Questions

• What can MT do for patent translation?

• What can patent translation do for MT?
MT for Patent Translation

- Properties of patent translations
  - large scale task: open domain, large texts
  - some formulaic language ("whereas...", "apparatus", etc.)
  - large specialized terminology
  - several distinct domains
  - structure: format of claims, references to prior patents, figures
MT for Patent Translation

- Properties of patent translations
  - large scale task: open domain, large texts
  - some formulaic language ("whereas...", "apparatus", etc.)
  - large specialized terminology
  - several distinct domains
  - structure: format of claims, references to prior patents, figures

- Some properties require special treatment

- Automatic learning methods offer good solutions
Patent Translation for MT

- **Open source**: contributing factor to the advance of SMT
  - tools: GIZA++, BLEU, Moses, Berkeley word aligner, ...
  - data: Canadian Hansards, Europarl, LDC corpora, UN, Acquis, ...
Patent Translation for MT

- **Open source**: contributing factor to the advance of SMT
  - tools: GIZA++, BLEU, Moses, Berkeley word aligner, ...
  - data: Canadian Hansards, Europarl, LDC corpora, UN, Acquis, ...

- Patents are a publicly available resource
Patent Translation for MT

- **Open source**: contributing factor to the advance of SMT
  - tools: GIZA++, BLEU, Moses, Berkeley word aligner, ...
  - data: Canadian Hansards, Europarl, LDC corpora, UN, Acquis, ...

- Patents are a publicly available resource

- Making translated patent corpora widely available
  - more researchers will take on this problem
  - unique challenges of patent translation will be addressed
  - potentially largest source of data for some languages (Korean, Japanese)