Leveraging Compounds to Improve Noun Phrase Translation from Chinese and German

Xiao Pu
Idiap Research Institute
1920 Martigny
Switzerland
xiao.pu@idiap.ch

Laura Mascarell
Institute of Computational Linguistics, U. of Zurich
8050 Zurich, Switzerland
mascarell@cl.uzh.ch

Andrei Popescu-Belis
Idiap Research Institute
1920 Martigny
Switzerland
apbelis@idiap.ch

Mark Fishel
Institute of Computational Linguistics, U. of Zurich
8050 Zurich, Switzerland
fishel@cl.uzh.ch

Ngoc-Quang Luong
Idiap Research Institute
1920 Martigny
Switzerland
nluong@idiap.ch

Martin Volk
Institute of Computational Linguistics, U. of Zurich
8050 Zurich, Switzerland
volk@cl.uzh.ch

Abstract

This paper presents a method to improve the translation of polysemous nouns, when a previous occurrence of the noun as the head of a compound noun phrase is available in a text. The occurrences are identified through pattern matching rules, which detect XY compounds followed closely by a potentially coreferent occurrence of Y, such as "Nordwand ... Wand". Two strategies are proposed to improve the translation of the second occurrence of Y: re-using the cached translation of Y from the XY compound, or post-editing the translation of Y using the head of the translation of XY. Experiments are performed on Chinese-to-English and German-to-French statistical machine translation, over the WIT3 and Text+Berg corpora respectively, with 261 XY/Y pairs each. The results suggest that while the overall BLEU scores increase only slightly, the translations of the targeted polysemous nouns are significantly improved.

1 Introduction

Words tend to be less ambiguous when considered in context, which partially explains the success of phrase-based statistical machine translation (SMT) systems. In this paper, we take advantage of this observation, and extend the disambiguation potential of n-grams to subsequent occurrences of their individual components. We assume that the translation of a noun-noun compound, noted XY, displays fewer ambiguities than the translations of its components X and Y. Therefore, on a subsequent occurrence of the head of XY, assumed to refer to the same entity as XY, we hypothesize that its previously-found translation offers a better and more coherent translation than the one proposed by an SMT system that is not aware of the compound.

Our claim is supported by results from experiments on Chinese-to-English (ZH/EN) and German-to-French (DE/FR) translation presented in this paper. In both source languages, noun-noun compounds are frequent, and will enable us to disambiguate subsequent occurrences of their head.

For instance, in the example in Figure 1, the Chinese compound 高跟鞋 refers to 'high heels', and the subsequent mention of the referent using only the third character (鞋) should be translated as 'heels'. However, the character 鞋 by itself could also be translated as 'shoe' or 'footwear', as observed with a baseline SMT system that is not aware of the XY/Y coreference.

Although the XY/Y configuration may not be very frequent in texts, errors in its translation are particularly detrimental to the understanding of a text, as they often conceal the coreference link between two expressions. Moreover, as we will show, such issues can be quite reliably corrected, and the proposed approach can later generalize to other configurations of noun phrase coreference.

Proceedings of the ACL-IJCNLP 2015 Student Research Workshop, pages 8–15,
Beijing, China, July 28, 2015.

Figure 1: Compound post-editing method illustrated on ZH/EN. The first translation of 高跟鞋 into 'heel' enables the correct translation of the subsequent occurrence of 鞋 as 'heel', by post-editing the baseline output 'shoes'.

The paper is organized as follows. In Section 2 we present the main components of our proposal: first, the rules for identifying XY/Y pairs, and then two alternative methods for improving the coherence of the translation of a subsequent mention Y, one based on post-editing and the other one based on caching, which builds upon initial experiments presented by Mascarell et al. (2014). In Section 3, we present our experimental setting. In Section 4, we evaluate our proposal on ZH/EN and DE/FR translation, demonstrating that the translation of nouns is indeed improved, mainly by automatic or human comparisons with the reference translation. We conclude with a brief discussion of related studies (Section 5) and with perspectives for future work (Section 6).

2 Description of the Method

2.1 Overview

We propose to use the translation of a compound XY to improve the translation of a subsequent occurrence of Y, the head of the XY noun phrase, in the following way, represented schematically in Figure 1 (details for each stage are given below).

First, the presence of XY/Y patterns is detected either by examining whether a compound XY is followed by an occurrence of Y, or, conversely, by examining for each Y candidate whether it appears as part of a previous compound XY. Distance constraints and additional filtering rules are implemented to increase the likelihood that XY and Y are actually co-referent, or at least refer to entities of the same type.

Second, each sentence is translated by a baseline SMT system, and the translation of the head Y of each compound XY is identified using the word alignment from the SMT decoder. This translation is used as the translation of a subsequent occurrence of Y either by caching the corresponding source/target word pair in the SMT or by post-editing the baseline SMT output. For instance, if the Chinese pair (蔬菜, 菜) is identified, where the first compound can unambiguously be translated into English by 'vegetable', then the translation of a subsequent occurrence of 菜 is enforced to 'vegetable'. This has the potential to improve over the baseline translation, because when considered individually, 菜 could also be translated as 'dish', 'greens', 'wild herbs', etc.

2.2 Identifying XY/Y Pairs

Chinese and German share a number of similarities regarding compounds. Although Chinese texts are not word-segmented, once this operation is performed, multi-character words in which all characters have individual meanings – such as the above-mentioned 蔬菜 ('vegetable') – are frequent. Similarly, in German, noun-noun compounds such as 'Bundesamt' ('Bund' + 'Amt', for Federal Bureau) or Nordwand ('Nord' + 'Wand', for North face) are frequent as well. While the identification of XY noun-noun compounds is straightforward with morpho-syntactic analysis

tools, the identification of a subsequent mention of the head noun, Y, and especially the decision whether this Y refers or not to the same entity XY, are more challenging issues. In other words, the main difficulty is to separate true XY/Y pairs from false positives.

To detect truly coreferent XY/Y pairs we narrow down the set of detected cases using handwritten rules that check the local context of Y. For example, only the cases where Y is preceded by demonstrative pronouns (e.g. 这 or 那 meaning 'this' and 'that' in Chinese, or 'diese' in German), possessive pronouns and determiners ('der', 'die', 'das' in German) are considered. Since other words can occur between the two parts (like classifiers in Chinese or adjectives), there are additional distance constraints: the pronoun or determiner must be separated by fewer than three words. Since the rules use morphological information and word boundaries, they are preceded by word segmentation[1] and tagging[2] for Chinese and morphological analysis for German.[3] For example, in the input sentence from Figure 1, we determine that the noun phrase 鞋 fits our condition for extraction as Y because as there are words before it which fulfill the condition for acceptance.

2.3 Enforcing the Translation of Y

Two language-independent methods have been designed to ensure that the translations of XY and Y are a consistent: post-editing and caching. The second one builds upon an earlier proposal tested only on DE/FR with subjective evaluations (Mascarell et al., 2014).

In the post-editing method, for each XY/Y pair, the translations of XY and Y by a baseline SMT system (see Section 3) are first identified through word alignment. We verify if the translations of Y in both noun phrases are identical or different. Both elements comprising the compound structure XY/Y are identified, for the standard cases, with only one possible XY referring to one Y. The translation of both words are provided by the baseline SMT system, and our system subsequently verifies if the translations of Y in both noun phrases are identical or different. We keep them intact in the first case, while in the second case we replace the translation of Y by the translation of XY or by its head noun only, if it contains several words. In the example in Figure 1, XY is translated into 'high heel' and Y into 'shoes', which is a wrong translation of 鞋 in this context. Using the consistency constraint, our method post-edits the translation of Y replacing it with 'heel', which is the correct word.

Several differences from the ideal case presented above must be handled separately. First, it may occur that several XY are likely co-referent with the same Y. In this case, if their translations differ, given that we cannot resolve the coreference, we do not post-edit Y.[4] If the translations of the several occurrences of XY are the same, but consist of one word, we still do not post-edit Y. We only change it if the translations consist of several words, ensuring that XY is a compound noun phrase. Second, if the compound XY is not translated (out-of-vocabulary word), we do not post-edit Y.[5] Third, sometimes the alignment of Y is empty in the target sentence (alignment error or untranslated word), in which case we apply post-editing as above on the word preceding Y, if it is aligned.

In the caching method (Mascarell et al., 2014), once an XY compound is identified, we obtain the translation of the Y part of the compound through the word alignment given by the SMT decoder. Next, we check that this translation appears as a translation of Y in the phrase table, and if so, we cache both Y and the obtained translation. We then enforce the cached translation every time a coreference Y to XY is identified. Note that this is different from the probabilistic caching proposed by Tiedemann (2010), because in our case the cached translation is deterministically enforced as the translation of Y.

3 Experimental Settings

The experiments are carried out on two different parallel corpora: the WIT[3] Chinese-English dataset (Cettolo et al., 2012) with transcripts of TED lectures and their translations, and the Text+Berg German-French corpus (Bubenhofer et al., 2013), a collection of articles from the year-

[1]Using the Stanford Word Segmenter available from http://nlp.stanford.edu/software/segmenter.shtml.

[2]Using the Stanford Log-linear Part-of-speech Tagger, http://nlp.stanford.edu/software/tagger.shtml.

[3]Using Gertwol (Koskeniemmi and Haapalainen, 1994).

[4]Upon manual examination, we found that using the most recent XY was not a reliable candidate for the antecedent.

[5]In fact, we can use the translation of Y as a translation candidate for XY. Our observations show that this helps to improve BLEU scores, but does not affect the specific scoring of Y in Section 4.

		Sentences	Tokens
ZH	Training	188'758	19'880'790
	Tuning	2'457	260'770
	Testing	855	12'344
DE	Training	285'877	5'194'622
	Tuning	1'557	32'649
	Testing	505	12'499

Table 1: Sizes of SMT data sets.

books of the Swiss Alpine Club. The sizes of the subsets used for training, tuning and testing the SMT systems are given in Table 1. The test sets were constructed by selecting all the sentences or fragments which contained the XY/Y pairs, identified as above, to maximize their number in the test data, given that they are not needed in the training/tuning sets, as the proposed methods are not based on machine learning.

The rules for selecting coreferent XY/Y pairs in Chinese identified 261 pairs among 192k sentences. The rather low rate of occurrence (about one every 700 sentences) is explained by the strict conditions of the selection rules, which are designed to maximize the likelihood of coreference. In German, less restrictive rules selected 7,365 XY/Y pairs (a rate of one every 40 sentences). Still, in what follows, we randomly selected 261 XY/Y pairs for the DE/FR test data, to match their number in the ZH/EN test data.

Our baseline SMT system is the Moses phrase-based decoder (Koehn et al., 2007), trained over tokenized and true-cased data. The language models were built using SRILM (Stolcke et al., 2011) at order 3 (i.e. up to trigrams) using the default smoothing method (i.e. Good-Turing). Optimization was done using Minimum Error Rate Training (Och, 2003) as provided with Moses.

The effectiveness of proposed systems is measured in two ways. First, we use BLEU (Papineni et al., 2002) for overall evaluation, to verify whether our systems provide better translation for entire texts. Then, we focus on the XY/Y pairs and count the number of cases in which the translations of Y match the reference or not, which can be computed automatically using the alignments.

However, the automatic comparison of a system's translation with the reference is not entirely informative, because even if the two differ, the system's translation can still be acceptable. Therefore, we analyzed these "undecided" situations manually, with three human annotators (among the authors of the paper). The annotators rated separately the system's translations of Y and the reference ones as 'good', 'acceptable' or 'wrong'.

4 Analysis of Results

4.1 Automatic Comparison with a Reference

The BLEU scores obtained by the baseline SMT, the caching and post-editing methods, and an oracle system are given in Table 2. The scores are in the same range as the baseline scores found by other teams on these datasets (Cettolo et al., 2012, Table 7 for ZH/EN), and much higher on DE/FR than ZH/EN.

Our methods have a small positive effect on ZH/EN translation, and a small negative effect on DE/FR one. Given the sparsity of XY/Y pairs with respect to the total number of words, hence the small number of changed words, these results meet our prior expectations. Indeed, we also computed the oracle BLEU scores for both language pairs, i.e. the scores when all Y members of XY/Y pairs are (manually) translated exactly as in the reference (last line of Table 2). These values are only slightly higher than the other scores, showing that even a perfect translation of the Y nouns would only have a small effect on BLEU.

	ZH/EN	DE/FR
BASELINE	11.18	27.65
CACHING	11.23	27.26
POST-EDITING	11.27	27.48
ORACLE	11.30	27.80

Table 2: BLEU scores of our methods.

We now turn to the reference-based evaluation of the translations of Y in the 261 XY/Y pairs, comparing the baseline SMT with each of our methods. These results are represented as four contingency tables – two language pairs and two methods against the baseline – gathered together as percentages in Table 3. Among these values, we focus first on the total of pairs where one of our systems agrees with the reference while the baseline system does not (i.e., improvements due to the system), and the converse case (degradations). The higher the difference between the two values, the more beneficial our method.

For ZH/EN and the post-editing system, among the 222 extracted pairs, there were 45 improvements (20.3%) of the system with respect to the

			CACHING		POST-EDITING	
			= ref	≠ ref	= ref	≠ ref
ZH/EN	BASELINE	= ref	59.3	*4.1*	42.3	*4.5*
		≠ ref	**13.8**	22.8	**20.3**	32.9
DE/FR	BASELINE	= ref	70.1	*10.3*	73.9	*5.0*
		≠ ref	**4.3**	15.2	**3.5**	17.5

Table 3: Comparison of each approach with the baseline, for the two language pairs, in terms of Y nouns which are identical or different from a reference translation ('ref'). All scores are percentages of the totals. Numbers in **bold** are improvements over the baseline, while those in *italics* are degradations.

baseline, and only 10 degradations (4.5%). There were also 94 pairs (42.3%) for which the baseline and the post-edited system were equal to the reference. The remaining 73 pairs (32.9%) will be analyzed manually in the next section. Therefore, from a pure reference-based view, the post-edited system has a net improvement of 15.8% (absolute) over the baseline in dealing with the XY/Y pairs.

A similar pattern is observed with the other method, namely caching, again on ZH/EN translation: 13.8% improvements vs. 4.1% degradations. The difference (i.e. the net improvement) is slightly smaller in this case with respect to the post-editing method.

For DE/FR translation, both methods appear to score fewer improvements than degradations. There are more than 70% of the pairs which are translated correctly by the baseline and by both systems, which indicates that the potential for improvement is much smaller for DE/FR than for ZH/EN.

While the pattern of improvement between ZH/EN and DE/FR is similar for post-editing and for caching, for both language pairs the post-editing method has a larger difference between improvements and degradations than the caching method. This can be explained by a lower coverage of the latter method, since it only enforces a translation when it appears as one of the translation candidates for Y in the phrase table (Mascarell et al., 2014).

4.2 Manual Evaluation of Undecided Cases

When both the baseline and one of our systems generate translations of Y which differ from the reference, it is not possible to compare the translations without having them examined by human subjects. This was done for the 73 such cases of the ZH/EN post-editing system. Three of the authors, working independently, considered each translation from each system (in separate batches) with respect to the reference one, and rated its meaning on a 3-point scale: 2 (good), 1 (acceptable) or 0 (wrong). To estimate the inter-rater agreement, we computed the average absolute deviation[6] and found a value of 0.15, thus denoting very good agreement. Below, we group '2' and '1' answers into one category, called "acceptable", and compare them to '0' answers, i.e. wrong translations.

When both the baseline and the post-edited translations of Y differ from the reference, they can either be identical (49 cases) or different (24). In the former case, of course, neither of the systems outperforms the other. The interesting observation is that the relatively high number of such cases (49) is due to situations where the reference translation of noun Y is by a pronoun (40), which the systems have currently no possibility to generate from a noun in the source sentence. Manual evaluation shows that the systems' translations are correct in 36 out of 40 cases. This large number shows that the "quality" of the systems is actually higher than what can be inferred from Table 3 only. Conversely, in the 9 cases when the reference translation of Y is not a pronoun, only about half of the translations are correct.

In the latter case, when baseline and post-edited translations differ from the reference *and* among themselves (24 cases), it is legitimate to ask which of the two systems is better. Overall, 10 baseline translations are correct and 14 are wrong, whereas 23 post-edited translations are correct (or at least acceptable) and only one is wrong. The post-edited system thus clearly outperforms the baseline in this case. Similarly to the observation above, we note that among the 24 cases considered here, almost all (20) involve a reference translation of Y by a pronoun. In these cases, the baseline

[6] Average of $\frac{1}{3}\sum_{i=1}^{3}|\text{score}_i - \text{mean}|$ over all ratings .

system translates only about half of them with a correct noun (9 out of 20), while the post-edited system translates correctly 19 out of 20.

5 Related Work

We briefly review in this section several previous studies from which the present one has benefited. Our idea is built upon the one-sense-per-discourse hypothesis (Gale et al., 1992) and its application to machine translation is based on the premise that consistency in discourse (Carpuat, 2009) is desirable. The initial compound idea was first published by Mascarell et al. (2014), in which the coreference of compound noun phrases in German (e.g. Nordwand/Wand) was studied and used to improve DE/FR translation by assuming that the last constituent of the compound Y should share the same translation as that of Y in XY.

Several other approaches focused on enforcing consistent lexical choice. Tiedemann (2010) proposed a cache-model to enforce consistent translation of phrases across the document. However, caching is sensitive to error propagation, that is, when a phrase is incorrectly translated and cached, the model propagates the error to the following sentences. Gong et al. (2011) later extended Tiedemann's proposal by initializing the cache with phrase pairs from similar documents at the beginning of the translation and by also applying a topic cache, which was introduced to deal with the error propagation issue. Xiao et al. (2011) defined a three step procedure that enforces the consistent translation of ambiguous words, achieving improvements for EN/ZH. Ture et al. (2012) encouraged consistency for AR/EN MT by introducing cross-sentence consistency features to the translation model, while Alexandrescu and Kirchhoff (2009) enforced similar translations to sentences having a similar graph representation.

Our work is an instance of a recent trend aiming to go beyond sentence-by-sentence MT, by using semantic information from previous sentences to constrain or correct the decoding of the current one. In this paper, we compared caching and post-editing as ways of achieving this goal, but a document-level decoder such as Docent (Hardmeier et al., 2012) could be used as well. In other studies, factored translation models (Koehn and Hoang, 2007) have been used with the same purpose, by incorporating contextual information into labels used to indicate the meaning of ambiguous discourse connectives (Meyer and Popescu-Belis, 2012) or the expected tenses of verb phrase translations (Loaiciga et al., 2014). Quite naturally, there are analogies between our work and studies of pronoun translation (Le Nagard and Koehn, 2010; Hardmeier and Federico, 2010; Guillou, 2012), with the notable difference that pronominal anaphora resolution remains a challenging task. Finally, our work and its perspectives contribute to the general objective of using discourse-level information to improve MT (Hardmeier, 2014; Meyer, 2014).

6 Conclusion and Perspectives

We presented a method to enforce the consistent translation of coreferences to a compound, when the coreference matches the head noun of the compound. Experimental results showed that baseline SMT systems often translate coreferences to compounds consistently for DE/FR, but much less so for ZH/EN. For a significant number of cases in which the noun phrase Y had multiple meanings, our system reduced the frequency of mistranslations in comparison to the baseline, and improved noun phrase translation.

In this work, we considered XY/Y pairs, hypothesizing that when they are coreferent, they should have consistent translations. In the future, we will generalize this constraint to complex noun phrases which are not compounds. More generally, we will explore the encoding of coreference constraints into probabilistic models that can be combined with SMT systems, so that coreference constraints are considered in the decoding process.

Acknowledgments

The authors are grateful for the support of the Swiss National Science Foundation (SNSF) through the Sinergia project MODERN: Modeling Discourse Entities and Relations for Coherent Machine Translation, grant nr. CRSII2_147653 (`www.idiap.ch/project/modern`).

References

Andrei Alexandrescu and Katrin Kirchhoff. 2009. Graph-based learning for statistical machine translation. In *Proceedings of the Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*, pages 119–127, Boulder, Colorado.

Noah Bubenhofer, Martin Volk, David Klaper, Manuela Weibel, and Daniel Wüest. 2013. Text+Berg-korpus (release 147_v03). Digitale Edition des Jahrbuch des SAC 1864-1923, Echo des Alpes 1872-1924 und Die Alpen 1925-2011.

Marine Carpuat. 2009. One Translation per Discourse. In *Proceedings of the Workshop on Semantic Evaluations: Recent Achievements and Future Directions (SEW)*, pages 19–27, Singapore.

Mauro Cettolo, Christian Girardi, and Marcello Federico. 2012. WIT^3: Web inventory of transcribed and translated talks. In *Proceedings of the 16th Conference of the European Association for Machine Translation (EAMT)*, pages 261–268, Trento, Italy.

William A Gale, Kenneth W Church, and David Yarowsky. 1992. One sense per discourse. In *Proceedings of the Workshop on Speech and Natural Language*, pages 233–237.

Zhengxian Gong, Min Zhang, and Guodong Zhou. 2011. Cache-based document-level statistical machine translation. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 909–919, Edinburgh.

Liane Guillou. 2012. Improving pronoun translation for statistical machine translation. In *Proceedings of EACL 2012 Student Research Workshop (13th Conference of the European Chapter of the ACL)*, pages 1–10, Avignon, France.

Christian Hardmeier and Marcello Federico. 2010. Modelling Pronominal Anaphora in Statistical Machine Translation. In *Proceedings of International Workshop on Spoken Language Translation (IWSLT)*, Paris, France.

Christian Hardmeier, Joakim Nivre, and Jörg Tiedemann. 2012. Document-Wide Decoding for Phrase-Based Statistical Machine Translation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing and Natural Language Learning (EMNLP-CoNLL)*, Jeju, Korea.

Christian Hardmeier. 2014. *Discourse in Statistical Machine Translation*. PhD thesis, Uppsala University, Sweden.

Philipp Koehn and Hieu Hoang. 2007. Factored translation models. In *Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing (EMNLP) and Computational Natural Language Learning (CONLL)*, pages 868–876, Prague, Czech Republic.

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 Constantin, and Evan Herbs. 2007. Moses: Open Source Toolkit for Statistical Machine Translation. In *Proceedings of 45th Annual Meeting of the Association for Computational Linguistics (ACL), Demonstration Session*, pages 177–180, Prague, Czech Republic.

Kimmo Koskeniemmi and Mariikka Haapalainen. 1994. Gertwol–lingsoft oy. *Linguistische Verifikation: Dokumentation zur Ersten Morpholympics*, pages 121–140.

Ronan Le Nagard and Philipp Koehn. 2010. Aiding pronoun translation with co-reference resolution. In *Proceedings of the Joint 5th Workshop on Statistical Machine Translation and Metrics (MATR)*, pages 258–267, Uppsala, Sweden.

Sharid Loaiciga, Thomas Meyer, and Andrei Popescu-Belis. 2014. English-French Verb Phrase Alignment in Europarl for Tense Translation Modeling. In *Proceedings of the 9th international conference on Language Resources and Evaluation (LREC)*, Reykjavik, Iceland.

Laura Mascarell, Mark Fishel, Natalia Korchagina, and Martin Volk. 2014. Enforcing consistent translation of German compound coreferences. In *Proceedings of the 12th Konvens Conference*, Hildesheim, Germany.

Thomas Meyer and Andrei Popescu-Belis. 2012. Using sense-labeled discourse connectives for statistical machine translation. In *Proceedings of the EACL 2012 Joint Workshop on Exploiting Synergies between IR and MT, and Hybrid Approaches to MT (ESIRMT-HyTra)*, pages 129–138, Avignon, France.

Thomas Meyer. 2014. *Discourse-level Features for Statistical Machine Translation*. PhD thesis, EPFL, Lausanne.

Franz Josef Och. 2003. Minimum Error Rate Training in Statistical Machine Translation. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 160–167, Sapporo, Japan.

Kishore Papineni, Salim Roukos, Todd Ard, and Wei-Jing Zhu. 2002. BLEU: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*.

Andreas Stolcke, Jing Zheng, Wen Wang, and Victor Abrash. 2011. SRILM at Sixteen: Update and Outlook. In *Proceedings of the IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, Waikoloa, Hawaii.

Jörg Tiedemann. 2010. Context adaptation in statistical machine translation using models with exponentially decaying cache. In *Proceedings of the 2010 Workshop on Domain Adaptation for Natural Language Processing*, pages 8–15, Uppsala, Sweden.

Ferhan Ture, Douglas W. Oard, and Philip Resnik. 2012. Encouraging consistent translation choices. In *Proceedings of the 2012 Conference of the North*

American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), pages 417–426, Montréal, Canada.

Tong Xiao, Jingbo Zhu, Shujie Yao, and Hao Zhang. 2011. Document-level consistency verification in machine translation. In *Proceedings of the 13th Machine Translation Summit*, pages 131–138, Xiamen, China.