Dependency-Based Bilingual Language Models for Reordering in SMT

Ekaterina Garmash **Christof Monz**

University of Amsterdam Informatics Institute, University of Amsterdam

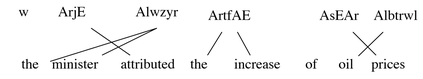
Contributions

- capture reordering as an order of a sequence of translation events
- characterize translation events with their source and target syntactic features
- simple alternative to tree-based models
- up to +0.98 BLEU improvement for Chinese-English and +0.4 BLEU improvement for Arabic-English over lexicalized BiLM

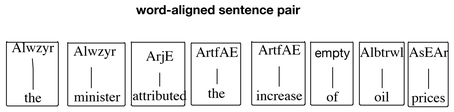
Motivation

Background: bilingual language models (BiLMs)

- n-gram model of sequences of elementary translation events
- elementary translation event — a pair of source and target words
- we adopt the definition of Niehues et al. (2011) of a **bilingual token**: (given word alignment) a target word and all the source words aligned to it



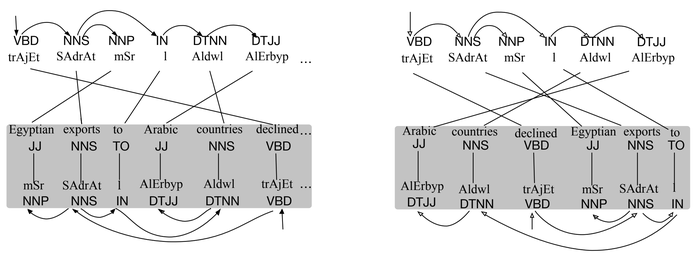
word-aligned sentence pair



corresponding sequence of bilingual tokens

Reordering with BiLMs

How well do various labelings differentiate between correct and incorrect reorderings?



correct reordering

incorrect reordering

expressiveness

- lexical-based BiLMs are too sparse to capture reordering regularities

vs.

generality

- Niehues et al. (2011): bilingual tokens substituted with the words' POS tags

Approach

Dependency-based BiLMs

- dependency grammar is commonly used in NLP to formalise role-based relations between words
- to label bilingual tokens, we try out combinations of different properties based on a source dependency parse
- generalized definition of a labeling for a bilingual token sequence $\mathrm{t}_1 \dots t_n$

$$\mathrm{t}_i = \langle e_i, \{ContF(f) | f \in A(e_i)\}\rangle$$

where e_i is the i-th target word, $\mathrm{A}: \mathrm{E} \rightarrow \mathcal{P}(F)$ is an alignment function, F and E — source and target sentences, ContF and ContE - **contextual functions**

Contextual functions

Return a word's sentential context (source or target).

We focus on ContF's, since they allow for a richer set of definitions in the MT setting (source side fully given) than ContE's.

Proposed contextual functions return:

- the word itself (designation: Lex)
- POS tag of the word (designation: Pos)
- POS tag of the word's parent (see below)
- POS tag of the word's grandparent (see below)

Notation

We use the proposed contextual functions in combinations to define individual BiLMs.

"●" horizontally connects source (left) and target (right) contextual functions

Lex●Lex is a BiLM with lexicalized tokens

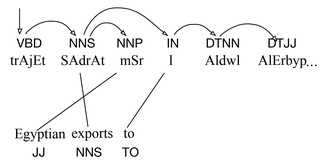
Pos●Pos is a BiLM with words in tokens substituteted with their POS tags

"→"connects parent (left) and child (right) from a dependency tree

Pos→Pos→Pos is a combination of functions returning the word's POS tag, its parent's POS tag and its grandparent's POS tag

Example

Sequence of bilingual tokens produced by a Pos→Pos→Pos●Lex after translating three words of the source sentence:



…

Implementation and Experiments

Implementation and integration into PBSMT decoder

- employed tools: Chinese Stanford dependency parser (Chang et a. 2009), Arabic Stanford constituency parser (Green and Manning, 2010) + extract dependencies based on Collins (1999), English Stanford POS-tagger (Toutanova et al., 2003)
- n-gram model training: 5-gram model with Kneser-Ney smoothing using SRILM (Stolke et al., 2011)
- dependency-based BiLMs are integrated as a feature in a log-linear model
- for each phrase pair, its most likely internal word alignment and target-side POS labelling is stored in the phrase table

Basic experimental setup

- phrase-based decoder
- distortion limit: 5
- lexicalized distortion model included in the log-linear interpolation
- compare performance of the original BiLMs (Niehues et al. 2011) and the dependency-based BiLMs

Statistical significance notation

▲ — significant improvement over Lex●Lex at $p < .01$

△ — significant improvement over Lex●Lex at $p < .05$

Arabic-English experiments

	MT08+MT09 test set BLEU		TER
PBSMT baseline	46.57		45.60
Lex●Lex	46.98		45.96
Pos→Pos●Pos	47.25△	+0.27	45.40▲
Pos→Pos→Pos●Pos	47.30△	+0.32	46.21▲
Lex●Lex + Pos→Pos→Pos●Pos	**47.38▲**	**+0.4**	45.63▲

Chinese-English experiments

	MT06+MT08 test set BLEU		TER
PBSMT baseline	28.99		59.14
Lex●Lex	29.69		58.72
Pos→Pos●Pos	29.78	+0.09	58.36▲
Pos→Pos→Pos●Pos	30.05▲	+0.36	58.54
Lex●Lex + Pos→Pos→Pos●Pos	**30.28▲**	**+0.59**	58.30▲

Reordering potential of the model

Distortion limit extended to 10 words.

	Arabic-English MT08+MT09 test set BLEU		Chinese-English MT06+MT08 test set BLEU	
		TER		TER
Lex●Lex	46.72	45.97	29.79	58.38
Pos→Pos→Pos●Pos	**47.12▲ +0.4**	45.52▲	**30.77▲ +0.98**	57.82▲

This research was funded in part by the Netherlands Organization for Scientific Research (NWO) under project numbers 639.022.213 and 612.001.218