HAL
archives-ouvertes.fr

Aligning through divergence

Johan Segura, Violaine Prince

▶ **To cite this version:**

Johan Segura, Violaine Prince. Aligning through divergence. SNLP-AOS'2011: Joint International Symposium on Natural Language Processing and Agricultural Ontology Service, Feb 2012, Phuket, Thailand. 1, pp.150-159, 2012. <lirmm-00839340>

HAL Id: lirmm-00839340
https://hal-lirmm.ccsd.cnrs.fr/lirmm-00839340

Submitted on 27 Jun 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Aligning Through Divergence

Johan Segura
LIRMM
segura@lirmm.fr

Violaine Prince
LIRMM
prince@lirmm.fr

***Abstract*—This document presents a bilingual phrase-based alignment method handling syntactic constituents (sub-sentential components) of parallel sentences. The method rely on an asymmetrical parsing of both languages: Light part-of-speech tagging for the target language, syntactic tree building for the 'source' language and the complexity of each is studied. The models and methods can be seen as a subclass of Example Based Machine Translation.**

I. Introduction

Automatic sub-sentential alignment is one of the basic tasks preceding machine translation (MT). It is performed to enhance its efficiency, by increasing translation memories and resources with human translated data. It is seen as a cornerstone in MT. Sub-sentential alignment needs parallel bilingual corpora. It aims at automatically providing translation links between sentences *constituents*, i.e., words or multiword expressions, smaller than a sentence, within a pair of parallel sentences. Two items are crucial in such a task: Alignment relevance and alignment requirements (paradigm, methods, resources). Both are related. Classical models, still representative, focus on word-to-word alignments. Late research in alignment tends to favor a granularity bigger than the single word (e.g. [6], [8]). Detecting relevant phrases for alignment can motivate the use for syntactical information. In representative rule-based systems, rules are either applied in a pre-determined fashion, or in a "first best-value" approach (statistically based, thus mixing statistical and symbolic methods). In different cases, rules overlapping conflicts are differently solved. Most of the time, such process relevance is less discussed than rule shapes. We will try here to discuss the role of segments shape through both problematics of tractability and linguistic relevance. We will also discuss the choices of the shapes proposed for the segments and the role they play in the alignment process. The methods described hereafter are example-based methods that use an 'alignment memory', which is a learned set of segments. These segments can be seen as bilingual phrase pairs presenting internal links. The process asynchronously combines alignment constraints in order to maximize coverage (in an EBMT style). The information acquisition process is facilitated by a graphical user interface. One of the original features of this method is that the process can align word segments as well as syntactic patterns. It relies on an asymmetrical effort in syntactic processing: A constituent and dependence parser is used for the source, and a POS tagger for the target. No dictionary or lexical resources are *a priori* required. The next section details related alignment methods. Section 3 presents our model, and section 4, a preliminary experiment with some results. Conclusion will shed light on the work extensions and further developments.

II. Related Works

The literature on alignment is abundant, and some work has already been mentioned in introduction. The founding work in alignment is attributed to Brown et al. at IBM [2]. The GIZA++ system [16] which is based on these IBM models, has evolved through time from a pure lexical to a sophisticated tool relying on a complex language model to account for translation divergence. It is still widely used in alignment literature (e.g. in [11], [6]). Syntactic trees as elements of the alignment process have appeared with [21]. Since then, hybrid systems, embedding syntactical information in a statistical model emerged as well as purely symbolic approaches. The use of structural information brought by syntax is claimed to be helpful for different reasons among which we can quote :

1) Preventing alignments violating linguistic structural properties (e.g.,[5])
2) Propagating alignments according to parent-child links (e.g., [13] [17])
3) Predicting an alignment with a POS tag, when the lexicon does not provide information [5]
4) Generating structures that accelerate the rule-base building process in a data driven approach (e.g. [12]).

Another aspect of this model tries to take advantage of the syntax: it is an example-based alignment model sharing common issues with example-based MT (EBMT). EBMT tries to imitate the human translation by analogy. It is an intuitive approach consisting in storing pieces of translations already met in the past, getting the relevant ones in a new situation, then combining the pieces to obtain a solution. The first suggestion of EBMT issues is attributed to Makato Nagao in [14]. He clearly defined the three important steps of an EBMT process:(1) Matching segments from a database,(2) filtering and (3) combining. Nagao claims that a human translation process does not involve a deep analysis structure but relies on analogy between generic segments. This idea motivates the whole example-based approach. EBMT literature agrees that segments size must be at the sub-sentential level for 'genericness' reasons [7] (Identical sentences occur only rarely), but raises in turn the issue of recombining segments in a way that preserves language structure and meaning [18]. Furthermore, it is known that linguistically motivated patterns are of a benefits [11]. For these reasons we thought it was

necessary to resort to deep syntactic informations to tackle the segmentation part. Then, when recombining segments from examples, one must choose a good matching measure. In [15], the author observes that "the simplest metric is a complete match" and proposes a heuristic: "Quality of a match is proportional to a measure of contiguity of matching". This classical argument in EBMT can also be found in SMT phrase-based methods like in deNero [8]. Our method sticks to this approach, although the recombining effort in an alignment method is quite different from an EBMT as we'll see in the next section. Finally, the shape the patterns should take in EBMT, is also motivated by a correct reuse. Efforts must be done to make the segments as generic as possible without losing consistency during the recombining process. The pattern generalization of Brown's method [3], which uses syntactic analysis to replace some words with their classes or categories, generalizes them to a much wider set of applications. This approach emphasizes the gain of generalization by showing an accelerating efficiency in the treatment. The methods detailed hereafter make an extensive use of the generalization with POS tags information. The aim of this work is to try to evaluate the viability of an original aligner close to EBMT paradigms, deterministic and asynchronous. As an early experiment, we reduced the use of lexical information to a strict minimum, then allowing to handle non-compositional translations and accelerating segment acquisition. It is certain that, in some future work, a word to word alignment based on lexical information will be considered since the model is meant to be embedded in some larger process. Thus, a crucial perspective of this work is to enrich a translation memory as a sort of 'super' lexicon of equivalent expressions, involving stylistic idiosyncrasies of both languages.

III. Model and Method

The pair of considered languages are respectively French and English (available parsing resources). The parsing of the French source sentence is carried out by SYGFRAN [4] which provides a deep syntactic tree. TreeTagger [19] is used for English POS tagging task. TreeTagger has not been used for French since it does not offer enough syntactic information (no deep tree structure). Therefore the method is asymmetrical.

A. Elements of the model

The model relies on a set of fragments, which are divided into two parts:
(1) The condition part: A condition on a word is a formula without negation involving POS-tags values.
(2) The application part: A set of alignment actions based on the condition checking.

1) Condition part: Let $(K_n)_{0\leq n\leq N_K}$ be a finite set of categories for source language and an other one for the target $(K'_m)_{0\leq m\leq N'_K}$. They can be instantiated by values from the two sets: $(v_n)_{n\in\mathbb{N}}$ and $(v'_n)_{n\in\mathbb{N}}$.
A **syntactic pattern** (recognized by the model) on a source term will be :

Tags belonging to SYGFRAN:
CAT: POS category
N: Noun
$SOUSN$: Nominal subcategory
$NCOM$: Common noun
$NPRO$: Proper noun
$DETERM$: Determinant
$SOUSD$: Subcategories of the Determinant type
$ARTD$: Determinate article (e.g. 'the' in English)
$ARTI$: Indeterminate article (e.g. 'a' in English)
$ADJOINT$: Adjoint type (adjectives, adverbs)
$SOUSA$: Subcategories of the adjoint type
$ADNOM$: Adjectives qualifying a noun
$PREP$: Preposition
$CATPREPSIMPLE$:Simple Preposition

Tags belonging to TreeTagger:
JJ: Adjective
NN: Common Noun, singular
IN:Preposition
NP:Proper Noun, singular

Fig. 1. Sets of tags

$$(K_{k_1} = v_{k_1,1} \vee \ldots \vee v_{k_1,n_1}) \wedge \ldots \wedge (K_{k_p} = v_{k_p,1} \vee \ldots \vee v_{k_p,n_p})$$

A syntactic pattern on a target term with the set K' and its values v' is defined in the same way.
Example:
The pattern below represents a word which analysis could not determine whether it is a common or a proper name (fig. 1 gives explanation for the tag sets used) :

$$(CAT = N) \wedge (SOUSN = NCOM \vee NPRO)$$

The admissible conditions recognized by the model deal with both the source and target sentences. A **well-formed condition** is when both source and target conditions are met in the **bi-sentence** on contiguous terms. Let $\Gamma S_1, \ldots, \Gamma S_n$ be a list of **conditions** for source terms and $\Gamma T_1, \ldots, \Gamma T_m$ for target terms. An **admissible condition** will be noted as follows :

$$\begin{cases} 1 : \Gamma S_1; \ldots; n : \Gamma S_n \\ 1 : \Gamma T_1; \ldots; m : \Gamma T_m \end{cases}$$

This condition will be **matched** in a bi-sentence if a contiguous list of terms from the source sentence respect each condition ΓS_i in the right order **and** if a list of contiguous terms from the target sentence respect each condition ΓC_j, also in the right order.
Example: The condition below would be matched on the pair *"un ciel bleu"-" blue sky"*, with the following definitions for the used POS tags.

$$\begin{cases} 1 : (CAT = DETERM) \wedge (SOUSD = ARTD \vee ARTI)\,; \\ 2 : (CAT = N) \wedge (SOUSN = NCOM)\,; \\ 3 : (CAT = ADJOINT) \wedge (SOUSA = ADNOM) \\ \\ a : (CATAng = JJ)\,; b : (CATAng = NN) \end{cases}$$

The **contiguity** hypothesis plays an important role in our method. The previous condition will not be matched on the pair *"un ciel très bleu"-"a very blue sky"*, that will be implemented in a larger pattern. So, the phrases concerned by the patterns:
(1) have an arbitrary length
(2) contains only contiguous words

2) Application of a segment: If a condition part is matched on a contiguous part of the bi-sentence, the application part provides a way of linking each term concerned by the condition. A segment can be applied if it does not violate a link already present in the bi-sentence. An edge is provided if a mapping is possible between an upper and a lower node, or a set of lower nodes. So, a correct alignment must result as the union of non-intersecting 'bicliques', that we assume to be a rather natural definition beyond which the notion of alignment would be meaningless.

Example:
An admissible segment to be applied on the pair: "*à la Cour*", "*at Court*" could be written as such:

$$\begin{cases} 1 : (CAT = PREP) \wedge (CATPREPSIMPLE = A) ; \\ 2 : (CAT = DETERM) \wedge (SOUSD = ARTD) ; \\ 3 : (CAT = N) \wedge (SOUSN = NCOM \vee NPRO) ; \\ \\ a : (CATAng = IN) ; b : (CATAng = NP) \end{cases}$$

$$\Longrightarrow a(1); b(2,3);$$

B. Segment memory

Alignments of quality were needed for this experimental study as well as a way to create and modify them easily. An online graphical interface was developped to accelerate the acquisition of examples[1]. The alignment tool(still under development), allows to constitute single-handedly, or in collaboration with partners, an aligned corpus of quality as it is necessary for a reference corpus. Segments are learned from hand-aligned or semi-automatically aligned data. The syntactic structure of the parsed source sentence allows one to cut out the total alignment into several relevant aligned bi-phrases producing valid segments. For instance, Figures 2 and 3 show a constituent tree for a source sentence, its leaves being source words, and how the target sentence words could be aligned according to a subtree division of the basic syntactic tree. The segment obtained from the first GN in figure 2 chunk is:

$$\begin{cases} 1 : (CAT = DETERM) \wedge (SOUSD = DEM) ; \\ 2 : (CAT = N) \wedge (SOUSN = NCOM) ; \\ 3 : (CAT = ADJOINT) \wedge (SOUSA = ADNOM) \\ \\ a : (CATAng = DT) ; b : (CATAng = JJ) ; c : (CATAng = NN) \end{cases}$$

$$\Longrightarrow a(1); b(2); c(3)$$

The segments consider only POS tags: Lexical resources are never used. This approach tends to rapidly create general segments applicable in many cases. One could object that the contiguity hypothesis weakens the segments generality, making it difficult to represent phenomena such as the French negation "*ne...pas*", but the segments shape has a precise algorithmic purpose and non-contiguous linguistic entities can be covered not by one, but by many segments. For instance, to be fully taken into account, "*ne...pas*" should be handled by segments such as : "ne [Verb] pas", "ne [Verb] [Adverb] pas", and so on. A segment can include several phrases when divergence is too high. In the next part, we comment

[1] *www.alignit.fr*

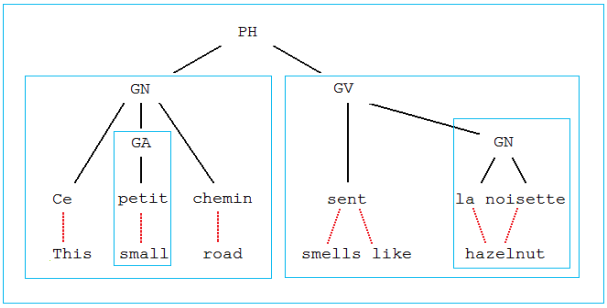

Fig. 2. Selecting Sub-sentential Segments

Ce	: CAT=DETERM;SOUSR=DEM;
petit	: CAT=ADJOINT; SOUSA=ADNOM;
chemin	: CAT=N; SOUSN=NCOM;
sent	: CAT=V;
la	: CAT=DETERM; SOUSD=ARTD;
noisette	: CAT=N; SOUSN=NCOM;
This	: CATAng=DT;
small	: CATAng=JJ;
road	: CATAng=NN;
smells	: CATAng=VBS;
like	: CATAng=VB;
hazelnut	: CATAng=NN;

Fig. 3. Labels From French and English Trees Leaves: Note That the tagging for 'like' is wrong!!

two different segmentation paradigms we used, with different recombining treatments.

C. Combining segments

Two ways of combining segments are here presented: Each of them depends on a different choice in segmentation. The complexity of proposed resolutions is also an asset. From a new bi-sentence to be aligned at the sub-sentential level, one has to collect compatible candidates in the partial segments database (the saving memory). *Combining them to obtain an optimum alignment consists in selecting a maximum covering set of compatible segments.* Many maximum-independent-set issues are known to be NP-hard (such as many alignment problems [8]). As an example, if we were to extend our set of segments to only complete sets of connected nodes, which is the most general possible shape, the combining process would lead to the NP-hard biclique decomposition problem [10].

1) Contiguous segments : The segments contiguous shape results from a need to use a lighter recombining process in a graph approach: Each segment is a weighted node, which weight is the coverage of the segment. There is an edge between two nodes if the associated segments are compatible. *Building an optimal alignment is seen as finding a maximum-weight clique in the compatibility graph.* Actually, even in this "simplified" framework, we encounter algorithmic difficulties which we shall detail here.

Two independent segments could be either *"crossing"* or *"following"* (cf fig.4). Looking for a maximal independent set of segments which are pairwise either crossing or following is still an NP-hard issue known as the "maximum weighted independent set of axis parallel rectangles" [1]. The segments

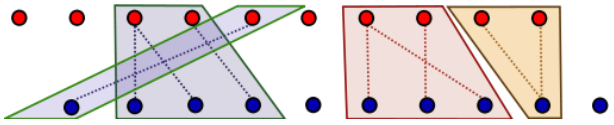

Fig. 4. Patterns presenting *a crossing* and *following* configuration

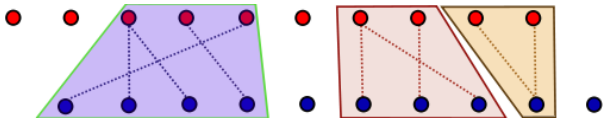

Fig. 5. Segments including *a crossing* configuration

shape used in this approach can lead to different issues: Apart from their linguistic justification, their impact on the recombination effort is substantial. As a preliminary experiment, we wanted to evaluate a heuristic recombining process over this approach (with the contiguous segments described before). A maximum covering set of *following patterns* (fig 4) could be built in a polynomial time [20]. Doing so, the solution proposed by the system will not present any crossing link. It means that every needed crossing configuration in the final alignment would lead, with this method, to a choice between links, thus generating holes, and then errors. An alternative consists in a pre-treatment among the set of compatible segments to deal with crossing links. When two segments present crossing links and form a larger contiguous pattern, as in figure 4, they are seen as a whole new segment and added to the database. This pre-treatment, when the number of contiguous-crossing configurations is reasonable, can lead to an exact recombining procedure, but most of the time, the combinatorial effort of dealing with crossing configurations is too heavy and one has to use a threshold or at least a filter. In this last case, the most common one on long sentences, the method is an approximation.

2) Contiguous segments capturing crossing configurations: As an alternative and a second experiment, we proposed another segmentation process, extending the first one and leading to a tractable compatibility resolution algorithm: We decided to capture the crossing configuration during the segmentation process to avoid the combinatorial cost of dealing with them. Indeed, compatible segments will be in *following configuration*. One can observe the new segment shape in figure 5. Of course, segments will provide us with less generic features (especially with great divergence). Recombining alignment among patterns thus formed is known as the maximum independent set problem for trapezoid graphs. Light combinatorial algorithms exist to solve this problem ($O(n\ log\ n)$ where n is the number of segments in [9]). Unless divergent behavior is pre-treated with, for example, a word-to-word alignment, this method will tend to favor a left to right alignment which appears to give good results for the French-English pair.

IV. A Preliminary Evaluation

A first set of $67,941$ pairs of sentences has been extracted from a journalistic corpus. French sentences present an average length of 27 words and English ones 23. In this section, we call *"contiguous segments"*, patterns obtained from the first segmentation described in figure 4. "crossing segments" will refer to the second ones from the extended segmentation including crossing configurations (figure 5). We hand-aligned 100 bi-sentences as a first training. The system segmented each memorized pair into generalized patterns. We ran several experiments. First, we tested our recombining process by trying to align those same 100 bi-sentences: The total alignment was memorized, but we inhibited large patterns during the database mining phase, susceptible to align in one shot, in order to test the recombination among short patterns (knowing every needed pattern was effectively in the base). The idea was to evaluate the recombining process over the two segmentation stages. In the table summarizing results, W_{100} stands for the recombining process over the *contiguous segments* and X_{100} over the *crossing segments*. We consolidated this alignment by using a pre-treatment cognates detection: Short patterns can lead to syntactical ambiguities sometimes quite frustrating when aligning a proper noun with an omitted uppercase first letter, with a common noun. Cognates detection was based on a Levenshtein measure and we noted an average number of 4 cognate pairs per bi-sentence (e.g.: *"musharraf"-"moucharraf", "judges"-"juges", "unpopularity"-"impopularité"*,...). $W_{100}^{Cog.}$ and $X_{100}^{cog.}$ are the same experiments using cognates alignment as reinforcement. Then results are much better when cognates are used as anchor. No mistakes were found in the cognate detecting process. Results have shown that recombining experiments are quite successful with the *crossing segments*, since the process has been tailored for their needs. Capturing crossing links during segmentation, has engulfed the main liability of the alignment process, thus leaving to recombination a minimal effort. It amounts to searching a database of already saved patterns and looking for following configurations. Then, we tried to align 100 fresh bi-sentences which were not from the training set. This time, alignment was performed with cognates reinforcement. W_{100}^{new} and X_{100}^{new} designate the aligning experiment on the 101-nd to 200-nd bi-sentences based on the training over the first 100, respectively for the contiguous and the crossing segments. Of course, the amount of data is insufficient to draw strong conclusion or to give predictions for the future evolution of the system, but we observe that the lack of generic features we feared for the *crossing segments*, does not impede the recombining process to reach results which quality is equivalent to the experiment with the *contiguous segments*. The two methods lead to an identical F-score, but the second method seems to have a lesser recall, thus corrupting its performance. This can be explained by the fact that the recombining process maximizes the bi-sentence coverage with following positions segments. This tends to create holes when two followings are not adjacent. The first method, a heuristic, tried to maximize coverage among the segments in following or crossing configurations. When adding crossing segments in the process, the second method reduces recall.

In order to measure the alignment quality, we had then to hand-align this 100 bi-sentences which provided us with an enriched database, so we ran over the first experiment consisting in evaluating the recombining process over the 200 bi-sentences already aligned with and without cognates thus trying to observe improvement or, on the contrary, a degradation. There were no significant differences. These experiments are referred to as W_{200} and X_{200}, $W_{200}^{Cog.}$ and $X_{200}^{cog.}$ in the results table. As a comparison, we gave Giza++ model results (from French to English) on the same pieces of the corpus (trained on the $67,941$ bi-sentences with the IBM model 4) although the two systems are definitely different: The sizes of needed training corpora, information used, and theories are hard to compare. No additional heuristic was used, the results are here as a baseline reference. In the table below, "P" stands for "precision", "R" stands for "recall", and "F" for the classically used F-measure. Let us note the very high values of the F-measure for the "X" based experiments, except "Al", which is invariant.

	W_{100}	$W_{100}^{Cog.}$	W_{200}	$W_{200}^{cog.}$	W_{100}^{new}	IBM4
P.	84%	92%	85%	91%	77%	75%
R.	82%	86%	83%	88%	52%	60%
F.	0.83	0.89	0.84	0.89	0.62	0.67

	X_{100}	$X_{100}^{Cog.}$	X_{200}	$X_{200}^{cog.}$	X_{100}^{new}
P.	98.7%	99.5%	97.9%	98.9%	82.3%
R.	97.2%	97.7%	97.1%	97.8%	49.9%
F.	0.98	0.99	0.98	0.98	0.62

V. Conclusion

In this paper, we have described two example-based aligning methods that almost exclusively uses syntactic information during the different steps of the process (the use of cognates is the only recourse to lexical information). Deep syntactic analysis was used to separate and collect segments from examples provided by users. Then again, these segments from bi-sentences were generalized using POS-tags. Alignment was performed between segments recognized from a database filled with syntactic correspondences. Two databases were built, suitable for two different process using different segmentations: The first, producing *contiguous segments* was followed by a heuristic recombining process, while the second, providing (*crossing segments*), led to an exact solution. The constrained form of the two segmentation processes we used, played an important role in the recombining effort based on coverage maximization. The shape of the memorized segments seems to play an important role in the recombining process. In such an approach, a trade-off should be found between the segments genericness and their combinatorial weight: The exact process using "crossing patterns" showed an almost perfect recombination of information. It pointed out the difficulty of crossing configurations in the alignment process, which should be carefully studied as a future work. Also, different heuristic resolutions should be tested on the contiguous segments memory. This first evaluation showed promising results, while quite a good precision was reached after a light training on only a hundred bi-sentences. With sufficient amount of data, the evolution of quality matching in the database size could be measured, and a more precise difference between the two approaches would then be better observed.

References

[1] V. Bafna, B. Narayanan, and R Ravi. Nonoverlapping local alignments (weigted independant sets of axis parallel rectangles). *Discrete Applied Mathematics*, 71:41–53, 1996.

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

[3] Ralf D. Brown. Brown-adding linguistic knowledge to a lexical example-based translation system. *Proceedings of the 8th International Conference on Theoretical and Methodological Issues in Machine Translation*, 1999.

[4] Jacques Chauché. Un outil multidimentionnel de l'analyse du discours. In *Coling*, 1984.

[5] Colin Cherry and Dekang Lin. A probability model to improve word alignment. In *41th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 88–95, July 2003.

[6] Colin Cherry and Dekang Lin. Inversion transduction grammar for joint phrasal translation modeling. In *NAACL-HLT*, 2007.

[7] Lambros Cranias, Harris Papageorgiou, and Stelios Piperidis. A matching technique in example-based machine translation. In *COLING*, pages 100–104, 1994.

[8] John DeNero and Dan Klein. The complexity of phrase alignment problems. In *ACL (Short Papers)*, pages 25–28, 2008.

[9] S. Felsner, L. Mller, and L Wernisch. Trapezoid graphs and generalizations, geometry and algorithms. *Discrete Applied Mathematics*, 74:13–32, 1997.

[10] H Fleischner, E Mujuni, D Paulusma, and S Szeider. Covering graphs with few complete bipartite subgraphs. In *27th FSTTCS, volume 4855 of Lecture Notes un computer Science*, 2007.

[11] Fabrizio Gotti, Philipphe Langlais, Elliott Macklovitch, Didier Bourigault, Benoit Robichaud, and Claude Coulombe. 3gtm: A third-generation translation memory. In *3rd computational Linguistics in the North-East (CLiNE) Workshop*, 2005.

[12] Mary Hearne and Andy Way. Seeing the wood for the trees : Data-oriented translation. In *MT Summit IX*, pages 165–172, 2003.

[13] Arul Menezes and Stephen D. Richardson. A best-first alignment algorithm for automatic extraction of transfer mappings from bilingual corpora. *DDMR Workshop, ACL*, 2003.

[14] Makoto Nagao. A framework of a mechanical translation between japanese and english by analogy principle. *Artificial and Human Intelligence: Edited Review Papers Presented at the International NATO Symposium on Artificial and Human Intelligence*, pages 305–332, 1984.

[15] Sergei Nirenburg. Two approaches to matching in example-based machine translation. In *Proceedings of TMI'93*, 1993.

[16] Franz Josef Och and Hermann Ney. A systematic comparison of various statistical alignment models. *Computational Linguistics*, 29(1):19–51, 2003.

[17] Sylvia Ozdowska. *ALIBI, un système d'ALIgnement BIlingue base de règles*. PhD thesis, Université de Toulouse 2, 2006.

[18] Satoshi Sato and Makoto Nagao. Toward memory-based translation. In *COLING*, pages 247–252, 1990.

[19] Helmut Schmid. Probabilistic part-of-speech tagging using decision trees. In *International Conference on New Methods in Language Processing*, pages 44–49, 1994.

[20] Stéphane Vialette. On the computational complexity of 2-interval pattern matching problems. *Theor. Comput. Sci.*, 312(2-3):223–249, 2004.

[21] Kenji Yamada and Kevin Knight. A syntax-based translation model. In *39th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 523–530, July 2001.