

Given **Best EN→CS systems on par with humans at seg. level.**

New **This includes discourse connectives and their AltLexes.**

New **Information structure (subjects in focus) also preserved.**

New **“Doc-level” systems not any better than seg-level ones.**

(IAA: 66–93%)

A Test Suite and Manual Evaluation of Document-Level NMT at WMT19



Kateřina Rysov, Magdalna Rysov, Tomš Musil, Lucie Polkov and Ondřej Bojar

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

Introduction

- We provided a test suite for WMT19 aimed at assessing discourse phenomena of English→Czech MT systems participating in the News Translation Task.
- We carried out an extensive manual annotation of the MT outputs and identified types of translation errors relevant to document-level translation.
- We also compared the systems’ performance with respect to the observed phenomena.

Manual Evaluation

- most of the errors in the translation cannot be found by a simple comparison with the reference translation, a bilingual evaluation is needed
- AltLexes are “multi-word discourse connectors” such as *for this reason* or *attributed that to*
- vast majority of AltLexes translated correctly, across all systems
- most discourse connectors also translated correctly
- the information structure is also mostly correct

Conclusion

- most errors (equally across systems) were detected in the area of word order and contextual (non-)boundness of the subjects
- for the observed phenomena, the systems performed with only a minor differences among each other and they reached the quality of the reference
- the two context-aware systems did not outperform the others in translating the followed document-level phenomena
- strong sentence-level systems seem to be performing slightly better than context-aware ones

Annotation Results

AltLexes

	adequate	missing	wrong
CUNI-Transf-2018	★★★★★	☆☆☆☆☆	★☆☆☆☆
CUNI-DocTransf-Marian	★★★★★	☆☆☆☆☆	★☆☆☆☆
online-B	★★★★★	☆☆☆☆☆	★☆☆☆☆
CUNI-DocTransf-T2T	★★★★★	☆☆☆☆☆	★☆☆☆☆
CUNI-Transf-2019	★★★★★	☆☆☆☆☆	★☆☆☆☆
reference	★★★★★	☆☆☆☆☆	★☆☆☆☆

Discourse Connectors

	adequate	wrong place	inadequate
CUNI-Transf-2018	★★★★★	☆☆☆☆☆	☆☆☆☆☆
CUNI-DocTransf-Marian	★★★★★	☆☆☆☆☆	☆☆☆☆☆
online-B	★★★★★	☆☆☆☆☆	☆☆☆☆☆
CUNI-DocTransf-T2T	★★★★★	☆☆☆☆☆	☆☆☆☆☆
CUNI-Transf-2019	★★★★★	☆☆☆☆☆	☆☆☆☆☆
reference	★★★★★	☆☆☆☆☆	☆☆☆☆☆

Information Structure

	context		rhema		placement	
	yes	no	yes	no	yes	no
CUNI-Transf-2018	11	1	2	0	14	0
CUNI-DocTransf-Marian	17	3	5	3	14	5
online-B	6	1	0	1	3	1
CUNI-DocTransf-T2T	17	1	1	2	13	3
CUNI-Transf-2019	6	1	0	1	6	0
reference	19	4	1	6	19	3

context – contextually non-bound subject in both English and Czech
rhema – subject is the focus proper in both English and Czech
placement – correct placement according to Czech word order rules

The full paper



We acknowledge support from the Czech Science Foundation project no. GA17-06123S (Anaphoricity in Connectives: Lexical Description and Bilingual Corpus Analysis) and the EU project H2020-ICT-2018-2-825460 (ELITR). This study has utilized language resources distributed by the LINDAT/CLARIN project of the Ministry of Education, Youth and Sports of the Czech Republic (project LM2015071). This research was partially supported by SVV project number 260 453.

Presented on August 1, 2019 at WMT19, collocated with ACL 2019 in Florence, Italy.