

Translation, Parallel Data and MWEs

– a follow-up project?

Eduard Bejček, Alexandr Rosen and Pavel Straňák

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

PARSEME 2016
7th final general meeting
Dubrovnik
26–27 September 2016

TRANSEME?

Multilinguality is a key feature of PARSEME:
common guidelines, taxonomy, methodologies



Opportunity for a theoretical & application-oriented **multilingual** follow-up

Possible goals

- Build parallel corpora, with MWEs partly hand-annotated
- Compare annotated MWEs
- Project MWE annotation
- Discover MWE in aligned parallel texts
- Focus on a specific domain, build parallel corpora, annotate terminology, link to encyclopaedia
- Improve translation using results from the tasks above