

Classification of Language Means Expressing Textual Relations in Czech: Difference between Discourse Connectives and Their Alternative Lexicalizations

Magdaléna Rysová

magdalena.rysova@post.cz

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

Aim of Research

To explore the differences between classic connectives (like *and*, *but*, *therefore* etc.) and Alternative Lexicalizations of Discourse Connectives (AltLex's) in Czech: (like *that is the reason why*, *the result is* etc.)

Methods and Material

- The analysis was based on the data from the *Prague Dependency Treebank* (PDT; 43,955 sentences).
- AltLex's are alternative lexicalizations of discourse connectives; they signal certain discourse relation between two verbal arguments but they differ from the classic connectives both lexically and syntactically
- An example of AltLex – *the reason is*
- An example of a classic connective – *therefore*

Results of Research

Classic Connectives	AltLexes
synsemantics	structures with autosemantic words
lexically frozen	open or fixed collocations
non-modifiable (with exceptions)	modifiable (with exceptions)
mainly one-word	mainly multiword
not sentence elements	sentence elements, clause modifiers or separate sentences
	uniqueness of some structures:
	a) syntactically higher than the 2 nd argument
	b) form of a separate sentence
	c) nominalization of the 2 nd argument

- 1) Synsemantics (*and*, *but*, *or*...) functioning as connectives on their own vs. Connecting Structures with Autosemantic Components (*the reason is* – *reason* is a noun; *this means* – *to mean* is a verb etc.)
- 2) Classic connectives are lexically frozen (i.e. grammaticalised), AltLexes form open or fixed collocations (that are not grammaticalised)
- 3) AltLexes mostly allow a free modification (*the main/only/first/important reason is* etc. vs. **generally and*); classic connectives are modifiable only exceptionally (*simply because*)

Ex.: *Another reason is more complex and in essence philosophical.*

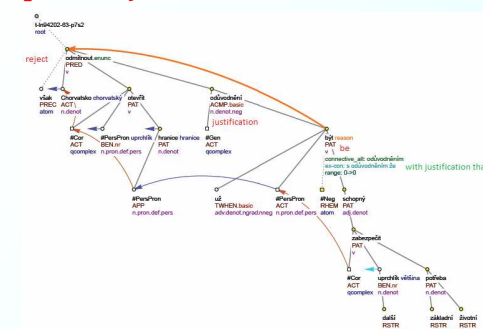
Classic connectives are modifiable only partly – e.g. *simply/mainly/easily because*, but not **philosophically because*.

- 4) Classic connectives are mainly one-word expressions, AltLexes form mostly multiword connecting structures
- 5) Classic connectives (as synsemantics like conjunctions *and*, *but* etc. or structuring particles – *too*, *only* etc.) are not sentence elements, AltLexes are sentence elements (like *because of this*), sentence modifiers (like *simply speaking*) or form a separate sentences (*the reason is simple*)

Uniqueness of some AltLex structures

- 1) AltLexes may form a separate sentence: *I will not go to school tomorrow. The reason is easy (= AltLex). I am ill.*
- 2) AltLexes may be syntactically higher than the second argument of the relation (i.e. the second argument is syntactically dependent on the AltLex expressed by the main clause) – *I will not go to school tomorrow. The reason is (= AltLex, main clause) that I am ill (= the second argument, nominal content subordinate clause).*
- 3) Some AltLexes allow nominalization of the second discourse argument: *The management of the club dismissed the coach. The reason are disappointing results (= NP).*

Example of AltLex Annotation in the Prague Dependency Treebank



Croatia, however, rejected to open their borders with justification that it is not able to provide other basic necessities to refugees. (AltLex = *with justification that*; discourse relation = *reason – result*)

Conclusion

The PDT material demonstrates that classic discourse connectives (like *and*, *but*, *or*) differ from the Alternative Lexicalizations of Connectives in many ways – in lexical, syntactic as well as semantic aspect. The AltLex structures are in many aspect unique in language, as they demonstrate a special behaviour which classic connectives can never do (e.g. to form a separate sentence etc.).

Acknowledgments

This paper was supported by the grant of Czech Science Foundation GA ČR (project n. P406/12/0658) and by the Ministry of Education, Youth and Sports of the Czech Republic (project n. LH14011).

Data sources

Bejček, E. et al. (2012). Prague Dependency Treebank 2.5 – extended version of PDT 2.0. In: *Proceedings of the 24th International Conference on Computational Linguistics (Coling 2012)*. Mumbai, India.