

# Book Review

## Syntax-Based Statistical Machine Translation

**Philip Williams, Rico Sennrich, Matt Post, Philipp Koehn**

(University of Edinburgh, University of Edinburgh, Johns Hopkins University, Johns Hopkins University)

Morgan & Claypool (Synthesis Lectures on Human Language Technologies, edited by Graeme Hirst, volume 33), 2016, xvii+190 pp; paperback, ISBN 978-1-62705-900-8; ebook, ISBN 978-1-62705-502-4; doi:10.2200/S00716ED1V04Y201604HLT033, \$70

*Reviewed by*

*Christian Hadiwinoto*

*National University of Singapore*

In its early development, machine translation adopted rule-based approaches, which can include the use of language syntax. The late 1980s and early 1990s saw the inception of the statistical machine translation (SMT) approach, where translation models can be learned automatically from a parallel corpus rather than created manually by humans. Initial SMT models were word-based and phrase-based, without the use of syntactic knowledge. In phrase-based SMT, a source sentence is first segmented into phrases and then translated phrase-by-phrase with some reordering of the translated phrases in the target sentence. This has posed challenges when translating between two syntactically different languages. Syntax-based SMT approaches take advantage of syntactic knowledge within the framework of SMT. This book provides an introduction to syntax-based SMT approaches. It is a valuable resource for those who are interested in syntax-based SMT.

The book consists of seven chapters. There is not an introduction chapter in this book, aside from the preface, which can be considered as a brief introduction. Readers are referred to Koehn (2010) for background knowledge. I think an introduction chapter categorized into sections would have been useful, before proceeding to describe the various models. The first two chapters provide principles applicable across various syntax-based SMT approaches. The next three chapters describe syntax-based SMT decoding in detail; this constitutes half of the book. Selected extended topics are provided in the next chapter, which is followed by a concluding chapter.

Chapter 1 describes the models and formalisms applicable to syntax-based SMT. The first section describes the phrasal translation units in phrase-based SMT, its limitations, and how tree structures address the limitations of the phrase-based approach. This explanation is useful as translation units are the key difference between the phrase-based and syntax-based SMT approaches. The next two sections describe the grammar formalisms and the statistical models that define syntax-based SMT. The section that covers the grammar formalisms (i.e., synchronous context-free grammar [SCFG] and synchronous tree-substitution grammar [STSG]), would have been clearer if their differences were presented in a side-by-side illustrating example. The remainder of the chapter discusses different categories of syntax-based SMT approaches and the history of these approaches, which include string-to-string, string-to-tree, tree-to-string, and

---

doi:10.1162/COLLr.00303

© 2017 Association for Computational Linguistics

Published under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) license

tree-to-tree SMT approaches. Although the syntax-based translation model in Galley et al. (2006) falls under the string-to-tree category, I wonder why hierarchical phrase-based SMT, or Hiero (Chiang 2007), is not explicitly put under the string-to-string category, since Hiero also uses “unlabeled hierarchical phrases where there is no representation of linguistic categories.”

Chapter 2 focuses on how the statistical framework of a syntax-based SMT approach learns its model from a word-aligned and parsed parallel text. The first section explains how phrase pairs are extracted as translation rules from a word-aligned sentence pair in phrase-based SMT (Koehn, Och, and Marcu 2003), highlighting the definition of a phrase as a sequence of words and the alignment-consistency property of a phrase pair as defined in Och and Ney (2004). The remainder of the chapter introduces three predominant instantiations of syntax-based models: hierarchical phrase-based SMT (Hiero) (Chiang 2007), which is a non-labeled syntax-based SMT approach arising from the phrase-based approach; syntax-augmented machine translation (SAMT), which introduces the notion of soft labels while keeping the nonlinguistic phrase notion; and GHKM (Galley et al. 2004), which only extracts translation rules consistent with constituency parse subtrees. This chapter is nicely organized and it is easy to follow the gradual evolution from phrase-based SMT to GHKM.

Chapter 3 introduces the decoding formalism in the form of a directed hypergraph, defined as a set of vertices and a set of directed hyperedges. The first section introduces the notion of a weighted parse forest represented in a weighted hypergraph, representing alternative parse trees of a sentence. I found it important to pay careful attention to this section, in order to understand the next section and the following chapters. The next section presents various algorithms on a hypergraph to translate a sentence in a hypergraph representation of possible tree derivations. Overall, I found this chapter to contain many technical details. The last section of this chapter provides historical notes on the sources of these concepts. This chapter needs to be read before the next chapter, which assumes understanding of the concepts introduced in Chapter 3.

Chapter 4 describes tree decoding—that is, decoding with the constituency parse tree of a source sentence as its input, focusing on the tree-to-string approach. The first two sections highlight decoding with local and non-local features, where non-local features accommodate  $n$ -gram language models and are more complex than local features. The next section is devoted to an in-depth description of a beam search algorithm on the parse tree of a source sentence. The description could have been improved if the running example showed the decoding steps. The next two sections present extensions to the concepts introduced in the earlier part of this chapter, by providing references to more efficient hypergraph operations. The content of this section requires readers who are interested in implementing an efficient tree-based algorithm to go through the cited references. Brief historical notes conclude this chapter nicely, by pointing to relevant materials for further reading.

Chapter 5 describes string decoding with a source sentence string as its input. The first two sections describe beam search decoding algorithms in a binary SCFG, namely, a maximum of two non-terminal symbols on the right-hand side of each rule, adopted in Hiero and SAMT. The algorithms covered are a basic algorithm and an optimized algorithm. The complexity comparison between the two is nicely presented here, emphasizing the complexity reduction achieved by algorithm optimization. The handling of non-binary rules is described in the following section, illustrated by GHKM rule extraction. A mid-chapter summary section divides this chapter into two parts: beam search decoding and parsing. The second part describes parsing algorithms in the context of shared-category SCFG, assuming the same set of non-terminal symbols

for the left-hand and right-hand sides of a rule, followed by a section extending the algorithm to STSG and distinct-category SCFG. The organization of this chapter is excellent. However, I feel that the inclusion of distinct-category SCFG decoding does not fit well into this chapter, as string decoding in string-to-tree SMT requires no knowledge of the source syntax. The historical notes also do not provide any references of prior work on string decoding using distinct-category SCFG.

Chapter 6 contains various selected topics on syntax-based SMT. The first section discusses tree transformations, which make translation rule learning more effective. The description of non-context-free models serves as a prelude to the next section on dependency-based SMT, which covers dependency treelet (equivalent to the tree-to-string approach) and string-to-dependency (equivalent to the string-to-tree approach). The next section focuses on the ability of syntax-based SMT to have a more grammatical output compared with phrase-based SMT, although there is still room for improvement, including the use of unification grammars and semantic properties. Finally, the last section of this chapter explains how MT evaluation benefits from syntax-based SMT principles. Overall, this chapter enriches readers' knowledge beyond basic syntax-based SMT in the earlier chapters. I would also suggest the inclusion of phrase-based decoding approaches that use syntax-based features (Cherry 2008; Chang et al. 2009).

Chapter 7 nicely concludes this book by discussing the comparison between phrase-based and syntax-based SMT approaches and proposing possible future developments of syntax-based SMT. The chapter also highlights that syntax-driven MT predates statistical MT, as I mentioned at the beginning of this review.

Overall, I found this book to be a useful reference book for those interested in syntax-based SMT. The book is well organized, which makes it easy for readers to refer to specific aspects of syntax-based SMT. An improvement can be made to the presentation of ideas in this book. Throughout the book, there are many technical keywords, resulting from the complexity of syntax-based SMT. It would be useful to highlight these keywords in a side bar to remind readers that they are important keywords. In addition, although examples are given throughout the book, it would be even more useful to use these examples to illustrate how the algorithms work, so that readers can gain a better understanding of the algorithms.

## References

- Chang, Pi Chuan, Huihsin Tseng, Dan Jurafsky, and Christopher D. Manning. 2009. Discriminative reordering with Chinese grammatical relations features. In *Proceedings of the Third Workshop on Syntax and Structure in Statistical Translation*, pages 51–59, Boulder, CO.
- Cherry, Colin. 2008. Cohesive phrase-based decoding for statistical machine translation. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 72–80, Columbus, OH.
- Chiang, David. 2007. Hierarchical phrase-based translation. *Computational Linguistics*, 33(2):201–228.
- Galley, Michel, Jonathan Graehl, Kevin Knight, Daniel Marcu, Steve DeNeefe, Wei Wang, and Ignacio Thayer. 2006. Scalable inference and training of context-rich syntactic translation models. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics*, pages 961–968, Sydney.
- Galley, Michel, Mark Hopkins, Kevin Knight, and Daniel Marcu. 2004. What's in a translation rule? In *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics: HLT/NAACL 2004*, pages 273–280, Boston, MA.
- Koehn, Philipp. 2010. *Statistical Machine Translation*. Cambridge University Press.
- Koehn, Philipp, Franz Josef Och, and Daniel Marcu. 2003. Statistical phrase-based

translation. In *Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, pages 48–54, Edmonton.

Och, Franz Josef and Hermann Ney. 2004. The alignment template approach to statistical machine translation. *Computational Linguistics*, 30(4): 417–449.

*Christian Hadiwinoto* is a research fellow at the National University of Singapore. He received his Ph.D. degree from the Department of Computer Science, National University of Singapore. His research interests are in statistical machine translation and natural language processing. Hadiwinoto's e-mail address is chrhad@comp.nus.edu.sg.