

**Focalizers and Discourse Relations**

Eva Hajičová, Jiří Mírovský, Barbora Štěpánková

Charles University, Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics

Abstract

The main concern of the present contribution is the relation between the focussing function of certain particles called focalizers and the relations in discourse. We focus our attention on the English focalizers *also*, *only*, *even*, and their Czech counterparts *také*, *jenom*, *dokonce*, and base our analysis on the data from the English–Czech annotated parallel corpus PCEDT. We attempt to find out in which respects and under which conditions the selected focalizers may be said to serve in a discourse as discourse connectives and which particular discourse relations are indicated by the focalizers in question. Our analysis confirms the hypothesis that the particles *also*, *only* and *even* as well as their Czech equivalents play basically a discursive role of explicit connectives, though in a different way and to a different extent.

1. Motivation and Research Questions

The analysis of the so-called focalizers, i.e. particles such as E. *also*, *only*, *even*, and their Czech counterparts *také/rovněž/těž/zároveň* for *also*, *jen/jenom/pouze* for *only* and *dokonce* for *even*, based on the data from the English–Czech parallel corpus PCEDT and studied from two aspects in Hajičová and Mírovský (in prep), namely (i) their position in the sentence surface word order, and (ii) their semantic scope, has demonstrated that the interpretation of the semantic scope of these particles is highly dependent on the previous context and in several respects has an important influence on the interpretation of discourse relations. In a certain way, this issue is closely connected also to the debate on the status of these particles in the word-class system in relation to conjunctions and adverbs (see Štěpánková, 2014). These observations have led us to formulate the following two research questions:

- (i) in which respects and under which conditions the selected focalizers may be said to serve in a discourse as discourse connectives,
- (ii) which particular discourse relations are indicated by the focalizers in question.

2. Data

We have based our analysis on the following data resources: (i) for Czech, the Prague Dependency Treebank of Czech (PDT 3.5, Hajič et al., 2018), containing documents of the total of about 50 thousand sentences annotated on the underlying syntactic layer also for information structure (topic–focus articulation, TFA, Hajičová et al., 1998) and containing also annotation of discourse relations (in a slightly modified PDTB style); (ii) for English, the Pennsylvania Discourse TreeBank (PDTB, ver. 2: Prasad et al., 2008, ver. 3: Prasad et al., 2019); (iii) for a comparison between Czech and English, the English–Czech parallel corpus (PCEDT, Hajič et al., 2012); (iv) the dictionary of Czech connectives (Mírovský et al., 2017; Synková et al., 2019).

We are aware that our analysis might have been influenced by the discourse genre of the annotated data of PDT and PCEDT (mostly journalistic style) but we assume that the phenomena under investigation, namely the discourse impacts of focalizers *also*, *only* and *even*, are general enough and that the genre in which they occur may have an impact only on their frequency.

3. Annotation of Underlying Syntactic and Discourse Relations

For our analysis, we have made use of the following features of the annotated data:

(a) underlying syntactic relations

The underlying layer sentence representations in the above mentioned PDT-based corpora (PDT 3.5 and in both the Czech and the English parts of PCEDT) have the form of dependency trees, with the PRED(icate) as the root of the tree corresponding to the main verb. Each node of the tree except for PRED is labeled among other features by a specification of the dependency relation (called a functor) such as ACT, PAT, ADDR, etc. (Hajič et al., 2017). One of these relations is the functor RHEM denoting the function of focalizer.

For illustration, we present in Figure 1 a dependency representation of the sentence *Only at the moment of maximum roll did I grasp what was going on.* (Czech translation: *Teprve ve chvíli největšího víření jsem pochopil, o co jde.*), where the functors printed in capitals stand for the following underlying syntactic relations: ACT for Actor, PAT for Patient (Objective), APP for Appurtenance, EXT for Manner-Extent and TWHEN for temporal modification.

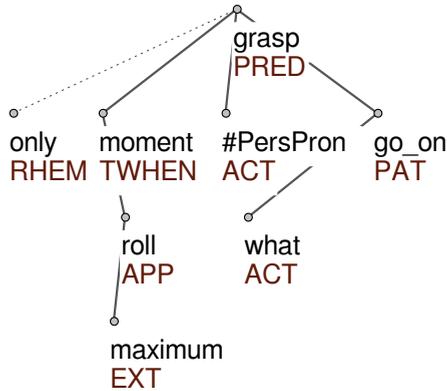


Figure 1. A simplified tectogrammatical representation of the sentence *Only at the moment of maximum roll did I grasp what was going on.*

(b) discourse relations

As for discourse relations, the annotation in both of these corpora is based on the Penn Discourse Treebank (PDTB) style. A discourse relation is understood to hold between two Arguments, Arg1 and Arg2, roughly speaking segments (adjacent sentences or in some cases between clauses within a compound sentences) including a verb as its core. The following types of relations are relevant for our discussion:¹

- (a) Explicit relation – discourse relation expressed by an explicit discourse connective (as in (1) below)
- (b) Implicit relation – a certain discourse relation can be inferred but cannot be identified to be expressed by an explicit discourse connective (as in (2))
- (c) EntRel – a discourse relation given by a coreference relation between entities that are a part of Arg1 and Arg2, respectively (as in (3))
- (d) NoRel – no discourse relation between Arg1 and Arg2 can be recognized (as in (4))
- (e) Hypophora: a coherence relation for Question-Answer pairs, where one argument (commonly Arg1) expresses a question and the other argument (commonly Arg2) provides an answer. As with Entity Relations, no explicit or implicit connective is identified and annotated.

¹ The examples are taken from the annotation manual of the PDTB 2.

Examples:

- (1) *The city's Campaign Finance Board has refused to pay Mr. Dinkins \$95,142 in matching funds <Explicit> because his campaign records are incomplete.*
- (2) *Motorola is fighting back against junk mail. So much of the stuff poured into its Austin, Texas, offices that its mail rooms there simply stopped delivering it. <Implicit=so> Now, thousands of mailers, catalogs and sales pitches go straight into the trash.*
- (3) *Pierre Vinken, 61 years old, will join the board as a nonexecutive director Nov. 29. <EntRel> Mr. Vinken is chairman of Elsevier N.V., the Dutch publishing group.*
- (4) *Mr. Rapanelli met in August with U.S. Assistant Treasury Secretary David Mulford. <NoRel> Argentine negotiator Carlos Carballo was in Washington and New York this week to meet with banks.*

For the purpose of our case study we do not distinguish between a relation expressed by a one-word connective and that expressed or implied by a complex connective called AltLex, such as the one in (5):

- (5) *After trading at an average discount of more than 20% in late 1987 and part of last year, country funds currently trade at an average premium of 6%. <AltLex> The reason: Shareprices of many of these funds this year have climbed much more sharply than the foreign stocks they hold.*

4. Data Analysis

4.1. Also

As for the focalizer *also*, we have searched for sentences without coordination in which *also* depends on the main verb labelled PRED. There were 1291 occurrences of this particle in the English part of PCEDT out of which 880 sentences were connected in some discourse relation with the immediately preceding sentence. As for the type of the relation, there were 828 cases annotated as an Explicit relation, 31 as an Implicit relation, 19 as an EntRel, 2 as NoRel. Out of the 828 Explicit relations there were 781 cases of the subtype Expansion.Conjunction. In 772 cases, the focalizer *also* was determined as the connective, i.e. as the indicator of the relation.

To find an answer to one of our research questions, namely whether the focalizer *also* may serve as an indicator of a certain discourse relation, we have focussed our attention on cases with *also* in which no Explicit discourse relation was annotated. There were 60 such cases in the PCEDT corpus which we have studied in relation to the preceding context. The following tendencies have been identified:

- (a) In most cases, we could assign an Explicit discourse relation of the type Expansion.Conjunction, see (6).

(6) *However, excluding the year-earlier charge for recall of steering gear, operating profit in the latest quarter declined 14%, reflecting higher start-up and product development expenses in passenger-restraint systems. – Materials and production costs **also** rose, TRW said.*

(b) Only in few cases, the discourse relation EntRel could be assigned based on a coreference relation, see (7).

(7) *State Farm Mutual Automobile Insurance Co., the largest home and auto insurer in California, believes the losses from the earthquake could be somewhat less than \$475 million in damages it expects to pay out for claims. – State Farm based in Bloomington, Ind, is **also** the largest writer of personal-property earthquake insurance in California.*

(c) There were also only few cases where no relation could be recognized between two adjacent sentences, see (8).

(8) *MCI has made hawks out of the upper echelon of AT&T, said T-2 PaineWebber's Mr. Grubman, who said he expected AT&T to become increasingly aggressive in dealing with longtime nemesis. – Julie Amparano Lopey in Philadelphia **also** contributed to this article.*

The statistical data quoted above and our analysis of the disputable examples has led us to the conclusion that the focalizer *also* plays a role of a connective expressing the relation of Explicit Expansion.Conjunction.

4.2. Only

The analysis of sentences containing the English focalizer *only* offers a much richer picture than those with the focalizer *also*, both as for the syntactic functions in which this particle occurs and as for the variety of Czech equivalents.

Concerning the functions the particle *only* obtains, the following four are prominent in the total of 1184 occurrences in the PCEDT (irrespective of its placement in the sentence):

- RHEM (focalizer): 750
- EXT (Extent as one of the functions of the modifier of Manner): 272
- CM (conjunction modifier): 81
- RSTR (restrictive modification of nouns, roughly speaking an attribute): 77

As for the relation of the particle *only* and the discourse relations, it should be noted that *only* serves only in 7 cases as a “pure” connective (indicating the Explicit relation of Expansion.Exception in 3 cases, of the relation Comparison.Concession in 2 cases, of the Expansion-Level-of-detail relation in 1 case and of the Comparison.Contrast in 1 case). However, there are 105 occurrences of *only* in multiword connectives (such as *not only but, only if*).

For the purpose of our analysis, the RHEM function is of importance, as in these cases the particle was classified by the annotators to function as a focalizer. In particular, we have been interested in cases where *only* depends on PRED and is placed before PRED so that it can be assumed that the whole predicative part of the sentence is in its scope. There were 61 such cases. After a closer inspection of these cases, only in 33 of them a discourse relation was found to hold between the sentence with *only* and the preceding sentence, the rest were sentences without such relations. Most relations were of the type Implicit (19), with only 7 Explicit ones, 5 of the type EntRel and 1 with NoRel type and 1 Hypophora. A closer look at the Implicit type has indicated that the presence of the focalizer *only* does contribute to a more detailed specification of the relation Expansion in the sense of a level of detail, see (9).

- (9) *Instead, they map out a strategy in several phases from now until 1995. Most of the measures would **only** start to have an effect on beleaguered Soviet consumers in two or three years at the earliest.*

In case of an implicit relation of Comparison, the presence of the focalizer *only* contributes to the implication of a contrast, see (10).

- (10) *For such products as canned vegetables and athletic shoes, devotion to a single brand was quite low, with fewer than 30% saying they usually buy the same brand. **Only** for cigarettes, mayonnaise and toothpaste did more than 60% of users say they typically stick with the same brand.*

We have also put under scrutiny those cases in which the underlying syntactic function of the particle was annotated as one of the modifications of Manner, namely EXT. In order to find out whether a presence of *only* may help to assign a particular discourse relation, we have searched for sentences in which *only.EXT* was present but which were not connected with the preceding sentence by any discourse relation. There were 76 such sentences in the PCEDT corpus. It came out that although *only* apparently does not by itself serve as a connective, its occurrence in the sentence influences the interpretation of the relation between the two adjacent sentences in a considerable way. The following tendencies have been identified:

- (i) The presence of *only* indicates an explanation, more precision, substantiation, see (11) and (12).

- (11) *Some even claim the group has become a lagging, not leading, indicator. The technology sector of the Dow Jones Equity Market Index has risen **only** about 6.24% this year, while the Nasdaq Composite Index has gained 18.35%.*
- (12) *But the last stock market boom, in 1986, seems small compared with the current rush to market. The \$6 billion that some 40 companies are looking to raise in the year ending March 31 compares with **only** \$2.7 billion raised on the capital market in the previous fiscal year.*

- (ii) The presence of *only* contributes to the inversion of the discourse relation, see (13).
- (13) *Toyota Motor Corp.'s Lexus division also provides specifications. But **only** two-thirds of Lexus dealers are constructing new buildings according to the Lexus specs.*
- (iii) The contrast is emphasized as in (14).
- (14) *The number one proposal for reducing crime in the New York survey was to put more police on foot or scooter patrol, suggested by more than two-thirds of the respondents. **Only** 22% supported private security patrols funded by the merchants themselves.*
- (iv) Indication of a comparison, see (15).
- (15) *The U.S. Bureau of Justice Statistics reports that almost 2% of all retail-sales workers suffer injuries from crime each year, almost twice the national average and about four times the rate for teachers, truck drivers, medical workers and door-to-door salespeople. **Only** a few other occupations have higher reported rates of criminal injury, such as police, bartenders and taxi drivers.*
- (v) An adversative relation is implied, see (16).
- (16) *Whether psyllium makes Sidhpur's fortune depends on cholesterol-fearing Americans, the U.S. Food and Drug Administration and, of course, the outcome of further research. **Only** one thing is certain here: Psyllium is likely to remain an export item from Sidhpur for a long time.*

As mentioned above, the parallel PCEDT corpus offered a variety of Czech equivalents of the particle *only* (besides the more straightforward translations *jenom*, *jen*, *pouze*, there occurred equivalents such as *až*, *ještě*, *dokonce* or *také*) and therefore we have also looked whether the Czech translation might help to recognize a more detailed specification of the discourse relation. However, we have not found any indications in the data of such a case.

4.3. Even

The frequency of the occurrence of the particle *even* (irrespective of its position in the sentence) analyzed as a focalizer was 653 times, that is much lower than that of the focalizers *also* and a little bit lower also than that of the focalizer *only*. However, a more striking fact was that in PDTB 3 *even* does not occur as a pure connective, it occurs only as a part of some multiword complex connectives such as *even if*, *even though*, *even as*, *even when* etc.

Therefore we have looked in more detail at the Czech translations of this particle to see if the Czech translations in the given contexts may offer a more varied picture. We have found 19 different Czech equivalents of *even*. RHEM, the most frequent of which was *dokonce* (242 times) and *ještě* (113 times).

Having these data at our disposal, we have decided to investigate whether the occurrence of *even*.RHEM translated as *dokonce* may influence the discourse relations, that is to say if it may play a role of a true connective. We have focussed our attention on the position of *even*.RHEM before the PRED (in non-coordinated constructions) and translated as *dokonce*, which occurred 98 times. Out of this number, there were 65 cases where a discourse relation to the previous sentence was annotated, 54 of which were marked as Implicit relations (of the type Expansion.Conjunction 32, other type of Expansion 14 and other Implicit 8); there were 8 Explicit relations (of the type Expansion.Conjunction 2, Comparison.Concession 4, Comparison.Contrast 1, and Temporal.Asynchronous 1), 2 relations were marked as EntRel and 1 as AltLex. None of the Explicit relations was marked by the focalizer *even*, the connectives were *but* (3), *and* (2), *however*, *still*, *even then*.

Looking at the Implicit relations in more detail, we have seen that in most cases marked as Expansion, there was a certain degree of gradation involved, see e.g. (17) and (18) with Expansion.Conjunction marked as “in fact”. The same is true with the relation annotated as Comparison.Concession and marked in as “nevertheless” in (19).

- (17) *All kinds of landmark Texas real estate has been snapped up by out-of-staters. **Even** the beloved Dallas Cowboys were bought by an Arkansas oil man.*
- (18) *Mr Hahn began selling non-core businesses, such as oil and gas and chemicals. He **even** sold one unit that made vinyl checkbook covers.*
- (19) *But that’s for the best horses, with most selling for much less. **Even** when they move outside their traditional tony circle, racehorse owners still try to capitalize on the elan of the sport.*

Also in case of an Explicit relation one can recognize a certain gradation, see e.g. (20) annotated as Expansion.Conjunction with the connective *and*:

- (20) *Press agents and public-relations practitioners are notorious name-droppers. And some **even** do it with malice afterthought.*

Our analysis of the interpretation of discourse relations between sentences the second of which contains the focalizer *even* has led to a proposal to introduce into the set of connectives the particle *even* for those relations of Expansion (and perhaps also of Comparison) that can be interpreted as gradation. It should be noted that the type gradation is not among the types of relations recognized by PDTB. Such a solution would comply with the treatment applied in the PDT, namely taking “*dokonce*” as a connective present in the relation of gradation (73 cases in total).

5. Conclusion and Summary

In the present case study, we have carried out an analysis of discourse relations between adjacent sentences (taken as discourse arguments) the second of which (ARG2)

contained one of the particles *also*, *only* or *even* in the (underlying syntactic) function of a focalizer (RHEM). Our analysis was based on the data from the annotated Czech–English parallel PCEDT and in the classification of the discourse relations we used basically the PDTB approach.

The statistical data quoted above and our analysis of the disputable examples has led us to the conclusion that the particle *also* as well as its Czech equivalents functions as a focalizer and plays basically a discursive role of an explicit “pure” connective expressing the relation of Expansion.Conjunction.

As for the particle *only*, the PCEDT data indicate that the prevalent underlying syntactic function of this particle is that of a focalizer and of a modification of Manner. In contrast to the focalizer *also*, the particle *only* serves as a “pure” discourse connective only in a negligible number of cases, relatively more frequently being a part of a multiword connective. However, the presence of this particle helps to understand a given discourse relation in a more specific way, for instance in the sense of a certain level of detail with the relation Expansion. With the relation of Comparison, the presence of the focalizer *only* implies a contrast. If *only* obtains the function of a modification of manner EXT, it contributes to the interpretation of the relation between two neighbouring sentences in a considerable way as well, strengthening a contrastive interpretation of this relation, indicating a comparison or a relation based on coreferential entities occurring in the two sentences.

A most interesting case is offered by the analysis of the focalizer *even*. Since it does not appear in the PDTB list of connectives (it occurs only as a part of some multiword complex connectives such as *even if*, *even though*, *even as*, *even when*), we have looked for the most frequent Czech equivalent of *even*.RHEM, namely *dokonce*, and considered its possible influence on the discourse relations, that is to say we wanted to find out if it may play a role of a true connective. In most cases the relation to the previous sentence was annotated as an Implicit Relation (mostly Expansion.Conjunction). A closer inspection of these examples has led to a recognition of a certain degree of gradation present and to a conclusion that the focalizer *even* may be understood as a connective with this meaning.

In a follow-up analysis of the data of PCEDT we want to include into our consideration other English focalizers as candidates for the role of connectives (e.g. *mainly*, *just*), investigate them in relation to their Czech translations and thus to analyze the role of focalizers as connectives in a broader perspective than allowed by our present case study.

Acknowledgements

We gratefully acknowledge support from the Grant Agency of the Czech Republic, project 20-09853S. The work described herein has been using resources provided by the LINDAT/CLARIAH-CZ Research Infrastructure, supported by the Ministry of Education, Youth and Sports of the Czech Republic (project no. LM2018101).

This paper is a full version of the contribution presented at the workshop on discourse markers at the 2020 annual conference of Societas Linguistica Europaea. Only a short abstract of the contribution was published in the electronic version of the conference proceedings.

Bibliography

- Hajič, Jan, Eduard Bejček, Alevtina Bémová, Eva Buráňová, Eva Hajičová, Jiří Havelka, Petr Homola, Jiří Kárník, Václava Kettnerová, Natalia Klyueva, Veronika Kolářová, Lucie Kučová, Markéta Lopatková, Marie Mikulová, Jiří Mírovský, Anna Nedoluzhko, Petr Pajas, Jarmila Panevová, Lucie Poláková, Magdaléna Rysová, Petr Sgall, Johanka Spoustová, Pavel Straňák, Pavlína Synková, Magda Ševčíková, Jan Štěpánek, Zdeňka Uřešová, Barbora Vidová Hladká, Daniel Zeman, Šárka Zikánová, and Zdeněk Žabokrtský. *Prague Dependency Treebank 3.5. Data/Software*, Univerzita Karlova, MFF, ÚFAL, Prague, Czech Republic, 2018.
- Hajič, Jan, Eva Hajicová, Jarmila Panevová, Petr Sgall, Ondrej Bojar, Silvie Cinková, Eva Fucíková, Marie Mikulová, Petr Pajas, Jan Popelka, Jiří Semecký, Jana Šindlerová, Jan Štěpánek, Josef Toman, Zdeňka Uřešová, and Zdeněk Žabokrtský. *Announcing Prague Czech–English Dependency Treebank 2.0*. In *LREC*, pages 3153–3160, 2012.
- Hajič, Jan, Eva Hajičová, Marie Mikulová, and Jiří Mírovský. *Prague Dependency Treebank*. Springer Verlag, Berlin, Germany, 2017. ISBN 978-94-024-0879-9.
- Hajičová, Eva, Barbara Hall Partee, and Petr Sgall. *Topic–Focus Articulation, Tripartite Structures, and Semantic Content*. Dordrecht: Kluwer, 1998.
- Hajičová, Eva and Jiří Mírovský. *Focalizers through the Lens of a Parallel English–Czech Corpus*. In *Submitted to COLING 2020*, in prep.
- Mírovský, Jiří, Pavlína Synková, Magdaléna Rysová, and Lucie Poláková. *CzeDLex – A Lexicon of Czech Discourse Connectives*. *The Prague Bulletin of Mathematical Linguistics*, 109(1):61–91, 2017.
- Prasad, Rashmi, Alan Lee, Nikhil Dinesh, Eleni Miltsakaki, Geraud Campion, Aravind Joshi, and Bonnie Webber. *Penn Discourse Treebank Version 2.0*. *Data/Software*, Linguistic Data Consortium, 2008. University of Pennsylvania, Philadelphia. LDC2008T05.
- Prasad, Rashmi, Bonnie Webber, Alan Lee, and Aravind Joshi. *Penn Discourse Treebank Version 3.0*. *Data/Software*, Linguistic Data Consortium, 2019. University of Pennsylvania, Philadelphia. LDC2019T05.
- Štěpánková, Barbora. *Aktualizátory ve výstavbě textu, zejména z pohledu aktuálního členění [Focalizers in the structure of text, esp. from the point of view of information structure]*. Ústav formální a aplikované lingvistiky, 2014.
- Synková, Pavlína, Lucie Poláková, and Jiří Mírovský. *CzeDLex 0.6*. *Data/Software*, LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University, 2019. <http://hdl.handle.net/11234/1-3074>.

Address for correspondence:

Eva Hajičová

hajicova@ufal.mff.cuni.cz

Institute of Formal and Applied Linguistics

Faculty of Mathematics and Physics, Charles University

Malostranské náměstí 25

118 00 Praha 1

Czech Republic