

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

Eleri Aedmaa
Krista Lih
Mark Fishel

Institute of CS,
Uni. Tartu

Oct. 6, 2016; BalticHLT

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!

FOR EXAMPLE :

Toetada tuleks arutleda, veebipäevikuid käsitlevat arutelu

An open debate on weblogs should be encouraged

RELATED WORK

* Manually designed rules

- e.g. Collins/Koehn/Kučerová'05

* Automatically learning to
reorder

- e.g. Genzel'10, Zhang/Zew/Ney'07

OUR MANUAL RULES

1. OVS $\rightarrow$ SVO

2. inf[.] fin[.] V, fin[.] V $\rightarrow$ fin[.] V, inf[.] V

3. Obj. after verb

EXPERIMENTS

* Phrase-based SMT

| * Europarl, Est $\rightarrow$ 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 patent



E.g. ...küsisima, 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

Malt Parser / Syntax Net

(LAS: 0.79 – 0.8)

ANALYSIS

2. Word alignment intersections — not a good optimization objective

— noisy

— likes monotone alignment

WORD ALIGNMENT

Toetada tuleks arutleda, veebipäevikuid käsitlevat arutelu.

An open debate on weblogs should be encouraged.

ANALYSIS

3. All rules have a low impact

- narrow influence

- very little to learn from

FUTURE

— post-reordering

— statistical models

— parser vs no parser

Gracias!