F-Structure Transfer-Based Statistical Machine Translation

Yvette Graham **Anton Bryl** **Josef van Genabith**
National Centre for Language Technology
School of Computing
Dublin City University
Dublin 9, Ireland
ygraham,abryl,josef@computing.dcu.ie

1 Abstract

Statistical Machine Translation (SMT) using deep linguistic representations for transfer is a relatively new and growing research area (Bojar and Hajič, 2008; Graham and van Genabith, 2008; Ding and Palmer, 2006; Riezler and Maxwell, 2006; Cmejrek et al., 2003; Eisner, 2003; Hajič et al., 2002; Alshawi et. al., 2000). In Transfer-Based SMT, among the different types of intermediate structures used for transfer are dependency structures. Bojar and Hajic (2008) use the Functional Generative Description (FGD) (Sgall et al., 1986) Tectogrammatical Layer (T-layer), labelled ordered dependency trees, while Riezler and Maxwell (2006) use the Lexical Functional Grammar (LFG) (Kaplan and Bresnan, 1982; Bresnan, 2001; Dalrymple, 2001) Functional Structure (F-structure), an attribute-value structure encoding of bilexical labelled dependencies and atomic valued features.

1.1 Transfer-Based SMT

Transfer-Based SMT is composed of three parts, i) parsing to linguistic structure, ii) transfer from SL linguistic structure to TL linguistic structure and iii) generation of TL sentence. Each of the three steps uses a statistical model to select the best or n-best output. For parsing, a disambiguation model is used to rank the parses and select the n-best output parses. Decoding (transfer) is then done on a parse structure (or n-best list of parse structures) via beam-search producing an n-best list of TL structures. For generation, the input is an n-best list of TL structures and all possible TL sentences are produced. The best TL sentence is then selected using an n-gram language model.

1.2 Fstructure-Based Transfer Rule Acquisition

In this paper, we describe a Transfer-Based SMT system that, like Riezler and Maxwell (2006), uses the LFG F-structure as the intermediate representation for transfer and is trained fully automatically on LFG-parsed bilingual corpora. Riezler and Maxwell (2006) used SMT phrases to hypothesize candidate transfer rules. We take a different approach to rule induction and hypothesize rules only using information from the f-structures, as opposed to using SMT phrases based on the surface form strings. We do this by firstly establishing a set of one-to-one correspondences between local f-structures, [1] before inducing all rules consistent with these alignments.

1.3 Experimental Evaluation

The transfer rule acquisition component uses context variables (Maxwell and Kaplan, 1991) to encode large numbers of rules in a compact represenation and has been fully implemented and tested on Europarl (Koehn et al., 2005) and Newswire corpora. We plan to include an experimental evaluation of the complete machine translation system in which we investigate the effects of using two different disambiguation models for parsing the data used by the SMT system. Riezler and Maxwell (2006) tested their machine translation system using an English disambiguation model for both sides of the training corpus. If a single disambiguation model is used, the f-structures of a given pair of training sentences are likely to be quite similar, and this may help the rule induction process. However, another approach is to use language-specific disambiguation models for parsing. In this case, it is more likely that the *actual* best f-structure for each sentence of the training data is selected. The rule induction algorithm must then induce rules from the output structures regardless of how dissimilar they are. We conduct an empirical investigation into which

[1] We currently carry out node alignment as follows: Giza++ (Och et al., 1999) is run in both language directions on a lemmatized version of the bitext training corpus and the intersection of the twos sets of alignments is gotten using Moses (Koehn et al., 2007).

approach achieves better machine translation output by training and testing the machine translation system using i) an English disambiguation model (Kaplan et al., 2004; Riezler et al., 2002) to select the best parse for both German and English sentences, and compare with results when ii) a German disambiguation model (Forst, 2007) is used for selecting the best German parse and an English disambiguation model (Kaplan et al., 2004; Riezler et al., 2002) is used to select the best parse for the English sentences. For the evaluation the system is trained on the Europarl (Koehn et al., 2005) and Newswire parallel corpora and tested on held-out data.

References

Hiyan Alshawi, Srinivas Bangalore and Shona Douglas. 2000. Learing Dependency Translation Models as Collections of Finite State Head Transducers. In *Computational Linguistics*, 26(1):45-60.

Ondřej Bojar and Jan Hajič. 2008. Phrase-Based and Deep Syntactic English-to-Czech Statistical Machine Translation. In *Proceedings of the third Workshop on Statistical Machine Translation*, Columbus, Ohio, June 2008.

Joan Bresnan. 2001. *Lexical-Functional Syntax.*, Blackweelm Oxford, 2001.

Miriam Butt, Helge Dyvik, Tracy H. King, Hiroshi Masuichi and Christian Rohrer. 2002. The Parallel Grammar Project. (grammar version 2005) In *Proceedings of the 19th International Conference on Computational Linguistics (COLING'02), Workshop on Grammar Engineering and Evaluation*, pages 1-7. Tapei, ROC.

Martin Cmejrek, Jan Curn and Jir Havelka 2003 Czech-English Dependency Tree-based Machine Translation In *Proceedings of EACL 2003*, pages 83-90.

Mary Dalrymple. *Lexical Functional Grammar*, Academic Press, San Diego, CA; London. 2001.

Yuan Ding and Martha Palmer. 2006. Better Learning and Decoding for Syntax Based SMt Using PSDIG. In *Proceedings of AMTA 2006*

Jason Eisner. 2003. Learning non-isomorphic tree mappings for machine translation. In *Proceedings of the 41st Annual Meeting of the Association of Computational Linguistics (ACL)*, pages 205-208, Sappora, July 2003.

Martin Forst. 2007. Disambiguation for a Linguistically Precise German Parser. University of Stuttgart.

Yvette Graham and Josef van Genabith. 2008. Packed Rules for Automatic Transfer Rule Induction. In *Proceedings of the European Association of Machine Translation Conference 2008*, Hamburg, Germany.

Jan Hajič, Martin Čmejrek, Bonnie Dorr, Yuan Ding, Jason Eisner, Daniel Gildea, Terry Koo, Kristin Parton, Gerald Penn, Dragomir Radev and Owen Rambow. 2002. Natural Language Generation in the Context of Machine Translation. Technical Report. *NLP WS'02*, final report.

Ronald M. Kaplan, Stefan Riezler, Tracy H. King, John T. Maxwell, and Alexander Vasserman. 2004 Speed and accuracy in shallow and deep stochastic parsing. In *Proceedings of Human Language Technology Conference/North American Chapter of the Association for Computational Linguistics Meeting*, Boston, MA, May 2-7 2004.

Ronald M. Kaplan, Tracy H. King and John T. Maxwell. 2002. Adapting existing grammars: the XLE experience. In *Proceedings of COLING 2002*, Taipei, Taiwan.

Ronald Kaplan and Joan Bresnan. 1982. Lexical Functional Grammar, a Formal System for Grammatical Represenation. In Bresnan, J. editor, *The Mental Representation of Grammatical Relations*, pages 173-281, MIT Press, Cambridge, MA.

Philipp Koehn, Franz Josef Och and Daniel Marcu. 2003. Statistical Phrase-based Translation. In *Proceedings of the HLT-NAACL 2003*, pages 48-54, Edmonton, May/June 2003.

Philipp Koehn. 2005. Europarl: A Parallel Corpus for Statistical Machine Translation. In *Proceedings of MT Summit 2005*

Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison Burch, Richard Zens, Alexandra Constantin, Marcello Federico, Nicola Bertoldi, Chris Dyer, Evan Herbst, Brooke Cowen, Wade Shen, Christine Moran and Ondřej Bojar. 2007. Moses: Open Source Toolkit for Statistical Machine Translation In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*

essing and Computational

John T. Maxwell III and Ronald M. Kaplan. 1991. A Method for Disjunctive Constraint Satisfaction. In *Current Issues in Parsing Technology*, Masaru Tomita editor, pages 173-190, Kluwer Aca demic Publishers.

Franz Josef Och, Christoph Tillmann Hermann and Ney. 2000. Improved Alignment Models for Statistical Machine Translation. In *Proceedings of the 1999 Conference on Empirical Methods in Natural Language Processing* (EMNLP'99). College Park, MD, pages 20-28.

Stefan Riezler, Tracy H. King, Ronald M. Kaplan, Richard Crouch, John T. Maxwell, and Mark Johnson. 2002. Parsing the Wall Street Journal using Lexical Functional Grammar and discriminitive estimation techniques . In *Proceedings of the 40th Annual Meeting of the Association of Computational Linguistics (ACL)*, Philadelphia, July 2002.

Stefan Riezler and John T. Maxwell III. 2006. Grammatical Machine Translation. In *Proceedings of HLT-ACL*, pages 248-255, New York.

Christian Rohrer and Martin Forst. 2006. Improving Coverage and Parsing Quality of a Large-scale LFG for German. In *Proceedings of the Language Resources and Evaluation Conference (LREC-2006)*, Genoa, Italy.

Petr Sgall, Eva Hajicova and Jarmilla Panevova. 1986. *The Meaning of the Sentence and its Semantic and Pragmatic Aspects*. Dordrecht: Reidel and Prague: Academia 1986.