Capturing Translational Divergences with a Statistical Tree-to-Tree Aligner

Mary Hearne, John Tinsley, Ventsislav Zhechev and Andy Way

National Centre for Language Technology
Dublin City University
Dublin, Ireland
{mhearne,jtinsley,vzhechev,away}@computing.dcu.ie

Abstract

Parallel treebanks, which comprise paired source-target parse trees aligned at sub-sentential level, could be useful for many applications, particularly data-driven machine translation. In this paper, we focus on how translational divergences are captured within a parallel treebank using a fully automatic statistical tree-to-tree aligner. We observe that while the algorithm performs well at the phrase level, performance on lexical-level alignments is compromised by an inappropriate bias towards coverage rather than precision. This preference for high precision rather than broad coverage in terms of expressing translational divergences through tree-alignment stands in direct opposition to the situation for SMT word-alignment models. We suggest that this has implications not only for tree-alignment itself but also for the broader area of induction of syntax-aware models for SMT.

1 Introduction

Previous work has argued for the development of parallel treebanks, defined as bitexts for which the sentences are annotated with syntactic trees and are aligned below clause level (Volk and Samuelsson, 2004). Such resources could be useful for many applications, e.g. as training or evaluation corpora for word and phrase alignment, as training material for data-driven MT systems and for the automatic induction of transfer rules, and for translation studies. Their development is particularly pertinent to the recent efforts towards incorporating syntax into data-driven MT systems, e.g. (Melamed, 2004), (Chiang, 2005), (Galley et al., 2006), (Hearne and Way, 2006), (Marcu et al., 2006), (Zollmann and Venugopal, 2006).

In this paper, we focus on how translational divergences are captured within a parallel treebank using a fully-automatic statistical tree-to-tree aligner.[1] In doing so, we take a somewhat different perspective on tree-alignment from that of e.g. (Wu, 2000; Wellington et al., 2006). We do not incorporate trees for the express purpose of constraining the word- and phrase-alignment processes, although this is certainly a consequence of using trees. Our purpose in aligning monolingual syntactic representations is to make explicit the syntactic divergences between sentence pairs rather than homogenising them. We are not seeking to maximise the number of links between a given tree pair, but rather to find the set of links which most precisely expresses the translational equivalences between that tree pair. How best to exploit such information through model induction for syntax-aware statistical MT remains an open question.

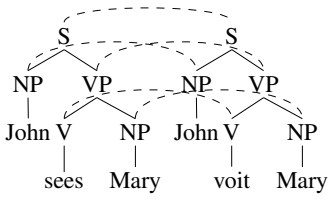
The remainder of this paper is organised as follows. In Section 2 we describe the tree-to-tree alignment process from a manual annotation per-

[1] Although the definition of a parallel treebank leaves room for a variety of types of tree structure, in this paper we focus on constituent structure trees only.

spective, outlining crucial ways in which it differs from the word-alignment process. We show how translational divergences are represented in an aligned parallel treebank in Section 3, giving insights into why such resources would be useful. In Section 4 we outline an automatic method for statistically inducing tree alignments between parsed sentence pairs – full details of the alignment algorithm are given in (Tinsley et al., 2007). In Section 5 we analyse the output to see how well translational divergences are captured. Finally, in Sections 6 and 7 we conclude and describe plans for future work.

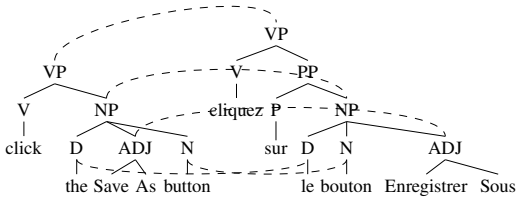
2 Manual Tree-to-Tree Alignment

The tree-to-tree alignment process assumes a parsed, translationally equivalent sentence pair and involves introducing links between non-terminal nodes in the source and target phrase-structure trees. Inserting a link between a node pair indicates that the substrings dominated by those nodes are translationally equivalent, i.e. that all meaning in the source substring is encapsulated in the target substring and vice versa. An example aligned English–French tree pair is given in (1). This example illustrates the simplest possible scenario: the sentence lengths are identical, the word order is identical and the tree structures are isomorphic.



(1)

However, most real-world examples do not align so neatly, as we will discuss in Section 3. The example given in (2) illustrates some important points. Not every node in each tree needs to be linked, e.g. *click* translates not as *cliquez*, but as *cliquez sur*. However, each node is linked at most once. Also, as we do not link terminal nodes, the lowest links are at the part-of-speech level. This means that multi-word units identified during parsing are preserved as such during alignment, cf. *Save As* and *Enregistrer Sous*.[2]



(2)

2.1 Tree Alignment vs. Word Alignment

When deciding how to go about linking a given tree pair, the logical starting point would seem to be with word alignment. However, some analysis reveals differences between the tasks of tree-alignment and word-alignment. We illustrate the differences by referring to the Blinker annotation guidelines (Melamed, 1998) which were used for the word alignment shared tasks at the workshops on *Building and Using Parallel Texts* at HLT-NAACL 2003[3] and ACL 2005.[4]

If a word is left unaligned in a sentence pair, it implies that the meaning it carries was not realised anywhere in the target string. On the other hand, if a node remains unaligned in a tree pair there is no equivalent implication. Because tree-alignment is hierarchical, many other nodes can carry indirect information regarding how an unaligned node (or group of unaligned nodes) is represented in the target string. Some consequences of this are as follows.

Firstly, the strategy in word-alignment is to leave as few words unlinked as possible "even when non-literal translations make it difficult to find corresponding words" (Melamed, 1998). Contrast this with the more conservative guideline for tree-alignment given in (Samuelsson and Volk, 2006): nodes are linked only when the substrings they dominate "represent the same meaning and ... could serve as translation units outside the current sentence context." This latter strategy is affordable because alignments at higher levels in the tree pair will account for the translation equivalence. Secondly, word-alignment allows many-to-many alignments at the word level but not phrasal alignments unless every word in the source phrase corresponds to every word in

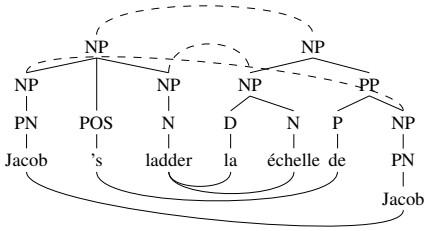
[2] Of course, an alternative parsing scheme which gives internal labelled structure in such phrases might permit further sub-tree links.

[3] http://www.cse.unt.edu/~rada/wpt/

[4] http://www.cse.unt.edu/~rada/wpt05/

the target and vice versa. Tree-alignment, on the other hand, allows each node to be linked only once but facilitates phrase alignment by allowing links higher up in the tree pair.

The contrasting effects of these guidelines are illustrated by the example given in (3)[5] where the dashed links represent tree-alignments and the solid links represent word-alignments. We see that the word-alignment must link *ladder* to both *la* and *échelle* whereas the tree-alignment specifies a single link between the nodes dominating the substrings *ladder* and *l'échelle*.

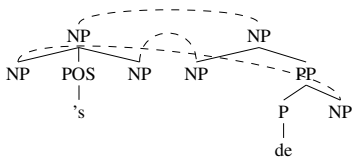


(3)

Note also that the word-alignment explicitly links *'s* with *de* whereas the tree-alignment does not; it is arguable as to whether these strings really represent precisely the same meaning. However, the relationship between these words is not ignored in the tree-alignment; rather it is captured by the link between the three NP links in combination.

In fact, many different pieces of information can be inferred from the tree-alignment given in (3) regarding the relationship between *'s* and *de*, despite the fact that they are not directly linked; examples exhibiting varying degrees of contextual granularity are given in (4).

$$
\begin{array}{ccc}
\textit{'s} & \longrightarrow & \textit{de} \\
\text{X } \textit{'s} \text{ Y} & \longrightarrow & \text{Y } \textit{de} \text{ X} \\
\text{NP}_1 \textit{ 's } \text{NP}_2 & \longrightarrow & \text{NP}_2 \textit{ de } \text{NP}_1 \\
\text{NP} \rightarrow \text{NP}_1 \textit{ 's } \text{NP}_2 & : & \text{NP} \rightarrow \text{NP}_2 \textit{ de } \text{NP}_1
\end{array}
\qquad (4)
$$



It is noteworthy, we feel, that the similarities between the 'rules' in (4) and templates in EBMT such as those in (Cicekli and Güvenir, 2003) are striking.

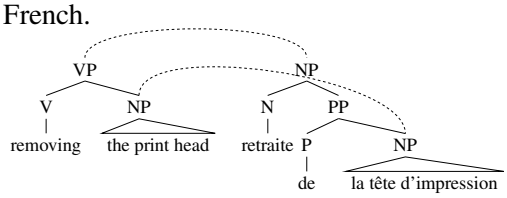
[5]The sentence pair and word alignments were taken directly from (Melamed, 1998).

3 Translational Divergences

Work such as that of e.g. (Lindop and Tsujii, 1992; Dorr, 1994; Trujillo, 1999) makes explicit the types of translational divergences which occur in real data. These divergences occur frequently even for language pairs with relatively similar surface word order, and generally prove challenging for MT models (Hutchins and Somers, 1992).[6] An important characteristic of parallel treebanks is that they provide explicit details, through tree-alignments, about the occurrence and nature of such divergences.

In this section, we examine how translational divergences are represented in the HomeCentre English–French parallel treebank. This corpus comprises a Xerox printer manual which was translated by professional translators and sentence-aligned and annotated at Xerox PARC. It contains 810 parsed, sentence-aligned English-French translation pairs. It was manually tree-aligned by one of the authors of this paper according to the guidelines outlined in Section 2.[7] As observed by (Frank, 1999), the HomeCentre corpus provides a rich source of both linguistic and translational complexity.

Instances of nominalisation are very frequent in the HomeCentre corpus. An example of a **simple nominalisation** is given in (5), where the English verb phrase *removing the print head* is realised as the noun phrase *retraite de la tête d'impression* in French.



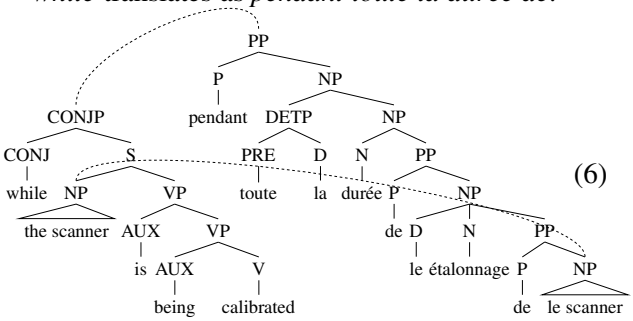
(5)

Instances of more **complex nominalisations** which incorporate further translational divergences are also common. Consider, for example, the translation pair given in (6). Firstly, we note the nominalisation: the English passive sentential form *the scanner is being calibrated* is realised as the French noun phrase *l'étalonnage*

[6]The picture is even more complex than we paint here; (Dorr et al., 2002) make the further observation that such 'hard' cases tend to co-occur much more often than might be expected.

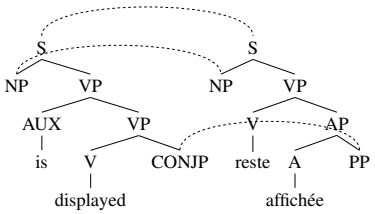
[7]As there was just a single annotator, inter-annotator agreement is obviously not a factor.

du scanner. However, we also observe the presence of **relation-changing**: the subject of this English sentential form, *the scanner*, functions as an oblique object in the French translation. In addition, this example exhibits stylistic divergence, as *while* translates as *pendant toute la durée de*.



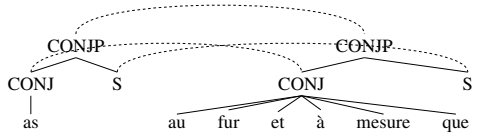
(6)

Another complex translation case which occurs in the HomeCentre corpus is that of **head-switching**, where the head word in the source language sentence translates as a non-head word in the target language realisation. An example of head-switching is given in (7). Here, the English verbal unit *is displayed* is realised in French as *reste affichée*; in this context, *reste* means (roughly) 'remains' and *display* is realised as the adverbial modifier *affichée*. Thus, the head of the English sentence, the verb *display*, corresponds to the French non-head word *affichée*.



(7)

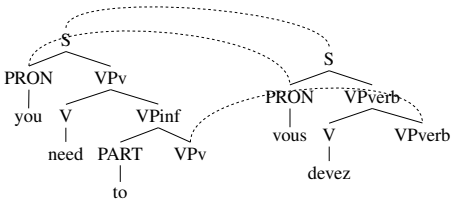
Of course, **lexical divergences** also occur frequently. In some instances, these divergences can be resolved in a straightforward manner. For example, we see in (8) that *as* in English can translate as *au fur et à mesure que* in French, but as the idiomatic reading of this French phrase is reflected in the parse assigned to the sentence, the overall shape of the sentence can remain the same despite the complexity of the translation.



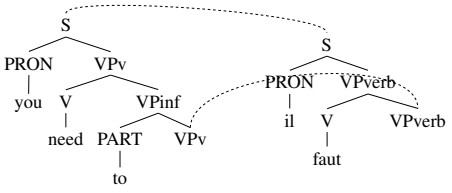
(8)

However, even for a relatively similar language pair, lexical divergence can cause source and target sentences expressing exactly the same concept to have completely **different surface realisations**. Consider, for example, the translation pair in Figure 1. As there is no French phrase which is directly equivalent to the English expression *null and void*, the given French sentence *toute intervention non autorisée invaliderait la garantie* – which translates roughly as 'any unauthorised action would invalidate the guarantee' – is entirely structurally dissimilar to its English counterpart.

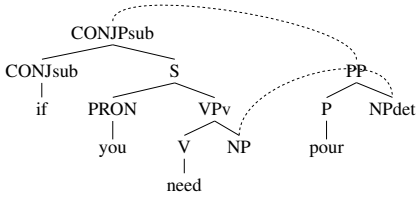
Finally, variation in how certain **frequently-occurring words** are translated, depending on the context in which the word appears, is also common. Examples (9) – (12) illustrate this phenomenon for the English verb *need*. *you need to X* can be realised as both *vous devez X* and *il faut X* in French, as shown in examples (9) and (10). The realisation differs, however, where the object is nominal rather than sentential: *if you need X* is shown in (11) to translate as *pour X*. Finally, we show in example (12) that the negative *you do not need to X* can translate as *il ne devrait pas être necessaire de X*, which literally means 'it should not be necessary to X' in English. We note that this is just a subset of the differing French realisations for the verb *to need* which occur in the HomeCentre corpus.



(9)



(10)



(11)

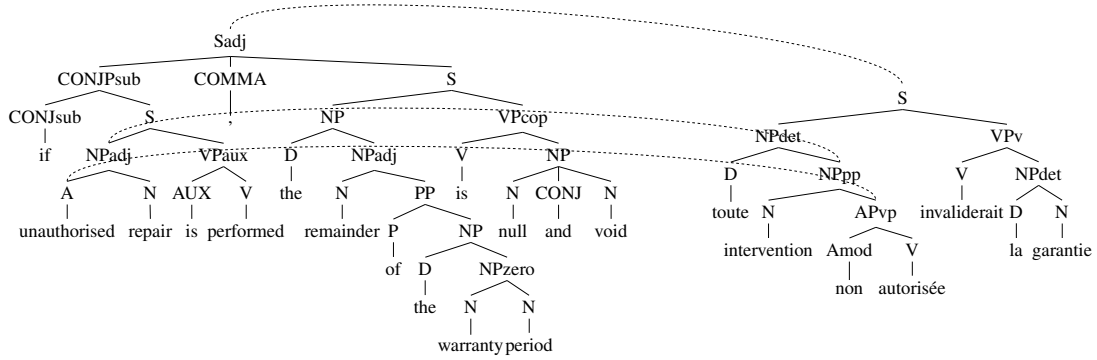
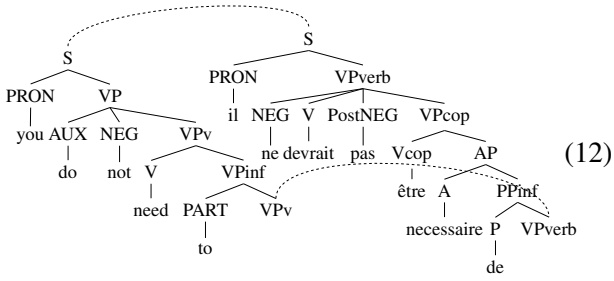


Figure 1: Completely different surface realisations can be seen even for language pairs with similar word order like English–French.



(12)

4 Automatic Tree-to-Tree Alignment

The tree-alignment algorithm briefly described here and detailed in (Tinsley et al., 2007) is designed to discover an optimal set of alignments between the tree pairs in a bilingual treebank while adhering to the following principles:

(i) independence with respect to language pair and constituent labelling schema;

(ii) preservation of the given tree structures;

(iii) minimal external resources required;

(iv) word-level alignments not fixed *a priori*.

4.1 Alignment Well-Formedness Criteria

Links are induced between tree pairs such that they meet the following well-formedness criteria:

(i) a node can only be linked once;

(ii) descendants of a source linked node may only link to descendants of its target linked counterpart;

(iii) ancestors of a source linked node may only link to ancestors of its target linked counterpart.

In what follows, a hypothesised alignment is ill-formed with respect to the existing alignments if it violates any of these criteria.

4.2 Algorithm

In this section we present how our alignment algorithm scores and selects links. We refer to the alternative methods by which decisions can be made at various points, and summarise the possible aligner configurations. (Tinsley et al., 2007) describes these variations in greater details and provides the motivation behind each variant.

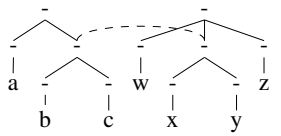
4.2.1 Selecting Links

For a given tree pair $\langle S, T \rangle$, the alignment process is initialised by proposing all links $\langle s, t \rangle$ between nodes in S and T as hypotheses and assigning scores $\gamma(\langle s, t \rangle)$ to them. All zero-scored hypotheses are blocked before the algorithm proceeds. The selection procedure then iteratively fixes on the highest-scoring link, blocking all hypotheses that contradict this link and the link itself, until no non-blocked hypotheses remain. These initialisation and selection procedures are given in **Algorithm 1** *basic*.

Algorithm 1 *basic*

Initialisation
  for each source non-terminal *s* **do**
    for each target non-terminal *t* **do**
      generate scored hypothesis $\gamma(\langle s, t \rangle)$
    end for
  end for
  block all zero-scored hypotheses
Selection *underspecified*
  while non-blocked hypotheses remain **do**
    link and block the highest-scoring hypothesis
    block all contradicting hypotheses
  end while

Hypotheses with equal scores: The **Selection** procedure given in **Algorithm 1** *basic* is incom-



s_l = b c
t_l = x y
$\overline{s_l}$ = a
$\overline{t_l}$ = w z

Figure 2: Values for s_l, t_l, $\overline{s_l}$ and $\overline{t_l}$ given a tree pair and a link hypothesis.

plete as it does not specify how to proceed if two or more hypotheses share the same highest score. When this case arises we invoke a method called *skip2*. Using this configuration, we skip over tied hypotheses until we find the highest-scoring hypothesis $\langle s, t \rangle$ with no competitors of the same score *and where neither s nor t has been skipped*.

Delaying lexical (span-1) alignments: It is sometimes the case that we want to delay the induction of lexical links in order to allow links higher up in the tree structures to be induced first. For this reason we have an optional configuration, *span1*. When this method is activated, it postpones links between any hypothesis $\langle x, y \rangle$, where either x or y is a constituent with a span of one, i.e. a lexical node. Only when all other possible hypotheses have been exhausted do we allow links of type $\langle x, y \rangle$.

4.2.2 Computing Hypothesis Scores

Inserting a link between two nodes in a tree pair indicates that (i) the substrings dominated by those nodes are translationally equivalent and (ii) all meaning carried by the remainder of the source sentence is encapsulated in the remainder of the target sentence. The scoring method we propose accounts for these indications.

Given tree pair $\langle S, T \rangle$ and hypothesis $\langle s, t \rangle$, we compute the following strings:

$$s_l = s_i ... s_{ix} \quad \overline{s_l} = S_1 ... s_{i-1} s_{ix+1} ... S_m$$
$$t_l = t_j ... t_{jy} \quad \overline{t_l} = T_1 ... t_{j-1} t_{jy+1} ... T_n$$

where $s_i ... s_{ix}$ and $t_j ... t_{jy}$ denote the terminal sequences dominated by s and t respectively, and $S_1 ... S_m$ and $T_1 ... T_n$ denote the terminal sequences dominated by S and T respectively. These string computations are illustrated in Figure 2.

The score for the given hypothesis $\langle s, t \rangle$ is computed according to (13).

$$\gamma(\langle s, t \rangle) = \alpha(s_l|t_l)\ \alpha(t_l|s_l)\ \alpha(\overline{s_l}|\overline{t_l})\ \alpha(\overline{t_l}|\overline{s_l}) \quad (13)$$

Individual string-correspondence scores $\alpha(x|y)$ are computed using word-alignment probabilities given by the Moses decoder[8,9] (Koehn et al., 2007). Two alternative scoring functions are given by *score1* (14) and *score2* (15).

Score *score1*

$$\alpha(x|y) = \prod_{j=1}^{|y|} \sum_{i=1}^{|x|} P(x_i|y_j) \quad (14)$$

Score *score2*

$$\alpha(x|y) = \prod_{i=1}^{|x|} \frac{\sum_{j=1}^{|y|} P(x_i|y_j)}{|y|} \quad (15)$$

4.3 Aligner Configurations

When configuring the aligner, we must choose *skip2* and we must choose either *score1* or *score2*. *span1* can be switched either on or off. The four possible configurations are as follows:

skip2_score1	skip2_score1_span1
skip2_score2	skip2_score2_span1

5 Alignment Evaluation and Analysis

In Section 5.1 we give an overview of aligner performance through two automatic evaluation methodologies. In Section 5.2 we then go on to describe the capture of translational divergences by manually analysing the aligner output.

5.1 Automatic Evaluation

We use two automatic evaluation methodologies in order to gain an overview of aligner performance: (i) we compare the links induced by the algorithm to those induced manually and compute precision and recall scores; (ii) we train a Data-Oriented Translation (DOT) system (Hearne and Way, 2006) on both the manually aligned data and the automatically aligned data and assess translation accuracy using the Bleu (Papineni et al., 2002), NIST (Doddington, 2002) and Meteor

[8] http://www.statmt.org/moses/

[9] Although our method of scoring is similar to IBM model 1, and Moses runs GIZA++ trained on IBM model 4, we found that using the Moses word-alignment probabilities yielded better results than those output directly by GIZA++.

	Alignment Evaluation						Translation Evaluation			
	all links		*lexical links*		*non-lexical links*		*(all links)*			
Configurations	Precision	Recall	Precision	Recall	Precision	Recall	Bleu	NIST	Meteor	Coverage
manual	–	–	–	–	–	–	0.5222	6.8931	71.8531	68.5417
skip2_score1	0.6162	0.7783	0.5057	0.7441	**0.8394**	0.7486	0.5091	6.9145	71.7764	71.8750
skip2_score2	0.6215	0.7876	0.5131	0.7431	0.8107	0.7756	**0.5333**	6.8855	**72.9614**	**72.5000**
skip2_score1_span1	**0.6256**	**0.8100**	0.5163	**0.7626**	0.8139	**0.8002**	0.5273	**6.9384**	72.7157	**72.5000**
skip2_score2_span1	0.6245	0.7962	**0.5184**	0.7517	0.8031	0.7871	0.5290	6.8762	72.8765	**72.5000**

Table 1: Evaluation of aligner performance using automatic metrics.

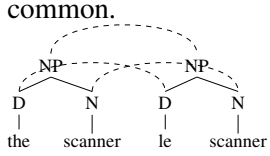
(Banerjee and Lavie, 2005) automatic evaluation metrics. The results of these evaluations are given in Table 1.

Looking firstly at overall alignment accuracy (the *all links* column), it is immediately apparent that recall is significantly higher than precision for all configurations. In fact, we have observed that all aligner variations consistently induce more links than exist in the manual version, with the average number of links per tree pair ranging between 10.4 and 11.0 for the automatic alignments versus 8.3 links per tree pair for the manual version. A clearer picture emerges when we differentiate between lexical and non-lexical links, where a link is non-lexical if both source and target nodes span more than one terminal. We see that, actually, precision is higher than recall for non-lexical links, and overall accuracy is higher for non-lexical links than for all links. In contrast, overall accuracy is much lower for lexical links than for all links, and the disparity between precision and recall is greater.

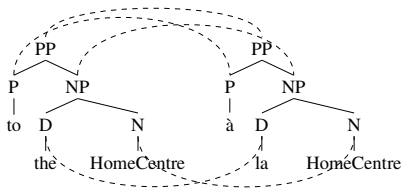
Turning our attention to translation accuracy, we observe that the scores for the automatic alignments are very encouraging: for all three evaluation metrics, at least two aligner configurations outperform the manual scores. Furthermore, all the automatically-aligned datasets achieve higher coverage than the manually-aligned run. It is perhaps somewhat surprising that the translation scores do not reflect the indication given by the alignment evaluation that word-level alignment precision is low compared to phrase-level precision. The explanation as to why the translation scores do not deteriorate may lie in how the MT system works: because DOT displays a preference for using larger fragments when building translations wherever possible, the impact of inconsistencies amongst smaller fragments (i.e. word-level alignments) is minimised. The reason for the improvement in scores lies in the increased coverage of the system trained on the automatic alignments.

5.2 Capturing translational divergences

Before looking at divergent cases, we first observe that the alignment algorithm generally produces accurate output for the simple translation cases. Examples (16) and (17) illustrate cases where the aligner correctly identifies equivalent constituents where length, word order and tree structure all match perfectly. For short phrases, such examples are relatively common.



(16)

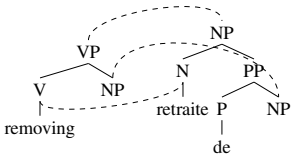


(17)

Lexical divergences which are of the form 1-to-many and many-to-1 occur frequently in the data and the aligner captures them with regularity. For example, the aligner output exactly matches the manual alignment for example (8). As mentioned in Section 4, when calculating the score for a particular hypothesis, we not only consider the translational equivalence of the dominated substrings but also the translational equivalence of the remainder of the source and target sentences. In this way, links can be inferred even when the constituent substrings are lexically divergent.

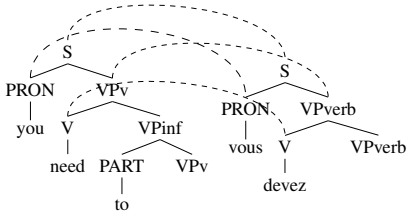
Instances of nominalisation are also commonly presented to the aligner. Consider, for example, the aligner output in (18) where the English verb phrase *removing the print head* is re-

alised as the French noun phrase *retraite de la tête d'impression*. As the aligner does not take into consideration the labels on the tree, but rather the likelihood that the surface strings are translations of each other, there is no impediment to the linking of the English VP to the French NP. Furthermore, the lower NP alignment is straightforward. Note, however, the (probably incorrect) link between the V *removing* and the N *retraite*. This link does not appear in the manual alignment (shown in (5)) as the annotator considered the meaning equivalence to be between *removing* and *retraite de*.

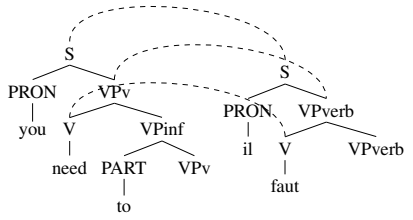


(18)

In Section 3 we noted that frequently-occurring words vary greatly in terms of how they are translated, as illustrated for the English verb *need* in examples (9) – (12). These examples are handled reasonably well by the aligner, again due to the strength of the equivalence between the object constituents. In (19) and (20) (for which the manual alignments were given in (9) and (10)), we again see lexical alignments in the automatic output which were not included in the manual versions; the annotator considered the equivalences to be (*need to*, *devez*) and (*you need to*, *il faut*). While the case for linking *need* with *devez* is arguable, the link between *need* and *faut* is incorrect.



(19)



(20)

The relation-changing and head-switching cases illustrated by (6) and (7) are not handled correctly by the aligner. However, in both cases poor choice of lexical alignments (for *being* and *reste* respectively) ruled out the possibly of correct higher-level alignments. Whether improved lexical choice will lead to the identification of the appropriate alignments in these cases remains to be seen.

6 Conclusions

We observe that while the algorithm performs well at the phrase level, performance on lexical-level alignments is relatively poor when we compare the aligner output to the manual alignments. This can be seen both in terms of precision and recall, where scores for phrase-level alignments are much higher than those for lexical ones, and through the manual evaluation where complex translation phenomena are identified correctly at a high level but then negated by inaccurate alignments at lexical level.

The lexical accuracy scores illustrate clearly that there is an imbalance between precision and recall: recall is consistently higher than precision across all variants of the alignment algorithm. The reason for this is based in the word-alignments used to seed our tree-alignment algorithm. We have adopted the widely used alignment tool GIZA++ (Och and Ney, 2003) (and, more recently, Moses (Koehn et al., 2007) which is based directly on GIZA++) which prioritises broad coverage rather than high precision (Tiedemann, 2004) and is appropriate to string-based SMT (Koehn et al., 2003). However, the work presented here indicates that the preference in terms of expressing translational divergences through tree-alignment is for the opposite – high precision rather than broad coverage – and this mismatch appears to impact on the overall quality of the alignments. We suggest that this has implications not only for tree-alignment itself but also for the broader area of induction of syntax-aware models for SMT.

Despite these observations, training our DOT system on automatically-aligned data gives slightly better translation performance than training on the manually-aligned data. The issue of coverage is key here. Crucially, the only model used by the system is the synchronous tree-substitution grammar induced directly from the parallel treebank. As the manual alignments con-

tain fewer links than the automatic alignments, the induced grammar achieves correspondingly lower coverage and, consequently, performance suffers. We conclude that it is appropriate for tree-alignment to prioritise precision in order to capture translational divergences as accurately as possible, and that MT systems making use of these alignments should employ them in conjunction with broad-coverage models (such as word- and phrase-alignments) in order to preserve robustness.

7 Future Work

In order to improve the accuracy of our tree-alignment algorithm, we plan to investigate alternative word-alignment techniques (e.g. (Tiedemann, 2004; Liang et al., 2006; Ma et al., 2007)) in order to establish which one is most appropriate for our task.

With regard to the broader area of parallel treebank construction and the use of statistical parsers such as those of Charniak (2000) and Bikel (2002), we would like to examine the impact of imperfect parse quality on the capture of translational divergences. We plan to extend our aligner so that it works with n-best parse forests on the source and/or target sides, thereby giving the aligner some (limited) influence over the configuration of the aligned parse trees.

Finally, we plan to investigate how best to incorporate the translation information encoded in parallel treebanks into existing data-driven MT systems, both indirectly in terms of complementary phrase/chunk extraction methods and directly in terms of inducing syntactic models of translation.

Acknowledgements

This work was generously supported by Science Foundation Ireland Grant No. 05/RF/CMS064 and the Irish Centre for High-End Computing.[10] We thank Khalil Sima'an, Declan Groves and the anonymous reviewers for their insightful comments.

References

Satanjeev Banerjee and Alon Lavie. 2005. METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments. In *Proceedings of Workshop on Intrinsic and Extrinsic Evaluation Measures for MT and/or Summarization at the 43th Annual Meeting of the Association of Computational Linguistics (ACL-05)*, pages 65–72, Ann Arbor, MI.

Daniel M. Bikel. 2002. Design of a Multi-lingual, Parallel-processing Statistical Parsing Engine. In *Proceedings of the 2nd International Conference on Human Language Technology Research*, pages 24–27, San Francisco, CA.

Eugene Charniak. 2000. A Maximum-Entropy-Inspired Parser. In *Proceedings of the 1st Conference on North American Chapter of the Association for Computational Linguistics*, pages 132–139, Seattle, Washington.

David Chiang. 2005. A Hierarchical Phrase-Based Model for Statistical Machine Translation. In *Proceedings of the 43nd Annual Meeting of the Association for Computational Linguistics (ACL'05)*, pages 263–270, Ann Arbor, MI.

Ilyas Cicekli and H. Altay Güvenir. 2003. Learning Translation Templates from Bilingual Translation Examples. In Michael Carl and Andy Way, editors, *Recent Advances in Example-Based Machine Translation*, pages 255–286. Kluwer Academic Publishers, Dordrecht, The Netherlands.

George Doddington. 2002. Automatic Evaluation of Machine Translation Quality Using N-gram Co-Occurrence Statistics. In *Human Language Technology: Notebook Proceedings*, pages 128–132, San Diego, CA.

Bonnie J. Dorr, Lisa Pearl, Rebecca Hwa, and Nizar Habash. 2002. DUSTer: A Method for Unravelling Cross-Language Divergences for Statistical Word-Level Alignment. In *Machine Translation: From Research to Real Users. Proceedings of the 5th Conference of the Association for Machine Translation in the Americas (AMTA 2002)*, pages 31–43, Tiburon, CA.

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

Anette Frank. 1999. LFG-based syntactic transfer from English to French with the Xerox Translation Environment. In *Proceedings of the ESSLLI'99 Summer School*, Utrecht, The Netherlands.

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 44th Annual Meeting of the Association for Computational Linguistics*, pages 961–968, Sydney, Australia.

[10] http://www.ichec.ie/

Mary Hearne and Andy Way. 2006. Disambiguation Strategies for Data-Oriented Translation. In *Proceedings of the 11th Conference of the European Association for Machine Translation (EAMT-06)*, pages 59–68, Oslo, Norway.

W. John Hutchins and Harold Somers. 1992. *An Introduction to Machine Translation*. Academic Press, London.

Philipp Koehn, Franz Josef Och, and Daniel Marcu. 2003. Statistical Phrase-Based Translation. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology (NAACL '03)*, pages 48–54, Edmonton, Canada.

Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondrej Bojar, Alexandra Constantin, and Evan Herbst. 2007. Moses: Open source toolkit for statistical machine translation. In *Annual Meeting of the Association for Computational Linguistics (ACL), demonstration session*, pages 177–180, Prague, Czech Republic.

Percy Liang, Ben Taskar, and Dan Klein. 2006. Alignment by Agreement. In *Proceedings of the 2006 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology (NAACL '06)*, pages 104–111, New York City, NY.

Jeremy Lindop and Jun-ichi Tsujii. 1992. Complex Transfer in MT: A Survey of Examples. Technical Report 91-5, Centre for Computational Linguistics, UMIST, Manchester.

Yanjun Ma, Nicolas Stroppa, and Andy Way. 2007. Bootstrapping Word-Alignment via Word Packing. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (ACL'07)*, pages 304–311, Prague, Czech Republic.

Daniel Marcu, Wei Wang, Abdessamad Echihabi, and Kevin Knight. 2006. SPMT: Statistical Machine Translation with Syntactified Target Language Phraases. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing (EMNLP-06)*, pages 44–52, Sydney, Australia.

I. Dan Melamed. 1998. Annotation Style Guide for the Blinker Project. Technical Report 98-06, IRCS, University of Pennsylvania, Philadelphia, PA.

I. Dan Melamed. 2004. Statistical Machine Translation by Parsing. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL'04)*, pages 653–660, Barcelona, Spain.

Franz Josef Och and Hermann Ney. 2003. A Systematic Comparison of Various Statistical Alignment Models. *Computational Linguistics*, 29(1):19–51.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. BLEU: a Method for Automatic Evaluation of Machine Translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL-02)*, pages 311–318, Philadelphia, PA.

Yvonne Samuelsson and Martin Volk. 2006. Phrase Alignment in Parallel Treebanks. In *Proceedings of the 7th Conference of the 5th Workshop on Treebanks and Linguistic Theories (TLT 2006)*, Prague, Czech Republic.

Jörg Tiedemann. 2004. Word to Word Alignment Strategies. In *Proceedings of the 20th International Conference on Computational Linguistics (COLING '04)*, pages 212–218, Geneva, Switzerland.

John Tinsley, Ventsislav Zhechev, Mary Hearne, and Andy Way. 2007. Robust Language Pair-Independent Sub-Tree Alignment. In *MT Summit XI*, Copenhagen, Denmark.

Arturo Trujillo. 1999. *Translation Engines*. Springer, London.

Martin Volk and Yvonne Samuelsson. 2004. Bootstrapping Parallel Treebanks. In *Proceedings of the 7th Conference of the Workshop on Linguistically Interpreted Corpora (LINC)* , pages 71–77, Geneva, Switzerland.

Benjamin Wellington, Sonjia Waxmonsky, and I. Dan Melamed. 2006. Empirical Lower Bounds on the Complexity of Translational Equivalence. In *Proceedings of the 44th annual conference of the Association for Computational Linguistics (ACL-06)*, pages 977–984, Sydney, Australia.

Dekai Wu. 2000. Bracketing and aligning words and constituents in parallel text using Stochastic Inversion Transduction Grammars. In Jean Veronis, editor, *Parallel Text Processing: Alignment and Use of Translation Corpora*, pages 139–167. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Andreas Zollmann and Ashish Venugopal. 2006. Syntax Augmented Machine Translation via Chart Parsing. In *Proceedings of the HLT-NAACL 2006 Workshop on Statistical Machine Translation*, pages 138–141, New York City, NY.