

# Group task: Creating an initial lexicon of Ukrainian connectives

Tatjana Scheffler, University of Potsdam, Germany, [tatjana.scheffler@uni-potsdam.de](mailto:tatjana.scheffler@uni-potsdam.de)

CompLing Workshop: Shallow Discourse Parsing, Grammarly, Kyiv, Ukraine, February 29, 2020

1. Determine whether the given candidate word (or phrase) is a connective: (Note: There are also some Russian tokens in the list of candidates, based on wrongly aligned material in the underlying corpora. Please just exclude any non-Ukrainian words.)
  - it is a fixed (closed class) expression that cannot be modified or conjugated
  - it semantically relates to arguments
  - the arguments are "abstract objects" (propositions, facts, events, ...)
  - the arguments are in principle expressible as clauses
2. After determining the connectives (from 1.), group them into semantic classes:
  - temporal
  - contingency (causal + conditional)
  - contrast (contrast and similarity)
  - expansion (additive relations)
3. Group connectives by substitution tests: which connectives can replace each other in the examples?
4. Determine the syntactic and semantic class of each connective and enter it in the Google sheet:  
[https://docs.google.com/spreadsheets/d/134\\_SxwVkoPpMlsyxZ2uThlaYBQ6rK0XxnmVpRKOjeNs/edit?usp=sharing](https://docs.google.com/spreadsheets/d/134_SxwVkoPpMlsyxZ2uThlaYBQ6rK0XxnmVpRKOjeNs/edit?usp=sharing)

*syntactic categories:*

- coord\_conj = coordinating conjunction
- subord\_conj = subordinating conjunction
- adverb
- preposition = pre- or postposition
- other = other, for example phrasal connectives

*semantic senses:*

We use the simplified list of PDTB3 senses in [pdtb3-senses-simplified.txt](#).

For more information, see the [PDTB3 Annotation Manual](#)

# Data

---

I've extracted the connective candidates from the bilingual (English-Ukrainian) dictionaries of the QED and OpenSubtitles corpora (see references below). The corpus data is available for **research purposes** only! Please see [OPUS](#) for more information.

As seeds, I have used the English connectives (heads only) extracted from the PDTB, as distributed in the [Eng-DiMLex lexicon](#).

Similarly, the corpus samples for each connective candidate have been extracted from the (EN-UK section of the) QED corpus.

# References

---

- A. Abdelali, F. Guzman, H. Sajjad and S. Vogel, "The AMARA Corpus: Building parallel language resources for the educational domain", The Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC'14). Reykjavik, Iceland, 2014. Pp. 1856-1862. Isbn. 978-2-9517408-8-4.
- Debopam Das, Tatjana Scheffler, Peter Bourgonje and Manfred Stede. Constructing a Lexicon of English Discourse Connectives. In: K. Komatani, D. Litman, K. Yu, A. Papangelis, L. Cavedon, M. Nakano (eds.): Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue. Melbourne, Australia , July 2018
- P. Lison and J. Tiedemann, 2016, OpenSubtitles2016: Extracting Large Parallel Corpora from Movie and TV Subtitles. In Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016) <http://www.opensubtitles.org/>
- J. Tiedemann, 2012, Parallel Data, Tools and Interfaces in OPUS. In Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012)

## Connective candidates (extracted automatically)

а	адже	вінець	одночасно
и	далі	додати	следующей
і	доки	допоки	следующий
в той час як	досі	згодом	відповідно
в цілому	если	колись	дальнейших
в іншому	коли	окремо	наступного
бо	крім	помимо	наступному
за	отже	потому	незважаючи
на	плюс	прежде	спеціально
но	пока	ранніх	СЛЕДОВАТЕЛЬН
ні	поки	раньше	О
та	разу	раніше	
то	тоді	інакше	
чи	тому	іншого	
ще	хотя	іншому випадку	
як	хоча	Замість	
же час	хіба	Зокрема	
до того	якби	Нарешті	
не менш	якщо	відразу	
по тому	інше	замість	
же время	перш ніж	зокрема	
не менше	якщо тільки	зрештою	
як тільки	Потім	коротше	
не зважаючи	Також	наконець	
або	время	нарешті	
але	затем	однажды	
ані	итоге	пізніше	
для	когда	скоріше	
еще	кроме	точніше	
ещё	однак	винятком	
или	окрім	временем	
как	перед	например	
пор	позже	наступне	
раз	поруч	наступні	
так	после	оскільки	
теж	потом	отдельно	
тож	потім	фактично	
хоч	проте	Наприклад	
все ж	після	загальним	
так як	радше	конкретно	
тим не	ранее	кінцевому	
час як	также	наприклад	
тем не менее	також	насправді	
той час	тогда	наступний	
тим самим	через	наступних	
тим часом	таким чином	настільки	
Коли	вместо	натомість	
Якщо	врешті	незалежно	

Level-1	Level-2	Level-3
TEMPORAL	SYNCHRONOUS	—
	ASYNCHRONOUS	PRECEDENCE
		SUCCESION
CONTINGENCY	CAUSE	REASON
		RESULT
		NEGRESULT
	CAUSE+BELIEF	REASON+BELIEF
		RESULT+BELIEF
	CAUSE+SPEECHACT	REASON+SPEECHACT
		RESULT+SPEECHACT
	CONDITION	ARG1-AS-COND
		ARG2-AS-COND
	CONDITION+SPEECHACT	—
	NEGATIVE-CONDITION	ARG1-AS-NEGCOND
		ARG2-AS-NEGCOND
	NEGATIVE-CONDITION+SPEECHACT	—
	PURPOSE	ARG1-AS-GOAL
		ARG2-AS-GOAL
COMPARISON	CONCESSION	ARG1-AS-DENIER
		ARG2-AS-DENIER
	CONCESSION+SPEECHACT	ARG2-AS-DENIER+SPEECHACT
	CONTRAST	—
	SIMILARITY	—
EXPANSION	CONJUNCTION	—
	DISJUNCTION	—
	EQUIVALENCE	—
	EXCEPTION	ARG1-AS-EXCPT
		ARG2-AS-EXCPT
	INSTANTIATION	ARG1-AS-INSTANCE
		ARG2-AS-INSTANCE
	LEVEL-OF-DETAIL	ARG1-AS-DETAIL
		ARG2-AS-DETAIL
	MANNER	ARG1-AS-MANNER
		ARG2-AS-MANNER
	SUBSTITUTION	ARG1-AS-SUBST
		ARG2-AS-SUBST

Table 1: PDTB-3 Sense Hierarchy. The leftmost column contains the Level-1 senses and the middle column, the Level-2 senses. For asymmetric relations, Level-3 senses are located in the rightmost column.