Automatic Translation Management System for Legal Texts

Atefeh Farzindar
NLP Technologies Inc.
3333 Queen Mary Road, suite 543
Montréal, Québec, Canada, H3V 1A2
farzindar@nlptechnologies.ca

Abstract

This paper presents the automatic translation management system known as *TransLI* developed for machine translation of Court judgments from English to French and from French to English. NLP's solution addresses our clientele's volume and time management issues. *TransLI*'s statistical algorithm has been trained on a large corpus of legal documents as to significantly reduce production time and costs. We describe the importance of review and post-editing mechanisms for judicial translations as part of the SMT-integrated workflow. This system is already being used by several government departments in Canada.

1 Introduction

The practice of law involves developing arguments based on the reasoning of courts in previous instances with similar circumstances. Said reasoning, known as precedent, is invoked to justify a legal standpoint or discredit the opposition's arguments. Precedent can be reinforced or dismissed by subsequent court decisions, and is therefore the fundamental point of debate in the application of the law. Consequently, it is imperative for legal practitioners to be well versed in legal precedents relevant to their area of practice, and remain up to date in their knowledge of such. The weight attributed to this information, found mainly through the study of previous cases, makes the ability to access the decisions essential. These processes are largely manual, imposing lengthy production delays on the industry. Furthermore, lawyers dedicate a significant portion of their time and energy to reading and analyzing decisions for the purposes of client representation and developing legal documents. Access to information is done with relative ease, as it is available through a number of publications including those of the courts. However given the inordinate volume of existing jurisprudence, filtering through the data to retrieve relevant information is time consuming. In addition, Canada's official languages policy requires that information be available in both English and French. In that respect, the burden of granting access to bilingual information falls on the government which must manage a continuous cycle of document translation needs. Automatic translation and summarization help legal practitioners drastically reduce the required efforts associated with the retrieval of pertinent materials. Nonetheless, being that translation is done manually; the delay associated can extend to several months from the date of publication by the court.

NLP Technologies[1] is an enterprise devoted to the adaptation of advanced information technologies for processing large volumes of documents, such as the needs reflected in the legal industry.. Its main focus is to provide innovative solutions to lawyers and other legal practitioners, associations, governments and courts. NLP's DecisionExpress™ summarization technology (Chieze et al., 2008) is a weekly bulletin of the latest case law

[1] http://www.nlptechnologies.ca

rendered by federal and provincial courts, and is currently being used by the Federal courts of Canada since 2004. NLP Technologies in collaboration with researchers from RALI at Université de Montréal have developed TransLI to automatically translate judgments from the Canadian Courts (Gotii et al., 2008).

Canadian jurisprudence is predominantly produced in English; 75% of decisions are written in English and 25% in French. According to legislation, the Federal Courts are required to provide a translated version of decisions in order to grant access to information in both official languages.

Quebec counts more than 30 administrative tribunals, the majority of which invest considerable resources in order to appease their translation needs.

The legal industry is marked by continuous publishing and translation cycles, large volumes of digital content and growing demand to distribute multilingual information. The capacity to administer high volumes of translation requests without delay thereby becomes imperative.

Currently, a certified translation of a legal judgment may take up to several months to complete. Consequently, there is a significant lapse of time between the publication of a judgment in its original language and the availability of a translated version in the other official language.

Initially, the project aim was to allow the federal court of Canada, during the few months when the official translation is pending, to publish automatically translated judgments and summaries. Once the official translation would become available, the Court would replace the machine translations by the official ones. However, given the quality of the machine translation system, developed and trained specifically on the Federal Courts corpora, NLP was presented with opportunities which are currently being investigated: machine translations could be considered as first drafts for official translations that would only require revision prior to publication. This procedure would thus reduce the delay between the publication of the decision in the original language and its official translation and provide translation cost savings.

We evaluated the French and English output and performed a more detailed analysis of the modifications made to the translations by the evaluators in the context of a pilot study conducted in cooperation with the Federal Courts (for details of this evaluation please see Farzindar and Lapalme, 2009).

An interesting aspect of our findings in the legal domain was that the review and the post-editing processes of judicial translations are an important part of an SMT-integrated workflow. Reviewers with subject knowledge need to have direct access to the translation process in order to provide a feedback loop to the SMT training process (Marcu, 2008).

To our knowledge, this is the first attempt to build a large-scale translation system of complete judgments for eventual publication. We also present new ways of combining the translation memory and SMT with post-editing tools in our platform.

2 Translation Management System

NLP's translation management system is comprised of several elements designed to facilitate both the translation and revision process as well as document transfer platform. These steps ensure production efficiency within NLP and for our clients.
Figure 1 shows the process of our translation management System.

1. A court user connects to our secure document management system to submit content to be electronically translated in a supported formatted document (for example in MSWord or HTML format) for processing.
2. NLP's system will create a project, parse each file, perform sentence segmentation and apply the translation memory
3. Each remaining segment is then systematically processed using the SMT engine that was trained with a very large quantity of paired sentences from the related domain, which returns the result in the second language in a timely fashion. The SMT pro-

ceed to translate the sentence one by one using a probabilistic method and the corpus to output a file containing the resulting translated text in a very similar presentation format.

4. The system then automatically initiates the post-editing process: the machine translated file is then analyzed by NLP's expert team of legal reviewers who are notified as soon as a new project requires their attention.
5. All linguists can login to the online system and complete the translation: the post-editing team can apply the corrections and confirm publishing of the translated content. In addition to the large existing training corpus, the corrections applied by the post-editing team will also be used to re-train the machine translator (Simard et al., 2007).
6. Furthermore, the approved content is also retained for the enhancement of the translation memory.
7. Once the reviewers has completed their corrections, they submit the final document through the document management system which also notifies the users when a new translation has been completed.
8. The court users will also be able to view and download the final translation results and submit new documents for translation.

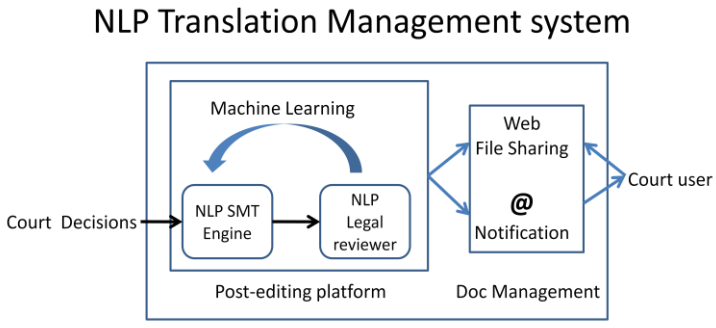

Figure 1: shows the process of our translation management System.

3 Statistical Machine Translation Engine

We have built a phrase-based statistical translation system, which takes as input the judgments and produces a translated file of the same judgment in Canada's other official language. The architecture of the system is shown in Figure 2.

The first phase (semantic analysis) consists in identifying various key elements within a decision, for instance the parties involved, the topics covered, the legislation referenced, whether the decision was in favour of the applicant, etc. This step also attempts to identify the thematic segments of a decision: **Introduction**, **Context**, **Reasoning** and **Conclusion**). During this phase, the original file is transformed into XML for internal use in order to produce DecisionExpress™ fact sheets and summaries. We extract the source text from these structured XML files in which sentence boundaries have already been identified. This is essential, since the translation engine works sentence by sentence.

The second phase translates the source sentences into the target language using SMT. The SMT module makes use of open source modules GIZA++ (Och and Ney, 2003) for creating the translation models and SRILM for the language models. We considered a few phrase-based translation engines such as Phramer (Olteanu et al, 2006), Moses (Koehn et al., 2007), Pharaoh (Koehn, 2004), Ramses (Patry et al., 2006) and Portage (Sadat et al., 2005). Moses was selected because we found it to be a state-of-the-art package with a convenient open source license for our testing purposes.

The last phase is devoted to rendering the translated decisions in the original format (for example HTML). Due to efficient record-keeping mechanisms, it is possible to merge the translation with the original XML file in order to yield a second XML file containing a bilingual version of each segment of text. This bilingual file can then be used to produce an HTML version of the translation, or for other types of processing, such as summarization.

Given that summaries of judgments produced by NLP Technologies are built by extracting the most

salient sentences from the original text, producing summaries in both languages should be as simple as selecting the translation of every sentence retained in the source-language summary.

Gotti et al. (2008) describe the development and the testing of the TransLI statistical machine translation system. The final configuration is a compromise between quality, ease of deployment and maintenance, and translation speed with the following features: a distance based reordering strategy, a tuning corpus based on recent decisions; a large training corpus and the integration of specialized lexicons.

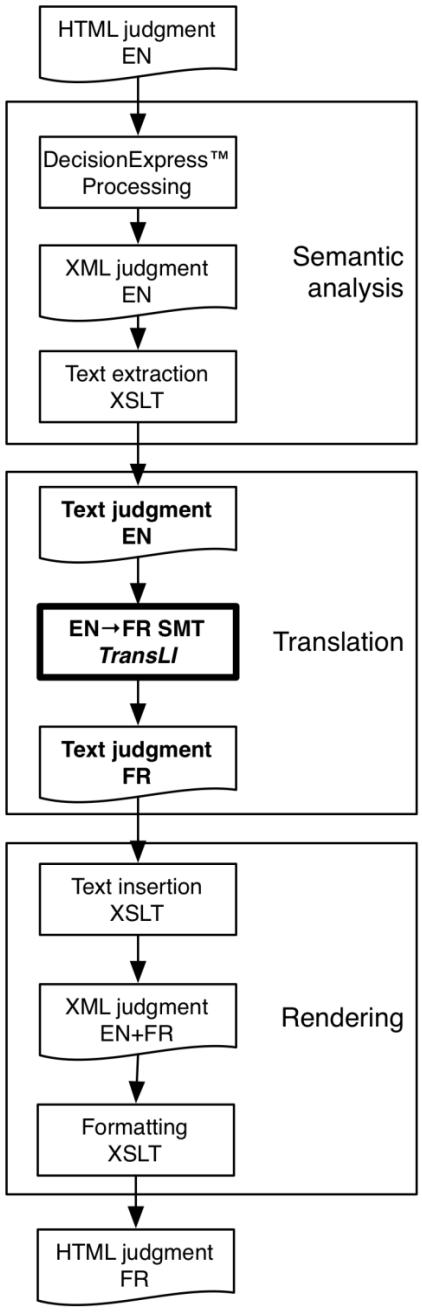

Fig 2: The translation pipeline translates an HTML court decision written in English into a French decision (also in HTML). A similar pipeline performs translations from French to English

Corpus name	# sent pairs	# en words	# fr words
principal	245k	6,510k	7,510k
train	244k	6,500k	7,500k
tune-1	300	8k	9k
test	1300	28k	33k
tune-recent	400	8k	10k
train-lexum	1,000k	22,340k	25,720k

Table 1: Corpora used for developing TransLI

Although these types of texts employ a specialized terminology and a specific cast of sentences, the availability of large amounts of high quality bilingual texts made it possible to develop a state-of-the-art SMT engine. These excellent results prompted us to perform a human evaluation also described in (Gotti et al. 2008) on 24 randomly selected sentences from our test set. This evaluation centered on the quality and fidelity of the produced translation, i.e. to what extent the SMT conveys all the semantic content of the original.

A key element in the success of an SMT system lies in the availability of large, high quality corpora. In the Canadian judicial domain, we are fortunate enough to have access to public web sites providing high quality translations for almost all judgments of the highest Canadian courts. For our work, we built a set of corpora, the characteristics of which are shown in Table 1.

`principal`: we downloaded 14,400 decisions in HTML from the Federal Court of Canada web site2 from which we extracted the text. Because many judgments did not have a translation or could not be parsed automatically with our tools due to inconsistent original formatting, we ignored them and we were left with 4500 valid judgment pairs. From these pairs, we extracted the sentences and aligned them to produce a bi-text of around 260,000 sentence pairs. A number of them had English citations in the French text and vice-versa. Once these cases were filtered out, we were left with 245,000 sentence pairs.

`train`: 99% of the sentences from `principal`, used to train the SMT system.

`tune-1`: 1% of principal used to adjust the parameters of the system. There is no overlap with `train`.

`test`: 13 recent decisions that were published after the decisions occurring in principal. This better simulates the application context for our system, which will be used for translating recent decisions.
`tune-recent`: 6 recent decisions that were published after the decisions in `principal`.

`train-lexum`: Since RALI has considerable experience in dealing with judicial texts in collaboration with the Faculty of Law at the Université de Montréal in the context of the TransSearch[3] system, we decided to add 750,000 bilingual sentence pairs from our existing bilingual text database. These sentences are taken from decisions by the Supreme Court, the Federal Court, the Tax Court and the Federal Court of Appeal of Canada.

As for language quality, we asked three evaluators to assign each of the 24 passages a score: 1 (unacceptable), 2 (bad), 3 (fair), and 4 (perfect), according to whether they found it to be correct and readable in the target language, independently of the source language. This would correspond to the case where a non-French speaking person wanted to consult an English translation of a French text. Our evaluators did not know which translations had been produced by a human or which were produced by a machine.

The same three evaluators were given groups of two or three sentences containing the source French text and the English translation produced either by or by a human translator (the reference text). The evaluators were asked to modify them in order to make them good enough for publication. Overall they took an average of 27 minutes to revise 8 texts (475 words), which corresponds to 1070 words/hour. That would sum up to 8000 words per day compared to the average of about 6000 often referenced in the industry for re-

[2] decisions.fct-cf.gc.ca/en/index.html

[3] www.tsrali.com

vision (4 times the productivity of 1500 words translated per day per translator).

4 Post editing platform

In order to upgrade the use of our services, we are also developing new interfaces to make the reviewer's job easier by providing new tools to review documents. The new interface will also provide statistics on the reviewing process such as the time and quantity of operations required by the reviewers to analyze a translation. With these statistics we will be able to pinpoint the most extensive parts of the process that can be optimized to further facilitate the reviewers' job. The cost savings are proportionate to the reduction of revision time. Translation workflow management is another benefit of this platform.

5 Evaluation of TransLI

Although still not of publishable quality, the translations produced by the system that we developed in this project can be readily used for human revision, with promising productivity gains. Following those encouraging results on a small sample of the sentences, we conducted a pilot study with the Federal Courts of Canada in which we translated a number of complete judgments from French to English and from English to French. We set out detailed evaluations of the revision process that was performed on a randomly selected set of 10 decisions (6 from French to English and 4 from English to French).

We also describe how we evaluate the quality of our current automatic judgment translations and the effort needed to revise them so that they can be published. As the summarization system of *NLP Technologies* already divides a judgment into four main thematic segments: **Introduction**, **Context**, **Reasoning** and **Conclusion**, we describe the evaluation using these divisions, in order to give an overview of the source text, of the raw SMT translation produced and of the revised output judged acceptable for publication.

The thematic segmentation is based on specific knowledge of the legal field. According to our analysis, legal texts have a thematic structure independent of the category of the judgment (Farzindar and Lapalme, 2004) Textual units dealing with the same subject form a thematic segment set. In this context, we distinguish four themes, which divide the legal decisions into thematic segments, based on the work of judge Mailhot (Mailhot ,1998):

Introduction describes the situation before the court and answers these questions: who did what to whom?

Context explains the facts in chronological order: it describes the story including the facts and events related to the parties and it presents findings of credibility related to the disputed facts.

Reasoning describes the judge's comments, the finding of facts, and the application of the law in light of the facts. This section of the judgment is the most important part for legal experts because it presents the solution to the problem between the parties and leads the judgment to a conclusion.

Conclusion expresses the disposition, which is the final part of a decision containing the information about what is decided by the court.

In order to evaluate the results of the automatic translation, we computed two automatic measures over the space-separated tokens of a sentence. A token is thus a word plus any accompanying punctuation or symbols. A token can also be any sequence of contiguous non-space characters:

Edit distance: the number of tokens that differ in the source and revised text as computed by the classical Levenshtein distance algorithm (Levenshtein, 1966).

Number of operations: the number of consecutive insertion, deletion and replacement operations to transform the source into the revised text. For ex-

ample, replacing 5 consecutive words would count as 5 in the edit distance but for only one operation. This measure approximates the number of cut and paste operations needed to revise an SMT translation.

The results we obtained show that for both translation directions, the number of editing operations is roughly equivalent to the number of tokens in each division in terms of ratio. The results also show that the global proportion of differences is similar for both translation directions. The results are slightly better on the French to English direction, which is expected due to the complexity of the French language (thus further complicating the machine translation process. When we compare the different themes, we see that the Introduction and Conclusion themes require significantly less (50% to 80%) editing than the Context or the Reasoning themes. The effort required for the Introduction is very similar for the Conclusion which is also true when we compare the Context and the Reasoning and this in both translation directions. These differences are partly explained by the type of text used in these themes. In the legal field, the sentences used for the Introduction and Conclusion of the judgments are often the same expressions while the Context and Reasoning contain more sentences which are seldom seen in multiple judgments. Sentences from the Context that explain the litigation events tend to vary more.

6 Conclusion

The volume of legal content is growing rapidly. This is especially problematic in Canada because documents are created in two languages and in different formats. As a result, the amount of data that must be translated in without delay has grown tremendously, making it difficult to translate and manage. Legal organizations need solutions that enable them to handle quickly a high volume of translations. Our goal was to study the ability to train translation systems on a specific domain or subject area like the legal field so as to radically increase translation accuracy. An interesting aspect of our findings is that review and post-editing of judicial translations are an important part of an SMT-integrated work flow. Reviewers with subject knowledge need to have direct access to the translation process in order to provide a feedback loop to the SMT training process.

In this work we have shown the components of a translation management system with the integration of a machine translation engine into a generic multi-lingual post editing content management system.

Ours is one of the first SMT systems that has been developed and evaluated specifically for judicial texts. We've demonstrated that an SMT engine trained on an appropriate corpus can produce a cost-effective revisable text.

7 Future Work

We will therefore continue further investigation into an optimization of the post-editing and reviewing process, specifically with a focus on quantifying the distance, measured in number of operations and edits, to arrive at a fully acceptable translation. Currently, we are optimizing our SMT's learning process, . Once the Engine has been trained, we can use it to translate the new documents. We are working on a new post-editing platform to allow a dynamic training and learning process for the engine. Most of the regular process will be the same but when our team of reviewer will submit their corrections, we will use these corrections to re-train the Engine which will eventually minimize the future post editing effort. We also intend to adapt the system to various domains.

Acknowledgments

We thank the Precarn Program for partially funding this work. We also thank Guy Lapalme, professor at the computer science department of Université de Montréal.

References

Atefeh Farzindar, Guy Lapalme. TRANSLI: *A System for Statistical Machine Translation of Legal Information.* Canadian AI09 The 22th Canadian Conference on Artificial Intelligence. *Kelowna, British Columbia, Canada, May 2009.*

Daniel Marcu. Language Translation Solutions for Community Content. *AMTA'2008 The Eighth Con-*

ference of the Association for Machine Translation in the Americas, Waikiki, Hawai'i, October 2008.

Michel Simard, Cyril Goutte, Pierre Isabelle. Statistical Phrase-based Post-editing. *Human Language Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT 2007).* Rochester, New York, USA. April 22-27, 2007. NRC 49288.

E. Chieze, A. Farzindar, and G. Lapalme, 2008, "Automatic summarization and information extraction from Canadian immigration decisions". In *Proceedings of the Semantic Processing of Legal Texts Workshop*, LREC 2008.

A. Farzindar and G. Lapalme. LetSum, an automatic Legal Text Summarizing system. In Thomas F. Gordon (editors), *Legal Knowledge and Information Systems, Jurix 2004: the Sevententh Annual Conference*, p. 11-18, IOS Press, Berlin, Dec 2004.

A. Farzindar, G. Lapalme and J.-P. Desclés, 2004, Résumé de textes juridiques par identification de leur structure thématique. In *Traitement automatique de la langue (TAL)*, vol. 45, number 1, p. 39-64.

F. Gotti, A. Farzindar, G. Lapalme and E. Macklovitch. Automatic Translation of Court Judgments. In *AMTA'2008 The Eighth Conference of the Association for Machine Translation in the Americas*, p. pp 1-10, Waikiki, Hawai'i, Oct 2008.

P. Koehn, 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, E. Herbst, 2007, Moses: Open Source Toolkit for Statistical Machine Translation In *Annual Meeting of the Association for Computational Linguistics (ACL)*, demonstration session, Prague, Czech Republic.

P. Koehn and J. Shroeder, 2007, Experiments in Domain Adaptation for Statistical Machine Translation. In *Proceedings of the 2nd Workshop on SMT*, Prague, Czech Republic.

P. Koehn, 2004, Pharaoh: A beam search decoder for phrase-based statistical machine translation models. In *Proceedings of AMTA 2004*, p. 115–124.

V. I. Levenshtein, Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady* 10 (1966):707–710.

M. Olteanu, C. Davis, I. Volosen and D. Moldovan, 2006, Phramer, An Open Source Statistical Phrase-Based Translator. In *Proceedings of the Workshop on Statistical Machine Translation*, p. 146–149.

L. Mailhot. Decisions, Decisions: a handbook for judicial writing. Editions Yvon Blais, Québec, Canada, 1998.

F. J. Och, H. Ney. A Systematic Comparison of Various Statistical Alignment Models, *Computational Linguistics*, volume 29, number 1, pp. 19-51 March 2003.

A. Patry, F. Gotti and P. Langlais, 2006, Mood at work: Ramses versus Pharaoh. In Workshop *on Statistical Machine Translation*, HLT-NAACL, New-York, USA.

F. Sadat, Johnson, H., Agbago, A., Foster, G., Kuhn, R., Martin, J., Tikuisis, A., Portage: A Phrase-based Machine Translation System, *ACL 2005 Workshop on Building and Using Parallel Texts: Data-Driven Machine Translation and Beyond.* Ann Arbor, Michigan, USA. June 29-30, 2005. pp. 133-136.