MultiMT

Ondřej Bojar, Aleš Tamchyna and Duc Tam Hoang
Charles University in Prague, Czech Republic

| any source | | a single target language |

From   Into

via: many pivots
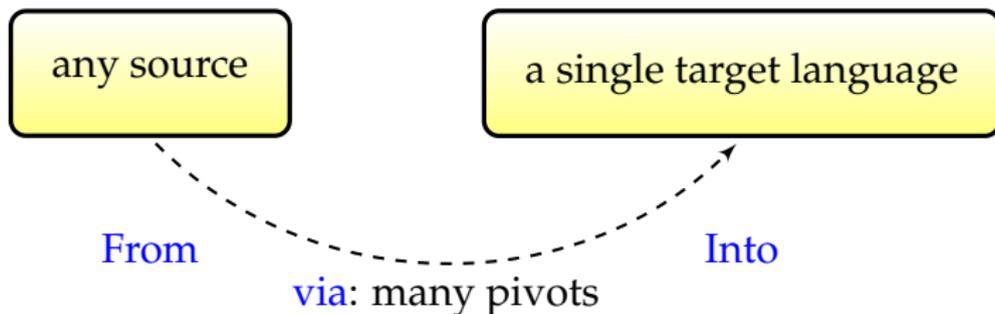
EXTEND MOSES TOOLKIT (A NEW TRAINING SCRIPT)

Assignments:

- Accept a directory with all multiparallel corpora, all pairwise word alignments, all target-language monolingual corpora.
- Construct translation models for a single Moses configuration, for example:
    - Produce: Czech, English
    - Accepts: Ukrainian, Russian, English
    - Pivots: English and German
- Support **Twitter Crowd Translation** but applicable itself

EXPERIMENT LANGUAGES
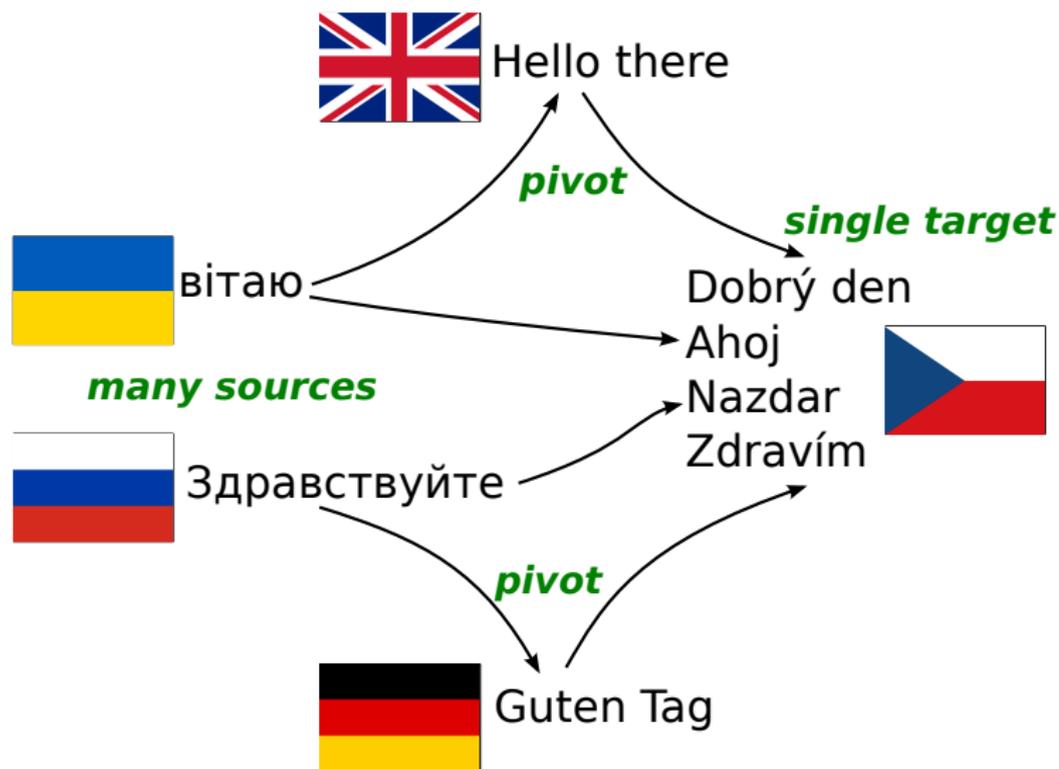
SCRIPTS

- A script to formalise phrase-tables of triangulation (pivot language)
    - Variables: A source-pivot phrase-table, a pivot-target phrase-table
    - Variables: Pruning options
    - Output: A source-target phrase-table
- A script to merge multiple phrase-tables into one phrase-table. (uniform distribution)

CONCLUSION

- ▶ The project is still under development.
- ▶ Requirements:
  - ▶ Perl
  - ▶ Moses toolkit experience