3

Aligning linguistically motivated phrases

Lieve Macken and Walter Daelemans
Ghent University College and University of Antwerp

Abstract

In this paper, we describe the architecture of a sub-sentential alignment system that links linguistically motivated phrases in parallel texts.

We conceive our sub-sentential aligner as a cascade model consisting of two phases. In the first phase, anchor chunks are linked on the basis of lexical correspondences and syntactic similarity. In the second phase, we will focus on the more complex translational correspondences based on observed translation shift patterns. The anchor chunks of the first phase will be used to limit the search space in the second phase.

We present the first results of our sub-sentential alignment system, which links linguistically motivated chunks. In our baseline system, the obtained recall scores range from 44% to 59% and precision scores from 90% to 98% depending on text type.

We experimented with two different types of bilingual dictionaries to generate the lexical correspondences: a handcrafted bilingual dictionary and probabilistic bilingual dictionaries. We demonstrate that although the handcrafted dictionary is twice the size of the probabilistic dictionary, the obtained recall scores are lower.

Proceedings of the 18th Meeting of Computational Linguistics in the Netherlands, pp. 37–52
Edited by: Suzan Verberne, Hans van Halteren, Peter-Arno Coppen.

3.1 Introduction

Sub-sentential alignments are used among other things to create phrase tables for statistical phrase-based machine translation (SMT) systems. In existing SMT systems, a phrase is not linguistically motivated. It can be any contiguous sequence of words. There is a strong intuition that the use of linguistically relevant phrases can improve the performance of phrase-based SMT systems. The experiments by Groves and Way (2006) for French-English confirm this assumption.

A stand-alone sub-sentential alignment module however, is also useful for human translators if incorporated in CAT-tools, e.g. sophisticated bilingual concordance systems, in sub-sentential translation memory systems (Gotti et al. 2005), or for bilingual terminology extraction (Macken et al. 2008).

Several researchers demonstrated that the addition of linguistic information can improve statistically-based word alignment systems. Tiedemann (2003) for example combines association measures with additional linguistic heuristics based on part-of-speech, phrase type, and string similarity measures. While Tiedemann makes use of chunk information, the alignment process remains word-based. In our approach, the whole alignment process is primarily chunk-driven.

We conceive our sub-sentential aligner as a cascade model consisting of two phases. In the first phase *anchor chunks*, i.e. chunks that can be linked with a very high precision based on lexical correspondences and syntactic similarity are retrieved. In the second phase, we will focus on the more complex translational correspondences based on observed translation shift patterns. The anchor chunks of the first phase will be used to limit the search space in this second phase. This paper describes the first phase, namely the alignment of anchor chunks.

3.2 Architecture

The global architecture of our system is visualized in figure 3.1. The sub-sentential alignment system takes as input sentence-aligned texts, together with additional linguistic annotations (part-of-speech codes and lemmas) for the source and the target text.

Although the global architecture of our sub-sentential alignment system is language-independent, some language-specific resources are used. In the first phase, two *external* language-specific linguistic resources are needed: first, a bilingual lexicon to generate the lexical correspondences; second, tools to generate additional linguistic information (PoS tagger, lemmatizer and chunker).

We will simulate the different steps of the sub-sentential alignment process for the following sentence pair:

En: Madam President, last week's attacks on innocent civilians in New York and Washington shocked and outraged the civilised people across the world.

Nl: Mevrouw de Voorzitter, de aanvallen op onschuldige burgers die vorige week hebben plaatsgevonden in New York en Washington hebben de beschaafde volkeren in de gehele wereld geschokt en verontwaardigd.

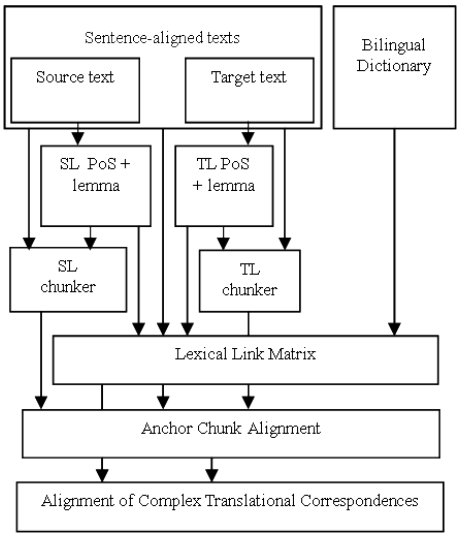

Figure 3.1: Outline of the full sub-sentential alignment system

In the first step of the process, the source and target sentences are divided into chunks based on PoS information, and lexical correspondences are retrieved from a bilingual dictionary. The chunk boundaries are visualized in figure 3.2 by means of horizontal and vertical lines. The lexical correspondences are marked by *x's*.

During anchor chunk alignment, the sub-sentential aligner links chunks based on lexical correspondences and chunk similarity. In figure 3.2, the anchor chunks are marked in light grey. In the example sentence, corresponding noun phrases, corresponding prepositional phrases and corresponding verb phrases are indicated.

In the second phase of the process, the sub-sentential aligner uses the anchor chunks of the first phase to retrieve the more complex translational correspondences. In figure 3.2, such a more complex translational correspondence is marked in dark grey. In the example sentence, the following translation shift can be observed: the English premodifier (*last week's*) is translated by a relative clause in Dutch (*die vorige week hebben plaatsgevonden*).

3.3 Bilingual lexicon

As explained in section 3.2, a bilingual dictionary is used in the first phase of the sub-sentential alignment process to indicate lexical correspondences in source and target sentences.

We experimented with two different types of bilingual dictionaries: a hand-crafted bilingual dictionary and probabilistic bilingual dictionaries, automatically extracted from bilingual corpora.

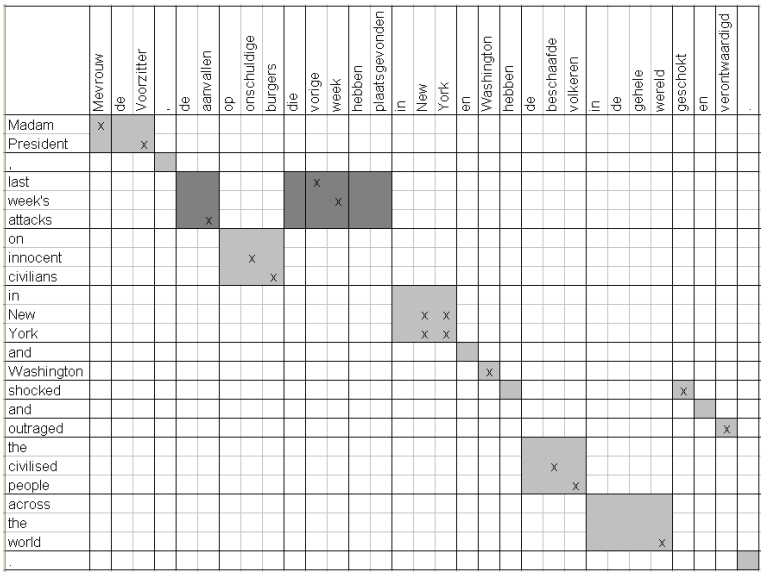

Figure 3.2: Simulation of the different alignment steps: chunk boundaries (horizontal and vertical lines), lexical correspondences (x's), anchor chunks (light grey) and complex translational correspondence (dark grey)

3.3.1 Handcrafted bilingual lexicon

The handcrafted bilingual lexicon was derived from the English-Dutch and Dutch-English Nl-Translex lexica[1](Goetschalckx et al. 2001). The Nl-Translex lexica contain apart from one-to-one correspondences (e.g. *force - kracht*) and compounds (e.g. *market sector - marktsegment*), also a large number of phrasal correspondences (e.g. *come into force - van kracht worden*, *agreement on government procurement - overeenkomst inzake overheidsopdrachten*).

As the major challenge of our research project is the automatic alignment of those more complex phrasal correspondences, we wanted to exclude these phrasal correspondences. However, as it was not possible to automatically distinguish between compounds and phrasal correspondences, we only retained all one-to-one correspondences.

The resulting bilingual dictionary contains 58,970 English-Dutch word pairs. More details can be found in table 3.1.

3.3.2 Probabilistic bilingual lexicon

We used a statistical word alignment package to derive a bilingual dictionary from a parallel corpus. Statistical word alignment is a semi-supervised method, which

[1]This work was carried out in the framework of the STEVIN DPC project.

Table 3.1: Nl-Translex English-Dutch dictionary containing only one-to-one correspondences

	# Entries
English-Dutch word pairs	58,970
English words	43,914
Dutch words	43,292

Table 3.2: Formal characterstics of the En-Nl parallel corpora used for dictionary creation (upper part) and of the resulting dictionary (lower part)

	2.4M	5.6M	9.3M
Aligned sentences	50,000	116,912	202,289
Total tokens	2,416,719	5,636,468	9,323,898
En word forms freq > 2	15,295	22,261	27,398
Nl word forms freq > 2	20,362	31,506	42,455
English-Dutch word pairs	16,728	21,486	28,342
English words	10,109	12,500	15,716
Dutch words	12,939	16,132	21,373

means that it starts from unannotated (raw) data. Most methods need a large sentence-aligned corpus to reliably estimate a statistical word alignment model. Statistical word alignment is based on the assumption of co-occurrence: words that are translations of each other co-occur more often than random in aligned sentence pairs. The output of a statistical word alignment model is a large bilingual word list with probability estimations.

As statistical word alignment tools need large sentence-aligned corpora, we opted for using the Europarl corpus (Koehn 2005). We created bilingual dictionaries using different parts of the Europarl corpus. The selected parts of the Europarl corpus were aligned on sentence level using the alignment tool that was released with the Europarl corpus. The corpora were tokenized and converted to lowercase.

Table 3.2 gives an overview of the formal characteristics of the resulting sentence-aligned parallel corpora that were used for deriving the bilingual dictionaries and the formal characteristics of the resulting bilingual dictionaries.
The most widely used statistical word alignment models are the IBM translation models (Brown et al. 1993). The most simple IBM model - IBM Translation Model One – is a purely *lexical* model: it only takes into account word frequencies in source and target sentences[2]. We used the Perl implementation of IBM Model

[2]The higher numbered IBM Models build on IBM Model One and take into account word order (distortion) and model the probability that a source word aligns to n target words (fertility).

One that is part of the Microsoft Bilingual Sentence Aligner (Moore 2002).

The IBM models allow only 1:n word mappings, and are therefore asymmetric. To overcome this problem, we ran the model in two directions: from English to Dutch and from Dutch to English. To get high-accuracy links, only the word pairs occurring in both the English-Dutch and Dutch-English word lists were retained, and the probabilities were averaged. To get rid of the noise produced by the translation model, only the entries with an averaged value of at least 0.1 were retained. This value was set experimentally.

To reduce the number of values, the averaged values were multiplied by 10 and only the integer part was retained. The obtained values allow us to rank the different translations according to frequency.

A sample of the resulting dictionary is shown in table 3.3. As the model was trained on a corpus of word forms, the dictionary does not abstract over word forms, e.g. *affordable - betaalbaar* and *affordable - betaalbare* are two separate entries in the dictionary. Model One can generate multiple translations for one word form, e.g. *affected* has three possible translations *getroffen, beïnvloed* and *getroffenen*, with *getroffen* being the most frequent translation.

Table 3.3: Sample of the English-Dutch probabilistic dictionary

English word form	Dutch word form	Frequency class
affect	invloed	1
affected	beïnvloed	1
affected	getroffen	4
affected	getroffenen	1
affection	genegenheid	1
afford	permitteren	3
afford	veroorloven	4
affordability	betaalbaarheid	2
affordable	betaalbaar	3
affordable	betaalbare	5

3.4 Additional linguistic annotations

Part-of-speech tagging and lemmatization for English was performed by the combined memory-based PoS tagger/lemmatizer, which is part of the MBSP tools (Daelemans and Van den Bosch 2005). Part-of-speech tagging and lemmatization for Dutch was performed by TADPOLE (Van den Bosch et al. 2007).

Although the MBSP toolkits contain chunking for English and Dutch, we opted for the development of two rule-based chunkers, the reason being that the English and Dutch shallow parsers adopt a different chunk definition. For example, adjacent verbs are clustered in one verbal group in the English memory-based shallow parser, but regarded as separate chunks in the Dutch memory-based shallow parser.

The rule-based chunkers for Dutch and English contain distituency rules. These rules add a chunk boundary when two consecutive part-of-speech codes cannot occur in the same constituent, e.g. between two finite verbs.

All Dutch and English texts of the test corpus (described in Section 3.6) were manually chunked, and the rule-based chunkers were evaluated by running the CoNLL-evalscript developed by Tjong Kim Sang, E.F. and Buchholz, S. (2000) on the test files. Precision scores of 93% (English) and 94% (Dutch) and recall scores of 95% (English and Dutch) were obtained.

3.5 Algorithm

As explained in section 3.2, we conceive our sub-sentential alignment system as a cascade model consisting of two phases. The objective of the first phase is to link *anchor chunks*, i.e. chunks that can be linked with a very high precision. Those anchor chunks are linked based on lexical clues and chunk similarity.

In order to link chunks based on lexical clues and chunk similarity, the following steps are taken for each sentence pair:

1. Creation of the lexical link matrix
2. Linking chunks based on lexical correspondences and chunk similarity
3. Linking adjacent function word chunks and final punctuation

The different steps are described in more detail below.

3.5.1 Creation of the lexical link matrix

Prior to creating the lexical link matrix, all possible translations for each word in the source and target sentence are retrieved from the bilingual dictionary. As explained in section 3.3, the probabilistic bilingual dictionary contains English-Dutch word pairs with numeric values that denote the frequency class of the word pair[3].

For each source and target word, all translations for the word form and the lemma are retrieved from the bilingual dictionary. If only the lemma of the source or target word is found in the bilingual dictionary, the resulting frequency weight is cut in half.

In the process of building the lexical link matrix, function words are neglected. Given the frequency of function words in a sentence, linking function words based on word alignment information alone often results in erroneous alignments. For that reason no lexical links are created for the following word classes: determiners, prepositions, coordinating conjunctions, possessive pronouns and punctuation symbols.

For all content words, if a source word occurs in the set of possible translations of a target word, or if a target word occurs in the set of possible translations of

[3]As in the Nl-Translex dictionary no frequency information is available, all word pairs get the same value.

the source words, a lexical link is created. Identical strings in source and target language are also linked.

The resulting lexical link matrix for our example is shown in figure 3.3.

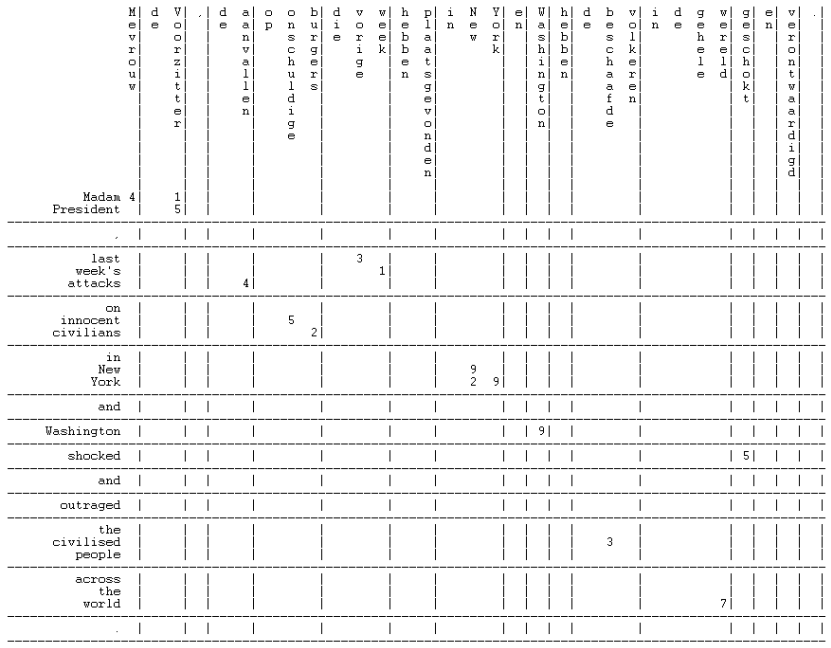

Figure 3.3: Lexical link matrix containing frequency weights

3.5.2 Linking anchor chunks

The problem of linking chunks based on lexical correspondences and similar chunks can be decomposed in two subproblems:

1. Selecting candidate anchor chunks
2. Testing chunk similarity of the candidate anchor chunks

Selecting candidate anchor chunks

The candidate anchor chunks are selected based on the information available in the lexical link matrix. For each source chunk a **contiguous** candidate target chunk is constructed. The contiguous candidate target chunk is built by concatenating all target chunks from a *begin index* until an *end index*. The begin index points to the first target chunk with a lexical link to the source chunk under consideration. The end index points to the last target chunk with a lexical link to the source chunk

under consideration. Possible intermediate chunks can contain additional lexical links, but this is not necessarily the case. If a source word contains more than one lexical link, the lexical link with the highest frequency weight is used.
In this way, the following contiguous 1:1 and 1:n candidate target chunks are built for our example:

Madam President	Mevrouw | de Voorzitter
last week's attacks	de aanvallen | op onschuldige burgers | die vorige week
on innocent civilians	op onschuldige burgers
in New York	in New York
Washington	Washington
shocked	geschokt
the civilised people	de beschaafde volkeren
across the world	in de gehele wereld

For some source chunks, it is also useful to build a **non-contiguous** candidate target chunk. The non-contiguous candidate target chunks are built by concatenating all target chunks with a lexical link to the source chunk under consideration. In our example, only one non-contiguous target chunk is constructed:

last week's attacks de aanvallen . . . vorige week

The process of selecting candidate chunks as described above, is performed twice: a first time starting from the source sentence; a second time starting from the target sentence. The second time, only those chunks for which no similarity test was performed are taken into consideration.

Testing chunk similarity

For each selected candidate pair, a *similarity test* is performed. Chunks are considered to be similar if at least a certain percentage of words of source and target chunk(s) are either linked by means of a lexical link or can be linked on the basis of corresponding part-of-speech codes.

All word classes can be linked based on PoS codes. In the candidate anchor chunk *the civilised people - de beschaafde volkeren*, one lexical clue (*civilised - beschaafde*) is sufficient to pass the similarity test as *the* and *de*, and *people* and *volkeren* are linked based on corresponding PoS codes.

The percentage of words that have to be linked was empirically set at 80% for contiguous chunks and 100% for non-contiguous chunks. The percentage of linked words is calculated as follows:

$$\frac{\text{\# words linked of source chunk + \# words linked of target chunk}}{\text{Total \# source chunk words + total \# target chunk words}}$$

The candidate anchor chunk *across the world - in de gehele wereld* contains one lexical link *world - wereld* and two PoS-links *across - in* and *the - de*. Hence the percentage of linked words $= (3 + 3)/(3 + 4) = 0.86$.

If a candidate anchor chunk passes the similarity test, the information in the matrix is updated as follows:

- All lexical links inside an anchor chunk are marked with the label **S** (Sure lexical links)
- Words linked based on corresponding part-of-speech codes are marked with the label **p** (PoS links)
- The label **r** is used to mark lexical links that are removed. If a source or target word had multiple lexical links, all lexical links other than the one(s) in the anchor chunk get the label **r**.

3.5.3 Linking adjacent function word chunks

In a final step, chunks consisting of one function word – mostly punctuation marks and conjunctions – can be linked based on corresponding part-of-speech codes if their left or right neighbours on the diagonal are anchor chunk. Corresponding final punctuation marks are also linked.

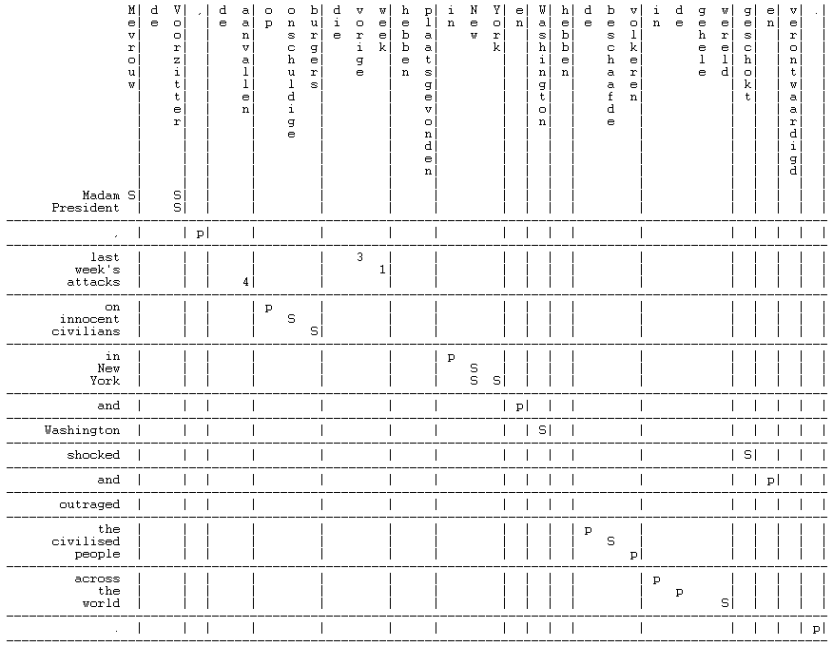

Figure 3.4: Matrix containing anchor chunks (**S** and **p** labels) and remaining lexical links with frequency weights

The resulting matrix for our example is shown in figure 3.4. In the example,

the following anchor chunks were retrieved:

Madam President	Mevrouw | de Voorzitter
,	,
on innocent civilians	op onschuldige burgers
in New York	in New York
and	en
Washington	Washington
shocked	geschokt
and	en
the civilised people	de beschaafde volkeren
across the world	in de gehele wereld

All retrieved anchor chunks but one can be considered to be entirely correct: *shocked* should have been linked to *hebben ... geschokt*, so the anchor chunk *shocked - geschokt* is only partially correct. In section 3.6, we describe how we evaluated the performance of the system.

3.6 Experimental results

A manual reference corpus was created that includes three different text types: user manuals, press releases and proceedings of plenary debates. Three different types of links were used: regular links for straightforward correspondences (e.g. *innocent - onschuldige*, *New York - New York*), fuzzy links for translation-specific shifts of various kinds (e.g. *last week's - die vorige week hebben plaatsgevonden*), and null links for words for which no correspondence could be indicated (deletions or additions). In the manual reference corpus, different units could be linked (words, word groups, paraphrased sections, punctuation). More details on the creation of the manual reference corpus can be found in (Macken 2007).

To evaluate the system's performance, the links created by the system were compared with the links in the manual reference files. Table 3.4 gives an overview of the number of words and documents used for testing the system.

To be able to compare the alignments of the system with the reference alignments, all phrase-to-phrase alignments were converted into word-to-word alignments by linking each word of the source phrase to each word of the target phrase

Table 3.4: En-Nl Test data

Text type	# Words	# Texts
Proceedings EP	3,139	7
Press Releases	4,926	4
User Manuals	4,010	2
Total	12,075	13

(all-pairs heuristic).

3.6.1 Metrics

The results of the experiments were evaluated in terms of precision and recall, which are widely used in the context of information retrieval. If we would calculate precision and recall on all word-to-word links, all links would be equally important. However, as Melamed (2001) pointed out, an evaluation metric that treats all links as equally important would place undue importance on words that were linked more than once (e.g. all word-to-word links resulting from the phrasal alignments). Therefore, a weight is assigned to each word-to-word link, and precision and recall are calculated on the weights of the word-to-word links.

We use the weighting method developed by Davis et al. (2007), which is a refinement of the weighting principles introduced by Melamed (2001). In this weighting scheme, every word contributes 0.5 to the total weight. In case of interlinked word-to-word links from the phrasal alignments, each link is assigned the total weight of the phrasal alignment divided by the number of word-to-word links. In the example of *last week's - die vorige week hebben plaatsgevonden*, the total weight of the phrasal alignment is 3.5 ($(2+5) \times 0.5$), and each word-to-word link gets a weight of 0.35 ($3.5/(2 \times 5)$). In the case of a regular word-to-word link (e.g. *innocent - onschuldige*), the resulting weight of the word-to-word link is 1 ($((1+1) \times 0.5)/(1 \times 1)$).

Precision and recall are then calculated on the weights, by using the following equations:

$$Precision = \frac{\text{total system weight of corresponding word-to-word links}}{\text{total system weight}}$$

$$Recall = \frac{\text{total reference weight of corresponding word-to-word links}}{\text{total reference weight}}$$

3.6.2 Results

The results presented here are the results of a system that uses the probabilistic dictionary trained on the 9.3M word corpus.

We considered all word-to-word links at two different stages in the system. A first time after dictionary lookup, and a second time after the alignment of chunks based on syntactic similarity. However, as it is also interesting to assess the reliability of the alignments of the anchor chunks, precision and recall are also calculated on only the word-to-word links that are part of the aligned anchor chunks. Table 3.5 contains the results per text type for all the texts of the test corpus.

In the columns under the heading *DCT*, the results after dictionary lookup are displayed. It it worth noticing that although the bilingual lexicon was trained on Europarl data, the coverage is quite good on other domains. The high figure for Press Releases can be explained by the high number of technical terms that are

Table 3.5: Normalized precision and recall on all word-to-word links at different stages in the system per text type

	DCT		DCT + AC		AC	
	Prec	Rec	Prec	Rec	Prec	Rec
Proceedings EP	.78	.28	.90	.44	.92	.38
Press Releases	.90	.34	.98	.59	.98	.54
User Manuals	.85	.27	.95	.46	.97	.40

identical in source and target text (the Press Releases test corpus contained 12% identical strings).

The columns under the heading *DCT + AC* contain the results after the alignment of syntactically similar chunks. During the alignment process, extra word-to-word links can be created for words belonging to the anchor chunks based on corresponding PoS codes (e.g. function words, adjectives). This explains the increase in recall. On the other hand, some disambiguation takes place for words that were linked several times. This explains the increase in precision.

In the last columns under the heading *AC*, precision/recall results are given for only the word-to-word links belonging to the chunks that were aligned based on syntactic similarity. The obtained precision scores (between .92 and .98) seem high enough to use the aligned chunks as anchors in the second phase of the alignment process.

We also investigated the impact of the size of the training corpus used for dictionary creation on the test results. We compared the obtained precision and recall scores at the different stages in the system on all the test files. As can be seen in figure 3.5, the size of the training corpus – and hence the size of the resulting dictionary – has a positive impact on recall at all stages in the system. No difference in precision was observed.

It is also interesting to compare the performance of the handcrafted bilingual lexicon with the performance of the probabilistic bilingual lexicon. Although the Nl-Translex dictionary is twice the size of the probabilistic dictionary trained on the 9.3M word corpus, the obtained recall scores are lower at all stages of the system[4]. The alignments retrieved by the system using the Nl-Translex system are more precise after dictionary lookup. But no difference in precision is observed if we only take into account the retrieved anchor chunks.

As explained in section 3.3.2, the probabilistic dictionaries were automatically extracted from a corpus containing word forms. We examined the impact of lemmatizing the training corpus prior to dictionary creation on the test results. By lemmatizing the training corpus, we expect that abstracting over word forms will

[4]It is also worthwhile to mention that the Nl-Translex dictionary and the probabilistic dictionary (trained on the lemmatized corpus) contain different word pairs: only 21% of the entries of the probabilistic dictionary are part of the NL-Translex dictionary.

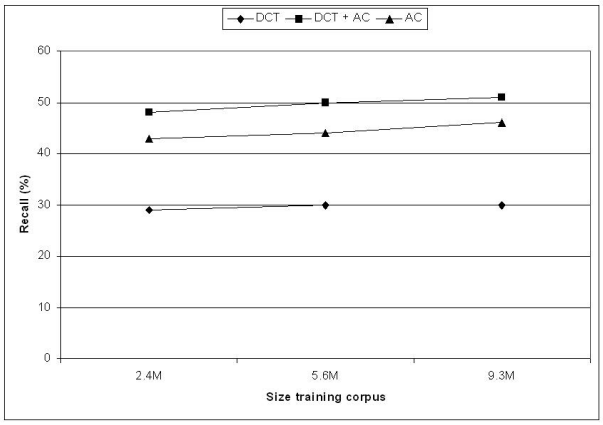

Figure 3.5: Impact of size training corpus on recall

increase the overall recall scores.

As the size of the dictionary might influence the impact of lemmatization, we trained probabilistic dictionaries on four different corpora: the lemmatized and word form version of both the 2.4M and the 9.3M word corpus respectively[5].

Lemmatization has a positive impact on recall for the system using the 2.4M word corpus. But the recall improvement is less clear for the system using the 9.3M word corpus. The reason being that the coverage of the 9.3M word forms dictionary is quite high[6]. The precision scores are slightly better for the word forms corpora.

3.7 Conclusion

In this paper, we have described the global architecture of our sub-sentential alignment system. We conceive our sub-sentential aligner as a cascade model with two phases. In the first phase *anchor chunks*, i.e. chunks that can be linked with a very high precision based on lexical correspondences and syntactic similarity are retrieved. In the second phase, we will focus on the more complex translational correspondences based on observed translation shift patterns. The anchor chunks of the first phase will be used to limit the search space in the second phase.

The objective of the first phase was to link *anchor chunks*, i.e. chunks that can be linked with a very high precision. In our baseline system, on average 40-50% of the words can be linked with a precision ranging from 90% to 98%. The obtained

[5]In the lemmatized systems, the retrieval of the lemmatized form is not penalized by reducing the frequency weight.

[6]We lemmatized the resulting dictionary extracted from the 9.3M word forms corpus off-line and compared it with the dictionary extracted from the lemmatized 9.3M corpus. An overlap of 95% was obtained. The overlap on the resulting dictionaries trained on the 2.3M word corpus was 85%

precision scores seem high enough to use the aligned chunks as anchors in the second phase of the alignment process.

We experimented with two different types of bilingual dictionaries to generate the lexical correspondences: a handcrafted bilingual dictionary and probabilistic bilingual dictionaries. We demonstrate that although the handcrafted dictionary is twice the size of the probabilistic dictionary, the obtained recall scores are lower. No difference in precision is observed for the retrieved anchor chunks.

We demonstrated that lemmatizing the training corpus prior to dictionary extraction can increase recall for small training corpora. As expected, increasing the size of the training corpora has a positive impact on the overall recall scores.

References

Brown, P.F., V.J. Della Pietra, S.A. Della Pietra, and R.L. Mercer (1993), The Mathematics of Statistical Machine Translation: Parameter Estimation, *Computational Linguistics* **19** (2), pp. 263–311.

Daelemans, W. and A. Van den Bosch (2005), Application to shallow parsing, *in* Daelemans, Walter and Antal Van den Bosch, editors, *Memory-based language processing*, Cambridge University Press, Cambridge, United Kingdom, pp. 85–103.

Davis, P.C., Z. Xie, and K. Small (2007), All Links are not the Same: Evaluating Word Alignments for Statistical Machine Translation, *in* Maegaard, Bente, editor, *Machine Translation Summit XI*, European Associaton for Machine Translation, Copenhagen, Denmark, pp. 119–126.

Goetschalckx, J., C. Cucchiarini, and J. Van Hoorde (2001), Machine Translation for Dutch: the NL-Translex Project.

Gotti, F., P. Langlais, E. Macklovitch, D. Bourigault, B. Robichaud, and C. Coulombe (2005), 3GTM: a third-generation translation memory, *3rd Computational Linguistics in the North-East (CLiNE) Workshop*, Gatineau, Québec.

Groves, D. and A. Way (2006), Hybridity in MT: Experiments on the Europarl corpus, *11th Conference of the European Association for Machine Translation*, Oslo, Norway.

Koehn, P. (2005), Europarl: a parallel corpus for statistical machine translation, *Tenth Machine Translation Summit*, Phuket, Thailand, pp. 79–86.

Macken, L. (2007), Analysis of translational correspondence in view of sub-sentential alignment, *METIS-II Workshop on New Approaches to Machine Translation*, Leuven, Belgium, pp. 97–105.

Macken, L., E. Lefever, and V. Hoste (2008), Linguistically-based sub-sentential alignment for terminology extraction from a bilingual automotive corpus, *22nd International Conference on Computational Linguistics (Coling 2008)*, Manchester, United Kingdom.

Melamed, D.I. (2001), Manual annotation of translational equivalence, *in*

Melamed, Dan I., editor, *Empirical methods for exploiting parallel texts*, MIT Press, Cambridge, Massachusetts, pp. 65–77.

Moore, R.C. (2002), Fast and accurate sentence alignment of bilingual corpora, *5th Conference of the Association for Machine Translation in the Americas*, Machine Translation: from research to real users, Tiburon, California, pp. 135–244.

Tiedemann, J. (2003), Combining Clues for Word Alignment, *10th Conference of the European Chapter of the ACL (EACL03)*, Budapest, Hungary.

Tjong Kim Sang, E.F. and Buchholz, S. (2000), Introduction to the CoNLL-2000 Shared Task: Chunking, *CoNLL-2000 and LLL-2000*, Lisbon, Portugal, pp. 127–132.

Van den Bosch, A., B. Busser, W. Daelemans, and S. Canisius (2007), An efficient memory-based morphosyntactic tagger and parser for Dutch, *Computational Linguistics in the Netherlands 2006*, Selected Papers of the 17th Computational Linguistics in the Netherlands Meeting, Leuven, Belgium, pp. 191–206.