PBML

The Prague Bulletin of Mathematical Linguistics
NUMBER 101 APRIL 2014 43–54

Visualization, Search and Analysis of Hierarchical Translation Equivalence in Machine Translation Data

Gideon Maillette de Buy Wenniger, Khalil Sima'an

Institute for Logic Language and Computation (ILLC), Faculty of Science, University of Amsterdam

Abstract

Translation equivalence constitutes the basis of all Machine Translation systems including the recent hierarchical and syntax-based systems. For hierarchical MT research it is important to have a tool that supports the qualitative and quantitative analysis of hierarchical translation equivalence relations extracted from word alignments in data. In this paper we present such a toolkit and exemplify some of its uses. The main challenges taken up in designing this tool are the efficient and compact, yet complete, representation of hierarchical translation equivalence coupled with an intuitive visualization of these hierarchical relations. We exploit a new hierarchical representation, called Hierarchical Alignment Trees (HATs), which is based on an extension of the algorithms used for factorizing n-ary branching SCFG rules into their minimally-branching equivalents. Our toolkit further provides a search capability based on hierarchically relevant properties of word alignments and/or translation equivalence relations. Finally, the tool allows detailed statistical analysis of word alignments, thereby providing a breakdown of alignment statistics according to the complexity of translation equivalence units or reordering phenomena. We illustrate this with an empirical study of the coverage of inversion-transduction grammars for a number of corpora enriched with manual or automatic word alignments, followed by a breakdown of corpus statistics to reordering complexity.

1. Introduction

What kind of translation equivalence occurs in the word aligned data of a language pair? Are the word alignments yielding them correct and what kind of requirements for effective translation models is implied by their complexity? Most practitioners of

 Corresponding author: gemdbw AT gmail.com
Cite as: Gideon Maillette de Buy Wenniger, Khalil Sima'an. Visualization, Search and Analysis of Hierarchical Translation Equivalence in Machine Translation Data. The Prague Bulletin of Mathematical Linguistics No. 101, 2014, pp. 43–54. doi: 10.2478/pralin-2014-0003.

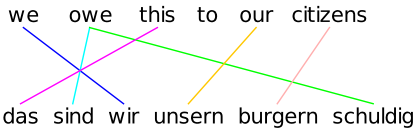

Figure 1: Example of alignment visualization for an aligned sentence pair (Europarl)

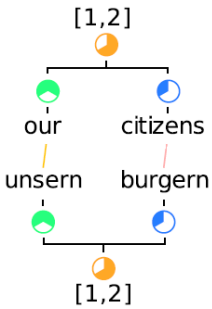

Figure 2: Example recursive composition of HATs

Machine Translation have some qualitative appreciation of the difference in similarity across language pairs. For example, it is a well known fact that for Chinese-English translation long range reordering is the norm, and as a result phrase-based translation does not work well for this language pair. Similarly it is known, that reordering is more local for Arabic-English than for Chinese-English, so that phrase-based translation gives relatively good performance for this language pair. Qualitative knowledge of the data is much more scarce for less popular language pairs and less famous but nevertheless valid reordering constructions; and in fact hard statistics about alignment complexity are mostly lacking for all language pairs. Crucial for the success of machine translation are the quality of the aligned corpus as operationalized by induced translation equivalence, and its complexity and compatibility with a translation model. A tool that facilitates more knowledge of these factors and that gives better qualitative understanding of translation equivalence occurring in real data can be valuable. A crucial aspect of such a visualization tool is an efficient and compact representation of hierarchical translation equivalence coupled with an intuitive visualization of the hierarchical relations between the translation equivalents. We employ Hierarchical Alignment Trees (HATs) (Sima'an and Maillette de Buy Wenniger, 2013) as the representation of choice for our *HATs toolkit*.

The word alignment in Figure 1 exemplifies the fact that manual or mental reconstruction of a representation of hierarchical translation equivalence over word alignments can be impractical for more than a few word long sentences with complex alignments, which constitutes the actual scenario when doing analysis of word alignments for representative, large data.

Besides visualization, one other application of the HATs toolkit is *search and data analysis*. Based on properties of the HATs representation it is easy to answer a question like "are their many valid/correct complex alignment constructions that cannot be covered by Inversion Transduction Grammars (ITG) in big automatically aligned corpora?". Previous work on alignment coverage (Zens and Ney, 2003; Wellington

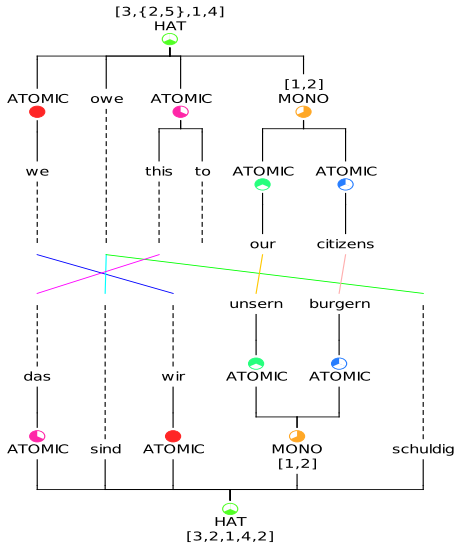

Figure 3: Visualization aspects: Filling (and color) of the circles indicates the equivalence between the top source-to-target HAT and the mirrored target-to-source HAT displayed below it. Labels, such as [3,{2,5},1,4] at the top node, denote permutation-set reordering operations at nodes. The labels *ATOMIC*, *MONO* and *HAT* in the HAT visualization indicate broad complexity categories for reordering.

et al., 2006; Søgaard and Wu, 2009; Søgaard and Kuhn, 2009; Søgaard, 2010) has been mainly involved with hand-aligned corpora, and as such has not given much direct insight into the validity of claims like "many of the non-binarizable alignments in automatically aligned data correspond to alignment errors". Our toolbox provides a method for automatically compiling a database of alignment properties for an input parallel corpus, so that alignments with certain properties can be effectively queried and retrieved. This makes it very simple to select for example all non-binarizable alignments. Based on such selections and by a quick visual inspection of the associated HATs, researching the validity of claims about the complexity of word alignments in real data becomes far simpler than building specific tools for ITG (or any other specific synchronous grammar formalism). Quantitative analysis in the form of statistics like alignment coverage for different types of permissible reordering complexity is another important component in the more global analysis of translation equivalence properties. For such quantitative analysis, keeping track of the reordering operations at the nodes while at the same time computing the HATs in an efficient and compact way will prove essential. We will exemplify this by reporting extensive empirical analysis of the coverage of ITG on a range of parallel corpora for manual as well as automatic word alignments. To the best of our knowledge, these results are the first on *large, automatically aligned* parallel corpora. In the remainder of this paper we first introduce HATs and discuss some of their visualization properties, then briefly discuss the methods available in our tool and move quickly to the empirical study regarding binarizable word alignments that can be covered by ITG.

2. Hierarchical Alignment Trees (HATs)

An efficient and compact representation of hierarchical phrase pair translation equivalence is **Hierarchical Alignment Trees (HATs).** A **HAT** (Sima'an and Maillette de Buy Wenniger, 2013) is a hierarchical representation/factorization of the phrase pairs in a word alignment; A HAT compactly represents a synchronous tree pair: a source and target tree pair with a node alignment relation between them. The recursive structure in a HAT shows the build up of phrase pairs from embedded phrase pairs. Hence, every node in a HAT represents the phrase pair at the fringe of the subtree under that node.

In Figure 3 we show what a visualization of a *Hierarchical Alignment Tree* for the example of Figure 1 looks like. There are actually *two* HATs displayed here. The upper tree shows the mapping from source to target while the bottom tree shows the mapping from target to source. Between the two trees we display the word alignments, making it directly clear how certain parts of the word alignment yield corresponding parts in the HATs. The filling and color/shade of the nodes clearly represents the translation equivalence between the source and target side of phrases.

Like **Normalized Decomposition Trees (NDTs)** (Zhang et al., 2008), HATs are *minimally branching factorization* of word alignments into phrase pairs, i.e., every HAT node covers a phrase pair and it dominates the smallest number of translation units (the child nodes) that the phrase pair decomposes into. In Figure 3, the alignment underlying the phrase pair *our citizens unsern burgern* decomposes down minimally to two phrase pair nodes *our unsern* and *citizens burgern*. More intricate HATs may arise due to complex word alignments that involve many-to-many and discontiguous translation units.

Discontiguous translation units One important property of HATs is their explicit representation of discontiguous translation units. In Figure 3 the root node on the English/German side dominates, among others, two terminal nodes: this explicitly represents the discontiguous unit given by the alignment between *sind + schuldig* (positions 2 and 5) on the German side with a single English word *owe* (position 2). More generally, to differentiate between phrase pairs and separate parts of discontiguous translation units in the HAT representation, the latter are depicted as *terminal nodes* (nodes labeled with words without any children), whereas the former (phrase pairs) are represented as non-terminal nodes (circles with filling dominating a subtree).

Reordering operators Crucially, HATs extend **NDTs** by providing explicit representation of the reordering of the children under every node by a transduction operator, called a *set-permutation*, as well as the internal word alignments for atomic (non-decomposable) phrase pairs. Hence, a HAT is a decorated tree where the nodes

are decorated with *set-permutations*. A set-permutation decorating a node in source side HAT is a list of integer sets denoting a transduction operation that applies to the children of that node to obtain the target side phrase reordering. We will immediately exemplify HATs and set-permutations before we proceed further with discussing the properties of HATs.

Set-permutations such as [3,{2,5},1,4] (see the root node in Figure 3) denote reordering operations occurring under these nodes. In this example the source children 1, 2, 3, 4 map to target children (relative order) 3, {2,5}, 1, 4 respectively, where {2, 5} represents the fact that the second child is linked with two on the other side in positions 2 and 5. In the simpler monotone mapping *our citizens | unsern burgern* the set-permutation label is [1,2]. We also see the coarse reordering categories *ATOMIC*, *MONO* and *HAT* in this figure, which are discussed next.

Complexity categories As mentioned above, every node is decorated with a set-permutation, specifying the relative mapping occurring directly below it. In the case of bijective mappings this describes a permutation. In the general case of arbitrary m-n mappings there are recurring target position in the mapping set of different source positions and/or multiple target positions occurring in the mapping set(s) of some source positions. Hence, the set-permutations can be grouped into coarser categories of mapping complexity. We distinguish the following five cases, ordered by increasing complexity:

1. *Atomic*: If the alignment does not allow the existence of smaller (child) phrase pairs: a subset of alignment positions that is not connected to the other positions while also forming a contiguous sequence on the source and target does not exists.
2. *Monotonic*: If the alignment can be split into two monotonically ordered parts.
3. *Inverted*: If the alignment can be split into two inverted parts.
4. *PET (Permutation Tree)*: If the alignment can be factored as a permutation of more than 2 parts.
5. *HAT (Hierarchical Alignment Tree)*: If the alignment cannot be factored as a permutation of parts, but the phrase does contain at least one smaller phrase pair.

Typically there are multiple HATs for a word alignment, corresponding to different possible minimally branching factorizations into phrase pairs. These alternative HATs can be efficiently computed and stored as a chart using a CYK-parser like chart parsing algorithm that parses the alignment and builds a Hypergraph of HATs in the process.[1]

[1]This is exactly what is done by our program. Note that in certain cases the number of HATs per alignment can become big, in particular for alignments that contain many monotone parts. One optimization we use in our algorithm is reasoning about null-aligned words outside the main algorithm. Computation is typically fast, provided enough memory is available. Rendering all HATs is done by

A categorization of the complexity of the HAT as a whole is determined based on the complexity categories of the alignment mappings at its nodes. *Binary Inversion-Transduction Trees (BITTs)* is the least complex class consisting of only binary HATs that can be built for binarizable permutations (Huang et al., 2009), any HAT that contains only Monotonic and/or Inverted nodes belongs to this class. If a HAT contains at least one PET node but no HAT nodes it belongs to the category called PETs corresponding to general permutations (Zhang et al., 2008; Satta and Peserico, 2005). Finally the occurrence of at least one HAT node implies the set *HATs* which captures all possible many-to-many mappings.

Having broadly explained what Hierarchical Alignment Trees (HATs) are about and what kind of information about hierarchical translation equivalence they give, the next question is what other things we can do with HATs apart from this most basic form of visualization.

3. Empirical analysis of word alignments in parallel corpora

In this section we exemplify one use of our toolkit for analysis of translation units (TUs) and word alignments in parallel corpora. We first look at the percentage of word alignments covered by Inversion-Transduction Grammar (ITG) (Wu, 1997), which are the cases of fully binarizable word alignments. Subsequently we provide a breakdown of word alignments into subclasses of increasing complexity.

Word alignments are considered the initial point for extracting translation units. In our analysis we differentiate between two cases of translation units (TUs) that can be extracted from a word alignment:

Contiguous TUs Only phrase pairs (called Contiguous translation units TUs)

Discontiguous TUs All contiguous + discontiguous TUs

When quantifying the coverage of ITG for word alignments, we explicitly make a difference between these two cases. As expected, it is usually more difficult to provide ITG derivations for discontiguous TUs in word alignments than for contiguous TUs. We define *ITG coverage* as the percentage of word alignments that can be covered by some ITG. A word alignment is covered by ITG if and only if it is fully binarizable, i.e., all nodes in the HATs are binary branching. This is a specific class of HATs called Binarizable Inversion Transduction Trees (BITTs).

Besides BITTs, word alignments can be grouped into subclasses of HATs according to the complexity of the reordering operators – set-permutations – on the nodes in these HATs. Apart from the BITT case (derivable by ITGs), we also define two more subclasses:

---

enumerating all of them from the Hypergraph and writing their tree structures to a textfile, then reading this unpacked forest from the textfile by the tree visualization component. As this can become somewhat slow in case of many HATs, rendering rendering all HATs can be turned on in the GUI, but only showing the first one is used as the default option.

PETs Percentage of word alignments covered by different Permutation Trees (PETs) that are beyond BITTs, i.e., these are bijective word alignments that are non-binarizable,

HATs All remaining HATs that are beyond PETs, i.e., non-bijective cases consisting of many-to-many word alignments and possibly discontiguous TUs.

3.1. Word alignments covered by ITG: BITTs

Data Sets We use manually and automatically aligned corpora. Manually aligned corpora come from two datasets. The first (Graça et al., 2008) consists of six language pairs: Portuguese–English, Portuguese–French, Portuguese–Spanish, English–Spanish, English–French and French–Spanish. These datasets contain 100 sentence pairs each and distinguish *Sure* and *Possible* alignments. Following Søgaard and Kuhn (2009), we treat these two equally. The second manually aligned dataset (Padó and Lapata, 2006) contains 987 sentence pairs from the English-German part of Europarl annotated using the Blinker guidelines (Melamed, 1998). The automatically aligned data comes from Europarl (Koehn, 2005) in three language pairs (English–Dutch, English–French and English–German). The corpora are automatically aligned using GIZA++ (Och and Ney, 2003) in combination with the grow-diag-final-and heuristic. With sentence length cutoff 40 on both sides these contain respectively 945k, 949k and 995k sentence pairs.

ITG Coverage is defined as the percentage word alignments (sentence pairs) in a parallel corpus that can be covered by an instance of ITG. Clearly, coverage depends on the chosen semantic interpretation of word alignments: contiguous translation units (phrase pairs) or discontiguous translation units.[2]

Results Table 1 shows the coverage of ITG for the different corpora dependent on the two alternative definitions of *translation equivalence*. The first thing to notice is that there is just a small difference between the Grammatical Coverage scores for these two definitions. The difference is in the order of a few percentage points, the largest difference is seen for Portuguese–French (79% v.s 74% Grammatical Coverage), for some language pairs there is no difference. For the automatically aligned corpora the absolute difference is on average about 2%. We attribute this to the fact that there are only very few discontiguous TUs that can be covered by ITG in this data.

[2] A note here on the computation of coverage for the different subclasses of HATs such as ITG Coverage in case of BITTs. As mentioned before every alignment typically yields a set containing multiple alternative HATs, corresponding to the different possible minimal factorizations of phrase pairs. Each of these HATs in the set however is by itself sufficient to determine the complexity for the whole set, since it always holds that all HATs in the set have the same complexity. This property is hence used for the efficient computation of the coverage statistics, based on just the first HAT from the computed Hypergraph of HATs for an alignment.

Alignments Set	Coverage contiguous	Coverage discontiguous
Hand aligned corpora		
English–French	76.0	75.0
English–Portuguese	78.0	78.0
English–Spanish	83.0	83.0
Portuguese–French	78.0	74.0
Portuguese–Spanish	91.0	91.0
Spanish–French	79.0	74.0
LREC Corpora Average	80.8±5.5	79.2±6.7
English–German	45.4	45.3
Automatically aligned Corpora		
English–Dutch	45.5	43.6
English–French	52.8	50.0
English–German	45.6	43.7
Automatically aligned corpora average	48.0±4.20	45.8±3.6

Table 1: The ITG coverage for different corpora dependent on the interpretation of word alignments: contiguous only or including discontiguous translation units

The second thing to notice is that the scores are much higher for the corpora from the LREC dataset than they are for the manually aligned English–German corpus. The approximately double source and target length of the manually aligned English–German corpus, in combination with somewhat less dense alignments makes this corpus much harder than the LREC corpora. Intuitively, one would expect that more alignment links make alignments more complicated. This turns out to not always be the case. Further inspection of the LREC alignments also shows that these alignments often consist of parts that are *completely linked*. Such completely linked parts are by definition treated as atomic TUs, which could make the alignments look simpler. This contrasts with the situation in the manually aligned English–German corpus where on average less alignment links exist per word.

When we look at the results for the automatically aligned corpora at the lowest rows in the table, we see that these are comparable to the results for the manually aligned English–German corpus (and much lower than the results for the LREC corpora). This could be explained by the fact that the manually aligned English–German is not only Europarl data, but possibly also because the manual alignments themselves were obtained by initialization with the GIZA++ alignments. In any case, the manually and automatically acquired alignments for this data are not too different from the perspective of ITG. Further differences might exist if we would employ another class of grammars, e.g., full SCFGs.

One the one hand, we find that manual alignments are well but not fully covered by ITG. On the other, the automatic alignments are not covered well but ITG. This suggests that these automatic alignments are difficult to cover by ITG, and the reason

Kind of HATs	English-Dutch	English-French	English-German
BITTs (Binarizable permut.)	45.5%	52.8%	45.6%
PETs (Permut.)	52.6%	56.6%	52.6%
HATs (Set-permut.)	100.0%	100.0%	100.0%

Table 2: The ratio of the different subsets of HATs in the corpus: BITTs, PETs and HATs

could be that these alignments are built heuristically by trading precision for recall (cf. Och and Ney, 2003). Søgaard (2010) reports that full ITG provides a few percentage points gains over ITG.

Overall, we find that our results for the LREC data are far higher than Søgaard's results but lower than the upperbounds of Søgaard and Wu (2009). A similar observation holds for the English–German manually aligned EuroParl data, albeit the maximum length (15) used in (Søgaard and Wu, 2009; Søgaard, 2010) is different from ours (40). We attribute the difference between our results and Sogaard's approach to our choice to adopt lexical productions of ITG that contain own internal alignments (the detailed version) and determined by the atomic TUs of the word alignment. Our results differ substantially from Søgaard and Wu (2009) who report upperbounds (indeed our results still fall within these upperbounds for the LREC data).

3.2. Breakdown according to reordering complexity

Table 2 shows the breakdown statistics of word alignments according to complexity class for three of the automatically aligned EuroParl corpora. The first line with percentages for BITTs corresponds to the "Coverage Contiguous" scores for the automatically aligned corpora, at the bottom lines in the middle column in Table 1. There is an obvious difference between the percentage of BITTs for English-French vs. English-German and English-Dutch. This difference carries over to PETs, fully bijective word alignments. The percentages for PETs clearly show that approximately half of the word alignments in the data are beyond the bijective case.

4. Related Work

General tools have been created for the visualization of basic word alignment (Smith and Jahr, 2000; Germann, 2008) as well as for the manual annotation of sentence pairs. For the alignment of syntactic trees another available toolkit is the Stockholm Tree Aligner (Volk et al., 2007). As a somewhat related problem, Maillette de Buy Wenniger et al. (2010) take on visualization of the coherence of tree-based reordering with word alignments.

Normalized Decomposition Trees (NDTs) were introduced in Zhang et al. (2008), giving an efficient tree-based hierarchical representation of nested phrase pairs. Combined with these structures a very efficient phrase extraction algorithm with

linear time complexity was proposed. In these representations, the relative reordering operations taking place at the nodes is left implicit, as it is not relevant from the point of view of plain phrase extraction. Indeed for phrase extraction NDTs are a superior tool. On the other hand, for effective visualization and quick insight into the structure, explicit reordering labels play an important role. For efficient and intuitive reordering complexity analysis this is equally true. Hierarchical Alignment Trees can be seen as an extension of Normalized Decomposition Trees, that makes the reordering relations between translation equivalents explicit and also keeps track of the alignment relations within phrase pairs. The latter guarantees that for every word alignment there is a mapping to a set of HATs, such that the set of HATs truly captures all information and structure of the word alignment. This is not completely the case for NDTs.

The array of work described in Zens and Ney (2003); Wellington et al. (2006); Søgaard and Wu (2009); Søgaard and Kuhn (2009); Søgaard (2010) concentrates on methods for calculating *upperbounds* on the alignment coverage for all ITGs, including NF-ITG. See Maillette de Buy Wenniger and Sima'an (2013) for a more complete overview. Some of this work also uses alignment parsing to compute more exact scores (Søgaard, 2010), recently extended by Kaeshammer (2013) which adds new flavor to this discussion by looking at alignment reachability for Synchronous Linear Context-Free Rewriting Systems.

5. Conclusions

We introduced a toolkit for the visualization, search and analysis of hierarchical translation equivalence. We have shown how this toolkit can help to get a better qualitative as well as quantitative understanding of translation equivalence relations as induced by word alignments for real big translation data. The software is distributed under the LGPL license and can be downloaded from:

`https://bitbucket.org/teamwildtreechase/hatparsing`

Acknowledgements

This work is supported by The Netherlands Organization for Scientific Research (NWO) under grant nr. 612.066.929. The authors would like to thank Remko Scha for his valuable feedback on the manuscript. The authors would furthermore like to thank the anonymous reviewers for all their helpful and detailed suggestions and comments.

Bibliography

Germann, U. Yawat: yet another word alignment tool. In *Proc. of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies (ACL-HLT)*, pages 20–23, 2008.

Graça, João, Joana Pardal, Luísa Coheur, and Diamantino Caseiro. Building a golden collection of parallel multi-language word alignment. In *LREC'08*, Marrakech, Morocco, 2008. European Language Resources Association (ELRA).

Huang, Liang, Hao Zhang, Daniel Gildea, and Kevin Knight. Binarization of synchronous context-free grammars. *Computational Linguistics*, 35(4):559–595, 2009.

Kaeshammer, Miriam. Synchronous linear context-free rewriting systems for machine translation. In *Proceedings of the Seventh Workshop on Syntax, Semantics and Structure in Statistical Translation*, pages 68–77, 2013. URL `http://www.aclweb.org/anthology/W13-0808`.

Koehn, Philipp. Europarl: A parallel corpus for statistical machine translation. In *Proc. of MT Summit*, 2005.

Maillette de Buy Wenniger, Gideon and Khalil Sima'an. A formal characterization of parsing word alignments by synchronous grammars with empirical evidence to the itg hypothesis. In *Proceedings of the Seventh Workshop on Syntax, Semantics and Structure in Statistical Translation*, pages 58–67, 2013. URL `http://www.aclweb.org/anthology/W13-0807`.

Maillette de Buy Wenniger, Gideon, Maxim Khalilov, and Khalil Sima'an. A toolkit for visualizing the coherence of tree-based reordering with word-alignments. *The Prague Bulletin of Mathematical Linguistics*, pages 97–106, 2010.

Melamed, Dan. Annotation style guide for the blinker project, version 1.0. Technical Report IRCS TR #98-06, University of Pennsylvania, 1998.

Och, Franz Josef and Hermann Ney. A systematic comparison of various statistical alignment models. *Computational Linguistics*, 29(1):19–51, 2003.

Padó, Sebastian and Mirella Lapata. Optimal constituent alignment with edge covers for semantic projection. In *ACL-COLING'06*, ACL-44, pages 1161–1168, Stroudsburg, PA, USA, 2006. Association for Computational Linguistics.

Satta, Giorgio and Enoch Peserico. Some computational complexity results for synchronous context-free grammars. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 803–810, 2005.

Sima'an, Khalil and Gideon Maillette de Buy Wenniger. Hierarchical alignment trees: A recursive factorization of reordering in word alignments with empirical results. ILLC Prepublication series: Technical Report PP-2011-38, 2013. URL `http://staff.science.uva.nl/~simaan/D-Papers/HATsReport2013.pdf`.

Smith, Noah A. and Michael E. Jahr. Cairo: An alignment visualization tool. In *Proc. of the 2nd Conference on Language Resources and Evaluation (LREC)*, page 549–551, 2000.

Søgaard, Anders. Can inversion transduction grammars generate hand alignments? In *Proccedings of the 14th Annual Conference of the European Association for Machine Translation (EAMT).*, 2010.

Søgaard, Anders and Jonas Kuhn. Empirical lower bounds on alignment error rates in syntax-based machine translation. In *SSST '09*, pages 19–27, Stroudsburg, PA, USA, 2009. Association for Computational Linguistics.

Søgaard, Anders and Dekai Wu. Empirical lower bounds on translation unit error rate for the full class of inversion transduction grammars. In *Proceedings of the 11th International Conference on Parsing Technologies*, pages 33–36, 2009.

Volk, M., J. Lundborg, and M. Mettler. Alignment tools for parallel treebanks. In *In Proc. of The Linguistic Annotation Workshop at the Association for Computational Linguistics (LAW-ACL)*, 2007.

Wellington, Benjamin, Sonjia Waxmonsky, and I. Dan Melamed. Empirical lower bounds on the complexity of translational equivalence. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*, 2006.

Wu, Dekai. Stochastic inversion transduction grammars and bilingual parsing of parallel corpora. *Computational Linguistics*, 3(23):377–403, 1997.

Zens, Richard and Hermann Ney. A comparative study on reordering constraints in statistical machine translation. In *Proceedings of the Annual Meeting of the ACL*, pages 144–151, 2003.

Zhang, Hao, Daniel Gildea, and David Chiang. Extracting synchronous grammar rules from word-level alignments in linear time. In *Proceedings of the 22nd International Conference on Computational Linguistics - Volume 1*, pages 1081–1088, 2008.

Address for correspondence:
Gideon Maillette de Buy Wenniger
gemdbw AT gmail.com
Institute for Logic Language and Computation
Faculty of Science,
University of Amsterdam
Science Park 107
1098 XG Amsterdam, Netherlands