Hybrid Domain Adaptation for a Rule Based MT System

Petra Wolf
Lucy Software and Services
Neumarkter Str. 81
81673 Munich, Germany
petra.wolf@lucysoftware.com

Ulrike Bernardi
Lucy Software and Services
Neumarkter Str. 81
81673 Munich, Germany
ulrike.bernardi@lucysoftware.com

Abstract

This study presents several experiments to show the power of domain-specific adaptation by means of hybrid terminology extraction mechanisms and the subsequent terminology integration into a rule based machine translation (RBMT) system, thus avoiding cumbersome human lexicon and grammar customization. Detailed evaluation reveals the great potential of this approach: Translation quality can be improved substantially in two domains.

1 Introduction

Adaptation to new domains is crucial to achieve high quality machine translation results. For RBMT systems, such adaptation is mainly performed by manual coding of domain specific terminology. Here, we present a hybrid design for domain adaptation of an RBMT system using an intertwined net of traditional linguistic methods together with statistics-driven techniques.

The remainder of this paper is organized as follows: Section 2 gives an overview of related work on domain adaptation and terminology extraction. Details on the baseline MT system are provided in Section 3 followed by a description of the adaptations performed in Section 4. Section 5 deals with the adapted system followed by a description of the results of translation quality evaluations in Section 6. Section 7 summarizes the key findings and outlines open issues for future work.

2 Related work

Hybrid machine translation approaches become more and more popular in order to overcome the drawbacks of RBMT or statistical machine translation (SMT) alone by a combined approach. Here, we want to evaluate now the potential of a proven hybrid system (Wolf et al., 2011) for domain adaptation. The weak points of RBMT which during adaptation to new domains have an even higher adverse effect are the lexical selection, transfer mapping and transformations, as several evaluations revealed (Thurmair, 2009; Chen et al., 2009). This study concentrates on automatic enlargements of RBMT lexicons going yet one step further by using statistically gained knowledge already during analysis. In this way, not only lexical target selection during transfer is improved, but also parsing within analysis is facilitated, since more naturally sounding material can be accessed.

Research in the field of domain adaptation for RBMT so far focused on statistically based post-editing mechanisms (Isabelle et al., 2007; Dugast et al., 2009) without any interference in the proper translation process. Here, we integrate the statistical knowledge during the translation phase, thus raising the overall quality. At the same time, we can avoid manual system tuning often applied for domain adaptation of RBMT systems, as we rely on the knowledge-augmented information automatically gathered by statistical extraction.

Within SMT, domain adaptation is even more crucial, since the translation quality highly depends on the similarity of development and test data and since it is often difficult to obtain enough data for specific domains. Therefore, SMT systems are developed with general domain data and then adapted to the specific domain with a small amount of in-domain data, as described for example in (Pecina et al., 2012).

There is a huge amount of research on terminology extraction, but only very few researchers deal

Sima'an, K., Forcada, M.L., Grasmick, D., Depraetere, H., Way, A. (eds.)
Proceedings of the XIV Machine Translation Summit (Nice, September 2–6, 2013), p. 321–328.

with extracted terms as input for machine translation, such as (Thurmair and Aleksic, 2012). They extract terms directly from a phrase table and import them in an RBMT system. Our approach is similar, but is more tailored to the specific RBMT system, since this RBMT system is already used during term extraction.

3 Baseline MT System

The baseline MT system is a transfer-based RBMT system with the three phases analysis, transfer and generation (cf. (Alonso and Thurmair, 2003)). It contains sophisticated transfer mechanisms to test on specific contexts so that the English word *brute* can be translated into German by *roh*, if the context contains *force* for example. Also transformations are in place for prepositions to map the source preposition to the appropriate target preposition.

	German → English	**English → German**
SYS-LEX	223,601	89,716
EXT-LEX	245,150	111,392

Table 1: Lexicon Size for the Baseline System

During transfer the baseline MT system already offers the possibility to select a specific domain to disambiguate readings. The bilingual system lexicons contain various domains, such as technical vs. general vocabulary. This study extends the coverage of two specific domains now by hybrid terminology extraction, thus achieving more accurate domain-dependent translations. To investigate the influence of the baseline lexicon, we performed experiments with two baseline systems containing different lexicons (cf. Table 1):

- ***SYS-LEX***: baseline lexicon with broad coverage in various domains
- ***EXT-LEX***: extended baseline lexicon with about 20,000 additional automotive entries for both language directions

4 Adaptations

We use *LiSTEX* (Linguistically augmented Statistical Terminology Extraction) in order to extract term pairs by means of statistical algorithms from existing bilingual corpora (Wolf et al., 2011) and extended it by an additional named entity recognizer and more filtering mechanisms to improve the quality of the extracted terminology. In the following *LiSTEX* will be explained showing the whole process from *Term Acquisition* via *Term Filtering* and *Translation Preparation* up to the final *Translation Phase* (cf. Figure 1).

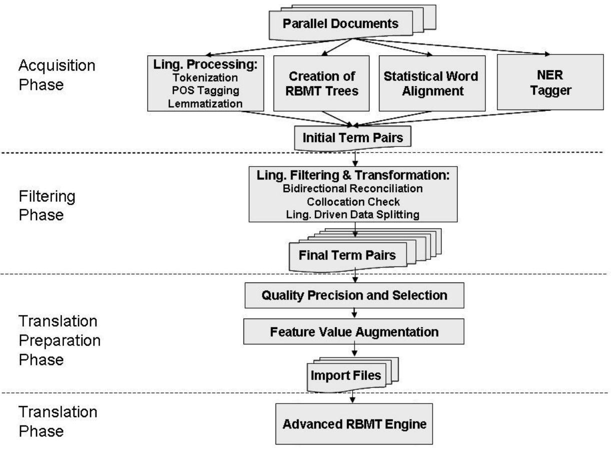
Parallel Documents
Acquisition Phase
Ling. Processing: Tokenization POS Tagging Lemmatization
Creation of RBMT Trees
Statistical Word Alignment
NER Tagger
Initial Term Pairs
Filtering Phase
Ling. Filtering & Transformation: Bidirectional Reconciliation Collocation Check Ling. Driven Data Splitting
Final Term Pairs
Translation Preparation Phase
Quality Precision and Selection
Feature Value Augmentation
Import Files
Translation Phase
Advanced RBMT Engine

Figure 1: *LiSTEX* Workflow

4.1 Term Acquisition Phase

During *Term Acquisition*, the bilingual texts are tokenized, lemmatized, tagged by part of speech information and statistically word aligned. The corpus texts are also processed by the baseline RBMT system and the resulting trees are aligned to the sentences. In order to identify possible term candidates, we select all terms with a translation equivalent in the RBMT trees different from the available human translation in the corpus. In this way, the system lexicon of the baseline system influences the extraction process so that only new term pairs are extracted.

Named entities are recognized by means of the Stanford Named Entity Recognizer (Finkel et al., 2005) and delivered as separated output, subclassified as personal names, locations and organizations for the later feature generation. Since for German gender information for proper nouns is crucial for correct translation, it was guessed based on some heuristics using this semantic subclassification.

4.2 Term Filtering Phase

Term Filtering consists of various filtering mechanisms of the initial term candidates in order to sort out wrong terms:

- **frequency filter**: all pairs which just appear once in the corpus are discarded.
- **bilingual filter**: only terms which were extracted for both language directions are kept.

- **single word filter**: all single words which only appear in multiword expressions and not on their own are sorted out.
- **alternative filter**: only the two most frequent alternative translations for a given source term are kept, since it turned out that with more alternatives we get more ballast, but not a higher precision.
- **spelling filter**: only the most frequent lower vs. upper case writing of equal terms stays in the final term pairs.

Finally, the term pairs are split according to linguistic criteria, such as part of speech of source and target terms, multiwords vs. single words.

4.3 Translation Preparation Phase

In this step, the term pairs are prepared for the actual RBMT system. Improbable multiword to single word transitions, such as German multiword to English single word, and term pairs with category changes are sorted out. For every term, the correct base form is generated. For multiwords, the agreement in gender and number is taken into account so that we generate *elementarer Rechtsgrundsatz der gemeinsamen Fischereipolitik* from the extracted string *elementar Rechtsgrundsatz d gemeinsam Fischereipolitik* by correctly inflecting the variable parts of the multiword.

The so far given shallow data structures are parsed and augmented by automatically generating the linguistic feature value pairs. Not only morpho-lexical information is synthesized, such as declension class, gender, sexus, multiword structure, but also semantico-syntactic information, such as kind of noun (abstract vs. concrete), obligatory for the RBMT system is generated and stored.

4.4 Evaluation of Terms

In order to identify the potential of the *LiSTEX* approach for domain-specific adaptation, we performed three extraction runs: One with the huge corpus of European parliamentary speeches and two with a much smaller, but more domain-specific corpus from the automotive domain:

1. We extract political terms from the big European Parliament corpus (Koehn, 2005). During *Term Acquisition*, we used *SYS-LEX* for creation of comparison trees.
2. We extract automotive terms from a small automotive corpus. During *Term Acquisition*, for creation of comparison trees we used *SYS-LEX* again.
3. The same automotive corpus is used, but during *Term Acquisition*, we used *EXT-LEX*, the extended lexicon already containing some automotive terminology.

In this way, we were not only able to evaluate the effects of large, general vs. small, domain-specific corpora for terminology extraction, but we could also analyze the role the base lexicon and its amount of intersections with the domain in question play. We evaluated a random assessment of 5% of the extratced terminology to calculate the error rate: the source and target canonical forms together with their category assignment were checked for correctness.

4.4.1 Terms Political Domain

From the *Europarl* corpus with more than a million lines around 13,800 term pairs and 2,600 proper noun pairs were evaluated to be usable for automatic import (cf. Figure 2). For terms, we have an error rate of about 6%, whereas the error rate for proper nouns is much higher, as expected.

	Terms	**Proper Nouns**
Term pairs	13,850	2,655
Error rate	5.94%	18.16%

Table 2: Political Domain: Evaluation of Terms

The augmented proper nouns were manually inspected and imported, since it became obvious that gender and type assignment which is ever so crucial for automatic feature augmentation could not be handled automatically: The gender assignment cannot be trusted without careful manual revision. Also the semantic subclassification leaves room for further improvement, since persons are often wrongly assigned to location or organization.

4.4.2 Terms Automotive Domain

In the automotive domain, we took the *Bordbuch*, a translation memory of about 117,000 German-English sentence pairs used to create car manuals. The *Bordbuch* was prepared for term extraction and cleaned in order to be used by GIZA++ (Och and Ney, 2000) for word alignment. After cleaning, 112,645 sentence pairs are

Base	***SYS-LEX***		***EXT-LEX***	
	Terms	**Proper Nouns**	**Terms**	**Proper Nouns**
Term pairs	404	67	419	69
Error rate	34.73%	22.50%	33.87%	22.35%

Table 3: Automotive Domain: Evaluation of Terms

left. Since this corpus was much smaller than *Europarl*, we added that corpus to the *Europarl* corpus for the computation of word alignments. Nevertheless the word alignment issues remained critical, since both text types are quite different in sentence length and kind of writing (nominal vs. verbal writing style).

The low number of extracted data allowed us to take a detailed look at all terms in order to check their quality. As Table 3 shows, the term error rates in both extraction runs - with *SYS-LEX* vs. with *EXT-LEX* - are very similar, but higher than in the political domain: About 400 terms and nearly 70 proper nouns are extracted.

The low overall yield of extracted terms may be again due to the insufficient alignment. This is confirmed by the errors we found while evaluating the terminology for correctness: Whereas in the *Europarl* corpus, we found a high percentage of wrong terms caused by generation errors (eg. *Frankfurterer Flughafen - Frankfurt airport*), in the *Bordbuch* corpus we now had more alignment-specific mistakes, i.e. the two terms are not related at all (eg. *hintere Sitzbank - airflow*).

Interestingly enough, the error rate for proper nouns is even lower than the one for terms. This might also be due to alignment problems, since misaligned proper nouns are not recognized by the NER tagger at all and thus remain as normal terms. The interesting question to be answered now is whether extraction by means of *SYS-LEX* with little intersections shows more changes in translation and results in better translation quality.

5 Adapted system

The import files generated in *Translation Preparation* are imported into the RBMT system without any manual interference. Conflicting entries with the same source and target lexemes already in the lexicon are stored as additional entries. This is of special interest for lexicon entries containing tests and transformations. In this case, the system lexicon entry with a test on a specific context will still be accessed and preferred, when the context condition is fulfilled, although it is in another domain. But when this context condition is not fulfilled, the specific newly imported term is used.

The import mechanism assures that the new terms are available in all lexicons used during analysis, transfer and generation. Monolingual source and target language entries are defaulted, if not existing. In this way, during translation, long multiword entries may facilitate the analysis by appropriate parsing of complex expressions automatically extracted beforehand. For example, the noun phrase *the fall of the wall and the reunification of Germany* is syntactically ambiguous:

[[the fall of the wall] and [the reunification of Germany]]
[the fall [of [[the wall] and [the reunification of Germany]]]]

In the baseline MT system, the second reading is preferred, whereas in the adapted system, with the help of the extracted terms *fall of the wall* and *reunification of Germany* the first correct reading is selected.

During MT transfer, entity-bound transformations need to be in place for dealing with the whole range of multiword expressions from fixed idioms, such as *Abu Dhabi*, up to syntactically free collocations, such as *brain drain - Abwanderung von Spitzenkräften* cf. section 6.3).

In line with our three extraction runs, we created three adapted systems:

1. ***POL***: Terms and proper nouns extracted from *Europarl* are imported as political terminology in *SYS-LEX*.

2. ***AUTO-1***: Terms and proper nouns extracted from *Bordbuch* while using *SYS-LEX* during terminology extraction are imported as automotive terminology in *SYS-LEX*.

3. ***AUTO-2***: Terms and proper nouns extracted from *Bordbuch* while using *EXT-LEX* during terminology extraction are imported as automotive terminology in *EXT-LEX*.

	German → English	**English → German**
Translated TUs	2,000	2,000
Different translations	87.30%	88.95%
Evaluated differences	217	200
Better	28.57%	39.0%
Equal	48.85%	43.5%
Worse	22.58%	17.5%
Overall Improvement	**5.99%**	**21.5%**

Table 4: Translation Quality Evaluation: Baseline vs. POL-Adapted RBMT System

As expected, the number of conflicting entries during the import is substantially lower in *AUTO-1* (4%) than in *AUTO-2* (12%), since the intersections between the *SYS-LEX* and the *Bordbuch* corpus are smaller than the ones between the extended lexicon, *EXT-LEX*, and this corpus.

6 Translation Quality Evaluation (TQE)

6.1 TQE Political Domain

As a test set, we choose the ACL WMT 2008 test set which contains the Q4/2000 portion of the EuroParl data and performed a comparative translation quality evaluation with the baseline MT system vs. the adapted MT system. The BLEU scores show small improvements of 0.3 (cf. Table 5). But since BLEU's correlation with human judgments has already been drawn into question (cf. (Callison-Burch et al., 2006)), we also performed a manual evaluation of the translations according to predefined comparative evaluation criteria *BETTER*, *EQUAL*, *WORSE*. In all cases where we found alternative translations within a translation unit, we evaluated the translation unit as better, as soon as the better alternative is among the set of alternatives.

This translation quality evaluation revealed a high number of differently translated sentences in both language directions: approx. 87-88% (cf. Table 4). Even within each of these sentences we find several distinctions so that we concentrated on about 10% of the differing sentences in our manual evaluation assuming that the rest will consist of similar phenomenons for the given data.

The translation quality shows substantial improvements: In German to English, we have an overall improvement of 5.99%; English to German is even better with an overall improvement of 21.5% which might be explained by the fact that the *SYS-LEX* in English to German is much smaller and therefore the effect of the new terminology is more visible.

	Baseline RBMT	**Adapted RBMT**
German → English	16.38	16.61
English → German	11.04	11.31

Table 5: BLEU Scores: Baseline vs. POL-Adapted RBMT System

Since we evaluated translations as better, as soon as the better alternative is among the alternatives, we performed another experiment to evaluate, whether the first alternative is the correct one, if we use the frequency information from the corpus. We modified the RBMT transfer modules so that the information on the term frequency from the corpus will be accessed and used as a preferencing control element in target selection: The more frequent target will be the first alternative. We measured in a specific benchmark run the translation quality of these reordered alternatives in the adapted RBMT system. As Table 6 shows, the frequency information is very valuable in both language directions so that we have a significantly better positioning of the correct and adequate alternative, thus achieving a translation sounding more natural. This also qualifies the above mentioned results of the overall translation quality: If we only look at the first alternative and not at the whole set of alternatives, the overall improvement as reported in Table 4 would be smaller of course.

6.1.1 Translation Quality of Named Entities

Since named entities are challenging for automatic processing but at the same time their correct translation is crucial, we evaluated the effects of named entities in detail by comparing the translation quality of the translation with terms only vs. with terms *and* proper nouns. This shows clear

	German → English	**English → German**
Translated TUs	2,000	2,000
Diff. translations	812	922
Eval. translations	288	322
Better	79.16%	61.49%
Equal	11.46%	18.63%
Worse	9.38%	19.88%
Overall Improvement	**69.78%**	**41.61%**

Table 6: Political Domain: Effects of Most Frequent Alternatives on Translation Quality

	German→English	**English→German**
Translated TUs	2,000	2,000
Diff. translations	128	355
Better	67.19%	22.81%
Equal	14.06%	71.27%
Worse	18.75%	5.92%
Overall Improvement	**48.44%**	**16.89%**

Table 7: TQE Political Domain: Adapted RBMT with Terms vs. with Terms and Named Entities

improvements in both language directions (cf. Table 7), although the number of differently translated sentences is quite low: For English to German, 355 sentences from the total of 2000 sentences were different, for German to English even only 128 sentences. To conclude, the careful manual import of named entities is worth while, since it improves the translation quality substantially.

6.2 TQE Automotive Domain

As a test set, we used part of a car owner's manual in German and English and compared the translations of the baseline MT system vs. the adapted MT system. Both manuals have the same contents, but are no direct translations of each other because of different country specific regularities. Both of them contained more than 5,000 sentences for translation, the English manual is a little longer.

In line with the two extraction runs in the automotive domain, we performed two comparisons:

1. Comparison of translations with *SYS-LEX* without any automotive terms to the translations with the adapted system *AUTO-1*.

2. Comparison of translations with *EXT-LEX* to the ones with the adapted system *AUTO-2*.

We evaluated 101 of the translation differences revealing clear improvements (cf. Table 8): The translation quality of German to English improved by about 40% in both cases. For English to German, the overall quality gain was about 23%. Thus, our system adaptions could raise the German to English quality significantly. On the other hand, a higher English to German quality gain was hindered by the mentioned alignment-specific mistakes. Interestingly enough, although we found more translation differences, when there are less intersections between the baseline lexicon and the corpus used for extraction, as in *AUTO-2*, the translation quality is very similar in both cases.

6.3 Additional Experiments: Multiword Expressions

Since lexically bound collocations in contrast to fixed idioms maintain a large variability as to their morpho-syntactic behavior, we implemented a supplementary approach in our hybrid RBMT system: Grammar tests, operators and additional transformations were developed for identifying and transferring collocations.

The tree-like structures stored e.g. for the collocates *Abwanderung* and *Spitzenkraft* (cf. Figure 2) allow at run-time for several analysis and transformational operations which in case of met conditions lead to the English translation *brain drain*. It is not simple in the sense of a fixed idiom, this strategy allows for open syntactic processing in the whole MT workflow. Thus syntactic variability can be guaranteed so that after normal built-up by

	German → English		**English → German**	
	AUTO-1 *SYS-LEX*	*AUTO-2* *EXT-LEX*	*AUTO-1* *SYS-LEX*	*AUTO-2* *EXT-LEX*
Translated TUs	5,190	5,190	7,182	7,182
Diff. translations	54.59%	46.44%	57.66%	38.79%
Eval. differences	101	101	101	101
Better	49.50%	45.54%	31.68%	38.61%
Equal	42.57%	47.52%	59.41%	47.52%
Worse	7.92%	6.93%	8.91%	13.86%
Overall Improvement	**41.58%**	**38.61%**	**22.77%**	**24.75%**

Table 8: TQE Automotive Domain: Baseline vs. adapted RBMT *AUTO-1* and *AUTO-2*

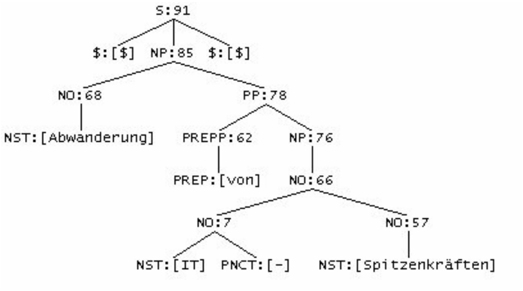

Figure 2: *Abwanderung von IT-Spitzenkräften*

the analysis grammar, a phrase like *Abwanderung von IT-Spitzenkräften* will be correctly recognized as a match to the stored collocates. The new modifier *IT* can be analyzed well as a free hyphenated nominal specifier to the head of the subordinated prepositional phrase *Spitzenkraft* in its obligatory plural form. A distinct transfer and generation processing can be initiated by the stored collocates: Complex transformations produce the structurally distinct transfer trees while managing the freely added *IT* and its correct syntactic and semantic generation.

With these collocation modules, the translation quality could be further raised, as a comparison between an import with the traditional multiword defaulting only and another import with defaulting multiwords and collocations in the political domain shows (cf. Table 9): We have an overall improvement of about 14%.

6.4 Summary of Domain Adaptation Effects

Comparing the whole range of experiments, it turned out that the alignment is the most crucial part of the workflow: The rate of extracted terms per line is the higher, the better the alignment in the corpus is. This is evident in the very well aligned *Europarl* corpus where about 1% of the lines produces a new term, whereas with the *Bordbuch* bilingual corpus only 0.37% of the lines results in a new term.

Also the quality of the extracted terms mainly depends on the quality of the word-to-word alignment of the corpus: The alignment errors in the automotive domain lead to a much higher term error rate than in the political domain. In contrast to the actual terms, the error rate for the proper nouns is quite stable: In all three test cases, the error rate stays at about 20% (cf. Table 4 and Table 8).

As to be expected, the bigger the corpus is and the less it interferes with the baseline lexicon, the higher is the number of translation differences. This is confirmed by the much smaller rate of translation differences in the automotive domain than in our tests in the political domain (cf. Table 4 and Table 8). Nevertheless, as our experiments with two kinds of baseline lexicons in the automotive domain show, this bigger amount of differences still results in comparable translation quality in both experiments.

All our experiments confirm a substantial translation quality gain after domain adaptation. Whereas the translation quality for English to German improves by about 20%, we have a more diverse picture in the other direction depending on how general vs. specific the vocabulary of the text is. This might be due to the substantially bigger German to English lexicon.

7 Conclusion and Future Work

In this paper, we present the validation and further development of the hybrid *LiSTEX* approach in the challenging area of domain adaptation. An intertwined net of linguistic and statistical adaptions have been implemented not only for the acquisition phase, but also in several RBMT mod-

	German→English	**English→German**
Translated TUs	2,000	2,000
Diff. translations	644	938
Eval. differences	151	150
Better	31.79%	29.33%
Equal	52.98%	53.33%
Worse	15.23%	17.33%
Overall Improvement	**16.56%**	**12%**

Table 9: Political Domain: Effects of Collocational Entities on Translation Quality

ules in order to cope with linguistic phenomenons of the new data structures processed at all different stages.

The evaluation within the two areas - European Parliamentary Speeches and the Automotive Domain - confirmed the big potential of the *LiSTEX* approach for adapting an RBMT system to a new domain: The translation quality can be improved substantially without manual tuning. Even with a small bilingual corpus, such as the automotive one, the translations get better. However, the word alignment quality of the bilingual corpus is absolutely vital: The better the texts are aligned, the more good entities can be extracted during *Term Acquisition and Filtering* and form correct input for the crucial feature value augmentation. Finally, this way the more valid deep linguistic data structures can be accessed and steer analysis and translation. For further enhancement of the advantages of *LiSTEX* for domain-specific adaptation, better alignment methods are crucial, especially given a small amount of in-domain data.

References

Alonso, J. and G. Thurmair. 2003. The Comprendium Translator system. In *Proceedings of the MT Summit IX*.

Callison-Burch, C., M. Osborne, and P. Koehn. 2006. Re-evaluating the role of bleu in machine translation research. In *Proceedings of the 11th EACL*, pages 249–256.

Chen, Y., M. Jellinghaus, A. Eisele, Yi Z., S. Hunsicker, S. Theison, Ch. Federmann, and H. Uszkoreit. 2009. Combining multi-engine translations with Moses. In *Proceedings of the 4th Workshop on Statistical Machine Translation*, pages 42–46.

Dugast, L., J. Senellart, and P. Koehn. 2009. Statistical post editing and dictionary extraction: Systran/Edinburgh submissions for ACL-WMT2009. In *Proceedings of the 4th Workshop on Statistical Machine Translation*, pages 110–114.

Finkel, J., T. Grenager, and C. Manning. 2005. Incorporating non-local information into information extraction systems by gibbs sampling. In *Proceedings of the 43nd Annual Meeting of the Association for Computational Linguistics*, pages 363–370.

Isabelle, P., C. Goutte, and M. Sinard. 2007. Domain adaptation of mt systems through automatic post-editing. In *Proceedings of the Machine Translation Summit XI*, pages 255–261.

Koehn, P. 2005. Europarl: A parallel corpus for statistical machine translation. In *Proceedings of the MT Summit X*, pages 79–86.

Och, F. J. and H. Ney. 2000. Improved statistical alignment models. In *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics*, pages 440–447.

Pecina, P., A. Toral, and J. van Genabith. 2012. Simple and effective parameter tuning for domain adaptation of statistical machine translation. In *Proceedings of the 24th International Conference on Computational Linguistics*, pages 2209–2224.

Thurmair, G. and V. Aleksic. 2012. Creating term and lexicon entries from phrase tables. In *Proceedings of the 16th EAMT*, pages 253–260.

Thurmair, G. 2009. Comparing different architectures of hybrid machine translation systems. In *Proceedings of the MT Summit XII*, pages 340–347.

Wolf, P., U. Bernardi, C. Federmann, and S. Hunsicker. 2011. From statistical term extraction to hybrid machine translation. In *Proceedings of the 15th EAMT*, pages 225–232.