Extended Tree Transducers in Natural Language Processing*

Andreas Maletti
Institut for Natural Language Processing, Universität Stuttgart
Pfaffenwaldring 5b, 70569 Stuttgart, Germany
maletti@ims.uni-stuttgart.de

Extended Abstract

Tree transducers are finite-state devices computing relations on trees. Their study was initiated by Thatcher (1970) and Rounds (1970), who established the classical top-down tree transducers that process the input tree from the root towards the leaves. Shortly afterwards, Baker (1973) introduced the bottom-up tree transducers that process the input tree from the leaves towards the root in analogy to the top-down and bottom-up tree automata (Thatcher, 1973). Due to applications in syntax-directed semantics (Fülöp and Vogler, 1998), tree transducers were extensively studied in the following years as detailed in (Gécseg and Steinby, 1984) and (Gécseg and Steinby, 1997). Notable extensions to the original top-down tree transducers include the top-down tree transducers with regular look-ahead by Engelfriet (1977), the attributed tree transducers of Fülöp (1981), and the macro tree transducers by Courcelle and Franchi-Zannettacci (1982) and Engelfriet and Vogler (1985).

In statistical machine translation (Koehn, 2009), syntax-based models (Chiang, 2010) [i.e., models that translate from or to syntax trees] have recently seen a lot of progress. It was identified already by Eisner (2003) that the classical linear top-down and bottom-up tree transducers cannot properly handle phenomena (such as rotation) that occur during the translation between natural languages. This result was presented for synchronous context-free grammars [SCFG] (Chiang, 2006), which is a formalism similar in spirit to (and essentially equally expressive as) the syntax-directed translation schemata by Aho and Ullman (1969), which were later refined to the more general bimorphism approach by Arnold and Dauchet (1982). Instead Eisner (2003) proposes synchronous tree substitution grammars [STSG], which are a restriction of the synchronous tree adjoining grammars [STAG] by Shieber and Schabes (1990). Indeed, STSG have a natural finite-state correspondence with the linear and complete bimorphisms (Arnold and Dauchet, 1982), which did not receive much interest from the theoretical computer science community in the sequel. Only with the advent of related models proposed by Galley et al. (2004) and corresponding translation systems following Graehl and Knight (2004), a systematic study of those finite-state tree transformers, now called extended [top-down] tree transducers, was initiated by Maletti et al. (2009).

Initial results for extended tree transducers showed that several results for classical tree transducers are no longer valid in the extended setting. For example, the composition results for classical tree transducers, derived by Engelfriet (1975) and Baker (1979), no longer hold for extended tree transducers, which was actually already confirmed in the seminal paper by Arnold and Dauchet (1982). However, essentially all reasonable extended tree transducers can capture important linguistic transformations (such as rotation) without the help of copying. Knight and Graehl (2005) provide a nice overview and establish requirements for an ideal syntax-based translation model. Following this call, extended tree transducers were thoroughly investigated and their basic properties have been established. In this invited talk, we recall the extended tree transducer model and summarize the results reported by Maletti et al. (2009) and Maletti (2011a) enriched with the recent composition closure results of Fülöp and Maletti (2013).

However, none of the extended tree transducer models that can handle rotation is closed under composition (Maletti et al., 2009). Consequently, May et al. (2010) developed an on-the-fly composition approach that can efficiently evaluate chains

*This research was financially supported by the German Research Foundation (DFG) grant MA / 4959 / 1-1.

of multiple extended tree transducers. With the hope that the composition closure is achieved at a low level, the composition hierarchy of most relevant extended tree transducer classes has been investigated in (Fülöp and Maletti, 2013), where we confirmed that several important composition hierarchies are actually infinite. As an alternative to extended tree transducers, we proposed another model originally proposed already by Arnold and Dauchet (1982). This model, nowadays known as multi bottom-up tree transducers [MBOT] following the nomenclature of Fülöp et al. (2004) and Fülöp et al. (2005), offers closure under composition (Engelfriet et al., 2009), which removes the need to evaluate chains of models. Maletti (2010) summarizes the advantages following the requirements set by Knight and Graehl (2005), and Maletti (2011b) provides an automatic extraction of an MBOT translation model from the usual training data for syntax-based translation models. This tree-to-tree model was subsequently implemented by Braune et al. (2013) in the syntax-based component (Hoang et al., 2009) of the statistical machine translation framework Moses (Koehn et al., 2007). In addition, a string-to-tree MBOT model was developed and evaluated by Seemann et al. (2015a). Finally, several related MBOT models were evaluated by Seemann et al. (2015b) and compared to state-of-the-art models in statistical machine translation. The invited talk will also recall those efforts and provide the latest developments and evaluations.

References

Alfred V. Aho and Jeffrey D. Ullman. 1969. Syntax directed translations and the pushdown assembler. *J. Comput. System Sci.*, 3(1):37–56.

André Arnold and Max Dauchet. 1982. Morphismes et bimorphismes d'arbres. *Theoret. Comput. Sci.*, 20(1):33–93.

Brenda S. Baker. 1973. Tree transductions and families of tree languages. In *Proc. 5th STOC*, pages 200–206. Association for Computing Machinery.

Brenda S. Baker. 1979. Composition of top-down and bottom-up tree transductions. *Inform. and Control*, 41(2):186–213.

Fabienne Braune, Nina Seemann, Daniel Quernheim, and Andreas Maletti. 2013. Shallow local multi bottom-up tree transducers in statistical machine translation. In *Proc. 51st ACL*, pages 811–821. Association for Computational Linguistics.

David Chiang. 2006. An introduction to synchronous grammars. In *Proc. 44th ACL*. Association for Computational Linguistics. Part of a tutorial given with Kevin Knight.

David Chiang. 2010. Learning to translate with source and target syntax. In *Proc. 48th ACL*, pages 1443–1452. Association for Computational Linguistics.

Bruno Courcelle and Paul Franchi-Zannettacci. 1982. Attribute grammars and recursive program schemes. *Theoret. Comput. Sci.*, 17(2–3):163–191 & 235–257.

Jason Eisner. 2003. Learning non-isomorphic tree mappings for machine translation. In *Proc. 41st ACL*, pages 205–208. Association for Computational Linguistics.

Joost Engelfriet and Heiko Vogler. 1985. Macro tree transducers. *J. Comput. System Sci.*, 31(1):71–146.

Joost Engelfriet, Eric Lilin, and Andreas Maletti. 2009. Composition and decomposition of extended multi bottom-up tree transducers. *Acta Inform.*, 46(8):561–590.

Joost Engelfriet. 1975. Bottom-up and top-down tree transformations—a comparison. *Math. Systems Theory*, 9(3):198–231.

Joost Engelfriet. 1977. Top-down tree transducers with regular look-ahead. *Math. Systems Theory*, 10(1):289–303.

Zoltán Fülöp and Andreas Maletti. 2013. Composition closure of ε-free linear extended top-down tree transducers. In *Proc. 17th DLT*, volume 7907 of LNCS, pages 239–251. Springer.

Zoltán Fülöp and Heiko Vogler. 1998. *Syntax-Directed Semantics—Formal Models Based on Tree Transducers*. EATCS Monographs on Theoret. Comput. Sci. Springer.

Zoltán Fülöp, Armin Kühnemann, and Heiko Vogler. 2004. A bottom-up characterization of deterministic top-down tree transducers with regular look-ahead. *Inf. Process. Lett.*, 91(2):57–67.

Zoltán Fülöp, Armin Kühnemann, and Heiko Vogler. 2005. Linear deterministic multi bottom-up tree transducers. *Theoret. Comput. Sci.*, 347(1–2):276–287.

Zoltán Fülöp. 1981. On attributed tree transducers. *Acta Cybernet.*, 5(3):261–279.

Michel Galley, Mark Hopkins, Kevin Knight, and Daniel Marcu. 2004. What's in a translation rule? In *Proc. NAACL*, pages 273–280. Association for Computational Linguistics.

Ferenc Gécseg and Magnus Steinby. 1984. *Tree Automata*. Akadémiai Kiadó, Budapest.

Ferenc Gécseg and Magnus Steinby. 1997. Tree languages. In Grzegorz Rozenberg and Arto Salomaa, editors, *Handbook of Formal Languages*, volume 3, chapter 1, pages 1–68. Springer.

Jonathan Graehl and Kevin Knight. 2004. Training tree transducers. In *Proc. NAACL*, pages 105–112. Association for Computational Linguistics.

Hieu Hoang, Philipp Koehn, and Adam Lopez. 2009. A unified framework for phrase-based, hierarchical, and syntax-based statistical machine translation. In *Proc. 6th IWSLT*, pages 152–159. National Museum of Emerging Science and Innovation, Tokyo, Japan.

Kevin Knight and Jonathan Graehl. 2005. An overview of probabilistic tree transducers for natural language processing. In *Proc. 6th CICLing*, volume 3406 of LNCS, pages 1–24. Springer.

Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondrej Bojar, Alexandra Constantin, and Evan Herbst. 2007. Moses: Open source toolkit for statistical machine translation. In *Proc. 45th ACL*, pages 177–180. Association for Computational Linguistics.

Philipp Koehn. 2009. *Statistical Machine Translation*. Cambridge University Press.

Andreas Maletti, Jonathan Graehl, Mark Hopkins, and Kevin Knight. 2009. The power of extended top-down tree transducers. *SIAM J. Comput.*, 39(2):410–430.

Andreas Maletti. 2010. Why synchronous tree substitution grammars? In *Proc. NAACL*, pages 876–884. Association for Computational Linguistics.

Andreas Maletti. 2011a. *Survey:* Weighted extended top-down tree transducers — Part II: Application in machine translation. *Fundam. Inform.*, 112(2–3):239–261.

Andreas Maletti. 2011b. How to train your multi bottom-up tree transducer. In *Proc. 49th ACL*, pages 825–834. Association for Computational Linguistics.

Jonathan May, Kevin Knight, and Heiko Vogler. 2010. Efficient inference through cascades of weighted tree transducers. In *Proc. 48th ACL*, pages 1058–1066. Association for Computational Linguistics.

William C. Rounds. 1970. Mappings and grammars on trees. *Math. Systems Theory*, 4(3):257–287.

Nina Seemann, Fabienne Braune, and Andreas Maletti. 2015a. String-to-tree multi bottom-up tree transducers. In *Proc. 53rd ACL*, pages 815–824. Association for Computational Linguistics.

Nina Seemann, Fabienne Braune, and Andreas Maletti. 2015b. A systematic evaluation of MBOT in statistical machine translation. In *Proc. 15th MT-Summit*. Association for Machine Translation in the Americas. to appear.

Stuart M. Shieber and Yves Schabes. 1990. Synchronous tree-adjoining grammars. In *Proc. 13th CoLing*, pages 253–258.

James W. Thatcher. 1970. Generalized^2 sequential machine maps. *J. Comput. System Sci.*, 4(4):339–367.

James W. Thatcher. 1973. Tree automata—an informal survey. In *Currents in the Theory of Comput.*, pages 143–172. Prentice Hall.