

Querying parallel corpora

Vojtěch Kovář

Lexical Computing Ltd.
`vojtech.kovar@sketchengine.co.uk`

SKEW 3
March 21, 2012

Outline

- 1 Parallel Corpora in the Sketch Engine
- 2 Querying Options
- 3 Future Directions

Parallel Corpora in the Sketch Engine

- Two (or more) separate corpora
 - separate data
 - display among other corpora in the list
 - aligned to each other
 - the “align” structure (usually sentence)
- Alignment enables special functionality
 - in the concordancer
 - no parallel sketches (yet)

Querying Options

- Query one part of the corpus
 - ordinary query
 - the system is able to display aligned segments in the concordance
- Querying both parts at the same time
 - e.g. “key” where the German equivalent is “Taste”
 - CQL special “within corpus: query” syntax
 - e.g. “key” within europarl5_de:”Taste”
 - more user-friendly interface?

Future Directions

- Parallel Word Sketches
 - guessing translations
 - matching probable translations of collocates
- More parallel corpora
 - currently in preparation