Building a Hierarchically Aligned Chinese-English Parallel Treebank

Dun Deng and **Nianwen Xue**

Computer Science Department, Brandeis University
415 South Street, Waltham MA, USA
ddeng@brandeis.edu, xuen@brandeis.edu

Abstract

We construct a hierarchically aligned Chinese-English parallel treebank by manually doing word alignments and phrase alignments simultaneously on parallel phrase-based parse trees. The main innovation of our approach is that we leave words without a translation counterpart (which are mostly language-particular function words) unaligned on the word level, and locate and align the appropriate phrases which encapsulate them. In doing so, we harmonize word-level and phrase-level alignments. We show that this type of annotation can be performed with high inter-annotator consistency and have both linguistic and engineering potentials.

1 Introduction

The value of human annotated syntactic structures for Statistical Machine Translation has been clearly demonstrated in string-to-tree (Galley et al., 2004; Galley et al., 2006; Huang et al., 2006), tree-to-string (Liu et al., 2006; Liu and Gildea, 2008), and tree-to-tree (Eisner, 2003; Liu et al., 2009; Chiang, 2010) models. One recurring issue which hampers the utility of syntactic structures is the incompatibility between word alignments and syntactic structures (Denero and Klein, 2007; Fossum et al., 2008; Pauls et al., 2010). The incompatibility arises because word alignments and syntactic structures are established independently of each other. In the case of tree-to-tree models, there is also the issue of incompatible parallel tree structures resulting from divergent syntactic annotation standards that have been independently conceived based on monolingual corpora (Chiang, 2010). In this paper, we report an effort in building a Hierarchically Aligned Chinese-English Parallel Treebank (HACEPT) where we manually do word-level and phrase-level alignments simultaneously on parallel phrase-based parse trees. In this process, we attempt to establish an annotation standard that harmonizes word-level and phrase-level alignments. We also analyze a common incompatibility issue between Chinese-English parallel parse trees exposed in the annotation process, with the goal of solving the issue by semi-automatically revising the trees.

In the rest of this paper, we describe how we construct the HACEPT and discuss issues arising in the construction process. In Section 2, we discuss the problems of word alignment done without considering its interaction with syntactic structures. In Section 3, we describe our annotation procedure where we perform word-level and phrase-level alignments simultaneously in a coordinated manner, and show how our approach is free of the problems discussed in Section 2. In Section 4, we report a common incompatibility issue between parse trees and propose a solution. We also compare the issue with translation divergence (Dorr, 1994) and show that they are different in nature and occurrence frequency. In Section 5, we present the results of two experiments we have done on our annotation to show the intuitiveness of our approach and the linguistic and engineering potentials of our corpus. We then describe related work in Section 6 and conclude our paper in Section 7.

Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers, pages 1511–1520, Dublin, Ireland, August 23-29 2014.

2 Incompatibilities between word alignments and syntactic structures

All the existing word alignment practice we know of treats word alignment as a stand-alone task without systematically considering its interaction with the syntactic structure of a sentence. The inevitable consequence of the practice is that both redundancies and incompatibilities between word alignments and syntactic structures will arise in many places. In this section, we illustrate the issues through language-particular function words, where the problems are most frequently found. Due to language-particular idiosyncrasy and lack of lexical content, these function words usually do not have a translation counterpart, which presents a great challenge to alignment annotation. There are two logical possibilities of dealing with these words, both of which are represented in existing annotation practice. The first is to leave them unaligned or link them to a fictitious NULL word (Ahrenberg, 2007; Brown et al., 1990), and the second, which also seems to be the more common pratice, is to attach these function words to a word that has a translation counterpart, and then align the function word and its host with the counterpart of the host (Melamed, 1998; Li et al., 2009). For ease of discussion, below we will refer to the latter pratice as the "glue-to-a-host" strategy (GTAHS). Both approaches are less than desirable: the former leaves the function words unaccounted for, and the latter leads to incompatibility issues we discuss in detail below.

First note that, by attaching language-particular function words to a host, the GTAHS creates redundancies between word alignments and syntactic structures since many of these function words have already been associated with a host within a constituent in the parse tree (e.g., the English determiner *the* is placed inside the projection of its host, namely an NP). A more serious issue is that the GTAHS creates spurious ambiguities. Lexical ambiguity is inevitable in translation. For instance, the English noun *bank* has more than one lexical meaning and each of the meanings corresponds to a different Chinese word. That fact aside, the GTAHS creates spurious ambiguities, which, in our view, would be harmful to Machine Translation (MT) if extracted as translation rules. Consider the following example, where the Chinese noun 苹果 is aligned to six English strings (aligned elements are underlined):

(1) a. eat apples <> 吃 苹果
  b. eat an apple <> 吃 苹果
  c. eat the apple <> 吃 苹果
  d. fond of apples <> 喜欢 苹果
  e. talk about apples <> 谈论 苹果
  f. provide them with apples <> 给 他们 苹果

The English *apple* and the Chinese 苹果 match in meaning and are both unambiguous. In cases where the English noun is used with a determiner as in (1b) and (1c), since Chinese has no determiners and the bare noun 苹果 can be the appropriate translation for either *an apple* or *the apple* given a context, the GTAHS attaches the determiner to *apple* and the whole string is aligned with 苹果. In other similar cases where an English element such as a preposition is absent in Chinese as in (1d), (1e) and (1f), the GTAHS glues the preposition to *apple* and the whole PP is aligned with 苹果. With the GTAHS, the unambiguous Chinese 苹果 ends up being aligned with more than one English string. This kind of spurious ambiguity is very common given the GTAHS.

The second issue is that, by attaching function words to a host, the GTAHS effectively creates rudimentary syntactic structures, which are often incompatible with the syntactic structures annotated based on existing treebanking annotation standards. For example, all the aligned multi-word strings underlined in (2) do not correspond to a constituent in a Penn TreeBank (Marcus et al., 1993) or Chinese TreeBank (Xue et al., 2005) parse tree:

(2) a. If I were him <> 如果 我 是 他 的话
  b. He is visiting Beijing <> 他 正 访问 北京

c. the beginning of the new year <> 新 年 伊始

d. to quickly and efficiently solve the problem <> 迅速 有效地 解决 问题

Given the incompatibilities between existing word alignments and syntactic structures, in the next section we describe an approach where we perform word-level and phrase-level alignments simultaneously on parallel phrase-based parse trees, attempting to construct a hierarchically aligned corpus where word alignments are harmonized with syntactic structures.

3 Annotation specification and procedure

The data we annotate is the Chinese-English portion of the Parallel Aligned Treebank (PAT) described in (Li et al., 2012). Our data consists of two batches, one of which is weblogs and the other of which is postings from online discussion forums. The English sentences in the data set are annotated based on the original Penn TreeBank (PTB) annotation stylebook (Bies et al., 1995) as well as its extensions (Warner et al., 2004), while the Chinese sentences in the data set are annotated based on the Chinese TreeBank (CTB) annotation guidelines (Xue and Xia, 2000) and its extensions (Zhang and Xue, 2012). The PAT only has word alignments, which are done under the GTAHS, and no phrase alignments.

The main departure of our approach is that we loosen the requirement that every word in a sentence pair needs to be word-aligned. On the word level, we only align words that have an equivalent in terms of lexical meaning and grammatical function. For words that do not have a translation counterpart, we leave them unaligned and locate the appropriate phrases in which they appear to be aligned. This way, we eliminate both the redundancies and spurious ambiguities discussed in Section 2. Since phrase alignment is done between syntactic nodes on parallel parse trees, we also eliminate the incompatibilities between word alignments and syntactic structures. See the discussion of the concrete example in Figure 1 below to see the points made here.

Next we discuss our annotation procedure in detail. Our annotators are presented with sentence pairs that come with parallel parse trees. The task of the annotator is to decide, first on the word level and then on the phrase level, if a word or phrase needs to be aligned at all, and if so, to which word or phrase it should be aligned. The decisions about word alignment and phrase alignment are not independent, and must obey well-formedness constraints as outlined in (Tinsley et al., 2007):

a. A non-terminal node can only be aligned once.
b. if Node n_c is aligned to Node n_e, then the descendants of n_c can only be aligned to descendants of n_e.
c. if Node n_c is aligned to Node n_e, then the ancestors of n_c can only be aligned to ancestors of n_e.

This means that once a word alignment is in place, it puts constraints on phrase alignments. A pair of non-terminal nodes (n_c, n_e) cannot be aligned if a word that is a descendant of n_c is aligned to a word that is not a descendant of n_e on the word level.

Let us use the concrete example in Figure 1 to illustrate the annotation process, which is guided by a set of detailed annotation guidelines. On the word level, only those words that are connected with a dashed line are aligned since they have equivalents. Note that the Chinese words 把 (a function word used to prepose the object to the left of the verb), 这样 (an adverb meaning "this way"), 可 (a modal meaning "can") and the English discourse connective *so that*, the auxiliary verb *is* and the preposition *from* are all left unaligned on the word level. Aligning these function words will generate artificial ambiguous cases and create incompatibilities between word alignments and parse trees that have already been illustrated and discussed in Section 2. For instance, if 把 is to be word-aligned, it would be glued to the noun 重力 and the whole string 把 重力 will be aligned to the English *gravity*. Note that both 重力 and *gravity* are unambiguous and form a one-to-one correspondence. With the word alignment between 把 重力 and *gravity*, we make the unambiguous *gravity* correspond to both 重力 and 把 重力 (and possibly

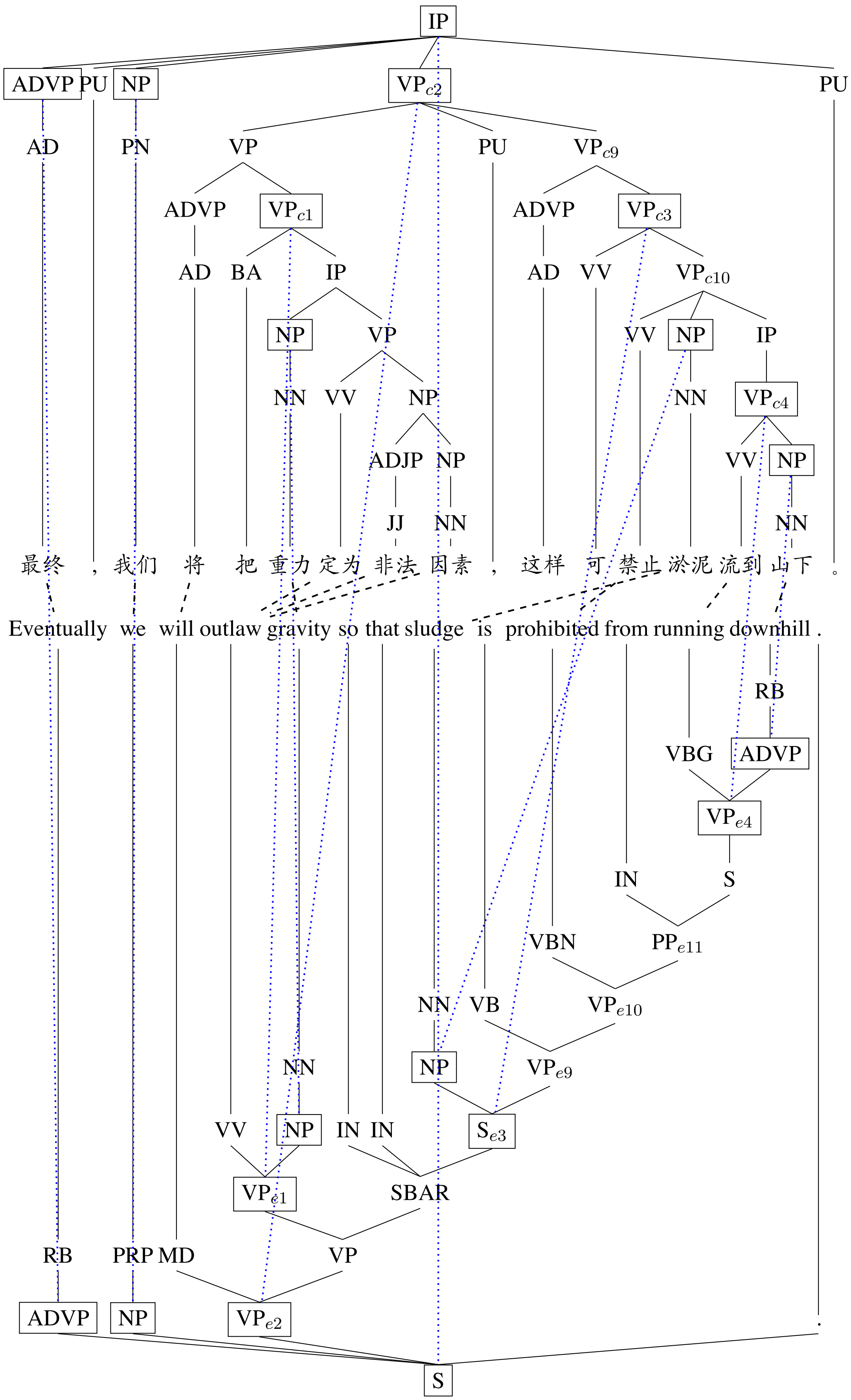


Figure 1: A hierarchically aligned sentence pair

more strings), thus creating a spurious ambiguity. Also note that the string 把 重力 does not form a constituent in the Chinese parse tree, so the word alignment is incompatible with the syntactic structure of the sentence. By leaving 把 unaligned, we avoid both the spurious ambiguity and the incompatibility.

With word alignments in place, next the annotator needs to perform phrase alignments. Note that word alignments place restrictions on phrase alignments. For instance, e_9 and e_{10} will be ruled out as possible alignments for c_{10}, because 淤泥, a descendant of c_{10}, is aligned to *sludge*, which is not a descendant of either e_9 or e_{10}. By contrast, e_3 is a possible alignment for c_{10} because the alignment does not violate the well-formedness constraints. The annotator then needs to decide whether this possible phrase alignment can be actually made. This is a challenging task since, for a given phrase, there usually are more than one candidate from which a single alignment needs to be picked. For instance, for e_3, there are in total three possible phrase alignments, namely c_{10}, c_3 and c_9, all of which obey the well-formedness constraints. Since a non-terminal node is not allowed to be aligned to multiple non-terminal nodes on the other side, the annotator needs to choose one among all the candidates. This highlights the point that the alignment of non-terminal nodes cannot be deterministically inferred from the alignment of terminal nodes. This is especially true given our approach where some terminal nodes are left unaligned on the word level. For instance, the reason why c_9 is a possible alignment for e_3 is because the word 这样 is left unaligned. If 这样 were aligned with *so that*, c_9 could not be aligned with e_3 since *so that* is not a descendant of e_3 and aligning the two nodes will violate Constraint *b*.

While Constraints *b* and *c* can be enforced automatically given the word alignments, the decisions regarding the alignment of non-terminal nodes which satisfy Constraint *a* are based on linguistic considerations. One key consideration is to determine which non-terminal nodes encapsulate the grammatical relations signaled by the unaligned words so that the alignment of the non-terminal nodes will effectively capture the unaligned words in their syntactic context. When identifying non-terminal nodes to align, we follow two seemingly conflicting general principles:

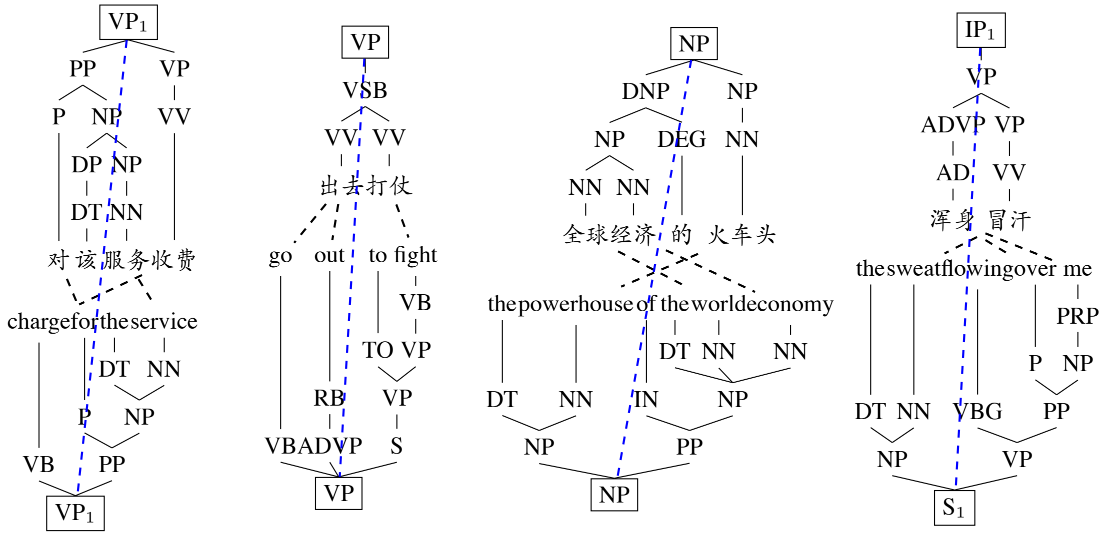
- Phrase alignment should not sever key dependencies involving the grammatical relation signaled by an unaligned word.
- Phrase alignment should be minimal, in the sense that the phrase alignment should contain only the elements involved in the grammatical relation, and nothing more.

The first principle ensures that the grammatical relation is properly encapsulated in the aligned non-terminal nodes. For example in Figure 1, if we attach the English preposition *from* to *running* and aligning them to 流到, we would fail to capture the fact that *from* signals a relation between *prohibit* and *running downhill*. Aligning VP_{c3} with S_{e3} captures this relation.

The first principle in and of itself is insufficient to produce desired alignment. Taken to the extreme, it can be trivially satisfied by aligning the two root nodes of the sentence pair. We also need the alignment to be minimal, in the sense that aligned non-terminal nodes should contain only the elements involved in the grammatical relation, and nothing more. These two requirements used in conjunction ensure that a unique phrase alignment can be found for each unaligned word. The phrase alignments (VP_{c1}, VP_{e1}), (VP_{c2}, VP_{e2}), (VP_{c3}, S_{e3}), as illustrated in Figure 1, all satisfy these two principles.

In addition to making phrase alignments, the annotator needs to assign labels to phrase alignments. We have four labels that are designed along two dimensions: the presence/absence of word order difference and the presence/absence of unaligned function words. The name and definition of each of the four labels are listed below, and an example for each label is given in Figure 2:

a REO, reordering that does not involve unaligned function words (Figure 2a)
b UFW, unaligned function words (Figure 2b)
c REU, reordering that also involves unaligned function words (Figure 2c)
d STD, structural divergence due to cross-linguistic differences (Figure 2d)



(a) REO (b) UFW (c) REU (d) STD

Figure 2: Phrase alignment types

Figure 2a is an example where there is a reordering of the immediate children of the aligned VP nodes. This is a very typical word order difference between Chinese and English. In Chinese, the PP modifier is before the verb while in English the PP modifier is after the verb. The phrase alignment illustrated by Figure 2b has an unaligned function word, namely the English infinitive marker *to*, which has no counterpart in Chinese. There are both reordering (difference in the relative order of *powerhouse* and *economy*) and unaligned function words (Chinese 的 and English *of*) in the phrase alignment in Figure 2c. Figure 2d provides an example where the aligned phrases have structural divergence caused by cross-linguistic differences between Chinese and English, which we will discuss in some detail in Section 4.

4 A common incompatibility issue between parse trees

During the annotation process, we encountered some incompatibility issues between parse trees. For a comprehensive and detailed discussion of the issues, see (Deng and Xue, 2014). Here we report the most common issue, which is caused by differences between treebank annotation guidelines. As already mentioned, the English parse trees we use are annotated based on the original PTB annotation stylebook (Bies et al., 1995) as well as its extensions (Warner et al., 2004), while the Chinese parse trees are annotated based on the CTB annotation guidelines (Xue and Xia, 2000) and its extensions (Zhang and Xue, 2012). Since PTB and CTB are independently annotated, there are some differences in how certain structures are annotated. The main issue is that certain structures are so flat as to make some nodes that should be aligned impossible to be aligned. In general, our alignment task favors deeper structures over shallower ones so that the annotator can have more choices. This is an issue for both Chinese and English parse trees. To get a concrete idea of the issue, take a look at Figure 3.

As shown by Figure 3, VP_{c1} and the English string *probably decrease rapidly with distance*, and VP_{e1} and the Chinese string 随 距离 而 快速 减少, cannot be aligned although they match in meaning and should be aligned. They cannot be aligned because there is no node for either of the two strings in the respective parse tree. Note that the incompatibility between the two trees here is due to a difference in annotation style but not a deep cross-linguistic difference. Both PTB and CTB simplified the annotation task by making the tree structures flatter to increase annotation speed, but the simplification does not always come from the same places. The consequence of these annotation decisions is that relevant structures are sometimes incompatible, which has negatively affected their utility for MT purposes (Chiang,

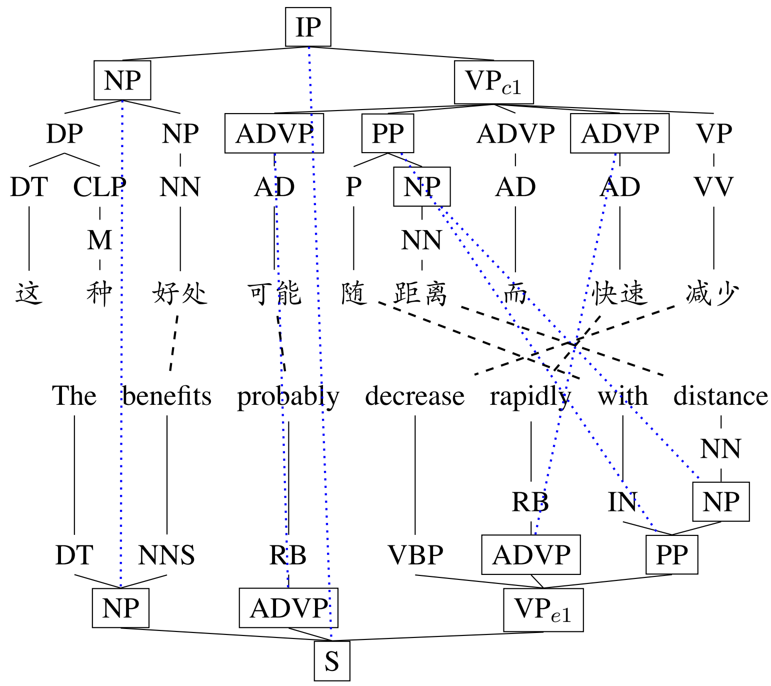


Figure 3: Unalignable nodes due to differences in tree representation

2010).

To solve this incompatibility issue, we need to create more structures through binarization, which can be done automatically. Still take Figure 3 for instance, on the English side, if we create a new VP by combining VP$_{e1}$ and its sister ADVP, the resulting VP can be aligned with VP$_{c1}$. On the Chinese side, if we do binarization to create a VP that dominates the string 随 距离 而 快速 减少, VP$_{e1}$ would have an alignment. Since changing tree structures has the potential risk of causing inconsistency with parse trees in the original treebanks and had better be done systematically after all the annotation is finished, we have not done binarization as of the writing of this paper. For the time being, we assign the label UA (short for Unalignable Node) to nodes which should be aligned but cannot be aligned so that we can gather some statistics on the extent of the problem. We will come back to revisit the nodes carrying UA such as VP$_{c1}$ and VP$_{e1}$ by proposing systematic changes to the original treebanks.

The UA case discussed above should not be confused with another case of incompatibility, namely structural divergence between parallel sentences in translation (Dorr, 1994). As shown above, UA is basically an artificial issue that is caused by difference in parsing guideline design and fixable through automatic binarization. Structural divergence arises mainly due to genuine cross-linguistic differences. We provide an example of structural divergence (STD) in Figure 2d. As shown in the figure, the two aligned phrases (VP and S) are structurally quite different: the English string is a clause with the NP *the sweat* as the subject and the VP *flowing over me* as the predicate (the example is taken out of the sentence *I felt the sweat flowing over me* to save space). The Chinese string is a simple verb phrase where the adverb 浑身 (literally whole-body) modifies the verb 冒汗 (literally emerge-sweat). In terms of meaning correspondence, 浑身 expresses the meaning of the English PP *over me* and the verb matches in meaning with *the sweat flowing*. We have run an experiment on STD and found that the STD cases are pretty rare (on average 5 instances in a file with 500 sentence pairs), indicating that the structural difference between Chinese and English is not so fundamental as to make a big impact on alignment annotation.

5 Annotation experiments

We did two experiments on our annotation. The first is about inter-annotator agreement (IAA), which is a way of both evaluating the annotation quality and judging the intuitiveness of the annotation task. An unintuitive annotation task would force the annotator to make subjective choices, which would result in low IAA. Since the annotation task involves parse trees, ideally we need annotators who are trained in syntax, but that would put a constraint on the pool of qualified annotators and make it difficult for the annotation to scale up. In our annotation experiments, we use four annotators who are fluent in both English and Chinese but have no prior linguistic training, led by a syntactician who performs the final adjudication.

As of this writing, we have completed the single annotation of 8,932 sentence pairs, 2,500 of which are double annotated. The IAA statistics presented in Table 1 are based on the double-annotated 2,500 sentence pairs, which are divided into 5 chunks of 500 sentence pairs each. The statistics are for phrase alignment only, and the micro-average for the 5 chunks is 0.87 (F1), indicating we are able to get good quality annotation for this task. In addition, the agreement statistics for the 5 chunks are very stable, even though they are performed by different pairs of annotators, indicating we are getting consistent annotation from different annotators.

Table 2 shows the result of the second experiment, namely the distribution of the different types of phrase alignment. It shows that alignments that contain unaligned function words outnumber those that do not, and that alignments that do not involve reordering outnumber those that do. It also shows that an overwhelming number of alignments that involve reordering also have unaligned function words. This means that the function words are potentially useful "triggers" for reordering, which is an important issue that MT systems are trying to address.

Chunk No.	precision	recall	F1-measure
1	0.91	0.86	0.89
2	0.92	0.80	0.86
3	0.89	0.89	0.89
4	0.88	0.88	0.88
5	0.89	0.89	.086
micro-average	0.90	0.85	0.87

Table 1: Statistics of IAA

	Annotator	+UFW	-UFW	total
+REO	1	6,473	379	6,852
	2	6,670	379	7,049
-REO	1	7,328	6,872	14,200
	2	7,797	7,334	15,131
total	1	13,801	7,251	21,052
	2	14,467	7,713	22,180

Table 2: Statistics of phrase alignment by types

6 Related work

Parallel treebanks are not something new. However, most of the existing parallel treebanks (Li et al., 2012; Megyesi et al., 2010) do not have phrase alignments. Some (Sulger et al., 2013; Kapanadze, 2012) do have phrase alignments, but neither discussion about the interaction between word-level and phrase-level alignments nor report of IAA is provided. There have been a few recent attempts at automatically aligning subtrees (comparable to our phrases) in the context of MT research, and the automatic alignments are evaluated against a small manually aligned data set. For example, (Tinsley et al., 2007) evaluated an unsupervised algorithm on 810 parsed English-French pairs annotated with subtree alignment. (Xiao and Zhu, 2013) also developed unsupervised subtree alignment methods (EM and Variational Bayes) and evaluated their automatic alignment model on 637 sentences from the Chinese TreeBank (and use the other 99 for tuning). (Sun et al., 2010b; Sun et al., 2010a) also report work on aligning subtrees and evaluate their impact on MT. However, we are not aware of any attempt to systematically harmonize word alignment with the alignment of phrases, or subtrees, or to systematically study the incompatibilities between parallel parse trees.

7 Conclusion

In this paper we report our effort on the construction of a Chinese-English parallel treebank with both word-level and phrase-level alignments. When constructing the treebank, we systematically consider the interaction between word alignments and phrase alignments, and try to harmonize the two kinds of alignments by removing redundancies and incompatibilities between them. We show that this type of annotation can be performed with high inter-annotator consistency. Given our intention for the treebank to be a resource for MT, the next step is to synchronize the parallel parse trees, and of course, to perform automatic hierarchical alignment experiments and MT experiments.

Acknowledgements

This work is supported by the IBM subcontract No. 4913014934 under DARPA Prime Contract No. 0011-12-C-0015 entitled "Broad Operational Language Translation". We would like to thank Libin Shen and Salim Roukos for their inspiration and discussion during early stages of the project, Abe Ittycheriah and Niyu Ge for their help with setting up the data, Loretta Bandera for developing and maintaining the annotation tool, and three anonymous reviewers for their helpful comments. We are grateful for the hard work of our four annotators: Hui Gao, Shiman Guo, Tse-ming Wang and Lingya Zhou. Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the sponsor or any of the people mentioned above.

References

Lars Ahrenberg. 2007. LinES: An English-Swedish parallel Treebank. In *Proceedings of Nodalida 2007*, pages 270--273, Tartu, Estonia.

Ann Bies, Mark Ferguson, Karen Katz, Robert MacIntyre, Victoria Tredinnick, Grace Kim, Mary Ann Marcinkiewicz, and Britta Schasberger. 1995. Bracketing guidelines for Treebank II style Penn Treebank project. Technical report, University of Pennsylvania.

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

David Chiang. 2010. Learning to translate with source and target syntax. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 1443--1452.

John Denero and Dan Klein. 2007. Tailoring word alignments to syntactic machine translation. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*, pages 17--24.

Dun Deng and Nianwen Xue. 2014. Aligning Chinese-English parallel parse trees: is it feasible? In *Proceedings of the 8th Linguistic Annotation Worshop (the LAW VIII)*.

Bonnie J. Dorr. 1994. Machine translation divergences: a formal description and proposed solution. *Computational Linguistics*, 20(4):597--633.

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

Victoria Fossum, Kevin Knight, and Steven Abney. 2008. Using syntax to improve word alignment precision for syntax-based machine translation. In *Proceedings of the Third Workshop on Statistical Machine Translation*, pages 44--52.

Michel Galley, Mark Hopkins, Kevin Knight, and Daniel Marcu. 2004. What's in a translation rule? In *HLT-NAACL*, pages 273--280.

Michel Galley, 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.

Liang Huang, Kevin Knight, and Aravind Joshi. 2006. Statistical syntax-directed translation with extended domain of locality. In *Proceedings of AMTA*, pages 66--73.

Oleg Kapanadze. 2012. Building parallel Treebanks for the lesser-resourced languages. Technical report, Universität des Saarlandes.

Xuansong Li, Niyu Ge, and Stephanie Strassel. 2009. Tagging guidelines for Chinese-English word alignment. Technical report, Linguistic Data Consortium.

Xuansong Li, Stephanie Strassel, Stephen Grimes, Safa Ismael, Mohamed Maamouri, Ann Bies, and Nianwen Xue. 2012. Parallel Aligned Treebanks at LDC: New Challenges Interfacing Existing Infrastructures. In *Proceedings of LREC-2012*, Istanbul, Turkey.

Ding Liu and Daniel Gildea. 2008. Improved tree-to-string transducer for machine translation. In *Proceedings of the Third Workshop on Statistical Machine Translation*, pages 62--69.

Yang Liu, Qun Liu, and Shouxun Lin. 2006. Tree-to-string alignment template for statistical machine translation. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics*, pages 609--616.

Yang Liu, Yajuan Lü, and Qun Liu. 2009. Improving tree-to-tree translation with packed forests. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, pages 558--566.

Mitchell P. Marcus, Beatrice Santorini, and Mary Ann Marcinkiewicz. 1993. Building a Large Annotated Corpus of English: The Penn Treebank. *Computational Linguistics*, 19(2):313--330.

Beata Megyesi, Bengt Dahlqvist, Eva A. Csato, and Joakim Nivre. 2010. The English-Swedish-Turkish Parallel Treebank. In *Proceedings of LREC-2010*, Valletta, Malta.

I. Dan Melamed. 1998. Annotation style guide for the Blinker project. Technical report, University of Pennsylvania.

Adam Pauls, Dan Klein, David Chiang, and Kevin Knight. 2010. Unsupervised syntactic alignment with inversion transduction grammars. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 118--126.

Sebastian Sulger, Miriam Butt, Tracy Holloway King, and Paul Meurer et.al. 2013. ParGramBank: The ParGram Parallel Treebank. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, pages 550--560.

Jun Sun, Min Zhang, and Chew Lim Tan. 2010a. Discriminative induction of sub-tree alignment using limited labeled data. In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 1047--1055.

Jun Sun, Min Zhang, and Chew Lim Tan. 2010b. Exploring syntactic structural features for sub-tree alignment using bilingual tree kernels. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 306--315.

John Tinsley, Ventsislav Zhechev, Mary Hearne, and Andy Way. 2007. Robust language pair-independent subtree alignment. In *Proceedings of Machine Translation Summit XI*.

Colin Warner, Ann Bies, Christine Brisson, and Justin Mott. 2004. Addendum to the Penn Treebank II style bracketing guidelines: BioMedical Treebank annotation. Technical report, University of Pennsylvania.

Tong Xiao and Jingbo Zhu. 2013. Unsupervised sub-tree alignment for tree-to-tree translation. *Journal of Artificial Intelligence Research*, 48:733--782.

Nianwen Xue and Fei Xia. 2000. The bracketing guidelines for Penn Chinese Treebank project. Technical report, University of Pennsylvania.

Nianwen Xue, Fei Xia, Fu dong Chiou, and Martha Palmer. 2005. The Penn Chinese Treebank: Phrase Structure Annotation of a Large Corpus. *Natural Language Engineering*, 11(2):207--238.

Xiuhong Zhang and Nianwen Xue. 2012. Extending and scaling up the chinese treebank annotation. In *Proceedings of the Second CIPS-SIGHAN Joint Conference on Chinese Language Processing*.