On Complex Word Alignment Configurations

Miriam Kaeshammer and Anika Westburg

University of Düsseldorf
Universitätsstraße 1, 40225 Düsseldorf, Germany
kaeshammer@phil.hhu.de, Anika.Westburg@hhu.de

Abstract

Resources of manual word alignments contain configurations that are beyond the alignment capacity of current translation models, hence the term *complex alignment configuration*. They have been the matter of some debate in the machine translation community, as they call for more powerful translation models that come with further complications. In this work we investigate instances of complex alignment configurations in data sets of four different language pairs to shed more light on the nature and cause of those configurations. For the English-German alignments from Padó and Lapata (2006), for instance, we find that only a small fraction of the complex configurations are due to real annotation errors. While a third of the complex configurations in this data set could be simplified when annotating according to a different style guide, the remaining ones are phenomena that one would like to be able to generate during translation. Those instances are mainly caused by the different word order of English and German. Our findings thus motivate further research in the area of translation beyond phrase-based and context-free translation modeling.

Keywords: word alignment, machine translation, complexity

1. Introduction and Motivation

A *word alignment* represents translational equivalence of words or phrases in a sentence pair (or paragraph pair) by a set of links between source and target words. It is a central concept of statistical machine translation (SMT), where translation models are commonly learned on the basis of word alignments (e.g. Koehn et al. (2003)), or they are induced jointly with the alignments (e.g. Marcu and Wong (2002)). The most prominent paradigms of translation models are phrase-based approaches (Koehn et al., 2003) and tree-based approaches that employ some form of a synchronous context-free grammar (SCFG) (Chiang, 2007; Zollmann and Venugopal, 2006; Hoang and Koehn, 2010), in particular inversion transduction grammar (ITG) (Wu, 1997). Such synchronous translation grammars also induce alignments between the words of a parallel text when applying a synchronous rule during parsing (Wu, 1997).

The space of alignments that can be generated with the mentioned translation grammars is limited. Inside-out alignments (IO), discussed in Wu (1997), are beyond the alignment capacity of SCFG of rank 2 (henceforth 2-SCFG) and ITG of any rank, but they can be generated by a standard phrase-based decoder due to its reordering component. Cross-serial discontinuous translation units (CDTU) (Søgaard and Kuhn, 2009) and bonbon configurations (Simard et al., 2005) can be induced by neither phrase-based nor SCFG-based translation systems. The underlying assumption thereby is that a *translation unit* (TU), containing the transitive closure of some set of nodes of the bipartite alignment graph, represents minimal translational correspondence. An adequate translation grammar should thus be able to generate each translation unit separately.

Figure 1 schematically depicts those three well-known complex alignment configurations. The inside-out alignment (i) consists of four translation units (a, b, c and d), while (ii) and (iii) are formed by two discontinuous translation units each that are intertwined.

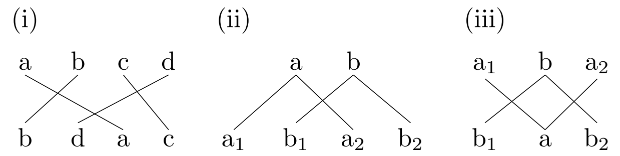

Figure 1: Complex alignment configurations: (i) IO; (ii) CDTU; (iii) bonbon. Source words are shown on the top, target words at the bottom. The configurations can also occur upside-down.

Knowing of the limitations of the alignment space of current translation models, their empirical adequacy has been put into question. In several studies, the empirical alignment capacity of various formalisms with respect to manually aligned data has been investigated in different setups (Wellington et al., 2006; Søgaard and Kuhn, 2009; Søgaard and Wu, 2009; Søgaard, 2010; Kaeshammer, 2013). They show that the complex alignment configurations occur relatively frequently. For instance, Wellington et al. (2006) found that 5% of their English-Chinese sentence pairs contain inside-out alignments, and Kaeshammer (2013) reports that 5.5% of the sentence pairs in an English-German data set and 9% in an Spanish-French data set cannot be induced with an SCFG of rank 2, which is the formalism behind the prevalent hierarchical phrase-based translation grammars (Chiang, 2007).

To be able to induce the complex alignment configurations, more expressive translation models have been proposed (Søgaard, 2008; Galley and Manning, 2010; Kaeshammer, 2013). They however come at the cost of higher decoding complexity, loss of translation speed or other complications, such as no tight probability estimators. A verification of the substantiality of the complex alignment configurations would serve as a justification for the use of more powerful translation models and further research in that area. To this end, it is necessary to examine instances of complex alignment configurations instead of merely relying on their frequency counts.

In this paper, we thus investigate the nature of the complex alignment configurations that occur in hand-aligned data. We approach this task by manual categorization of the complex alignment configurations. Our categories address the issue of whether the involved alignments adhere to the corresponding alignment guidelines or whether they are annotation errors. We furthermore identify alignments that are correct with respect to the guidelines of the data set, but which are disputable since other guidelines would align them in a different way, leading to fewer complex alignment configurations. We furthermore point out which linguistic phenomena cause the complex configurations in our data, and we address the question of how necessary they are for translation. To the best of our knowledge the available word alignment resources have not been studied with respect to these aspects so far.

2. Alignment Data Sets and Guidelines

While large-scale word alignments are created automatically, mostly in an unsupervised fashion (e.g. Och and Ney (2000)), a number of gold alignment data sets for several language pairs exist. They are manually created, high quality reference alignments that have emerged from various projects and shared tasks on word alignment. They vary in size, annotation methodology, alignment guidelines and original purpose.

In this investigation, we concentrate on data sets that have also been explored in previous studies on empirical alignment capacity: the manual alignments of 987 English-German (*en-de*) sentence pairs from Europarl, whose original purpose was the projection of semantic roles (Padó and Lapata, 2006), and the Europarl data sets described in Graça et al. (2008a) for the combinations of English (*en*), French (*fr*) and Spanish (*es*), each containing 100 sentence pairs.

Since annotating word alignments is a non-trivial task, different sets of alignment guidelines have been developed. The *en-de* data is aligned following the style guide of the Blinker Project, specified in Melamed (1998), which was concerned with English-French alignment. The *en-fr-es* data is aligned according to guidelines provided in Graça et al. (2008b). They are a refined version of the style guide for English-Spanish alignment by Lambert et al. (2005).

The basic assumptions in the guidelines are similar. Their goal are full-text alignments (as opposed to sample word alignments), and they strive to align units of the same meaning on both sides that are as small as possible, but that include as many words as necessary. While in the Blinker style guide only one type of alignment link is used, the *en-fr-es* data is aligned with S(ure)-links resp. P(ossible)-links if the correspondence is valid in every resp. some context. However, since in the previous experiments no distinction is made between the types of links, we will not differentiate between them either.

The style guides differ, of course, in many, sometimes language-specific, details. We only review those which we came across when studying the complex alignment configurations in the data.

Anaphora If a pronoun occurs in one sentence with its antecedent and does not have a translation in the other language, both the antecedent and the pronoun are aligned to the translation of the antecedent as specified in the Blinker guidelines. According to the Lambert guidelines however, such anaphoric links between a pronoun in one language and a co-referent noun or proper noun in the other language are not licensed because they cannot be considered translations of each other.

Repetitions According to the Blinker guidelines, for repetitions that occur only in one language, but not in the other, all instances of the repetition are linked to the one translation. This stands in contrast to the Lambert guidelines which state that only the first instance of the repetition is aligned while the subsequent ones are without correspondence.

Punctuation The style guides generally agree on how to align punctuation marks. However, the Blinker guidelines explicitly advise to align similar punctuation symbols which occur in different quantities on the two sides such that as few crossing links as possible arise. This can also mean that punctuation symbols remain unaligned. Even though the Lambert and Graça style guides do not contradict this guideline, the Graça data contains alignments that do not adhere to it.

3. Investigation

This section presents the details of our investigation of complex alignment configurations.

3.1. Complex Alignment Configurations

The alignment configurations of interest are those which cannot be induced with a 2-SCFG (or equivalently a 2-ITG without normal form constraint). Those are IO alignments, CDTUs, bonbon alignments and certain multigap DTUs.[1]

Interestingly, we found that there is another class of configurations that is beyond the alignment capacity of phrase-based and 2-SCFG-based translation models, which has not been reported before. They consist of a discontinuous TU with one gap (x in the following) and three other translation units (a, b and c). The four TUs are configured in a similar way as the inside-out alignment, i.e. no three of the four TUs form a continuous sequence in the source and target strings. More specifically, the configurations can be described by the following patterns: (i) one of the set $\{\mathrm{xabc}, \mathrm{abxc}\}$ on one side, and one of the set $\{\mathrm{bx_1acx_2}, \mathrm{x_1cax_2b}\}$ on the other side, or (ii) one of the set $\{\mathrm{axbc}, \mathrm{abcx}\}$ on one side and one of the set $\{\mathrm{bx_1cax_2}, \mathrm{x_1acx_2b}\}$ on the other side, assuming that same letters are aligned and therefore form a TU. We name the configurations in this class IO-DTUs. Figure 2 shows two examples. Some other combinations of the TUs a, b, c and x are also beyond the alignment capacity of 2-SCFG, but they coincide with the already known IO alignments.

[1]Namely those multigap TUs where the words in the gaps form more than two continuous sequences of aligned source and target words, e.g. DTUs with three or more gaps on one side.

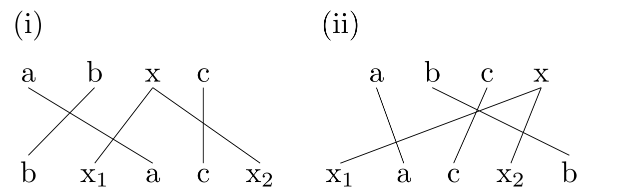

Figure 2: IO-DTU: new complex alignment configurations

Note that SCFGs of higher rank than 2 could induce IO alignments, IO-DTUs and multigap DTUs, but not CDTUs and bonbon alignments. However, higher ranks than 2 are disadvantageous for parsing/translation complexity. In contrast, all of the above mentioned complex alignment configurations, except a subset of the multigap DTUs, are beyond the alignment capacity of ITG in general (Søgaard and Wu, 2009). This is because ITG requires constituents on the target side to be in the same order as on the source side or exactly in reverse order. The finding of a new class of configurations beyond ITG, the IO-DTUs, means that lower bounds on translation unit error rate are higher than reported in Søgaard and Wu (2009).

We automatically identified the complex alignment configurations in our data, leaving aside unaligned words. To be considered as distinct, configurations of the same structure have to differ on all translation units, or, if they overlap, the overlap has to occupy a different part of the configuration.[2] See Table 1 for the statistics of the data. For visualization of the word alignments, we used the Tree-Alignment Visualizer (Maillette de Buy Wenniger et al., 2010).

3.2. Categories and Classification

We inspected each complex alignment configuration and classified it into one of the following categories:[3]

Annotation error A link is an obvious annotation error if we can neither find a guideline that justifies the link nor think of any context in which the indicated translational equivalence would hold. By removing it, the alignment configuration is not complex anymore. See Figure 3 for an example.[4]

Artifact of style guide A link is correct with respect to the style guide of the data (i.e. the Blinker guidelines for the *en-de* alignments). However, its existence is arguable since a different established style guide (i.e. the Lambert guidelines) would have aligned the phenomenon differently, thereby simplifying the configuration. The details about such differences in the style guides have been worked out in Section 2. Figure 4 shows an example of anaphor alignment.

Correct alignment All translation units that are part of the complex configuration are correctly aligned. Figure 5 shows an example.

Table 2 shows the classification results. If several overlapping configurations are caused by the same dubious link, we only count them once.

For all of the investigated data sets, no or very few complex alignment configurations are due to real annotation errors. In the *en-de* data, quite a large portion of complex configurations (32.8%) are caused by artifacts of the Blinker guidelines. Thereof 58.1% are due to anaphora, and 41.9% are due to repetitions. In the *en-fr-es* data, all complex configurations in the artifact category are due to questionable punctuation alignment.

The remaining complex alignment configurations, 55.7% in the *en-de* data and $>$ 83% in each of the *en-fr-es* data sets (see Table 2), are those which are correctly aligned and therefore interesting for translation modeling. We will therefore further examine them in the following section.

3.3. Phenomena and Translation

First, we shed more light on which linguistic phenomena elicit the complex alignment configurations. This is of course dependent on the language pair. The issue of how important those configurations are for translation will also be considered.

3.3.1. English-German

In the *en-de* data, most complex alignment configurations are caused by the different word orders of the English and German sentences. While English sentences follow SVO order, the order of German constituents in a sentence is less rigid, traditionally described within the topological field model (Höhle, 1983).[5] In a nutshell, the position of the verbs in a German sentence is fixed: the finite verb occupies the second position of the sentence, the *left bracket* (LB), in main clauses, or the final position, the *right bracket* (RB), in subordinate clauses. Non-finite verbs are also located in the right bracket, but left of the finite verb if there is one. Argument and modifier constituents of the sentence are located between this verbal frame, in the *middle field*; however, in verb-second clauses, the initial position (the *initial field*) is filled by one constituent. The preference for a specific ordering of the arguments and modifiers is influenced by many syntactic and non-syntactic factors, e.g. pronominalization and pragmatic constraints. Certain word orders lead to IO alignments, as will be exemplified in the following.

Figure 5 shows an example of a German main clause, in which both sentence brackets are filled, showing how the auxiliary-participle combination which is adjacent in English is placed in very distant parts of the German sentence.

[2]The reason is that the overlapping TUs might not be the discussable ones. This is slightly different than the methodology of Søgaard and Wu (2009) where configurations have to differ on all TUs to be counted as distinct configurations when they determine lower bounds on translation unit error rate.

[3]Note that none of the authors was part of the original annotation body of any of the data sets. We have drawn our knowledge from the published annotation style guides and patterns observed in the annotated data. Our decisions have been made to the best of our knowledge on the basis of this information.

[4]For the sake of clarity, all figures show only the links of the alignment which make up the complex alignment configuration. Assume the other words are aligned appropriately.

[5]Refer to Telljohann et al. (2012, Section 3.1) for a short introduction to the topological field model.

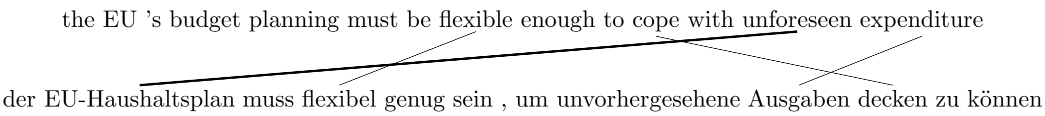

Figure 3: IO alignment due to an annotation error: *unforeseen - EU-Haushaltsplan* (*en-de*, sent. 183)

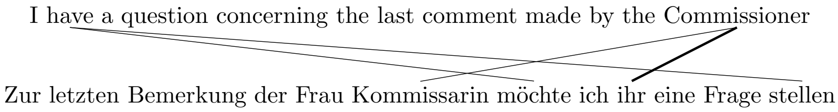

Figure 4: CDTU; *Commissioner - ihr* is correctly aligned according to the Blinker guidelines (*en-de*, sent. 887)

The subject *1813 Menschen* and the prepositional phrase *in 31 Ländern* are located in the middle field of the German sentence. Together with the verbs, they are ordered in such a way that an IO alignment is created. Figure 6 shows a similar example, but where the German sentence is verb-final since it is a subordinate clause. The complementizer is usually analysed as filling the left bracket. Here, the subject (*die israelischen Streitkräfte*), a prepositional phrase (*aus einer Stadt*) and a reflexive pronoun (*sich*), elicited by the fact that the English verb *withdraw* translates into the German reflexive verb *sich zurückziehen*, are located in the German middle field. Together with the finite verb, the IO alignment is created. Even though not shown in the examples, the German initial field can also be involved in creating IO alignments.

A related phenomenon is the translation of an English verb into a German separable particle verb. While the core of the German finite verb remains in the left sentence bracket in main clauses, the particle is found at the end of the clause in the right sentence bracket. Together with the English verb, this configurations forms a DTU which, in combination with the other arguments and modifiers of the clause, can lead to an IO alignment or IO-DTU.

While 56.5% of the aforementioned IO alignments and IO-DTUs occur with fairly literally translated TUs that differ in their ordering (as in the sentence pairs in Figure 5, 6 and 7), the rest is part of rather free translations or freely translated TUs, sometimes crossing clause boundaries. Figure 8 shows an example. Free translations also give rise to a few CDTUs and bonbon alignments.

A question that often arises when considering the complex alignment configurations is whether they are essential for translation. For the examples of IO/IO-DTU alignments due to different word orders of literally translated TUs on the clause level, we therefore additionally investigated the sentence pairs according to the following criterion: Given one sentence of the sentence pair (e.g. English) as input to a translation system, and given that the translation model yields the provided target (e.g. German) translations of each TU, is the provided target (e.g. German) sentence the only valid translation?

Consider Figure 5 as an example. When translating from English to German, a German sentence with the provided TUs but with a simpler alignment could be produced, e.g. *1999 wurden 1813 Menschen in 31 Ländern hingerichtet*. The same holds for the other direction, e.g. *in 31 countries, 1813 people were executed in 1999*. In contrast to that, given the German sentence in Figure 6 as input, and given that the model provides the shown English TUs, the English sentence in Figure 6 is the only valid translation, due to the strict English word order. Only the prepositional phrase *from a city* could potentially be placed at the beginning of the clause. However, it is very marked in this position and it therefore obtains a strong focus which it does definitely not have in the original sentence.

Figure 7 shows an IO alignment which involves a focus adverb (*also*/*auch*). The complicacy of this is that it is often impossible to unambiguously determine the focus denoted by the adverb in a single sentence (Sudhoff, 2010). Since our data sets neither include context beyond the aligned sentence pair nor intonation information, we are not able to judge whether a reordering of the involved constituents would lead to an equally good translation (without complex alignment). In the given sentence pair, for the direction *en*$\rightarrow$*de*, it is for instance grammatical to place *auch* at the very beginning or end of the middle field, which would simplify the configuration. However, this probably changes the focus, so we decide against making a statement about alternative translations via a different word ordering. A further complication is that in some sentence pairs, the focus denoted by a focus adverb in the English sentence is obviously different than in the German sentence. Negation particles and their scope are similar to the focus adverbs.

Under the given criterion, for 48.6% of the IO/IO-DTUs caused by word order, an equally good translation without complex configuration can be found when translating from English to German. For the remaining ones, a different translation would involve possible scope/focus changes. In the other direction (*de*$\rightarrow$*en*), 31.4% of the cases can be simplified, and 28.6% involve a possible change in scope/focus. Remarkably, in 40% of the cases, the word order that involves the IO/IO-DTU is the only possible one, given the criterion defined above.

Only 11 of the complex alignment configurations in category (III) are not due to the above explained word order phenomena. They arise from a variety of phenomena, including local nominal discontinuities on both sides which lead to a bonbon alignment, infinitive clauses where *um ... zu* in German and *to* in English form a DTU in combination with a verbal DTU such that *zu* is in its gap creating a CDTU, and coordinations where the order within the conjuncts is different in English and German creating an IO or

	en-de (30)	*en-fr*	*en-es*	*es-fr*
Sentence pairs	694	100	100	100
Sentence pairs with at least one compl. config.	116	5	4	9
IO	92	2	4	1
IO-DTU	38	1	0	2
CDTU	34	1	0	6
Bonbon	5	2	0	0
Multigap	8	0	0	0

Table 1: Data characteristics and frequency of the complex alignment configurations (number in parentheses: sentence length cut-off)

			en-de (30)	*en-fr*	*en-es*	*es-fr*
(I)	Error		0.115	0.000	0.000	0.000
(II)	Artifact		0.328	0.167	0.000	0.111
(III)	Correctly aligned		0.557	0.833	1.000	0.889
	from (III)	IO	0.822	0.200	1.000	0.125
		IO-DTU	0.096	0.200	0.000	0.125
		CDTU	0.068	0.200	0.000	0.750
		Bonbon	0.014	0.400	0.000	0.000

Table 2: Ratio of the classes of complex alignment configurations

IO-DTU together with the verb.

3.3.2. English-French

The *en-fr* data of course includes the often cited bonbon alignment due to the French two-part negation that frames the finite verb, and a verbal DTU with the gap on the English side that is usually elicited by the use of an auxiliary because of the negation. Figure 9 (top half) shows an example from the data set.

If the French negation occurs in combination with a complex French verbal unit, e.g. in a compound tense, a CDTU is formed, since *ne* precedes the verbal TU and *pas* interrupts it.

CDTUs and bonbon alignments caused by the French negation can generally not be resolved by reordering in analogy to the IO/IO-DTU alignments, because of the strict rules of placement of negation in relation to verbs in the involved languages. For generating these configurations, translations models beyond phrase-based and SCFG-based models are thus necessary. Further considerations about the translation of these structures are presented in Section 4.

The IO/IO-DTU configurations in the *en-fr* data are again caused by different word orders. They involve an adverb, which is placed at the beginning of the sentence in English, but between the finite and the non-finite verb in French compound tenses. They are either free translations, or, if not, a different word order without the complex configuration is possible.

3.3.3. English-Spanish

All complex alignment configurations in the *en-es* data are IO alignments. Most of them reside in rather free translations. The one that is created by a relatively literal translation is shown in Figure 10. It is caused by the different adjectival placement in English and Spanish (pre-nominal vs. post-nominal) and the fact that the adverbial phrase *to a certain extent* is located on the clause level in English, but within the noun phrase in Spanish. In this case, a more direct translation that does not require an IO alignment is possible in both translation directions.

3.3.4. Spanish-French

In the *es-fr* data, all CDTUs involve a French two-part negation. We observe two different phenomena. First, if the French verb is complex, e.g. in a compound tense, the second part of the negation (*pas*) is placed after the finite verb, thus creating a cross-serial DTU. The same phenomenon is also found in the *en-fr* data, see Section 3.3.2. Second, Spanish usually does not realize pronouns in subject position. The French pronoun is then aligned to the verbal TU. In combination with a negation, the first part of the negation (*ne*) interrupts this verbal TU, creating a CDTU. An example is shown in Figure 9 in the lower half.

Just as in the *en-fr* data, the IO/IO-DTU configurations are also due to different word orders involving adverbial phrases. In the French sentences, the adverbials are placed between the finite and the infinite verb, while in Spanish they are found at the beginning of the sentence or on the right of the verb. One of the configurations occurs in a rather free translation, the other in sentences that could be reordered for a simpler alignment, but possibly involving a change of focus.

4. Discussion

The results of the investigation show that, while only very few of the complex alignment configurations are true annotation errors, many can be argued away with reference to other annotation guidelines. The remaining ones are those of interest for translation modeling.

Within the limited context of our analysis, we can certainly not generally answer how crucial it is for an SMT system to be expressive enough to induce complex alignment con-

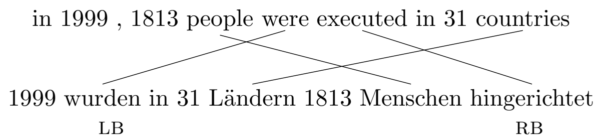

Figure 5: IO alignment due to different word orders; verb-second clause in German (*en-de*, sent. 119)

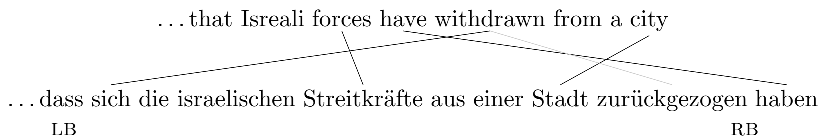

Figure 6: IO alignment due to different word orders; verb-final clause in German (*en-de*, sent. 306)

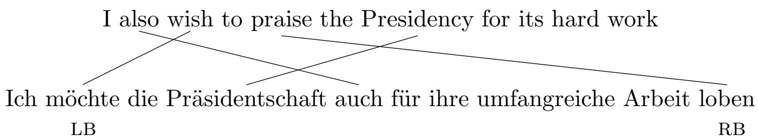

Figure 7: IO alignment due to different word orders, includes a focus adverb (*en-de*, sent. 389)

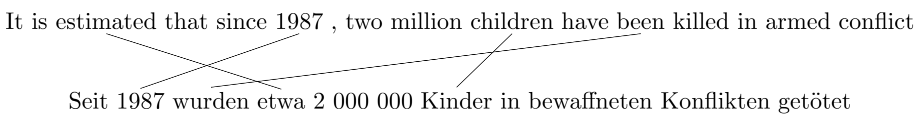

Figure 8: IO alignment due to different word orders in a rather free translation (*en-de*, sent. 876)

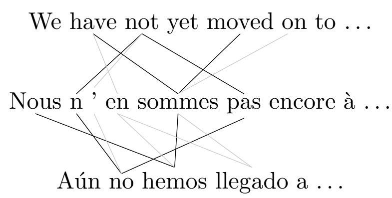

Figure 9: Bonbon alignment and CDTU due to French two-part negation (*en-fr-es*, sent. 13)

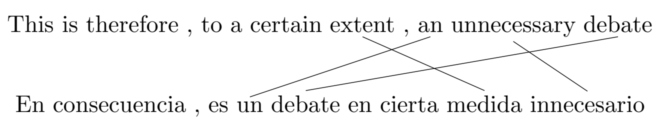

Figure 10: IO alignment due to modifier placement (*en-es*, sent. 59)

figurations. When thinking about this question, it is important to keep the following points in mind: Translation is concerned with producing a good/correct output string, not with producing a correct alignment. This means that generating a specific alignment is not necessarily important for generating a specific translation. It might even happen that a system produces a translation that corresponds to a complex alignment configuration on the surface, without actually having generated this complex configuration.

Furthermore, it is absolutely clear that, by paraphrasing, a good translation that is different from the one with the complex configuration present in our data can always be found. It is not possible to answer the question of whether the translation in our data is the best from all possible translation options. Deciding whether one translation is better than another one is highly subjective and depends on many factors outside the context of one sentence.

What we argue for is that there exist good translation options that involve complex alignment configurations (category III). As we have found, many of them are fluent and relatively literal translations. A machine translation system should thus not exclude them a priori from the translation search space. This entails the usage of more expressive translation models than those based on 2-SCFG. In particular, when translating from a language with a rather free word order to a language with a rather rigid word order, it seems important to be able to induce IO and IO-DTU alignments: for 40% of the *en-de* complex configurations caused by different word orders of literal TUs, the English word order present in the data is the only possible one when translating the German sentence. If focus and scope considerations are also taken into account, the number will certainly be higher.

For the *en→de* direction, we always found a reordered translation alternative without complex configuration (or a focus/scope ambiguity), which is not surprising due to the flexible word order of German. However, as already pointed out before, in a given context, pragmatic constraints can lead to a strong preference of one of the word orders. This can certainly be the one induced by the complex alignment configuration. The same holds for the instances of word order phenomena in the *en-fr-es* data sets: even though a translation with reordered TUs is possible, the other one might be the more canonical one in a certain situation. Thus, even for those language pairs and translation directions, there is reason to investigate more powerful translation models.

It should furthermore be noted that being able to induce the complex configurations is especially of importance if the involved construction is productive. This is certainly the case for the word order phenomena. If a construction is not productive, the translation model can just memorize the non-productive parts as a whole, without being able to induce each TU individually. As an example, consider the English-French bonbon alignments, caused by the French two-part negation and the requirement of an auxiliary verb by the English negation (Figure 9). While the two DTUs are certainly perfectly aligned in terms of the style guides and in terms of lexical translational correspondence, one could argue that an SMT model could memorize all non- or less productive parts, i.e., the negation TU together with the English auxiliary. Combining this with the productive main verbal TU does not involve a bonbon anymore. Such considerations, however, come at the price of a less modular translation model.

For correctly aligned, but relatively free translations, we did not make statements about how essential generating the complex configuration is. The reason is twofold: First, a more literal translation could be produced. Second, in a machine translation system, ideally, free translations are not generated by composing individual translation units, but as larger structures/blocks, since it is only together that they make sense. This means that, in those cases, being able to generate the complex alignment configuration is less important. However, especially if the free translation option involves productive parts, a perfect translation model should of course also be able to generate them.

In the end, experimental evidence will have to show how alignment capacity relates to machine translation quality. The results in Galley and Manning (2010) for Chinese-English translation indicate that more powerful translation models also lead to better translations.

5. Conclusion

The work at hand presents an investigation of the complex alignment configurations in existing manually aligned data sets. While some of those configurations are caused by annotation errors or artifacts of a specific annotation style guide, we found that more than half of the complex configurations in the *en-de* data and between 83% and 100% in the *en-fr-es* data sets are correctly aligned. Mostly, especially in the *en-de* data, it is the word order on the clausal level which leads to IO alignments and IO-DTUs. In the *en-fr* and *es-fr* data, the French two-part negation is often involved in CDTUs and bonbon alignments.

Even though the translations generated by the complex alignment configurations certainly do not represent the only translation options, they are correct, often fairly literal translations. Especially if they involve productive constructions, one should not exclude them a priori from the translation search space. This motivates further research in the area of translation modeling beyond phrase-based and context-free grammars.

Our investigation complements previous studies on the empirical alignment capacity of various formalisms. Since a growing number of manually aligned resources are available, e.g. English-Swedish (Holmqvist and Ahrenberg, 2011), English-Dutch (Macken, 2010) and Danish-German (Buch-Kromann et al., 2009) alignments, this study could be extended to other language pairs in the future.

6. Acknowledgements

This research was funded by the German Research Foundation as part of the project *Grammar Formalisms beyond Context-Free Grammars and their use for Machine Learning Tasks*.

7. References

Matthias Buch-Kromann, Iørn Korzen, and Henrik Høeg Müller. 2009. Uncovering the lost structure of translations with parallel treebanks. *Copenhagen Studies in Language*, 38:199–224.

David Chiang. 2007. Hierarchical phrase-based translation. *Computational Linguistics*, 33(2):201–228.

Michel Galley and Christopher D. Manning. 2010. Accurate non-hierarchical phrase-based translation. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 966–974.

João de Almeida Varelas Graça, Joana Paulo Pardal, Luísa Coheur, and Diamantino António Caseiro. 2008a. Building a golden collection of parallel multi-language word alignment. In *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC'08)*.

João de Almeida Varelas Graça, Joana Paulo Pardal, Luísa Coheur, and Diamantino António Caseiro. 2008b. Multi-language word alignments annotation guidelines. Technical report, Soken Language Systems Laboratory, Lisboa, Portugal.

Hieu Hoang and Philipp Koehn. 2010. Improved translation with source syntax labels. In *Proceedings of the Joint Fifth Workshop on Statistical Machine Translation and MetricsMATR*, pages 409–417. Association for Computational Linguistics.

Tilman Höhle. 1983. *Topologische Felder*. Ph.D. thesis, Universität zu Köln.

Maria Holmqvist and Lars Ahrenberg. 2011. A gold standard for English–Swedish word alignment. In *Proceedings of the 18th Nordic Conference of Computational Linguistics NODALIDA 2011*, pages 106–113.

Miriam Kaeshammer. 2013. 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. Association for Computational Linguistics.

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*, pages 48–54. Association for Computational Linguistics.

Patrik Lambert, Adrià Gispert, Rafael Banchs, and José B. Mariño. 2005. Guidelines for word alignment evaluation and manual alignment. *Language Resources and Evaluation*, 39:267–285.

Lieve Macken. 2010. An annotation scheme and gold standard for Dutch-English word alignment. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*.

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

Daniel Marcu and William Wong. 2002. A phrase-based, joint probability model for statistical machine translation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 133–139. Association for Computational Linguistics.

I. Dan Melamed. 1998. Annotation style guide for the Blinker Project. Technical Report IRCS-98-06, University of Pennsylvania.

Franz Josef Och and Hermann Ney. 2000. Improved statistical alignment models. In *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics*, pages 440–447. Association for Computational Linguistics.

Sebastian Padó and Mirella Lapata. 2006. Optimal constituent alignment with edge covers for semantic projection. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the ACL*, pages 1161–1168.

Michel Simard, Nicola Cancedda, Bruno Cavestro, Marc Dymetman, Eric Gaussier, Cyril Goutte, Kenji Yamada, Philippe Langlais, and Arne Mauser. 2005. Translating with non-contiguous phrases. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP)*, pages 755–762.

Anders Søgaard and Jonas Kuhn. 2009. Empirical lower bounds on alignment error rates in syntax-based machine translation. In *Proceedings of the Third Workshop on Syntax and Structure in Statistical Translation (SSST '09)*. Association for Computational Linguistics.

Anders Søgaard and Dekai Wu. 2009. 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.

Anders Søgaard. 2008. Range concatenation grammars for translation. In *Proceedings of Coling 2008: Companion volume: Posters*.

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

Stefan Sudhoff. 2010. *Focus Particles in German: Syntax, Prosody, and Information Structure*. Benjamins.

Heike Telljohann, Erhard W. Hinrichs, Sandra Kübler, Heike Zinsmeister, and Kathrin Beck. 2012. Stylebook for the Tübingen Treebank of Written German (TüBa-D/Z). Technical report, Seminar für Sprachwissenschaft, Universität Tübingen, January.

Benjamin Wellington, Sonjia Waxmonsky, and I. Dan Melamed. 2006. Empirical lower bounds on the complexity of translational equivalence. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, pages 977–984.

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

Andreas Zollmann and Ashish Venugopal. 2006. Syntax augmented machine translation via chart parsing. In *Proceedings of the Workshop on Statistical Machine Translation, HLT/NAACL*.