The Knowledge Engineering Review, Vol. 00:0, 1–24. © 2004, Cambridge University Press
DOI: 10.1017/S000000000000000 Printed in the United Kingdom

Domain Adaptation Strategies in Statistical Machine Translation: a brief overview

Marta R. Costa-jussà

Centro de Investigación en Computación, Instituto Politécnico Nacional
E-mail: marta@nlp.cic.ipn.edu

Abstract

Statistical Machine Translation (SMT) is gaining interest given that it can easily be adapted to any pair of languages. One of the main challenges in SMT is domain adaptation because the performance in translation drops when testing conditions deviate from training conditions. Many research works are arising to face this challenge. Research is focused on trying to exploit all kinds of material, if available. This paper provides an overview of research which copes with the domain adaptation challenge in SMT.

1 Motivation

The need for Machine Translation (MT) is continuously growing due to the evolution of information society. Today translation technology is trying to keep pace with the demand for high quality translations.

MT systems can be classified into rule-based and corpus-based approaches, in terms of their core technology. Recently, the statistical systems, which fall in the corpus-based category, are popular given that the knowledge is automatically extracted by analyzing translation examples from a parallel corpus built by human experts. The advantage is that, once the required techniques have been developed for a given language pair, Statistical Machine Translation (SMT) systems can be quickly developed for new language pairs using provided training data. For further information on SMT systems please refer to surveys such as Lopez (2008) and Koehn (2010) .

One of the main limitations of SMT as in many machine learning systems is domain adaptation. The task of domain adaptation consists of building models for some specific target domain, given that training data is generally extracted from various domains (including or not the target domain). For example, the Workshop on Statistical Machine Translation (WMT)[1] organizes an international evaluation campaign, with different shared tasks. Generally, it provides training data from different sources including: the European Parliament Plenary Speeches (EPPS), News and United Nations (UN). However, the test data is only from the News domain. Since EPPS and UN use different vocabulary words and a different language style than News, it may seem reasonable that several techniques have to be applied to give more importance to the News training data in order to obtain a better translation output.

The main challenges of performing out-of-domain translations are the out-of-vocabulary words, the unknown expressions and the syntactic constructions. These challenges have mainly been addressed either with limited or no[2] in-domain data by using morphology/linguistic rules or exploiting different kinds of in-domain material. The nature of in-domain material in SMT varies from bilingual or monolingual corpus, translation memories to lexicons.

[1] http://www.statmt.org/wmt14
[2] We refer to handling the problem when data is missing

This paper presents an overview on overcoming the challenge of nonexitent in-domain data or exploiting limited, large in-domain data to improve domain adaptation in SMT. Also we dedicate a section to mention how domain adaptation techniques are exploited on online scenarios. Table 1 shows a summary of this classification.

In-domain data	**Type**	**References**
Nonexistent	Linguistically motivated	Farrús *et al.* (2009); Formiga *et al.* (2012) Costa-jussá *et al.* (2013)
	Data driven	Hidelbrand *et al.* (2005); España *et al.* (2010) Carpuat *et al.* (2007); Ueffing *et al.* (2007) Haque *et al.* (2011) Henríquez *et al.* (2011); Formiga *et al.* (2013) Moore and Lewis (2010); Axelrod *et al* (2011) Pecina *et al.* (2011); Senrich (2012)
Limited		Schwenk (2007); Zens & Ney (2004) Foster & Kuhn (2007)
Large	Parallel corpus	Khalilov *et al.* (2008); Koehn & Schroeder (2007) Foster *et al.* (2006); Finch *et al* (2008) Foster *et al.* (2010); Civera & Juan (2007) Rogati (2009); Niehues & Waibel (2010) Ceausu *et al.* (2011)
	Comparable corpus	Daumé III & Jagarlamudi (2011) Sakadina *et al* (2012)
	Monolingual corpus	Eck (2004); Bulyko (2007); Wu (2008) Schwenk & Estève (2008) Bertoldi & Federico (2009)
	Lexicons	Farrús *et al.* (2009); Okuma (2008); Wu (2008)
	Translation Memories	Abekawa & Kageura (2007); Marcu (2001)
Online Scenarios		Hardt & Elming (2010) Ortíz-Martínez *et al.* (2010) Levenberg *et al.* (2010)

Table 1 Classification of domain adaptation approaches.

2 Domain adaptation with limited or nonexistent in-domain data

SMT is a corpus-based approach. Usually a large quantity of parallel sentences is required in order to build a high quality translation system. The more in-domain material, the better the translation. However, usually, in-domain material is limited or even nonexistent. Under these circumstances, to solve out-of-domain problems, researchers have to either overcome the challenge of nonexistent in-domain data or exploit limited resources by making use of linguistic knowledge or statistical techniques.

2.1 Adaptation with nonexistent in-domain data

There are several techniques that somehow focus on adapting data without resources. We are going to describe most relevant techniques in this area classified into linguistically motivated and data driven. The former is when techniques use some kind of linguistics inherent in the languages dealt with. The latter does not take into account any characteristic of the languages used.

2.1.1 Linguistically motivated techniques

The use of linguistic knowledge with language dependencies helps in domain adaptation. For example, in Farrús *et al.* (2009), authors develop linguistic rules to further improve the Catalan-Spanish translation. The main advantage of linguistic rules is that they generalize for any type of domain. Examples of linguistic rules in this work includes codifying time expressions, solving homonymy by attaching the Part-of-Speech to the surface word and post processing verbs and pronouns either to be attached with a hyphen (Catalan) or without (Spanish).

Other works such as Formiga *et al.* (2012) try to deal with the morphology challenge. Usually, a training set does not contain all morphology alternatives of a single word, specially when working with highly-inflected languages. In this approach authors propose considering Spanish morphology generation as a standard classification task. The idea is that one characteristic of verb forms like number is generalized. So, the translation system learns words without the number information. A posterior classifier is in charge of deciding whether the word is singular or plural. Therefore, after defining a set of simple relevant features, they apply machine learning techniques to classify this independent morphology generation task. The main drawback of this linguistic exploitation is that it is not directly generalizable to other pairs of languages. However, the linguistic rules approach is still investigated, specially working in hybrid MT approaches, see Costa-jussà *et al.* (2013).

2.1.2 Data driven techniques

Authors in Hidelbrand *et al.* (2005) try to exploit information retrieval techniques to retrieve sentence pairs from the training corpus that are relevant to the test sentences. Then, both the language and the translation models are retrained on the extracted data.

Other works try to exploit source similarities such as España *et al.* (2010). Authors built classifiers inspired by those used in word-sense disambiguation (WSD) in order to select translation phrases. In Carpuat *et al.* (2007), authors embed context-rich approaches from WSD methods into SMT systems. In Ueffing *et al.* (2007), authors use semi-supervised model adaptation in the source text. In Haque (2011), the author adds source-side contextual features by incorporating context-dependent phrasal translation probabilities learned using decision trees.

A recent approach by Henríquez *et al.* (2011) interpolates the phrase and lexical reordering tables with a new adapted translation model with derived units. Authors obtain the units identifying the mismatching parts between the non-adapted translation and the actual reference. This approach has successfully been tested in the context of the WMT evaluation Formiga *et al.* (2013).

There are several works that use different techniques to extract monolingual or parallel in-domain corpora, i.e. Moore and Lewis (2010), Axelrod *et al.* (2011), by using perplexity or cross-entropy methods. Pecina *et al* (2011) use web-crawled data.

2.2 Limited resources exploitation

Actually, in SMT, there are many works which try to exploit limited resources to their great advantage like using big target language models or factored language models. One interesting approach is using techniques of neural language modeling, see Schwenk *et al.* (2007). The basic idea of continuous space language models (CSLM), also called neural network language models, is to project the word indices onto a continuous space and to use a probability estimator operating on this space. Since the resulting probability functions are smooth functions of the word representation, better generalization to unknown n-grams can be expected. This is believed to be particularly important for tasks with limited resources. A neural network can be used to simultaneously learn the projection of the words onto the continuous space and to estimate the n-gram probabilities. This is still an n-gram approach, but the LM conditional probabilities are *interpolated* for any possible context of length n-1, instead of backing-off to shorter contexts. The main disadvantage of this approach is the high computational cost.

Another interesting approach when having limited resources is smoothing the translation probabilities. Zens & Ney (2004) suggested to use IBM-1 probabilities or other lexical translation probabilities. Another method of smoothing is the one presented in Foster & Kuhn (2007). Two types of phrase-table smoothing are compared: black-box and glass-box methods. Black-box methods do not look inside phrases but instead treat them as atomic objects and can use all the methods developed for language modeling.

3 Techniques for exploiting larger quantities of different types of in-domain data

SMT can be further developed by using available in-domain material which can be in the form of bilingual or monolingual corpus; translation memories and lexicons.

3.1 Exploiting in-domain parallel corpora

Exploiting in-domain parallel corpora has been recently addressed in the scientific community by the annual WMT international evaluation campaign.

The straightforward way to use in-domain parallel corpora is to simply concatenate it with the out-of-domain corpora available. Then, use the combined data for both translation model and language model training. However, in case in-domain data is much smaller than out-of-domain data, the gain expected through a simply concatenation is not much. The result can be worse than using only in-domain data. This normally happens when the in-domain data has a very particular style, as shown by Khalilov *et al.* (2008).

Another proposal is to implement a translation model interpolation strategy. Authors in Koehn & Schroeder (2007) show that the linear interpolation of translation models achieve gains in BLEU. Alternatively, a log-linear interpolation also improves translation, see Khalilov *et al.* (2008). Similarly, Ceausu *et al.* (2011) show that in-domain translation models and general language models perform better for patent translation than other combinations of in-domain and general data.

The technique of mixture models is explored in works such as Foster *et al.* (2006), Finch *et al.* (2008) and Senrich (2012). Mixture modeling is a standard technique for density estimation which is capable of learning specific probability distributions that better fit subsets of the training dataset. Extensions of these works can be found in Foster *et al.* (2010) and Civera & Juan (2007).

In Rogati (2009) the author adds in-domain corpora by automatically weighting and combining multiple translation resources, according to several criteria, in order to better match a target corpus or a specific domain sample. The criteria include lexical-level domain match, translation quality estimates, size, and taxonomy representation. The main drawback of these approaches is that they may not generalize for different types of corpus. Usually, it depends on the size of the in-domain corpora. In Niehues & Waibel (2010) authors use the corpus id as a target word factor in order to optimize a factored-based translation.

3.2 Exploiting comparable corpora

Comparable corpora are those texts that talk about the same topic in different languages, but do not have corresponding translations as parallel texts. There is plenty of work in the explotation of comparable corpora from different perspectives. Most of them focus on extracting parallel sentences from comparable corpora. If we focus on domain adaptation, we found the work of Daumé III & Jagarlamudi (2011) that use the specific setting in which they have plentiful parallel (labeled) data in a source domain and plentiful comparable (unlabeled) data in a target domain. They can use the unlabeled data in the target domain to build a good language model. Finally, they assume access to a very small amount of parallel (labeled) target data, but only enough to evaluate or run weight tuning. All knowledge about unseen words is extracted from the

comparable data. As a related project we found ACCURAT[3] that particularly targets a number of under-resourced languages and evaluates applicability of comparable corpora for adapting MT to specific narrow domains. The techniques that have been used to improve MT with this additional corpora include a recommendation system towards better translations, see Sakadina *et al* (2012).

3.3 Exploiting in-domain monolingual corpora

The exploitation of in-domain monolingual corpora have been mainly limited to the target language model, see Eck (2004). In Bulyko (2007), they explored discriminative estimation of language model weights by directly optimizing MT performances. In Wu (2008), they investigate linear and log-linear interpolation of in-domain and out-of-domain language models which is still highly used in state-of-the-art systems given its simplicity and usefulness.

In Schwenk & Estève (2008), authors present how the CSLM technique, seen in section 2.2, scales in larger sets. Bertoldi & Federico (2009) have used in-domain monolingual corpora to synthesize a bilingual corpus by translating the monolingual adaptation data into the counterpart language. Translation, re-ordering and language models were estimated after translating in-domain texts with the baseline system.

3.4 Exploiting lexicons

The WMT evaluation, which evaluates domain adaptation, does not provide lexicons. In fact, from the SMT perspective, there are not many works in this direction. We can summarize the related studies as follows.

There are several straightforward approaches. One of them is to introduce the lexicon as phrases in the phrase-based table as in Farrús *et al.* (2009). Unfortunately, the word has no context information. Another approach is to introduce them to substitute the unknown words in the translation, which have the same problem as before. Okuma (2008) presents a more sophisticated approach where the lexicon words are introduced in the training corpus to obtain their corpus. The criteria that they use is basically a Name Entity Recognition classification. Notice that their lexicon is based on person and places Name Entities. Wu (2008) uses domain dictionaries as a translation table. To solve the challenge of assigning a probability to each dictionary entry, they used alternatively uniform, constant or monolingual corpus probabilities.

3.5 Exploiting translation memories

Translation Memories (TMs) are widely used to enhance human translation efforts as in Abekawa & Kageura (2007). A typical TM system has a database of translated sentence pairs, called translation units, and exploits a well-designed matching algorithm to retrieve the most relevant translation units as translation references for a given source sentence to be translated.

Up to this point, works mentioned above do not use SMT methods together with TMs. As far as we know, there are not many works that use in-domain TMs to improve SMT. Despite the fact that TM is widely used to improve human translation, it has yet to be much applied to improve SMT. Indeed, there appears to be just one paper on this topic by Marcu (2001). In this paper the authors extract a TM from a parallel corpus. This automatic TM is used together with a phrase-based system to improve the translation output. Basically, they modify the decoder so that it attempts to find good translation starting from two distinct points in the space of possible translations: one point corresponds to a phrase-table; the other point corresponds to a translation that resembles most closely translations stored in the TM.

More recently, Google translator tools provide the possibility to translate texts by using user's translation materials such as TM's and glossaries[4]. The tool searches all available translation

[3]http://www.accurat-project.eu
[4]http://translate.google.com/toolkit

databases for previous human translations of each segment. If any previous human translations of the segment exist, the tool picks the highest-ranked search result and 'pretranslate' the segment with that translation. If no previous human translation of the segment exists, the tool uses SMT.

3.6 Online Scenarios

Finally, this subsection reviews the emerging topic of incremental (or online) adaptation. Incremental adaptation fits naturally in those translations scenarios where human translators supervise the output of MT systems and relevant works are cited as follows. Hardt & Elming (2010) show an application of online learning techniques to process training data in a sentence-wise manner. Ortíz-Martínez *et al.* (2010) show an application similar to the previous work, but removing the heuristics that they introduce to estimate the phrase-based model. This is achieved by means of the application of the incremental Expectation-Maximization (EM) algorithm. Levenberg *et al.* 2010 show an application of online learning techniques to process large chunks of training data using stepwise EM algorithm.

4 Remarks

One relevant challenge in SMT is domain adaptation, because the more similar is the training material to the test material, the better the translation quality. The increase in the amount of available data has pushed research in domain adaptation. In addition to allowing for a better translation quality, research in domain adaptation may allow to develop SMT systems of better quality to offer to professional translators. Domain adaptation may be really helpful in the recent initiative of Interactive Machine Translation (IMT), where the system aids the translator by interactively making suggestions for completing the translation. The IMT idea have been developed in the TransType and TransType2 European projects, see Barrachina (2009) and Esteban *et al* (2004). The continuation of the mentioned projects is the Casmacat European Project[5]. Out of these projects, a relevant online IMT (i.e. Caitra) system is now available[6].

Acknowledgements

This work has been funded by the Seventh Framework Program of the European Commission through the International Outgoing Fellowship Marie Curie Action (IMTraP-2011-29951), the Spanish Ministerio de Economía y Competitividad, contract TEC2012-38939-C03-02, and the European Regional Development Fund (ERDF/FEDER).

References

Abekawa, T. and Kageura k. A translation aid system with a stratified lookup interface. In *ACL*. The Association for Computer Linguistics, 2007.

Axelrod, A., He, X. and Gao, J. Domain adaptation via pseudo in-domain data selection. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP '11)*. pages355-362. 2011.

Barrachina, S., Bender, O., Casacuberta, F., Civera, J., Cubel, E., Khadivi, S., Lagarda, A., Ney, H., Toms, J., Vidal, E. Statistical approaches to computer-assisted translation. *Computational Linguistics*, 35(1):3–28, 2009.

Bertoldi, N. and Federico, M.. Domain adaptation for statistical machine translation with monolingual resources. In *Proceedings of the Fourth Workshop on Statistical Machine Translation*, pages 182–189, Athens, Greece, March 2009. Association for Computational Linguistics.

Bulyko, I., Matsourkas, S., Schwartz, R., Nguyen, L. and Makhoul, J. Language model adaptation in machine translation from speech. In *Proceedings of the 32nd International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 117–120, 2007.

Carpuat, M. and Wu, D. Improving statistical machine translation using word sense disambiguation. In *Empirical Methods in Natural Language Processing (EMNLP)*, pages 61–72, Prague, June 2007.

[5]http://www.casmacat.eu
[6]http://www.caitra.org

Ceausu, A., Tinsley, J., Zhang, J. and Way, A. Experiments on domain adaptation for patent machine translation in the PLuTO project. In *Procedings of the EAMT*. 2011.

Civera, J. and Juan, A. Domain adaptation in statistical machine translation with mixture modelling. In *Proceedings of the Second Workshop on Statistical Machine Translation (StatMT '07)*, pages 177-180, 2007

Costa-jussà, M.R., Banchs, R.E., Rapp, R., Lambert, P., Eberle, K., and Babych, B. Workshop on hybrid approaches to translation: Overview and developments. In *Proceedings of the ACL Second Workshop on Hybrid Approaches to Translation (HyTra)*. Association for Computational Linguistics, 2013.

Daum III, H. and Jagarlamudi, J.. Domain adaptation for machine translation by mining unseen words. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers - Volume 2 (HLT '11)*, Vol. 2. pages 407-412. 2011.

Skadin I., Aker, A., Mastropavlos, N., Su, F., Tufis, D., Verlic, M., Vasijev, A., Babych, B., Clough, P., Gaizauskas, R., Glaros, N., Lestari Paramita, M. and Pinnis, M. In *Proceedings of the Language and Resources Evaluation Conference*, 2012

Eck, M., Vogel, S. and Waibel, A.. Language model adaptation for statistical machine translation based on information retrieval. In *Proceedings of the LREC*, pages 327–330, 2004.

España-Bonet, C., Giménez, J. and Màrquez, L.. Discriminative phrase-based models for arabic machine translation. *ACM Transactions on Asian Language Information Processing Journal (TALIP)*, 8:1–20, March 2010.

Esteban, J., Lorenzo, J., Valderrábanos, A. S. and Lapalme, G. TransType2 - An Innovative Computer-Assisted Translation System. In *The Companion Volume to the Proceedings of 42st Annual Meeting of the Association for Computational Linguistics*, pages 94-97, 2004.

Farrús, M., Costa-jussà, M.R., Hernández, A., Hneríquez, C., Mariño, J.B. and Fonollosa, J.A.R. On the enhancement of catalan-spanish ngram-based translation by using human evaluation. *Submitted to Language Resources and Evaluation*, 2009.

Finch, A. and Sumita, E.. Dynamic model interpolation for statistical machine translation. In *Proceedings of the Third Workshop on Statistical Machine Translation*, pages 208–215, 2008.

Formiga, L., Costa-jussà, M.R., Mariño, J.B., Fonollosa, J.A.R., Barrón-Cedeño, A. and Màrquez, L. The TALP-UPC phrase-based translation systems for WMT13: System Combination with Morphology Generation,Domain Adaptation and Corpus Filtering. In *Proceedings of the Eighth Workshop on Statistical Machine Translation*. Association for Computational Linguistics, 2013.

Formiga, L., Hernández, A., Mariño, J.B. and Monte, E.. Improving english to spanish out-of-domain translations by morphology generalization and generation. In *Proceedings of the AMTA Monolingual Machine Translation-2012 Workshop*, 2012.

Foster, G. and Kuhn, R. Mixture-model adaptation for SMT. In *Proceedings of the Second Workshop on Statistical Machine Translation*, pages 128–135, 2007.

Foster, G., Kuhn, R. and Johnson, H. Phrasetable smoothing for statistical machine translation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 53–61, Sydney, Australia, 2006.

Foster, G., Goutte, C. and Kuhn, R. Discriminative instance weighting for domain adaptation in statistical machine translation. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 451-459, 2010

Haque, R. Integrating Source-Language Context into Log-linear Models of Statistical Machine Translation. *PhD thesis*, Dublin City University, 2011.

Hardt, D. and Elming, J.. Incremental re-training for post-editing smt. In *Proceedings of the 9th Annual Conference of The Association for Machine Translation in the Americas*, Denver, Colorado, 2010.

Henríquez, C.A., Mariño, J.B. and Banchs, R.E. Deriving translation units using small additional corpora. In *Proceedings of the 15th Conference of the European Association for Machine Translation*, 2011.

Hildebrand, A.S., Eck, M., Vogel, S. and Waibel, A. Adaptation of the translation model for statistical machine translation based on information retrieval. In *Proceedings of EAMT*, pages 133–142, 2005.

Khalilov, M., Costa-Jussà, M.R., Henríquez, C.A., Fonollosa, J.A.R., Hernández, A., Mariño, J.B., Banchs, R.E., Chen, B., Zhang, M., Aw, A. and Li, H. The TALP & I2R SMT Systems for IWSLT 2008. In *Proceedings of the International Workshop on Spoken Language Translation*, pages 116–123, Hawaii, USA, 2008.

Koehn, P. Statistical Machine Translation. *Cambridge University Press*, 2010.

Koehn, P. and Schroeder, J. Experiments in domain adaptation for statistical machine translation. In *Annual Meeting of the Association for Computational Linguistics: Proceedings of the Second Workshop on Statistical Machine Translation (WMT)*, pages 224–227, Prague, June 2007.

Levenberg, A., Callison-Burch, C. and Osborne, M. Stream-based translation models for statistical machine translation. In *Proceedings of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies*, pages 394402, Los Angeles, California, 2010.

López, A. Statistical machine translation. *ACM Computing Surveys*, 40(3):1–49, 2008.

Marcu, D. Towards a unified approach to memory- and statistical-based machine translation. In *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics*, pages 386–393, Morristown, NJ, USA, 2001.

Moore, R.C. and Lewis, W. Intelligent selection of language model training data. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics Short Papers,* pages 220-224. 2010.

Niehues, J. and Waibel, A. Domain adaptation in statistical machine translation using factored translation models. In *Proceedings of EAMT*. 2010.

Okuma, H., Yamamoto, H. and Sumita, E. Introducing a translation dictionary into phrase-based smt system. *IEICE TRANSACTIONS on Information and Systems*, E91-D(7):2051–2057, 2008.

Ortíz-Martnez, D., García-Varea, I. and Casacuberta, F. Online learning for interactive statistical machine translation. In *Proceedings of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies*, pages 546554, Los Angeles, 2010.

Pecina, P., Toral, A., Way, A., Papavassiliou, V., Prokopidis, P. and Giagkou, M. Towards using web-crawled data for domain adaptation in statistical machine translation. In *Procedings of the EAMT*. 2011.

Rogati, M. Domain Adaptation of Translation Models for Multilingual Applications. *PhD Thesis*, Carnegie Mellon University, 2009.

Schwenk, H., Costa-jussà, M. R., and Fonollosa, J.A.R. Smooth bilingual translation. In *In Proceedings of the Empirical Methods in Natural Language Processing*, pages 430–438, Prague, 2007.

Schwenk, H. and Estève, Y. Data selection and smoothing in an open-source system for the 2008 NIST machine translation evaluation. In *Proceedings of the Interspeech*, Brisbane, Australia, 2008.

Sennrich, R. Perplexity minimization for translation model domain adaptation in statistical machine translation. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics* pages 539-549. 2012.

Ueffing, N., Haffari, G., and Sarkar, A. Semi-supervised model adaptation for statistical machine translation. *Machine Translation Journal*. 2008.

Wu, H., Wang, H. and Zong, C. Domain adaptation for statistical machine translation with domain dictionary and monolingual corpora. In *Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1*, pages 993–1000, 2008.

Zens, R. and Ney, H. Improvements in phrase-based statistical machine translation. In *Proceedings of the Human Language Technology Conference*, pages 257–264, 2004.