Whether to English more alike reordered suits Estonian input for SMT better?

Eleri Aedmaa
Krista Lin
Mark Fishel

Institute of CS, Uni. Tartu

Oct. 6, 2016; Baltic HLT
IDEA: PRE-REORDERING

- put input tokens in the output order
- do it to train/dev/test/new input
- less reordering for SMT to learn

... Profit!
An open debate on weblogs should be encouraged.
RELATIVE WORK

* Manually designed rules
  - e.g. Collins/Koehn/Kačerová '05

* Automatically learning to reorder
  - e.g. Genzel '10, Zhang/Zern/Nej '07
**Our Manual Rules**

1. OVS $\rightarrow$ SVO

2. infinV, finV $\rightarrow$ finV, infinV

3. Obj. after verb
EXPERIMENTS

* Phrase-based SMT
* Europarl, Est→Eng
* Baseline: 34.9 BLEU
RESULTS

* BLEU scores 0.3 - 0.8 lower

* Translation quality, order not better (checked manually)

* Single rules / combinations
WHY ???

- parser not reliable
- rule influence very narrow

What to do?
AUTOMATIC RULES

Inspired by TBL:

* try all rules like
  - bring child X before/after parent Y
  - drag grandchildren etc. with it

* choose the rule that results in least word alignment intersections
LIKE:

* put **SUBJ** before its **Vmod** parent

E.g. ...küsima, kuidas peab muutuma meie suhtumine...

Gloss: ask, how must change our attitude...
RESULTS

- BLEU scores 0.1 - 0.9 lower

- Manual analysis: some better, some worse (more)

- WHY ?????
ANALYSIS

1. Bad parser
   - constraint grammar
     - LAS: 0.67

Future: will use MaltParser/SyntaxNet
   (LAS: 0.79 – 0.8)
2. Word alignment intersections – not a good optimization objective
   - noisy
   - likes monotone alignment
An open debate on weblogs should be encouraged.
3. All rules have a low impact
   - narrow influence
   - very little to learn from
Future

- post-reordering
- statistical models
- parser vs no parser
Gracias!