Linguistically-based sub-sentential alignment for terminology extraction from a bilingual automotive corpus

Lieve Macken and **Els Lefever** and **Veronique Hoste**
Language and Translation Technology Team
Ghent University College
Belgium
{Lieve.Macken, Els.Lefever, Veronique.Hoste}@hogent.be

Abstract

We present a sub-sentential alignment system that links linguistically motivated phrases in parallel texts based on lexical correspondences and syntactic similarity. We compare the performance of our sub-sentential alignment system with different symmetrization heuristics that combine the GIZA++ alignments of both translation directions. We demonstrate that the aligned linguistically motivated phrases are a useful means to extract bilingual terminology and more specifically complex multiword terms.

1 Introduction

This research has been carried out in the framework of a customer project for PSA Peugeot Citroën. The final goal of the project is a reduction and terminological unification process of PSA's database, which contains all text strings that are used for compiling user manuals. French being the source language, all French entries have been translated to some extent into the twenty different languages that are part of the customer's portfolio. Two sub-projects have been defined:

1. automatic terminology extraction for all languages taking French as the pivot language

2. improved consistency of the database entries, e.g. through the automatic replacement of synonyms by the preferred term (decided in (1))

This paper presents a novel terminology extraction method applied to the French-English part of the database.

There is a long tradition of research into bilingual terminology extraction (Kupiec, 1993), (Gaussier, 1998). In most systems, candidate terms are first identified in the source language based on predefined PoS patterns – for French, *N N*, *N Prep N*, and *N Adj* are typical patterns. In a second step, the translation candidates are extracted from the bilingual corpus based on word alignments. In recent work, Itagaki et al. (2007) use the phrase table derived from the GIZA++ alignments to identify the translations.

We use a different and more flexible approach. We developed a sub-sentential alignment system that links linguistically motivated phrases in parallel texts based on lexical correspondences and syntactic similarity. Rather than predefining terms as sequences of PoS patterns, we first generate candidate terms starting from the aligned phrases. In a second step, we use a general purpose corpus and the n-gram frequency of the automotive corpus to determine the specificity of the terms.

The remainder of this paper is organized as follows: Section 2 describes the corpus. In Section 3, we present our linguistically-based sub-sentential alignment system and in Section 4 we describe how we use the aligned phrases for terminology extraction.

2 Automotive corpus

For developing our terminology extraction module, we have used the French-English sentence-aligned database that contains 363,651 entries. These entries can be full sentences, parts of sentences, as well as isolated words and are aligned across languages by means of a unique ID. The

Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008), pages 529–536
Manchester, August 2008

	PoS tagging error rate	Lemmatisation error rate	PoS after update	Lemmas after update
French	4.50 %	2.29 %	1.92 %	1.22 %
English	5.16 %	3.13 %	2.66 %	3.03 %

Table 1: Part-of-Speech tagging and lemmatisation error rate on the test sentences

average sentence length of a database entry is 9 words.

2.1 Linguistic annotation

In order to ensure consistent processing of the languages in the corpus (e.g. Italian, Spanish, German), we have used the freely availabe TreeTagger tool (Schmid, 1994) for performing tokenisation, part-of-speech tagging and lemmatisation of the corpus. In order to evaluate the domain-adaptability of the tagger, we have manually validated the quality of the TreeTagger output for a training set of 12,200 tokens (about 1,200 sentences). We have used this validated set to derive a list of hard coded PoS tags (e.g. the French word *vis* can be a noun or verb, but is always a noun in our corpus) as well as post-processing rules for remediating erroneous PoS tags. We additionally annotated 350 test sentences (about 3,500 tokens). Table 1 shows the error rate figures for PoS-tagging and lemmatisation before and after updating the default output of the TreeTagger tool.

We further enriched the corpora with chunk information. During text chunking, syntactically related words are combined into non-overlapping chunks based on PoS information. We developed rule-based chunkers for English and French. The rule-based chunkers contain distituency rules, i.e. the rules add a chunk boundary when two part-of-speech codes cannot occur in the same constituent. The following example shows a French-English sentence pair divided in non-overlapping chunks:

> Fr: valable | uniquement | pour la ceinture | de sécurité avant latérale | du côté passager
> En: applies | only | to the outer seat belt | on the passenger side

We manually indicated chunk boundaries in the 350-sentences test corpus and evaluated the rule-based chunkers by running the CoNLL-evalscript (Tjong Kim Sang and Buchholz, 2000). We obtained precision scores of 89% and 87% and recall scores of 91% and 91% for French and English respectively.

	# Words	# Sentence pairs
Short (< 8 words)	4,496	404
Medium (8-19 words)	4,493	212
Long (> 19 words)	4,498	97
Total	13,487	713
Development corpus	4,423	231

Table 2: Number of words and sentence pairs in the three test corpora and the development corpus

2.2 Test corpora

As we expect that sentence length has an impact on the alignment performance, we created three test corpora with varying sentence length. We distinguished short sentences (2-7 words), medium-length sentences (8-19 words) and long sentences (> 19 words). Each test corpus contains approximately 4,500 words.

We also compiled a development corpus containing sentences of varying sentence length to debug the system and to determine the value of the thresholds used in the system. The formal characteristics of the test corpora and the training corpus are given in Table 2.

3 Sub-sentential alignment

Sub-sentential alignments – and the underlying word alignments – are used in the context of Machine Translation to create phrase tables for phrase-based statistical machine translation systems (Koehn et al., 2007). 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, or in sub-sentential translation memory systems (Gotti et al., 2005). A quite obvious application of a sub-sentential alignment system is the creation of bilingual dictionaries and terminology extraction from bilingual corpora (Melamed, 2000), (Itagaki et al., 2007).

In the context of statistical machine translation, GIZA++ is one of the most widely used word alignment toolkits. GIZA++ implements the IBM models and is used in Moses (Koehn et al., 2007) to generate the initial source-to-target and target-to-source word alignments after which some symmetrization heuristics combine the alignments of both translation directions.

We present an alternative – linguistically-based – approach, that starts from a lexical probabilistic bilingual dictionary generated by IBM Model One.

3.1 Architecture

The basic idea behind our approach is that – at least for European languages – translations conveying the same meaning use to a certain extent the same building blocks from which this meaning is composed: i.e. we assume that to a large extent noun and prepositional phrases, verb phrases and adverbial phrases in one language directly map to similar constituents in the other language[1]. The extent to which our basic assumption holds depends on the translation strategy that was used. Text types that are typically translated in a more literal way (e.g. user manuals) will contain more direct correspondences.

We conceive our sub-sentential aligner 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 syntactic similarity. In the second phase, we will try to model 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.

As the application at hand is terminology extraction, we are interested in alignments with very high precision. As the automotive corpus contains rather literal translations, we expect that a high percentage of anchor chunks can be retrieved using only the first phase of our approach.

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.

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. During anchor chunk alignment, the sub-sentential aligner links chunks based on lexical correspondences and chunk similarity.

3.2 Bilingual Dictionary

We used the Perl implementation of IBM Model One that is part of the Microsoft Bilingual Sentence Aligner (Moore, 2002) to derive a bilingual dictionary from a parallel corpus. IBM Model One is a purely *lexical* model: it only takes into account word frequencies of source and target sentences[2].

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 French to English and from English to French. To get high-accuracy links, only the words pairs occurring in both the French-English and English-French 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[3].

The resulting bilingual dictionary contains 28,990 English-French word pairs. The bilingual dictionary is used to create the lexical link matrix for each sentence pair.

3.3 Lexical Link Matrix

For each source and target word in each sentence pair, all translations for the word form and the lemma are retrieved from the bilingual dictionary.

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

3.4 Anchor chunks

Anchor chunk alignment comprises two steps. In a first step, we select candidate anchor chunks; in a second step we test the syntactic similarity of the candidate anchor chunks.

3.4.1 Selecting candidate anchor chunks

The candidate anchor chunks are selected based on the information available in the lexical link matrix.

[1] The more syntax-aware SMT systems assume that to a certain extent syntactic relationships in one language directly map to syntactic relationships in the other, which Hwa (2002) calls the *Direct Correspondence Assumption*.

[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).

[3] Lowering this threshold significantly decreased precision scores of the sub-sentential alignment system.

For each source chunk a candidate target chunk is constructed. The 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. In this way, 1:1 and 1:n candidate target chunks are built.

The process of selecting candidate chunks as described above, is performed a second time starting from the target sentence. In this way additional n:1 candidates are constructed.

3.4.2 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 addition to linking words based on PoS codes, a small set of predefined language-dependent rules were implemented to handle function words. For example:

- Extra function words (determiners and prepositions) in source or target language are linked together with their noun to the noun's translation.
- In French, the preposition *de* is contracted with the definitive articles *le* and *les* to *du* and *des* respectively. The contracted determiners are linked to an English preposition and determiner.

The percentage of words that have to be linked was empirically set at 85%.

3.5 Remaining chunks

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

In a final step, additional candidates are constructed by selecting non-anchor chunks in the source and target sentence that have corresponding left and right anchor chunks as neigbours. The anchor chunks of the first step are used as contextual information to link n:m chunks or chunks for which no lexical link was found in the lexical link matrix.

In Figure 1, the chunks [Fr: gradient] – [En: gradient] and the final punctuation mark have been retrieved in the first step as anchor chunk. In the last step, the n:m chunk [Fr: de remontée pédale d' embrayage] – [En: of rising of the clutch pedal] is selected as candidate anchor chunk because it is enclosed within anchor chunks.

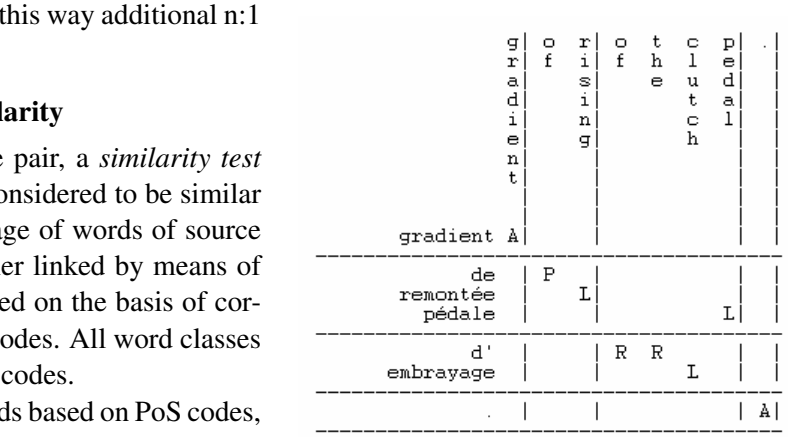

Figure 1: n:m candidate chunk: 'A' stands for anchor chunks, 'L' for lexical links, 'P' for words linked on the basis of corresponding PoS codes and 'R' for words linked by language-dependent rules.

As the contextual clues (the left and right neigbours of the additional candidate chunks are anchor chunks) provide some extra indication that the chunks can be linked, the similarity test for the final candidates was somewhat relaxed: the percentage of words that have to be linked was lowered to 0.80 and a more relaxed PoS matching function was used:

- Verbs and nouns can be linked
  Fr: pour permettre de **vidanger** proprement le circuit
  En: to permit clean **draining** of the system
- Adjectives and nouns can be linked
  Fr: l' **entrée** d' air
  En: **incoming** air
- Past participles can be linked to past tense[4]

3.6 Evaluation

All translational correspondences were manually indicated in the three test corpora (see section 2.2).

[4]The English PoS tagger often tags a past participle erroneously as a past tense.

We adapted the annotation guidelines of Macken (2007) to the French-English language pair, and used three different types of links: *regular* links for straightforward correspondences, *fuzzy* links for translation-specific shifts of various kinds, and *null* links for words for which no correspondence could be indicated. Figure 2 shows an example.

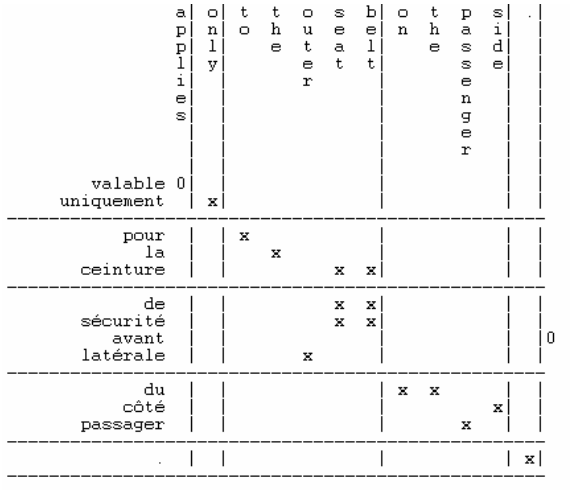

Figure 2: Manual reference: regular links are indicated by x's, fuzzy links and null links by 0's

To evaluate the system's performance, we used the evaluation methodology of Och and Ney (2003). Och and Ney distinguished *sure* alignments (S) and *possible* alignments (P) and introduced the following redefined precision and recall measures:

$$precision = \frac{|A \cap P|}{|A|}, recall = \frac{|A \cap S|}{|S|} \quad (1)$$

and the alignment error rate (AER):

$$AER(S, P; A) = 1 - \frac{|A \cap P| + |A \cap S|}{|A| + |S|} \quad (2)$$

We consider all regular links of the manual reference as *sure* alignments and all fuzzy and null links as *possible* alignments to compare the output of our system with the manual reference.

We trained statistical translation models using Moses. Moses uses the GIZA++ toolkit (IBM Model 1-4) in both translation directions (source to target, target to source) and allows for different symmetrization heuristics to combine the alignments of both translation directions. We used three different heuristics: *grow-diag-final* (default), *intersection* and *union*.

	SHORT			MEDIUM			LONG		
	p	r	e	p	r	e	p	r	e
∩	.99	.83	.10	.98	.73	.16	.99	.77	.13
∪	.95	.92	.07	.91	.86	.11	.91	.89	.10
Gdf	.95	.91	.07	.93	.85	.11	.94	.88	**.09**
Ling.	.96	.93	**.06**	.94	.88	**.09**	.92	.87	.10

Table 3: Precision (p), recall (r) and alignment error rate (e) for three symmetrization heuristics based on the GIZA++ alignments (intersection(∩), union (∪), Grow-diag-final (Gdf)) vs the linguistically-based system (Ling.) for the three test corpora

Table 3 compares the alignment quality of our linguistically-based system with the purely statistical approaches. Overall, the results confirm our assumption that shorter sentences are easier to align than longer sentences. As expected, the *intersection* heuristic aligns words with a very high precision (98-99%). We further observe that the alignment error rate of the linguistically-based system is the lowest for the short and medium-length sentences, but that on the long sentences the default symmetrization heuristic yields the best results. Manual inspection of the alignments revealed that in some long sentences, the linguistically-based system misaligns repeated terms in long sentences, a phenomenon that occured frequently in the long sentence corpus. As expected, the linguistically-based system scores better on function words.

Overall, on this data set, the linguistically-based system yields results that are comparable to the results obtained by the complex and computationally expensive chain of IBM models.

4 Terminology extraction

As described in Section 1, we generate candidate terms starting from the aligned anchor chunks. In a second step, we use a general purpose corpus and the n-gram frequency of the automotive corpus to determine the specificity of the terms.

4.1 Generating candidate terms

English and French use a different compounding strategy. In English, the most frequently used compounding strategy is the concatenation of nouns, while in French prepositional phrases are concatenated. The following example illustrates the different compounding strategy:

Fr: une procédure d' initialisation du calculateur de boîte de vitesses automatique

En: an automatic gearbox ECU initialisation procedure

We start from the anchor chunks as they are the minimal chunks that could be linked together. We implemented two heuristics to generate additional French candidate terms: a first heuristic strips off adjectives and a second heuristic considers each N + PP pattern as candidate term.

For each French candidate term, the English translation is constructed on the basis of the word alignments. The following candidate terms are generated for our example:

1	procédure d' initialisation du calculateur de boîte de vitesses automatique	automatic gearbox ECU initialisation procedure
2	procédure d' initialisation du calculateur de boîte de vitesses	gearbox ECU initialisation procedure
3	procédure d' initialisation	initialisation procedure
4	initialisation du calculateur	ECU initialisation
5	calculateur de boîte de vitesses	gearbox ECU
6	boîte de vitesses automatique	automatic gearbox
7	boîte de vitesses	gearbox
8	procédure	procedure
9	initialisation	initialisation
10	calculateur	ECU
11	automatique	automatic

4.2 Filtering of candidate terms

As our terminology extraction module is meant to generate a bilingual automotive lexicon, every entry in our lexicon should refer to a concept or action that is relevant in an automotive context. This also means we want to include the minimal semantical units (e.g. seat belt) as well as the larger semantical units (e.g. outer front seat belt) of a parent-child term relationship. In order to decide on which terms should be kept in our lexicon, we have combined two algorithms: Log-Likelihood for single word entries and Mutual Expectation Measure for multiword entries.

4.2.1 Log-Likelihood Measure

In order to detect single word terms that are *distinctive* enough to be kept in our bilingual lexicon, we have applied the Log-Likelihood measure (LL). This metric considers frequencies of words weighted over two different corpora (in our case a technical automotive corpus and the more general purpose corpus "Le Monde"), in order to assign high LL-values to words having much higher or lower frequencies than expected. Daille (1995) has determined empirically that LL is an accurate measure for finding the most *surprisingly* frequent words in a corpus. Low LL values on the other hand allow to retrieve common vocabulary with high frequencies in both corpora. We have created a frequency list for both corpora and calculated the Log-Likelihood values for each word in this frequency list. In the formula below, N corresponds to the number of words in the corpus, whereas the *observed values* O correspond to the real frequencies of a word in the corpus. The formula for calculating both the expected values (E) and the Log-Likelihood have been described in detail by (Rayson and Garside, 2000).

$$E_i = \frac{N_i \sum_i O_i}{\sum_i N_i} \tag{3}$$

We used the resulting Expected values for calculating the Log-Likelihood:

$$-2ln\lambda = 2\sum_i O_i ln(\frac{O_i}{E_i}) \tag{4}$$

Manual inspection of the Log-Likelihood figures confirmed our hypothesis that more domain-specific terms in our corpus got high LL-values. As we are mainly interested in finding distinctive terms in the automotive corpus, we have only kept those terms showing positive Expected Values in our domain-specific corpus combined with user-defined Log-Likelihood values. Examples of French-English translation pairs that are filtered out using the LL values are:

Fr: tout – En: entire
Fr: propre – En: clean
Fr: interdits – En: prohibited
Fr: nombre – En: number

4.2.2 Mutual Expectation Measure

Dias and Kaalep (2003) have developed the Mutual Expectation measure for evaluating the degree of cohesiveness between words in a text. We have applied this metric on our list of multiword terms, to exclude multiword terms which components do not occur together more often than expected by chance. In a first step, we have calculated all n-gram frequencies (up to 8-grams) for our English and French sentences. We use these frequencies to derive the Normalised Expectation (NE) values for all multiword entries, as specified by the formula of Dias and Kaalep:

$$NE = \frac{prob(n-gram)}{\frac{1}{n}\sum prob(n-1-grams)} \quad (5)$$

The Normalised Expectation value expresses the cost, in terms of cohesiveness, of the possible loss of one word in an n-gram. The higher the frequency of the n-1-grams, the smaller the NE, and the smaller the chance that it is a valid multiword expression. As simple n-gram frequency also seems to be a valid criterion for multiword term identification (Daille, 1995), the NE values are multiplied by the n-gram frequency to obtain the final Mutual Expectation (ME) value.

We have calculated Mutual Expectation values for all French and English multiword terms and filtered out incomplete or erroneous terms having very low ME values. The following example has been filtered out:

> Fr: permettant d'alimenter le circuit d'eau arrière
> En: to supply the rear water circuit
> *Incomplete term:*
> eau arrière - rear water (should be Fr: circuit d'eau arrière - En: rear water circuit)

4.3 Evaluation of the Terminology Extraction Module

To evaluate the terminology extraction module, we used all sentences of the three test corpora (see Section 2.2). We compared the performance of our algorithm with the output of a commercial state-of-the-art terminology extraction program SDL MultiTerm Extract[5]. MultiTerm first extracts source language terms and identifies in a separate step the term translations. MultiTerm makes use of basic vocubulary lists to exclude general vocabulary words from the candidate term list. We ran MultiTerm Extract with the default settings on 70,000 aligned sentences[6] of the automotive corpus. The extracted terms of our system have been filtered by applying Log-Likelihood thresholds (for single word terms) and Mutual Expectation thresholds (for multiword terms). Tabel 4 shows the number of terms after each reduction phase.

The output of both systems has been manually labeled taking into account the following guidelines:

[5] www.translationzone.com/en/products/sdlmultitermextract

[6] 70,000 sentences was the maximum size of the corpus that could be processed within MultiTerm Extract.

	# extracted entries	# entries after ME filtering	# entries after LL filtering
Anchor chunk approach	2778	2688	2549
Multiterm Extract	1337	N/A	N/A

Table 4: Figures after Log-Likelihood and Mutual Expectation reduction

Anchor chunk approach	Correct	Not correct	Maybe correct
Multiwords	78.5%	19%	2.5%
Single words	89.5%	9.5%	1%
All terms	83%	15%	2%
Multiterm Extract	**Correct**	**Not correct**	**Maybe correct**
Multiwords	51%	48.5%	0.5%
Single words	83%	16%	1%
All terms	66%	33.5%	0.5%

Table 5: Results Term Extraction Module

- judge the quality of the bilingual entry as a whole, meaning that the French and English terms should express the same concept
- each entry should form a semantic unit and refer to an existing concept or action in the automotive context

During manual validation, the following three labels have been used: *OK* (valid entry), *NOK* (not a valid entry) and *MAYBE* (when the annotator was not sure about the correct label). Table 5 lists the results of both our system and MultiTerm Extract and illustrates that our linguistically based alignment approach works particularly well for the extraction of more complex multiword expressions.

Error analysis on the errors made by the anchor chunk approach revealed the following error types:

1. compounds that are only partially retrieved in one of the two languages:

ceinture (valable uniquement pour la **ceinture de sécurité** avant latérale)	outer seat belt (applies only to the outer **seat belt**)

2. fuzzy word links (different grammatical and syntactical structures, paraphrases etc) that result in bad lexicon entries:

fusibles (montage avec vide-poches inférieur fixe **sans rangement des fusibles**)	no fuse (fitting with fixed lower storage compartment with **no fuse storage**)

3. translation errors in the parallel corpus:

automatique (tableau de commande climatisation **automatique**)	additional (**additional** air conditioning unit control panel)

4. ambiguous words that cause PoS and chunking errors (in the corpus *avant* is usually used as an adjective, but in the example it has a prepositional function as *avant de*):

câbles avant	cables
(repérer la position des câbles **avant de** les déclipper)	(mark the position of the cables **before** unclipping them)

5 Conclusions and future work

We presented a sub-sentential alignment system that links linguistically motivated phrases in parallel texts based on lexical correspondences and syntactic similarity. Overall, the obtained alignment scores are comparable to the scores of the state-of-the-art statistical approach that is used in Moses.

The results show that the aligned linguistically motivated phrases are a useful means to extract bilingual terminology for French-English. In the short term, we will test our methodology on other language pairs, i.e. French-Dutch, French-Spanish and French-Swedish. We will also compare our work with other bilingual term extraction programs.

6 Acknowledgement

We would like to thank PSA Peugeot Citroën for funding this project.

References

Daille, B. 1995. Study and implementation of combined techniques for automatic extraction of terminology. In Klavans, J. and P. Resnik, editors, *The Balancing Act: Combining Symbolic and Statistical Approaches to Language*, pages 49–66. MIT Press, Cambridge, Massachusetts; London, England.

Dias, G. and H. Kaalep. 2003. Automatic Extraction of Multiword Units for Estonian: Phrasal Verbs. *Languages in Development*, 41:81–91.

Gaussier, E. 1998. Flow Network Models for Word Alignment and Terminology Extraction from Bilingual Corpora . In *36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics (Proceedings of COLING-ACL '98)*, pages 444–450, Université de Montréal, Montreal, Quebec, Canada.

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

Hwa, R., P. Resnik, A. Weinberg, and O. Kolak. 2002. Evaluating translational correspondence using annotation projection. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 392–399, Philadelphia, PA, USA.

Itagaki, M., T. Aikawa, and X. He. 2007. Automatic Validation of Terminology Consistency with Statistical Method. In Maegaard, Bente, editor, *Machine Translation Summit XI*, pages 269–274, Copenhagen, Denmark. European Associaton for Machine Translation.

Koehn, P., H. Hoang, A. Birch, C. Callison-Burch, M. Federico, N. Bertoldi, B. Cowan, W. Shen, C. Moran, R. Zens, C. Dyer, O. Bojar, A. Constantin, and E. Herbst. 2007. Moses: Open Source Toolkit for Statistical Machine Translation. In *Proceedings of the ACL 2007 Demo and Poster Sessions*, pages 177–180, Prague, Czech Republic.

Kupiec, J. 1993. An algorithm for finding noun phrase correspondences in bilingual corpora. In *Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics*.

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

Melamed, I. Dan. 2000. Models of translational equivalence among words. *Computational Linguistics*, 26(2):221–249.

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

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

Rayson, P. and R. Garside. 2000. Comparing corpora using frequency profiling. In *Proceedings of the workshop on Comparing Corpora, 38th annual meeting of the Association for Computational Linguistics (ACL 2000)*, pages 1–6.

Schmid, H. 1994. Probabilistic part-of-speech tagging using decision trees. In *International Conference on New Methods in Language Processing*, Manchester, UK.

Tjong Kim Sang, Erik F. and Sabine Buchholz. 2000. Introduction to the CoNLL-2000 Shared Task: Chunking. In *CoNLL-2000 and LLL-2000*, pages 127–132, Lisbon, Portugal.