Probabilistic Synchronous Tree-Adjoining Grammars for Machine Translation: The Argument from Bilingual Dictionaries

Stuart M. Shieber
School of Engineering and Applied Sciences
Harvard University
Cambridge, MA 02138
shieber@seas.harvard.edu

Abstract

We provide a conceptual basis for thinking of machine translation in terms of synchronous grammars in general, and probabilistic synchronous tree-adjoining grammars in particular. Evidence for the view is found in the structure of bilingual dictionaries of the last several millennia.

1 Introduction

In this paper, we provide a conceptual basis for thinking of machine translation in terms of synchronous grammars in general, and probabilistic synchronous tree-adjoining grammars in particular. The basis is conceptual in that the arguments are based on generalizations about the translation relation at a conceptual level, and not on empirical results at an engineering level. Nonetheless, the conceptual idea is consistent with current efforts in MT, and in fact may be seen as underlying so-called syntax-aware MT.

We will argue that the nature of the translation relation is such that an appropriate formalism for realizing it should have a set of properties — expressivity, trainability, efficiency — that we will characterize more precisely below. There may be multiple formalisms that can achieve these ends, but one, at least, is probabilistic synchronous tree-adjoining grammar, and to our knowledge, no other qualitatively distinct formalism has been argued to display all of the requisite properties.

Below, we will discuss the various properties, with particular attention to an examination of a particular source of data about the translation relation, namely bilingual dictionaries. Multilingual lexicography has a history of some four millennia or more. In that time, a great deal of knowledge about particular translation relations has been explicitly codified in multilingual dictionaries. More interestingly for our present purposes, multilingual dictionaries through their own structuring implicitly express information about translation relations in general.

In Section 2, we introduce the Construction Principle, a property of the translation relation implicit in the structure of bilingual dictionaries throughout their four millennium history. Section 3 provides a review of synchronous tree-adjoining grammars showing that this formalism directly incorporates the Construction Principle and allows the formal implementation of bilingual dictionary relations. In Section 4, we argue that the probabilistic variant of STAG (PSTAG) inherits the expressivity advantages of STAG while adding the trainability of statistical MT. Section 5 concerns the practical efficacy of STAG. We conclude (Section 6) with an overall proposal for the use of PSTAG in a statistical MT system. By virtue of its fundamentality to the modeling of the translation relation, PSTAG or its formal relatives merits empirical examination as a basis for statistical MT.

2 Expressivity

Of course, a formalism for describing the translation relation must be able to capture the relations between words in the two languages: *acqua* means *water*, *dormire* means *sleep*, and so forth. Indeed, the stereotype of a bilingual dictionary is just such a relation; the HarperCollins Italian College Dictionary (HCICD) (Clari and Love, 1995) contains en-

Proceedings of SSST, NAACL-HLT 2007 / AMTA Workshop on Syntax and Structure in Statistical Translation, pages 88–95,
Rochester, New York, April 2007. ©2007 Association for Computational Linguistics

tries ⟨*acqua / water*⟩$_{10}$ and ⟨*dormire / sleep*⟩$_{191}$.[1] This property doesn't distinguish among any of the formal means for capturing these direct lexical relationships. Finite-state string transducers naturally capture these simple relationships, but so do more (and less) expressive formalisms.

Simple word-by-word replacement is not a viable translation method; this was noted even as early as Weaver's famous memorandum (Weaver, 1955). Systems based on word-to-word lexicons, such as the IBM systems (Brown et al., 1990; Brown et al., 1993), incorporate further devices that allow reordering of words (a "distortion model") and ranking of alternatives (a monolingual language model). Together, these allow for the possibility that

The Word Principle:

> Words translate differently when adjacent to other words.

This property of the translation relation is patently true.

Even a word-to-word system with the ability to reorder words and rank alternatives has obvious limitations, which have motivated the machine translation research community toward progressively more expressive formalisms. Again, we see precedent for the move in bilingual dictionaries, which provide phrasal translations in addition to simple word translations: ⟨*by and large / nel complesso*⟩$_{86}$, ⟨*full moon / luna piena*⟩$_{406}$. The insight at work here is

The Phrase Principle:

> Phrases (not words) translate differently when adjacent to other phrases.

And again, we see this insight informing statistical machine translation systems, for instance, in the phrase-based approaches of Och (2003) and Koehn et al. (2003). These two principles, while true, do not exhaust the insights implicit in the structure of bilingual dictionaries. A fuller view is accomplished by moving from words and phrases to constructions.

2.1 The construction principle

The phenomenon that underlies the use of synchronous grammars for MT is simply this:

The Construction Principle:

> Words and phrases translate differently in construction with other words.

The notion of *in construction with* is a structural notion. A word is in construction with another if they are related by a structural relation of some sort dependent on the identity or role of the word.

For example, the English word *take* is prototypically translated with a form of the Italian *prendere* ⟨*take / prendere*⟩$_{661}$. But when its object is a *bath*, as in the sentence "I like to take several long bubble baths every day", the word is translated with a form of *fare*. More accurately, the *construction* typified by the phrase *take a bath* is translated by the corresponding construction typified by the phrase *fare un bagno* (⟨*take a bath / fare un bagno*⟩$_{662}$).

One may think that we are still in the realm of the Phrase Principle; the phrase *take a bath* translates as the phrase *fare un bagno*. But the generalization is clearly much more general than that in several ways.

First, the notion of *in construction with* does not necessarily lead to contiguous phrases because of *variability* within the constructions. Bilingual dictionaries have developed notational conventions for such cases. When freely variable objects can intervene between the words in construction, a kind of variable word is used in dictionary entries, such as SB (*somebody*), STH (*something*), QN (*qualcuno*), QC (*qualcosa*). The word *take* participates in another construction ⟨*take* SB *by surprise / cogliere [literally "catch"]* QN *di sorpresa*⟩. The phenomenon is widespread. We find entries for light verb phrases such as *take* SB *by surprise*, idiomatic constructions such as ⟨*pull* SB*'s leg / prendere in giro* QN⟩$_{507}$, and particle constructions such as ⟨*call* SB *up / chiamare* QN⟩$_{86}$. These variable notations not only stand in for variable textual material and categorize that material (as specifying an entity (QC) or human (QN)) but also provide links between the portions of the two constructions. Whatever lexical material instantiates a SB variable on the English side, its translation instantiates the QN in the Italian. Thus translations require not only structure in the monolingual representations, but structure bilingually across them.[2]

[1] Throughout, we notate entries in HCICD with the notation ⟨*entry form / translation form*⟩$_{page}$, providing the Italian and English forms, along with the page number of the cited entry.

[2] The linking of the subject roles in these constructions is typically left implicit in these entries, following from an as-

Second, even constructions that are in and of themselves contiguous may become discontiguous by *intervention* of other lexical material: modifiers, appositives, and the like. An example has already been seen in the example "I like to take several long bubble baths every day". There is no contiguity between *take* and *bath* here. A formalism based purely on concatenation of contiguous phrases will be unable to model such constructions.

These two aspects of variability and intervention within and between constructions preclude simple concatenative formalisms such as finite-state or context-free formalisms.

2.2 Prevalence of bilingual constructions

A natural question arises as to the prevalence of such nontrivial bilingual constructions. Presumably, if they are sufficiently rare and exotic, it may be acceptable, and in fact optimal, from an engineering point of view to ignore them and stay with simpler formalisms.

We can ask the prevalence question at the level of types or tokens. At the type level, a simple examination of a comprehensive modern bilingual dictionary reveals a quite high frequency of non-word-for-word translations. Analysis of a small random subsample of HCICD yielded only 34% of entries of the $\langle$*acqua / water*$\rangle_{10}$ sort. In contrast, 52% were contiguous multi-word translations, e.g., $\langle$*guarda caso / strangely enough*$\rangle_{100}$. An additional 11% of entries had variable content, split about equally between entries with overt marking of variability ($\langle$*prendere* QN *in castagna / to catch* SB *in the act*$\rangle_{100}$) and implicit variability ($\langle$*hai fatto caso al suo cappello? / did you notice his hat?*$\rangle_{100}$, in which the $\langle$*suo cappello / his hat*$\rangle$ pair serves as a placeholder for other translates. (The remaining 3% is accounted for by entries providing monolingual equivalences and untranslated proper names.) The line between implicit variability and multi-word translations is quite permeable, so that many of the 54% of entries classified as the latter might in fact be better thought of as the former, and in any case many of the multi-word entries would be subject to noncontiguity through insertion of other lexical material. At the type level, then, there is plenty of evidence for the Phrase Principle and the Construction Principle.

At the token level, the general interest in so-called syntax-aware statistical MT approaches is itself evidence that researchers believe that the tokens accounting for the performance gap in current systems based on the Word and Phrase Principles transcend those principles in some way, presumably because they manifest the Construction Principle.[3] Only time will tell if such syntax-aware systems are able to display performance improvements over their nonstructural alternatives. Successful experiments such as those of Chiang (2005) using synchronous context-free grammar are a good first start.[4]

2.3 Heritage of the construction principle

We have argued that a formalism expressive enough to model the translation relation implicit in bilingual dictionaries must be based on relations over constructions, the primitive relations found in such bilingual dictionaries and founded by the Construction Principle. The fundamentality of this principle is evidenced by the fact that it has informed bilingual dictionaries literally *since their inception*. The earliest known bilingual dictionaries are those incorporated in the so-called lexical texts of ancient Mesopotamia from four millennia ago. Even there, we find evidence of the Construction Principle in entries that describe translation of words dependent upon words they are in construction with. Civil (1995) cites an example of the Akkadian word *nakāpu* (to gore, to knock down) whose translation into Sumerian is given differentially dependent on the nature of "grammatical constructions with particular subjects or objects":

sumption that subjects are typically linked across these languages. Where this assumption fails, however, explicit marking is found in the dictionary, either by using a passive alternation $\langle$*piacere a* QN */ to be liked by* SB$\rangle_{424}$, or implicit linking $\langle$*mi piace / I like it*$\rangle_{424}$.

[3] A reviewer objects that this point is vacuous: "Is the fact that researchers aren't building large-scale statistical semantic transfer models evidence for the fact that they don't believe in semantics?" This is an instance of the logical fallacy of denying the antecedent. If researchers act on a premise, they believe the premise. From this it does not follow that if they fail to act on a premise, they deny the premise.

[4] It would be more convincing to have empirical token-level statistics on the prevalence of constructions found in bilingual dictionaries. Unfortunately, this would require much of the effort of building an MT system on a construction basis itself.

Translation	When said of
sag-ta-dug$_4$-ga	the head
du$_7$	oxen
ru$_5$	rams
si-tu$_{10}$	oxen/bulls
kur-ku	a flood
ru-gú	a finger
si-ga	a garment

3 Synchronous Grammars Reviewed

To summarize, the translation relation in evidence implicitly in bilingual dictionaries requires a formalism expressive enough to directly represent *relations* between *constructions*, appropriately *linked*, and to do so in a way that allows these constructions to be realized *noncontiguously* by virtue of *variability* and *intervention*. As we will show, the former requirement is exactly the idea underlying synchronous grammars. The latter requirement of noncontiguity in its two aspects further implicates operations of substitution and adjunction (respectively) to combine constructions. The requirements lead naturally to a consideration of synchronous tree-adjoining grammar as the direct embodiment of the bilingual dictionaries of the last four millennia.

A synchronous grammar formalism is built by synchronizing grammars from some base formalism. A grammar in the base formalism consists of a set of elementary tree structures along with one or more combining operations. All of the familiar monolingual formalisms—finite-state grammars, context-free grammars, tree-substitution and -adjoining grammars, categorial grammars, inter alia—can be thought of in this way. A synchronous grammar consists of a set of pairs of elementary trees from the base formalism together with a linking relation between nodes in the trees at which combining operations can perform. Derivation proceeds as in the base formalism, whatever that is, except that a pair of trees operate at a pair of linked nodes in an elementary tree pair. An operation performed at one end of a link must be matched by a corresponding operation at the other end of the link. For example, the tree pair in Figure 1 might be appropriate for use in translating the sentence *Eli took his father by surprise*. The links between the NP nodes play the same role as the linked variables SB and QN in the bilingual dictionary entry. They allow for substitution of tree pairs for *Eli* and its translation and *his father* and its. The additional links allow for further modification, as in *Eli recently took his father by surprise by preparing dinner*, the modifiers *recently* and *by preparing dinner* adjoining at the VP and S links, respectively.

Expressing this relation in other frameworks involves either limiting its scope (for instance, to particular objects and intervening material), expanding its scope (by separating the translations of the contiguous portions of the constructions), or mimicking the structure of the STAG (as described at the end of Section 5).

The basic idea of using synchronous TAG for machine translation dates from the original definition (Shieber and Schabes, 1990), and has been pursued by several researchers (Abeille et al., 1990; Dras, 1999; Prigent, 1994; Palmer et al., 1999), but only recently in its probabilistic form (Nesson et al., 2006). The directness with which the formalism follows from the structure of bilingual dictionaries has not to our knowledge been previously noted. It leads to the possibility of making direct use of bilingual dictionary material in a statistical machine translation system.[5] But even if the formalism is not used in that way, there is import to the fact that its expressivity matches that thought by lexicographers of the last several millennia to be needed for capturing the translation relation; this fact indicates at least that STAG's use as a substrate for MT systems may be a promising research direction to pursue, should other necessary properties be satisfiable as well. We turn next to two of these properties: trainability and efficiency.

4 Trainability

The mere ability to formally represent the contents of manually developed bilingual dictionaries is not sufficient for the building of robust machine translation systems. The last decade and a half of MT research has demonstrated the importance of trainability of the models based on statistical evidence found in corpora. Without such training, manually

[5] For construction-based MT, reconstruction of tree alignments from data is much more difficult than for phrase-based MT, and hence extracting them from a dictionary becomes much more appealing.

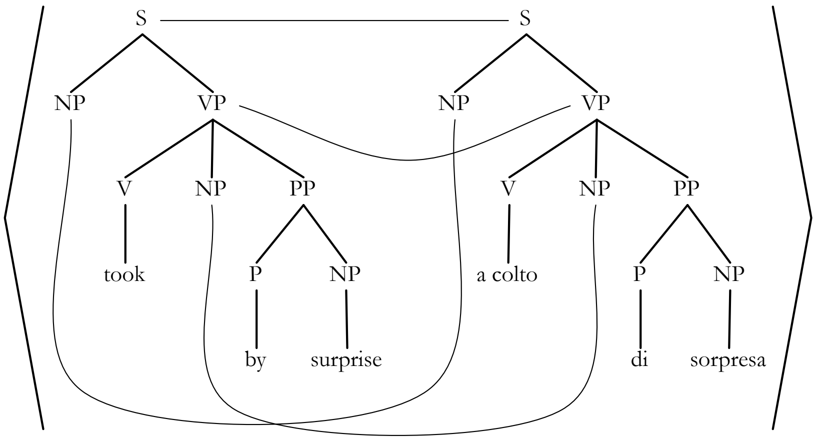

Figure 1: A synchronous tree pair.

developed models are too brittle to be seriously considered as a basis for machine translation.

It may also be the case that *with* such training, the manually generated materials are redundant. Certainly, it has been difficult to show the utility of manually generated annotations in improving MT performance. But this may be because the means by which the materials are represented is not yet appropriate; it does not articulate well with the statistical substrate used by the training methodology.

A further property, then, for the formalism is that it be trainable based on bilingual corpora. Consider training of the sort that underlies the IBM-style word models and their phrase-based offshoots, or statistical parsing based on probabilistic CFGs (Lari and Young, 1990) or other generative formalisms. Such methods use an underlying probabilistic formalism, typically structuring the parameters based on a universal parametric normal form (as *n*-gram probabilities are for finite-state grammars and Chomsky-normal form is for PCFGs), and applying an efficient training algorithm to set values for the parameters.

A full system based on STAG would use the formalism to express both the detailed bilingual constructional relationships as found in a bilingual dictionary and a backbone in the form of the universal normal form. Trained together, the normal form would serve to smooth the brittle construction-specific part, while the construction-specific part would relieve the burden on the universal learned portion to allocate parameters to rare constructions.

How do synchronous tree-adjoining grammars fare in this area? Do they admit of the kind of universal normal-form training that might serve as a smoothing method for the more highly articulated but brittle lexicographic relation?

A probabilistic variant of synchronous TAG is straightforward to specify, given that the formalism itself has a natural generative interpretation (Shieber, 1994). A universal parametric normal form has been provided by Nesson et al. (2006) (see Figure 2), who show that, at least on small training sets, a synchronous TAG in this normal form performs at a level comparable to standard word- and phrase-based systems. Synchronous TAGs thus seem to have the best of both worlds: They can directly express the types of ramified bilingual constructions as codified in bilingual dictionaries, and they can also express the types of universal assumption-free normal forms that underlie modern statistical MT. Importantly, they can do so *at one and the same time*, as both types of information are expressed in the same way, as sets of tree pairs. Both can therefore be trained together based on bilingual corpora.

We emphasize that the advantage that we find for STAGs in displaying well the necessary properties for statistical machine translation systems implicit in bilingual dictionaries is not that they are able to code efficiently all generalizations about the translation relation. Indeed, STAG is not able to do so (Shieber, 1994), which has motivated more expressive exten-

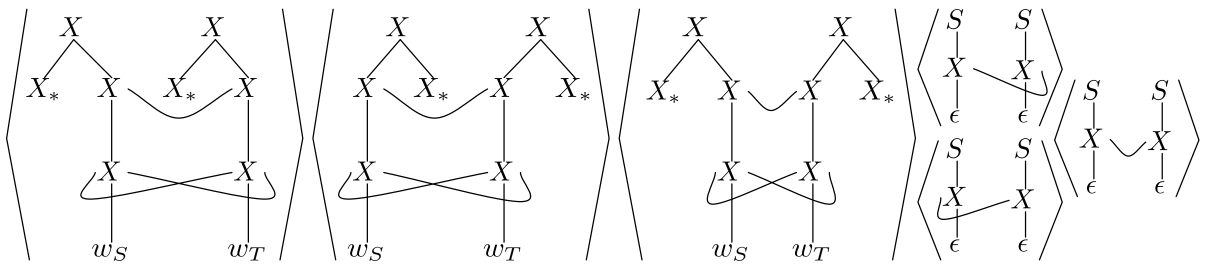

Figure 2: A normal form for synchronous tree-insertion grammar. (Reproduced from Nesson et al. (2006).)

sions of the formalism (Chiang et al., 2000). For example, STAG might express the construction relation ⟨*attraversare* QC *di corsa* / *run across* ST⟩ and similar relations between Italian verbs of direction with modifiers of motion and English verbs of motion with directional modifiers. However, the generalization that directional verbs with motion-manner adverbials translate as motion-manner verbs with directional adverbials is not expressed or expressible by STAG. Each instance of the generalization must be specified or learned separately.[6] Nonetheless, we are content (in the spirit of statistical MT) to have lots of such particular cases missing a generalization, so long as the parts from which they are constructed are pertinent, that is, so long as we do not need to specify ⟨*attraversare la strada di corsa* / *run across the road*⟩$_{51}$ separately from all of the other things one might run across.

5 Efficiency

A final set of considerations has to do with the efficiency of the formalism. Is it practical to use STAG for the purposes we have outlined? It is important not to preclude a formalism merely based on impracticality of its current use (given the constant increases in computer speed), but inherent intractability is another matter.[7]

Here, the STAG situation is equivocal. Bilingual parsing of a corpus relative to an STAG is a necessary first step in parameter training. The recognition problem for STAG, like that for synchronous context-free grammar (SCFG) is NP-hard (Satta and Peserico, 2005). Under appropriate restrictions of binarizability, SCFG parsing can be done in $O(n^6)$ time, doubling the exponent of CFG parsing. Similarly, STAG parsing under suitable limitations (Nesson et al. (2005)) can be done in $O(n^{12})$ time doubling the exponent of monolingual TAG parsing. On the positive side, recent work exploring the automatic binarization of synchronous grammars (Zhang et al., 2006) has indicated that non-binarizable constructions seem to be relatively rare in practice. Nonetheless, such a high-degree polynomial makes the complete algorithm impractical.

Nesson et al. (2006) use synchronous tree-insertion grammar (STIG) (Schabes and Waters, 1995) rather than STAG for this very reason. STIG retains the ability to express a universal normal form, while allowing $O(n^6)$ bilingual parsing. (Again, limitations on the formalism are required to achieve this complexity.) Even this complexity may be too high. Methods such as those of Chiang (2005) have been proposed for further reducing the complexity of SCFG parsing; they may be applicable to STIG (and STAG) parsing as well.

The STIG formalism can be shown to be expressively equivalent to synchronous tree-substitution grammar (STSG) and even SCFG. Does this vitiate the argument for STIG as a natural formalism for MT? No. The reductions of STIG to these other formalisms operate by introducing additional nodes

[6] Palmer et al. (1999) provide an approach to STAG that attempts to address this particular problem as does the extension of Dras (1999). It is unclear to what extent such extensions are amenable to trainable probabilistic variants.

[7] Of course, too much might be made of this question of computational complexity. The algorithms used for decoding of statistical MT systems almost universally incorporate heuristics for efficiency reasons, even those that are polynomial. One reviewer notes that “the admittedly perplexing reality is that exponential decoders run much faster than polynomial ones, presumably because they prune more intelligently.”

in the elementary trees that extend the size of those trees and hence the complexity of their parsing, unless subtle tricks are used to take advantage of the special structure of these added nodes. These tricks essentially amount to treating the formalism as an STIG, not an SCFG. That is, even if an SCFG were to be used, its structure would best be built on the observations found here.

For example, the method of Cowan et al. (2006) synchronizes elementary trees of a prescribed form to handle translation of clauses (verbs plus their arguments) essentially implementing a kind of STSG. However, because modifiers can make these trees discontiguous, they augment the model by allowing for free insertion of modifiers in certain locations. One view of this is as an implementation of the principle that motivates adjoining, without using adjoining itself. Thus, systems that are designed to take account of the principles adduced in this paper are likely to be implementing aspects of STAG implicitly, even if not explicitly.

Similarly, recent research is beginning to unify synchronous grammar formalisms and tree transducers (Shieber, 2004; Shieber, 2006). There may well be equally direct transducer formalisms that elegantly express construction-based translation relations. This would not be a denial of the present thesis but a happy acknowledgment of it.

6 Conclusion

We have argued that probabilistic synchronous TAG or some closely related formalism possesses a constellation of properties—expressivity, trainability, and efficiency—that make it a good candidate at a conceptual level for founding a machine translation system. What would such a system look like? It would start with a universal normal form subgrammar serving as the robust "backoff" relation to which additional more articulated bilingual material could be added in the form of additional tree pairs. These tree pairs might be manually generated, automatically reconstructed from repurposed bilingual dictionaries, or automatically induced from aligned bilingual treebanks (Groves et al., 2004; Groves and Way, 2005) or even unannotated bilingual corpora (Chiang, 2005). In fact, since all of these sources of data yield interacting tree pairs, more than one of these techniques might be used. In any case, further training would automatically determine the interactions of these information sources.

The conclusions of this paper are admittedly programmatic. But plausible arguments for a program of research may be just the thing for clarifying a research direction and even promoting its pursual. In that sense, this paper can be read as a kind of manifesto for the use of probabilistic synchronous TAG as a substrate for MT research.

Acknowledgments

We thank Rani Nelken, Rebecca Nesson, and Alexander Rush for helpful discussion and the anonymous reviewers for their insightful comments. This work was supported in part by grant IIS-0329089 from the National Science Foundation.

References

Anne Abeille, Yves Schabes, and Aravind K. Joshi. 1990. Using lexicalized tags for machine translation. In *Proceedings of the 13th International Conference on Computational Linguistics*.

Peter F. Brown, John Cocke, Stephen Della Pietra, Vincent J. Della Pietra, Frederick Jelinek, John D. Lafferty, Robert L. Mercer, and Paul S. Roossin. 1990. A statistical approach to machine translation. *Computational Linguistics*, 16(2):79–85.

Peter F. Brown, Stephen Della Pietra, Vincent J. Della Pietra, and Robert L. Mercer. 1993. The mathematics of statistical machine translation: Parameter estimation. *Computational Linguistics*, 19(2):263–311.

David Chiang, William Schuler, and Mark Dras. 2000. Some remarks on an extension of synchronous TAG. In *Proceedings of the 5th International Workshop on Tree Adjoining Grammars and Related Formalisms (TAG+5)*, Paris, France, 25–27 May.

David Chiang. 2005. A hierarchical phrase-based model for statistical machine translation. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*, pages 263–270, Ann Arbor, Michigan, June. Association for Computational Linguistics.

Miguel Civil. 1995. Ancient Mesopotamian lexicography. In Jack M. Sasson, editor, *Civilizations of the Ancient Near East*, volume 4, pages 2305–14. Scribners, New York.

Michela Clari and Catherine E. Love, editors. 1995. *HarperCollins Italian College Dictionary*. HarperCollins Publishers, Inc., New York, NY.

Brooke Cowan, Ivona Kucerov, and Michael Collins. 2006. A discriminative model for tree-to-tree translation. In *Proceedings of EMNLP 2006*.

Mark Dras. 1999. A meta-level grammar: Redefining synchronous TAG for translation and paraphrase. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, pages 80–87, Morristown, NJ, USA. Association for Computational Linguistics.

Declan Groves and Andy Way. 2005. Hybrid example-based SMT: the best of both worlds? In *Workshop on Building and Using Parallel Texts: Data-Driven Machine Translation and Beyond*, Ann Arbor, MI, June. ACL '05.

Declan Groves, Mary Hearne, and Andy Way. 2004. Robust sub-sentential alignment of phrase-structure trees. In *COLING '04, Geneva Switzerland*.

Philipp Koehn, Franz Josef Och, and Daniel Marcu. 2003. Statistical phrase-based translation. In *Proceedings of HLT/NAACL*.

Karim Lari and Steve J. Young. 1990. The estimation of stochastic context-free grammars using the inside-outside algorithm. *Computer Speech and Language*, 4:35–56.

Rebecca Nesson, Alexander Rush, and Stuart M. Shieber. 2005. Induction of probabilistic synchronous tree-insertion grammars. Technical Report TR-20-05, Division of Engineering and Applied Sciences, Harvard University, Cambridge, MA.

Rebecca Nesson, Stuart M. Shieber, and Alexander Rush. 2006. Induction of probabilistic synchronous tree-insertion grammars for machine translation. In *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas (AMTA 2006)*, Boston, Massachusetts, 8-12 August.

Franz Josef Och. 2003. *Statistical Machine Translation: From Single-Word Models to Alignment Templates*. Ph.D. thesis, Technical University of Aachen, Aachen, Germany.

Martha Palmer, Joseph Rosenzweig, and William Schuler. 1999. Capturing motion verb generalizations in synchronous tree-adjoining grammar. In Patrick Saint-Dizier, editor, *Predicative Forms in Natural Language and in Lexical Knowledge Bases*. Kluwer Press.

Gilles Prigent. 1994. Synchronous TAGs and machine translation. In *Proceedings of the Third International Workshop on Tree Adjoining Grammar and Related Formalisms (TAG+3)*, Université Paris 7.

Giorgio Satta and Enoch Peserico. 2005. Some computational complexity results for synchronous context-free grammars. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing (HLT/EMNLP 05)*, pages 803–810, Morristown, NJ, USA. Association for Computational Linguistics.

Yves Schabes and Richard C. Waters. 1995. Tree insertion grammar: A cubic time, parsable formalism that lexicalizes context-free grammars without changing the trees produced. *Computational Linguistics*, 21(3):479–512.

Stuart M. Shieber and Yves Schabes. 1990. Synchronous tree-adjoining grammars. In *Proceedings of the 13th International Conference on Computational Linguistics*, volume 3, pages 253–258, Helsinki, Finland.

Stuart M. Shieber. 1994. Restricting the weak-generative capacity of synchronous tree-adjoining grammars. *Computational Intelligence*, 10(4):371–385, November. Also available as cmp-lg/9404003.

Stuart M. Shieber. 2004. Synchronous grammars as tree transducers. In *Proceedings of the Seventh International Workshop on Tree Adjoining Grammar and Related Formalisms (TAG+ 7)*, Vancouver, Canada, May 20-22.

Stuart M. Shieber. 2006. Unifying synchronous tree-adjoining grammars and tree transducers via bimorphisms. In *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL-06)*, Trento, Italy, 3–7 April.

Warren Weaver. 1955. Translation. In W.N. Locke and A. D. Booth, editors, *Machine Translation of Languages: Fourteen Essays*, pages 15–23. Technology Press of the Massachusetts Institute of Technology, Cambridge, Massachusetts.

Hao Zhang, Liang Huang, Daniel Gildea, and Kevin Knight. 2006. Synchronous binarization for machine translation. In *Proceedings of the Conference on Human Language Technology and Annual Meeting of the North American Chapter of the Association of Computational Linguistics (HLT/NAACL 2006)*, pages 256–263, Morristown, NJ, USA. Association for Computational Linguistics.