Dependency-Based Phrase Alignment
Radu Ion, Alexandru Ceaușu, Dan Tufiş
Romanian Academy Research Institute for Artificial Intelligence
13, Calea 13 Septembrie, 050711, Bucharest
{radu, aceasu, tufis}@racai.ro

Introduction
Derived from the IBM influential work (Brown et al. 1993) in Statistical Machine Translation, the Weighted Finite State Translation Models (Och, Ney 2002, Kumar, Byrne, 2003) have a series of drawbacks: their low expressive power became visible when the parameter estimation has limited training data and computing resources.

Phrase alignment can be hardly achieved without a minimal knowledge on syntax and lexical collocation in a language. Of course, such system is still limited, it will not account for phenomena as relative order of noun phrases and adjectival phrases. For this kind of information, the language model should account for.

The dependency-based phrase alignment system we present in this paper does not require dependency annotation. Furthermore, it can work only on word forms but it performs better when annotation of part-of-speech is available.

Two stages are required for the completion of dependency-based phrase alignment: (i) the annotation of dependencies and (ii) dependency alignment.

Monolingual dependency annotation
The definition of the phrase as “group of adjacent words” is not suited for a dependency-based system that has to recover phrases that frequently contain non-adjacent words.

Unlike the chunking methods that limit themselves to finding labelled groups of adjacent words without specifying their internal structure, our system is committed to discover some (very regular) kind of group structure. When POS annotation is available, the system can also name the groups by the name of their heads: groups governed by a noun will be called noun phrases (NP), those governed by a preposition will become the prepositional phrases (PP), and so on.

The dependency discovery algorithm runs on a tokenized, part-of-speech (POS) annotated and lemmatized text. Specifically, we used the IBM model 1 to find undirected links between pairs ⟨POS, lemma⟩ (notation ⟨p,l⟩) of the same sentence. Training IBM1 on a large POS-annotated and lemmatized corpus, resulted in the POS and lemmas translation tables (t-tables). The Viterbi alignment of a new input sentence with respect to itself was computed by maximizing the sum $t(p_i \mid p_j) + t(l_i \mid l_j), \quad i \neq j$ for all pairs of wordforms $w_i$ and $w_j$ of each sentence, constraining the alignments to generate projective (Melčuk 1988) dependency trees. To build a proper tree structure we need to assign orientations to the undirected links found by the Viterbi alignment. We do that by recognizing some very general (linguistically universal) governor-dependant relations such as noun-adjective, preposition-noun, adjective-adverb, and so on. Of all the trees possible, we choose those with maximum alignment probabilities to give us the input sentence with its words grouped in tree structures.

Related work includes (Eisner 1996) and (Paskin 2001) which induce dependency linkages from annotated text and, (Riccardi et al. 1999) which uses chunking heuristics to
produce the same kind of linkage. Link grammar (Sleator, Temperley 1991) is a
formalism that uses named but undirected links between word forms in a sentence.

Our dependency finder differs from the ones mentioned above in that it does not aim
at a parse tree for the entire sentence (the IBM1 alone would not suffice for this job) and
in that it uses POS tags and word lemmas to find dependency links in word groups.
Furthermore, it makes use of raw text to estimate the model parameters and no grammar
induction is involved.

**Dependency alignment**

For phrase alignment, we differentiate between two kinds of features characterizing a
link: context independent and context dependent features. Context independent features,
such as translation equivalence or cognate scores, rely only on the lexical tokens (words,
phrases) paired by an alignment link. Context dependent features for a candidate link
contain information from the surrounding links (locality, links crossed, relative
position/distortion, etc.).

One of the features required by a word aligner is the translation equivalence of two
tokens. Word translation equivalence by itself does not cover phenomena as collocation,
verb phrase composition, etc. To cope with these requirements, the COWAL aligner
(Tuğş et al. 2005) defined the collocation feature, but only with respect to adjacent
lexical tokens. However, if one considers the lexical tokens, as well as the dependency
relations among them as being the basic units subject to the alignment process, they,
become altogether the atoms of the parameters estimation procedure. As such, the role of
the collocation parameter of COWAL or of the fertility parameter introduced in the IBM
models 3-5, is nicely taken over by the dependency relations. The computational
complexity of the usual N-M word aligners is severely reduced because the 1-1 alignment
of the dependency links, together with the 1-1 lexical tokens alignment links encode the
information necessary to compute the phrase alignments (the N-M alignment in the word
alignment jargon).

The aligner was trained on a 1.6 million words news corpus (TM1) and on a 16
million words subset of the Acquis Communautaire corpus (TM2). For the evaluation
of the system, we used the translation model TM1 and the gold standards provided in the
ACL (2003 and 2005) Romanian-English word alignment competitions.

The full paper will provide quantitative and qualitative evaluation results showing
superior performance versus the COWAL aligner, the best-rated word-aligner in the
ACL2005 Romanian-English shared task (Martin et al., 2005).

**Conclusions**

We described a dependency-based phrase aligner which at the price of additional memory
requirements for storing the dependencies translation equivalence table (approximately
twice larger than the table for word translation equivalence) needs much less training
time as IBM models 4/5 and produces better results than our previous word aligners.

**References:**

1. [Och, Ney 2002] F. J. Och, H. Ney. Discriminative Training and Maximum Entropy
Models for Statistical Machine Translation. Proceedings of the 40th Annual Meeting
of the Association for Computational Linguistics", pp. 295-302, Philadelphia, PA, July 2002

2. [Shankar, Byrne 2003] K., Shankar and W. Byrne. 2003. A Weighted Finite State Transducer Implementation of the Alignment Template Model for Statistical Machine Translation. Proceedings of the Human Language Technology Conference and the North American Association for Computational Linguistics (HLT/NAACL), pages 63-70, Edmonton, Canada.

3. [Gildea, 2003] D. Gildea. 2003. Loosely Tree-based Alignment for Machine Translation. In Proceedings of the 41th Annual Conference of the Association for Computational Linguistics (ACL-03), pages 80-87, Sapporo, Japan.

4. [Brown et al. 1993] P. F. Brown, S. A. D. Pietra, V. J. D. Pietra, and R. L. Mercer. The Mathematics of Statistical Machine Translation: Parameter Estimation. Computational Linguistics, vol. 19, pp. 263-311, June 1993.

5. [Eisner 1996] J. Eisner. Three New Probabilistic Models for Dependency Parsing: An Exploration. In Proceedings of COLING-96, pages 340-245, Copenhagen, Denmark, 1996.

6. [Martin et al., 2005] J. Martin, R. Mihalcea, T. Pedersen. Word Alignment for Languages with Scarce Resources, Proceedings of the ACL Workshop on Building and Using Parallel Texts, June, 2005, Ann Arbor, Michigan, Association for Computational Linguistics, pages 65-74

7. [Melčuk 1988] I. A. Melčuk. Dependency Syntax: Theory and Practice. State University of New York Press, Albany, NY, 1988.

8. [Paskin 2001] M. Paskin. Grammatical Bigrams. In T. Dietterich, S. Becker, and Z. Gharahmani eds., Advances in Neural Information Processing Systems 14. MIT Press: Cambridge, MA.

9. [Riccardi et al. 1999] G. Riccardi, S. Bangalore, P.D. Sarin, Learning Head-Dependency Relations from Unannotated Corpora. Proceedings of the IEEE Automatic Speech Recognition and Understanding Workshop, Colorado, 1999.

10. [Sleator, Temperley 1991] D. D. Sleator and D. Temperley. Parsing English with a Link Grammar. In Proceedings of the Third International Workshop on Parsing Technologies, pages 277-292.

11. [Tușiș et al. 2005] D. Tușiș, R. Ion, Al. Ceaușu, D. Ștefănescu, Combined Word Alignments, Proceedings of the ACL Workshop on Building and Using Parallel Texts, June, 2005, Ann Arbor, Michigan, Association for Computational Linguistics, pages 107-110.