Parallel data

Parallel corpora:

- Europarl
- Movie subtitles
- Translated news, books
- Wikipedia (comparable)
- http://opus.lingfil.uu.se/

Lot's of problems with data:

- Noisy
- Specific domain
- Rare language pairs
- Not aligned, not enough