Identification and Translation of Significant Patterns for Cross-Domain SMT Applications

Han-Bin Chen[1], **Hen-Hsen Huang**[1], **Jengwei Tjiu**[2], **Ching-Ting Tan**[2], and **Hsin-Hsi Chen**[1]
[1] Department of Computer Science and Information Engineering
National Taiwan University
[2] National Taiwan University Hospital
Taipei, Taiwan
{hbchen, hhhuang}@nlg.csie.ntu.edu.tw; {p99748036, tanct5222, hhchen}@ntu.edu.tw

Abstract

Adaptation of statistical machine translation (SMT) systems from generic to specific domains is challenging due to the lack of training data. In this paper we propose a framework for domain adaptation by exploiting a large monolingual in-domain corpus. We identify the significant patterns to capture the domain specific writing styles. The patterns are then translated with the involvements of domain experts. The major issue of our framework is to reduce the cost of the experts and better allocate their efforts. The experimental results show the proposed methods are effective, in terms of the significance and diversity of the patterns. The approaches to integrate the mined patterns into background SMT are also discussed.

1 Introduction

The language usages in different domains vary significantly. The differences come from different linguistic aspects such as lexical choice, writing style, and so on. These varieties affect the term distribution in corpora used for training, and thus greatly change the statistical model for a specific domain. A straightforward way to deal with the domain-specific problem is to train the model using in-domain data. However, such a domain-dependent corpus is not always available. This problem is much more serious in cross-language cross-domain applications, in particular, for those applications in highly specific domains such as biochemistry and medical science.

Consider a typical example in hospitals. Doctors summarize the statuses of patients after they leave hospitals. On the one hand, medical summary keeps present illness of a patient in a hospital, which makes doctors quickly understand the past history of an incoming patient. On the other hand, patients have rights to know the treatments during their stay in a hospital by reading these documents. However, a medical summary is always written in English in some countries where the official languages are not English. That becomes barriers for patents to know what treatments have been done. Machine translation (MT) systems, which translate documents in one language into another, may play important roles in medical summary translation.

Many translation models such as phrase-based model, syntax-based model and example-based model have been proposed in the past. In addition, some typical search engine portals also provide translation services like Google Translate and Yahoo Babelfish. However, these MT models or services are not suitable for medical summary translation because of the specific medical domains. Consider an example. Sentences like "Port-A implantation was performed on 2009/10/9" are frequently used in medical summaries. Google Translate reports its Chinese translation as "港口是一個植入上執行 2009 年 10 月 9 日". Compared to the Chinese reference translation "在 2009 年 10 月 9 日執行人工血管置放術", there are several translation and reordering errors.

In this paper, we develop an English-Chinese medical summary translation system to tackle this problem. In an SMT model, an English-Chinese parallel corpus is indispensable for training the translation model (TM). However, only English medical summary corpus is available in this domain. Here, we first develop a general English-Chinese SMT system with Moses toolkit (Koehn et al., 2007) and a general English-Chinese parallel corpus. Next, we adapt this SMT system with patterns learned from an English medical summary corpus. The problem is that these patterns are still monolingual. It is necessary to have domain expert involved in setting up bilingual patterns.

The cost of domain experts is the major concerns. Therefore, to identify significant patterns from a monolingual in-domain corpus, to find their coverage relationships, to decide which patterns should be translated by experts, and to introduce these patterns to the general MT system are research issues in this paper. Such a methodology can not only be contributed to design a medical summary translation system, but also be applied to other domains with similar resource poor problems.

This paper is organized as follows. Section 2 reviews the related works based on the types of available in-domain corpora and summarizes our major contributions. Section 3 gives an overview of our methodology. Sections 4, 5 and 6 describe pattern identification, translation and integration in detail, respectively. Section 7 evaluates the performance of these algorithms. Finally, Section 8 concludes the remarks.

2 Related Work

Domain adaptation in MT dates back to studies of sublanguage properties. Kittredge et al. (1985) described the importance of sublanguage analysis in automatic translation. In their work, sentence patterns were expressed in terms of words and domain word classes. Sager et al. (1989) adapted the Linguistic String Project - Medical Language Processor (LSP-MLP) from English to other languages. The LSP-MLP analyzes and organizes clinical data into structured information by using the established sublanguage linguistic patterns.

In recent years, training SMT system from large bilingual data has been a common practice. The parallel corpus used to train an MT system mainly comes from fixed domains such as parliamentary and news articles. The bilingual resources for a specific language pair or a specific domain usually come in small size, even unavailable. One of the challenging issues in a cross-domain MT application is to realize an in-domain MT model in such a resource-poor environment.

Depending on what kinds of in-domain resources are at hand, various adaptation techniques have been proposed. Foster and Kuhn (2007) proposed a mixture-model approach to deal with the case where bilingual in-domain text is available but in a relatively small size. A training corpus was divided into several components to train several models. Each model was weighted to estimate the similarity between components and in-domain development data. Based on this work, Foster et al. (2010) incorporated instance weighting that learned the weights of phrase pairs to capture the degree of relevance to the target domain. Similarly, a mixture-model approach was also applied in word-alignment task (Civera and Juan, 2007). In their work, domain related parameters were added in standard HMM training to derive an alignment model sensitive to the topic of each sentence.

In some applications, bilingual in-domain corpus is unavailable while monolingual one (either source or target side) is relatively easy to acquire. Zhao et al. (2004) combined the baseline language model (LM) with the in-domain LM, which was trained by retrieving documents from large text collections using query models. Besides LM, Bertoldi and Federico (2009) generated a synthetic bilingual corpus from a monolingual one to train a domain-specific TM.

Our work is close to the monolingual scenario. Provided with a monolingual in-domain corpus, we adapt our background MT into the one suitable for translating medical summaries. There are some major differences among our work and those proposed previously. First, we identify and translate significant patterns from in-domain corpus and introduce them into our SMT system. Instead, the related works exploited the entire in-domain training data to adapt the existing LM or TM by model mixture and parameter tuning. Second, the significant patterns are translated by domain experts to deal with the large domain difference between background training corpus and medical summaries. To reduce the cost of experts, filtering, clustering and ranking the patterns are the major issues.

3 Framework

Translating articles in a specific domain using a general MT system is challenging. In this paper, we aim to identify significant patterns embedded in domain-dependent source documents to deal with the translation problems. In the application of a medical domain, medical summaries are in special written styles and are usually short. In this study, our experimental dataset is selected from National Taiwan University Hospital (NTUH). The average length of a sentence is 10 words. In contrast, the background general domain corpus is Hong Kong Parallel Text purchased from LDC. The average sentence length is 29 words.

A number of patterns repeat in medical summaries. The sentence "Port-A implantation was performed on 2009/10/9" contains a frequent medical pattern shown as follows. It states a kind of surgery was performed on some particular date.

SURGERY was performed on **DATE** (1)

Here, SURGERY represents a class of medical terms denoting surgeries and DATE is a class of date expressions. Similarly, "paracentesis was performed on 2010-01-08" is also an instance of this pattern.

In this paper, we present a framework for pattern extraction and translation to tackle the translation problems in specific domains. Given a large collection of medical summaries *C*, we extract a set of significant patterns *P*, translate them and form bilingual patterns *B*. During the runtime translation, we apply *B* to translate each source language pattern occurring in an input medical summary.

The following shows the overall framework proposed in this paper. The details of each stage will be discussed in Sections 4-6, respectively.

(a) Pattern Identification

(1) Named Entity Classification

Recognize named entities including medical terms, hospital names, date/time expressions, etc., transform them into the corresponding classes and derive a new corpus *C*'.

(2) Highly Frequent Pattern Identification

Employ n-gram models (n=2~5) to extract a set of highly frequent patterns in *C*'.

(3) Linguistic Completeness Verification

For each pattern, randomly sample *m* sentences having this pattern, parse these sample sentences, and keep the pattern if there exists at least one parsing sub-tree for it.

(4) Pattern Coverage Finding and Filtering

For the remaining patterns, check coverage relationship among higher order patterns and lower order patterns, and remove those lower patterns being covered.

(5) Pattern Clustering

Cluster the remaining patterns of the same order, and select the representative patterns from each cluster for pattern translation.

(b) Pattern Translation

(1) Translation by Google

Send each pattern not consisting of any medical class to Google Translate.

(2) Translation by Domain Experts

Translate those patterns consisting of at least one medical class by domain experts (i.e., doctors).

(3) Review by Domain Experts

All the translated patterns are reviewed by doctors and revised if necessary.

(c) Pattern Integration

The reviewed bilingual patterns are integrated into a general SMT system.

4 Pattern Identification

Provided with a large in-domain monolingual corpus, we aim to (1) extract the significant patterns to capture the domain specific writing styles as much as possible; (2) refine and reduce the size of the pattern set to minimize the cost of expert involvements in reviewing and translating the patterns. Stages (a.1)-(a.3) in our framework deal with the first issue, i.e., to extract patterns of high qualities. Stages (a.4)-(a.5) touch on the second issue, i.e., to select patterns of high diversities.

4.1 Named Entity Classification

As illustrated, a pattern may include classes representing the general concept of a group of terms. Thus, given an in-domain plain text, identifying domain specific terms and classifying them into suitable classes is the first step toward the extraction of significant patterns.

In this study, we employ n-grams to represent patterns. Length of n-grams shows some limitation on their usages in translation. On the one hand, lower order n-grams only capture local cues in a restricted scope only. On the other hand, we need more training data to achieve reliable statistics of higher order n-grams. Recognizing word strings of specific semantics and replacing them with classes is useful to resolve the locality issue of n-grams. Thus, patterns are in terms of combinations of words and classes rather than words only. That will enlarge the scopes of patterns in some senses.

Named entities such as medical terms, hospital names, date/time expressions, etc. are our targets. Recognition of traditional named entities like organization names and date/time expressions has been discussed intensively before, so that they are neglected in this paper. Here, we focus on the classification of medical terms only.

In our application, we are provided with lists of terms frequently used by several hospital departments at NTUH. Terms of a variety of subjects, such as diagnosis, surgery, pharmacy, and laboratory medicine are used in these departments. They thus form the basic medical classes in our patterns: DIAGNOSIS, SURGERY, DRUG and TEST.

In addition, we incorporate larger amount of public resources in the Internet to further extend our knowledge base for medical term classification. While online medical dictionaries are free to consult, they are mainly built for explaining the meaning of medical terms, without explicit information of general concept that a medical term represents.

Here, we apply the resources from the Unified Medical Language System (UMLS) maintained by National Library of Medicine. The UMLS covers a wide range of terms in medical domain, and relations between these medical terms. Among these resources, the Metathesaurus organizes medical terms into groups of concepts. Moreover, each concept is assigned at least one Semantic Type. Semantic Types provide categorization of concepts at a more general level, and therefore are well-suited to be incorporated.

Merging existing ontologies is another research issue. In this paper we propose a mapping from 133 Semantic Types of UMLS to our 4 medical classes. To identify and classify medical terms in our domain specific corpus, we examine each sentence from left to right and adopt a longest-first strategy to replace medical terms with classes. In this way, a set of medical summaries are transformed into a new corpus.

4.2 Highly Frequent Pattern Identification

To address the domain adaptation problem in MT, we extract patterns from an in-domain corpus to capture domain specific writing styles. These patterns are translated and will be applied in the run-time translation. Accordingly, we prefer the format of patterns that is easy to be integrated into an SMT system for our target application.

The phrase-based model (Koehn et al., 2003; Koehn, 2004) is one of the state of the art TMs, in terms of both accuracy and speed. The phrase-based system translates source phrases into target ones with phrase table, which consists of bilingual phrases and feature functions. Since a phrase (i.e., a string of consecutive words) is served as the basic unit of translation, integrating n-gram based patterns into the background phrase-based SMT system is a natural choice.

We enumerate all n-grams from sentences of our in-domain corpus that contains words and medical classes. In this way, two kinds of patterns are extracted: (1) **class patterns** that contain at least one medical class and (2) **lexical patterns** that contain only words. Note that both patterns are easy to be integrated into a phrase-based SMT system, by either embedding them into the phrase table, or starting from the partial hypothesis where medical patterns have already been translated ahead.

4.3 Linguistic Completeness Verification

The main role of domain experts in our framework is to translate patterns extracted by our algorithms. This includes reviewing the patterns, neglecting insignificant ones, and translating the patterns considered important. However, more than 7M distinct patterns were extracted from medical summaries consisting of 1.8M sentences. It is infeasible to judge the significances through this enormous number of patterns by doctors. Therefore, filtering the patterns to an acceptable size is necessary before the expert involvements.

The linguistic meaningfulness of patterns is proposed to judge their significance. For example, "SURGERY was performed" is a linguistic constituent, while "SURGERY was performed on" is not complete. Accordingly, we filter out patterns that do not meet the requirements of complete linguistic constituents. A parser is adopted to determine the linguistic completeness of patterns.

A cross-domain issue arises when applying a general purpose parser to a domain specific corpus, because the parser built from a general domain training set may suffer from parsing the text with domain specific terms, such as diagnoses and drug names, especially when a term spans across multiple words. Here we take advantage of named entity classification introduced in Section 4.1. For each named entity in a sentence, we replace it with a common word in favor of our general purpose parser. For instance, we replace a complicated diagnosis "primary biliary cirrhosis" with the simpler one "disease". In this way, we reduce not only the OOV words, but also the length of sentences, and thereby facilitate the parsing procedure.

For each extracted pattern, we select *m* distinct sentences in which it occurs. These sentences are then analyzed by a parser and *m* parsing trees are produced. The pattern is considered as a significant candidate, if it is a syntactic constituent in any one of these parsing trees. In this paper we apply Stanford Parser (Klein and Manning, 2003) and set *m* to 10 in consideration of the parsing speed.

4.4 Pattern Coverage Finding and Filtering

The involvement of domain experts often guarantees the quality of annotation, but much higher cost is introduced at the same time. In this paper we try to further reduce the efforts made by doctors in translating the patterns, while keeping the diversities of the translated patterns to cover the in-domain writing styles as much as possible.

A higher order pattern A may be composed of two lower order patterns B and C. We call A covers B and C if all of them are linguistically complete. Consider an example. Pattern (1) in Section 3 is a concatenation of patterns "SURGERY was performed" and "on DATE". After pattern (1) is translated, we can derive the translations of their lower order composing components without translations by experts. By this coverage relation, we keep only pattern (1) and omit its trigram and bigram components. In our experiments, four kinds of relations are defined for 5-grams including "1+4", "2+3", "3+2" and "4+1" shown in Table 1. Translating the higher order patterns not only extends the translations of its components, but also gives the correct ordering of their combination. Thus, keeping the covering patterns and ruling out the covered ones reduce the size of patterns and preserve their integrity at the same time.

4.5 Pattern Clustering

Pattern clustering partitions a set of patterns into subgroups. This process reduces the cost of expert involvements in the pattern translation further. Given a cluster similar patterns, translating the most representative pattern may imply the translations of the others in the same cluster. An example of a cluster of similar patterns is illustrated below:

Coverage Relation	Examples
1+4	she is admitted for **SURGERY**
	she
	is admitted for **SURGERY**
2+3	Lab data showed no **DIAGNOSIS**
	Lab data
	showed no **DIAGNOSIS**
3+2	**TEST** on **DATE** showed **DIAGNOSIS**
	TEST on **DATE**
	showed **DIAGNOSIS**
4+1	Past Surgical History : **SURGERY**
	Past Surgical History :
	SURGERY

Table 1. Four kinds of coverage relations for 5-grams.

he received **SURGERY** on **DATE**
he received **TEST** on **DATE**
he underwent **SURGERY** on **DATE**
he underwent **TEST** on **DATE**

If the first pattern is translated by an expert, the translations of the others are easy to be inferred without doctors' help. Consequently, we reduce the cost of the experts from translating similar patterns, and thus enrich the diversity of their efforts.

In clustering, we define the similarity between two n-gram patterns to be the number of identical words in identical positions. Two n-grams are placed into the same cluster if their similarity is not less than n-1. Single-link clustering is adopted.

To achieve the diversity of patterns, we present them in a round-robin style among the groups generated by the clustering algorithm. Due to the large number of groups and the limited human resources, we present the patterns in a specific order by measuring the inter-group and intra-group scores. On the one hand, groups are ranked by sum of frequencies of their patterns. On the other hand, patterns are ranked by their frequencies in each group. In this manner, we focus on translating the most significant patterns among the groups.

5 Pattern Translation

This section introduces translation resources for building bilingual significant patterns. Domain experts are involved in translating class patterns, while lexical patterns are translated by free online translator first, and then corrected by doctors.

5.1 Translation by Google

In Stage (a.1) of our framework, domain specific terms are identified and transformed into medical classes. As a result, the lexical patterns extracted from the transformed corpus contain only common words, and can be translated by MT systems without the OOV problem.

We use Google Translate to translate the lexical patterns. Then human experts review and correct these translations. Building the bilingual patterns from the existing MT system can save much more time from the experts, compared to starting from scratch with only monolingual patterns.

5.2 Translation by Domain Experts

We deploy our experts to translate the class patterns, which contain medical classes and require in-domain knowledge from doctors. The translation of pattern (1) is shown below as an example.

Source: **SURGERY** was performed on **DATE**
Target: 在 **DATE** 施行 **SURGERY**

We design a Web UI for the experts. To focus on the translation quality, we make efforts on the friendly interface to reduce editing steps of translations, and to help the experts understand the meanings of the patterns. For example, the lexical part of the target language pattern ("在" and "施行") are edited by the doctors. On the other hand, the class part ("DATE" and "SURGERY") is output by mouse clicks on the corresponding classes of source language pattern to save editing time. Some patterns are relatively hard to understand and translate, and we present the experts with several instances of the patterns. For each source pattern in our UI, we give up to 10 sample sentences where the pattern occurs, and highlight the pattern.

5.3 Review by Domain Experts

For lexical patterns translated by Google Translate, the translation quality may be sabotaged due to the domain specific usages. For example, word sense disambiguation (WSD) problem often causes translation errors by such a general purpose translation system, as illustrated in Table 2.

Each lexical pattern is reviewed by doctors and corrected to the domain specific usage. Modifying these patterns is faster than translating class patterns, since the experts make corrections only on the error parts, and keep the others untouched. The results will be analyzed in detail in Section 7.

Source Pattern	Translated by Google / Corrected by Doctor
She was then **referred** to	然後她被**稱爲**
	然後她被**轉介**到
Feeding : on full diet	**餵養**：全飲食
	進食：普通飲食
to our ward for **operation**	我們的病房**操作**
	我們病房接受**手術**

Table 2. Bilingual lexical patterns with translation errors due to WSD problems. Words in bold show these errors.

6 Pattern Integration

Based on the acquired bilingual patterns, we attempt to achieve domain adaptation by integrating them into the background SMT system. Since we use n-gram patterns, the integration can be carried out without major changes to the phrase-based system. Lexical patterns can serve as a separate phrase table, as proposed in Bertoldi and Federico (2009), to provide in-domain translation options. Because class patterns are mostly used in a specific domain and their translations by domain experts are unlikely to be ambiguous, we adopt the translations of these class patterns in each input sentence, and start decoding from the partial hypothesis. This is feasible with the support of some advanced functions, such as XML markup and continuing partial translation, in the current version of Moses.

In addition to phrase-based model, pattern integration is also another choice in structured SMT models such as Hiero (Chiang, 2005) and BTG (Xiong et al., 2006) systems. In their work, CKY decoders are implemented and phrases are translated at a time. Therefore, bilingual patterns can be considered as one of the translations rules in these models.

7 Experiments

We evaluate the performance of our pattern identification from two aspects, i.e., significance (Section 7.2) and diversity (Section 7.3), and then discuss the quality of translated lexical patterns.

N-gram	NSP Stage (a.2)	Linguistic Stage (a.3)	Coverage Stage (a.4)
5-gram	2,642,714	7,627	7,627
4-gram	2,310,139	14,744	10,827
Trigram	1,557,801	19,117	12,586
Bigram	699,307	15,803	9,244

Table 3. Numbers of remaining patterns after Stage (a.2) – (a.4).

N-gram	#Patterns	#Clusters	Avg. Cluster Size
5	4,634	2,149	2.17
4	6,229	1,957	3.18

Table 4. Clusterings of 5-gram and 4-gram patterns.

7.1 Data Description

The in-domain corpus is selected from the NTUH medical summaries written from January to June, 2010. It is composed of 60,448 medical summaries with 1.8M sentences and 18M words.

After named entity classification, we use Ngram Statistics Package (NSP) (Banerjee and Pedersen, 2003) to enumerate patterns, calculate their frequencies, and examine their linguistic completeness and coverage relations.

Table 3 shows the number of n-grams after each stage. The size is drastically reduced in the Linguistic stage to below 3% of those extracted by NSP. Most of the patterns filtered by Stanford Parser contain conjunctions, prepositions or adjectives at their end. Note that 5-gram patterns remain unchanged in the Coverage stage, since 5 is the highest order in the extracted patterns. After State (a.4), we translate the class patterns and lexical patterns to obtain bilingual ones.

For class patterns, we sample top 5-grams and 4-grams for translations. Stage (a.5) is performed to cluster these patterns and present them to the doctors. The statistics of class patterns and derived clusters are shown in Table 4. Each cluster contains only 2.17 and 3.18 patterns on the average for 5-grams and 4-grams, respectively.

We ask 32 NTUH residents to translate class patterns in larger clusters to achieve diversity. The domain experts are instructed by an on-site tutorial. Then, they examine each pattern in the order we present. Based on their expertise, common patterns are chosen and translated.

N-gram	Translate	Discard	Accuracy
5-gram	642	432	59.78%
4-gram	348	152	69.60%

Table 5. Results of translating class patterns.

N-gram	Doctor	Coverage	Cluster	Total
5-gram	642	+0	+1,628	2,270
4-gram	348	+1,208	+2,238	3,794

Table 6. Extended bilingual patterns from coverage relations and clustering.

For lexical patterns, we sample 5-grams translated by Google Translate for reviewing by one NTUH visiting staff. These 5-grams are either accepted or modified based on the doctor's expert knowledge.

7.2 Evaluating Pattern Significance

The identified class patterns are either translated or discarded by the doctors. The experts consider the former as significant patterns and translate them into Chinese. In contrast, the latter that cannot be translated are non-significant. Table 5 shows the accuracy is 59.78% and 69.60% for 5-gram and 4-gram patterns, respectively. It demonstrates the effectiveness of our strategy to select linguistically complete patterns. Among the discarded patterns, some have misclassified words, due to noisy data in our knowledge base. Parsing errors also cause some non-linguistic n-grams.

7.3 Evaluating Pattern Diversity

We evaluate the diversity of the translated class patterns, by extending them based on the coverage relations and pattern clustering in Stages (a.4) and (a.5). For each translated 5-gram pattern, we produce a new 4-gram bilingual pattern if such a coverage relation exists. For each cluster with at least one pattern translated by a doctor, we uncover the translations of other similar patterns based on the expert translation. Table 6 reports the results of our extension methods, showing the newly discovered patterns at Coverage and Cluster stages.

The 5-gram and 4-gram patterns after the extensions are 3.54 and 10.9 times larger than those translated by the doctors. This suggests we better allocate our expert efforts and achieve high diversity among the translated patterns.

N-gram	Accept	Modify	Accuracy
5	354	820	30.15%

Table 7. Results of reviewing the translated lexical patterns.

7.4 Analyzing Errors of Google Translate

We examine the effectiveness of a general purpose MT when applied to the specific domain. For the lexical patterns translated by Google, Table 7 gives the statistics of acceptances and modifications by the doctor. Only 30.15% of the translations are left unchanged, while the others are corrected by the expert. Further analysis on these 820 corrected translations reveals that 50% of them have WSD errors as illustrated in Section 5.3. Disagreements with writing styles and reordering errors account for 25% and 20% respectively.

8 Conclusion and Future Work

We proposed a framework to identify and translate significant patterns for domain adaption in SMT. The main concern throughout the proposed framework is to reduce the cost of domain experts. We identified and arranged the significant patterns with high quality and diversity. We designed a user friendly interface and applied an online translator to save the translation time of the experts. The experiments were performed on monolingual in-domain corpus. The results showed the significance of the presented patterns, and the diversity of the translated bilingual patterns.

In future work, we will build a medical summary SMT system, based on the acquired bilingual patterns. We will also investigate ways for tuning the system by supervised learning techniques, with the continuous help from the domain experts.

Acknowledgments

We are very thankful for NTUH doctors' helps on annotating the medical summary corpus.

References

Satanjeev Banerjee and Ted Pedersen. 2003. The design, implementation, and use of the Ngram Statistics Package. In *Proceedings of the Fourth International Conference on Intelligent Text Processing and Computational Linguistics*, pages 370–381.

Nicola Bertoldi and Marcello Federico. 2009. Domain adaptation for statistical machine translation with monolingual resources. In *Proceedings of the Fourth Workshop on Statistical Machine Translation*, pages 182-189.

David Chiang. 2005. A hierarchical phrase-based model for statistical machine translation. In *Proceedings of ACL 2005*, pages 263–270.

Jorge Civera and Alfons Juan. 2007. Domain adaptation in statistical machine translation with mixture modeling. In *Proceedings of the Second Workshop on Statistical Machine Translation*, pages 177–180.

George Foster and Roland Kuhn. 2007. Mixture model adaptation for SMT. In P*roceedings of the Second Workshop on Statistical Machine Translation*, pages 128–135.

George Foster, Cyril Goutte, and Roland Kuhn. 2010. Discriminative instance weighting for domain adaptation in statistical machine translation. *In Proceedings of EMNLP 2010*, pages 451-459.

Richard I. Kittredge. 1985. The significance of sublanguage for automatic translation. In *Proceedings of the Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, pages 154-166.

Dan Klein and Christopher D. Manning. 2003. Accurate unlexicalized parsing. In *Proceedings of ACL 2003*, pages 423-430.

Philipp Koehn, Franz Joseph Och, and Daniel Marcu. 2003. Statistical phrase-based translation. In *Proceedings of HLT/NAACL 2003*, pages 127-133.

Philipp Koehn. 2004. Pharaoh: a beam search decoder for phrased-based statistical machine translation models. In *Proceedings of AMTA 2004*, pages 115-124.

Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan,Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondrej Bojar, Alexandra Constrantin, and Evan Herbst. 2007. Moses: Open source toolkit for statistical machine translation. In *Proceedings of ACL 2007, Demonstration Session*, pages 177–180.

Naomi Sager, Margaret Lyman, Leo J. Tick, Francois Borst, Ngo Thanh Nhan, Claude Reveillard, Yun Su and Jean-Raoul Scherrer. 1989. Adapting a medical language processor from English to French. In *Proceedings of the Sixth Conference on Medical Informatics*, pages 795-799.

Deyi Xiong, Qun Liu and Shouxun Lin. 2006. Maximum entropy based phrase reordering model for statistical machine translation. In *Proceedings of ACL-COLING 2006*, pages 521-528.

Bing Zhao, Matthias Eck, and Stephan Vogel. 2004. Language model adaptation for statistical machine translation via structured query models. In *Proceedings of COLING 2004*, pages 411–417.