texts *BulTreeBank*. In Proposition Bank, a special semantic role MOD for modal verbs was included (cf. [11]). In *BulTreeBank*, deontic and epistemic reading of modal verbs was distinguished (see [9]). Also [3], [13], [4] or [15] were concerned with (semi-)automatic detection of deontic and epistemic usage of modal verbs, especially in scientific articles.

5 Representing the primary function of the conditional mood at the tectogrammatical layer of PDT 2.0

5.1 Grammateme of factual modality

As sketched in Section 3, the conditional in its primary function is a semantically relevant means: by the use of the conditional an event is presented as hypothetical. In this function, the conditional is opposed to the indicative, by means of which events are presented as given. The semantic opposition which is expressed by the conditional and the indicative is thus to be captured when representing the meaning of the sentence, i.e., within the tectogrammatical tree in PDT 2.0.

Since verbal mood is a morphological category (similarly to that of tense etc.), we propose to capture the semantic opposition expressed by the conditional and the indicative by a grammateme which belongs to a verbal node of the given tectogrammatical tree. We introduce a new grammateme of factual modality (*factmod*). Besides the opposition of hypotheticality and assertion also

verbal mood	grammateme factmod	grammateme tense
past conditional	irreal	–
present conditional expressing potentiality	potential	–
present conditional expressing irreality	irreal	–
present conditional expressing potentiality or irreality (unresolved)	potential irreal	–
indicative expressing assertion	asserted	ant / sim / post
indicative instead of the conditional with the meaning of potentiality	asserted	ant / sim / post
indicative instead of the conditional with the meaning of irreality	irreal	ant / sim / post

Table 1. Rules for assigning the new grammateme **factmod**. In the 1st column, the mood of the verb form is given. According to the meaning of this form, the appropriate value of the **factmod** is chosen (the 2nd column). In the 3rd column, it is indicated whether one of the values of the grammateme **tense** is to be chosen (if not, – is filled in).

the difference between two types of hypothetical meanings – the potential and the irreal ones – has to be taken into consideration. Three values of the grammateme **factmod** are therefore proposed: **potential** for potential events, **irreal** for irreal events and **asserted** for given events.⁴

5.2 Annotation rules for assigning the grammateme of factual modality

When describing the assignment of the values of the proposed grammateme, we proceed from the surface structure of the sentence in a way similar to that in which a tectogrammatical representation is assigned to a sentence. A tectogrammatical node which represents an indicative verb form (the primary function of the indicative and the conditional is taken into consideration here and in the sequel) is typically assigned with the value **asserted** of the grammateme **factmod**. When a verb form of past conditional occurs, the value **irreal** is filled in the grammateme **factmod** at the tectogrammatical node representing this verb form.

Concerning the present conditional, one of the values of the grammateme **factmod** is to be chosen by the annotator on the basis of context, knowledge of situation etc.: if this verb form expresses an irreal event, the value **irreal** is assigned to the corresponding node; if a potential event is concerned, the value **potential** is the right one; in case the annotator is not able to decide between the two interpretations, both values are to be filled in (**potential|irreal**). As for

⁴ The solution to represent both oppositions by a single grammateme was preferred to the possibility to introduce two grammatemes just for the economy of the former solution.

the expressions of irreality, the past conditional and the present conditional are thus considered as synonymous means: for sentences which differ just in the conditional form used, identical tectogrammatical representations are supposed.

Besides these basic cases, we also deal with tectogrammatical representation of the sentences in which an indicative verb form is used instead of the conditional one (cf. the results of the substitution test in Section 3). Examples such as the governing clause in the sentence (7), in which the conditional can be substituted for more than one indicative form (cf. (7')), indicate that these sentence pairs are semantically close to each other, though not synonymous. Therefore, they have to be represented differently at the tectogrammatical layer. Rules concerning the assignment of the indicative which occurs instead of the conditional are specified in the next subsection.

(7) ... *jestliže bychom se rozhodli stát jadernou zemí, musela by následovat celá řada kroků ...* '... if we decide to become a nuclear country, a whole series of steps **would have to follow ...**' (PDT 2.0)

(7') ... *jestliže bychom se rozhodli stát jadernou zemí, musí následovat / bude muset následovat celá řada kroků ...* '... if we decide to become a nuclear country, a whole series of steps **has to follow / will have to follow ...**'

5.3 Annotation of the indicative in originally conditional contexts

We propose to represent the indicative in sentences rendering irreal events in the same way as the past conditional (or the present conditional with irreal meaning), i.e., by a tectogrammatical node with the value *irreal* in the grammateme *factmod*; additionally, the node is assigned with the grammateme *tense*, in which the temporal characteristic of the indicative form is captured. Unlike the annotation of the indicative in sentences expressing irreal events, the value *asserted* is chosen in the grammateme *factmod* and an appropriate value of the grammateme *tense* is assigned if the indicative is used instead of the present conditional with a potential meaning.

The reason for a different annotation of the indicative in irreal and potential contexts is the disparity of conditions under which the indicative can be used instead of the conditional in these two context types. In sentences rendering irreal events, the conditional can be substituted for the indicative only under relatively strong conditions (if the irreality is clearly expressed by some other means) while the conditional and the indicative are mostly freely interchangeable in potential contexts. As mentioned above, this interchangeability is related to the semantic closeness of the potentiality of the conditional and the future meaning of the indicative. If we decided to represent the indicative similarly to the conditional in sentences expressing potential events (by the value *potential* in the grammateme *factmod* in combination with a value of the grammateme *tense*), it would mean at the same time that we have to resolve at each future indicative form whether it expresses a future or a potential event, which we consider as impossible in many cases. Assignment rules for the grammateme *factmod* are resumed in Table 1.

ex. nr.	verb form	gramm. factmod	gramm. tense	ex. nr.	verb form	gramm. factmod	gramm. tense
(1)	<i>by oslavil</i>	irreal	–	(5)	<i>skončil by</i>	irreal	–
(2)	<i>by bylo bývalo</i>	irreal	–	(5')	<i>skončí</i>	irreal	post
(3)	<i>by bylo</i>	potential	–	(6)	<i>by vypukla</i>	potential	–
(3')	<i>bude</i>	asserted	post	(7)	<i>musela by následovat</i>	potential	–
(4)	<i>by byla</i>	potential	–	(7')	<i>musí následovat</i>	asserted	sim
	<i>by byla</i>	potential	–		<i>bude muset následovat</i>	asserted	post
(4')	<i>bude</i>	asserted	post				
	<i>bude</i>	asserted	post				

Table 2. Verb forms from the example sentences (1) to (7') and corresponding values of the grammatemes **factmod** and **tense** which are to be assigned to the tectogrammatical nodes representing the verb forms in question. Only verb forms which were written in bold in the examples are included in the table. Verb forms from the sentences marked with a hash mark (i.e., (1'), (2'), and (6')) were not assigned.

Examples of verb forms assigned with the values of the grammateme **factmod** (and grammateme **tense** if necessary) are given in Table 2.

6 Conclusions

The present paper focused on the Czech conditional mood when expressing hypothetical events (we talked about the primary function of this mood). It was illustrated with several examples that the conditional is a semantically relevant means in contemporary Czech although it could be substituted for the indicative in some of the studied contexts. The conditional in its primary function should thus be included in the tectogrammatical representation of PDT 2.0.

For this purpose, the grammateme **factmod** was proposed. Three values of this grammateme were suggested by means of which the difference between hypothetical and asserted events as well as the difference between two types of hypothetical events (the potential and irreal ones) are captured. Rules for manual assignment of the values were further described in the paper.

In the near future, rules for (at least) semi-automatic assignment of this grammateme have to be specified. Also tectogrammatical representation of other functions of the conditional and further modal means is to be elaborated.

Acknowledgements

I would like to thank Professor Jarmila Panevová for extensive discussion on the presented topic. I thank also Professor Eva Hajičová and Zdeněk Žabokrtský for valuable comments on the draft of this paper. The work reported on in this paper was supported by the project 1ET101120503.

Bibliography

- [1] Bybee, J. L. (1985). *Morphology: A Study of the Relation between Meaning and Form*. Benjamins, Philadelphia.
- [2] Daneš, F., Hlavsa, Z., Grepl, M. a kol. (1987). *Mluvnice češtiny 3*. Academia, Praha.
- [3] Danilava, S. and Schommer, C. (2008). A Semi-Automatic Framework to Discover Epistemic Modalities in Scientific Articles.
- [4] Gabrielatos, C. and McEnery, T. (2005). Epistemic Modality in MA Dissertations. *Lengua y Sociedad: Investigaciones recientes en lingüística aplicada. Lingüística y Filología*, 61:311–331.
- [5] Grepl, M. a Karlík, P. (1998). *Skladba češtiny*. Votobia, Olomouc.
- [6] Hajič, J., Panevová, J., Hajičová, E., Sgall, P., Pajas, P., Štěpánek, J., Havelka, J., Mikulová, M., Žabokrtský, Z., and Ševčíková Razímová, M. (2006). *Prague Dependency Treebank 2.0*. Linguistic Data Consortium, Philadelphia.
- [7] Karlík, P. (1983). K významu kondicionálu minulého. *Slovo a slovesnost*, 44:12–21.
- [8] Komárek, M., Kořenský, J., Petr, J., Veselková, J. a kol. (1986). *Mluvnice češtiny 2*. Academia, Praha.
- [9] Osenova, P. and Simov, K. (2003). The Bulgarian HPSG Treebank: Specialization of the Annotation Scheme. In *Proceedings of the Second Workshop on Treebanks and Linguistic Theories (TLT 2003)*, pages 129–140, Vaxjo, Sweden. Vaxjo University Press.
- [10] Palmer, F. R. (2001). *Mood and Modality*. Cambridge University Press, Cambridge.
- [11] Palmer, M., Gildea, D., and Kingsbury, P. (2005). The Proposition Bank: An Annotated Corpus of Semantic Roles. *Computational Linguistics*, 31:71–106.
- [12] Panevová, J., Benešová, E. a Sgall, P. (1971). *Čas a modalita v češtině*. Univerzita Karlova, Praha.
- [13] Piqué, J., Posteguillo, S., and Andreu-Besó, J. V. (2001). A Pragmatic Analysis Framework for the Description of Modality Usage in Academic English Contexts. *Estudios de Lingüística Inglesa Aplicada*, 2:213–224.
- [14] Sgall, P., Hajičová, E., and Panevová, J. (1986). *The Meaning of the Sentence in Its Semantic and Pragmatic Aspects*. D. Reidel Publishing Company, Dordrecht.
- [15] Thompson, P., Venturi, G., McNaught, J., Montemagni, S., and Ananiadou, S. (2008). Categorising Modality in Biomedical Texts. In *Proceedings of the LREC 2008 Workshop on Building and Evaluating Resources for Biomedical Text Mining*, pages 31–34, Marrakech, Morocco. ELRA.
- [16] Ševčíková, M. (2009). *Funkce kondicionálu z hlediska významové roviny. Disertační práce*. MFF UK, Praha.
- [17] Šmilauer, V. (1966). *Novočeská skladba*. SPN, Praha.