Proper Syntax is Hard

See slides by Chiang (2010):

- The source and target trees constrain too much.
  $\Rightarrow$ Too few rules extracted.
  $\Rightarrow$ Coverage lost.
- Labelled non-terminals do not always match.
  $\Rightarrow$ Some valid translations not admissible.

Relaxation of the hard constraints needed:

- Allow breaking trees into smaller fragments (e.g. binarization).
- Allow attachment of non-matching non-terminals.