Using BabelNet to Improve OOV Coverage in SMT

Jinhua Du,[1] **Andy Way,**[1] **Andrzej Zydron**[2]

[1]ADAPT Centre, School of Computing, Dublin City University, Dublin, Ireland
[2]XTM International Ltd., London, UK

jdu@computing.dcu.ie, away@computing.dcu.ie, azydron@xtm-intl.com

Abstract

Out-of-vocabulary words (OOVs) are a ubiquitous and difficult problem in statistical machine translation (SMT). This paper studies different strategies of using BabelNet to alleviate the negative impact brought about by OOVs. BabelNet is a multilingual encyclopedic dictionary and a semantic network, which not only includes lexicographic and encyclopedic terms, but connects concepts and named entities in a very large network of semantic relations. By taking advantage of the knowledge in BabelNet, three different methods – using direct training data, domain-adaptation techniques and the BabelNet API – are proposed in this paper to obtain translations for OOVs to improve system performance. Experimental results on English–Polish and English–Chinese language pairs show that domain adaptation can better utilize BabelNet knowledge and performs better than other methods. The results also demonstrate that BabelNet is a really useful tool for improving translation performance of SMT systems.

Keywords: BabelNet, SMT, unknown words, OOVs, domain adaptation

1. Introduction

OOVs – source words that have no translation in the phrase table – are a ubiquitous and difficult problem for SMT. Due to the fact that they are trained on pre-defined static data sets, SMT systems necessarily encounter OOVs when translating new documents. In such circumstances, there are two main strategies deployed: (i) to output the source word 'as is' on the target side; or (ii) to omit it altogether. Of course, in both cases, erroneous and disfluent translations are produced. The problem is exacerbated when bilingual data are scarce, or if the text to be translated is not from the same domain as the training data.

OOVs are often named entities, specialized terms and neologisms. For example, some person names or technical terms are phonetically transliterated from other languages. In the past, plenty of work has been done to alleviate the impact of OOVs, including orthographic (lexicon-induction-based) and morphosyntactic preprocessing (Popovic and Ney, 2004; Sadat and Habash, 2006; Habash, 2008; Garera et al., 2009), pivot languages (Callison-Burch et al., 2006), grapheme-based model for phonetic transliterations (Lehal and Saini, 2012; Luo et al., 2013), paraphrases (Habash, 2008; Marton et al., 2009; Du et al., 2010) and context-based semantic models (Haghighi et al., 2008; Daume-III and Jagarlamudi, 2011; Zhang et al., 2012).

In this paper, we propose to use BabelNet to handle OOVs. BabelNet is both a multilingual encyclopedic dictionary, with lexicographic and encyclopedic coverage of terms, as well as a semantic network comprising 14 million entries which connects concepts and named entities in a very large network of semantic relations. These are called Babel synsets, each of which represents a given meaning and contains all the synonyms which express that meaning in a range of different languages (Navigli and Ponzetto, 2010; Roberto Navigli, 2012; Navigli, 2012). The most recent version of BabelNet is 3.5 which integrates knowledge from WordNet, Wikipedia, Microsoft Terminology, ImageNet etc.[1]

BabelNet has been applied in many natural language processing tasks, such as multilingual lexicon extraction, crosslingual word-sense disambiguation, annotation, and information extraction, all with good performance (Elbedweihy et al., 2013; Jadidinejad, 2013; Navigli et al., 2013; Ehrmann et al., 2014; Moro et al., 2014). However, to the best of our knowledge, to date there is no comprehensive work applying BabelNet knowledge to SMT tasks. In this paper, we present three different strategies to utilize BabelNet resources in SMT systems, namely using direct training data, domain adaptation and OOV post-processing approaches. Specifically, the first strategy very straightforwardly appends the bilingual dictionary extracted from BabelNet to the initial training data, and then verifies the impact on translation performance; the second uses domain-adaptation methods to select in-domain entries from the extracted bilingual dictionary, which are then added to the initial training data in a similar manner; finally we directly call the BabelNet API to post-process OOVs contained in the translation of the source sentence. Experiments conducted on different language pairs show that the second and third strategies are more robust and effective than the first one in augmenting SMT systems.

The remainder of the paper is as follows. Section 2 reviews related work. In Section 3, we present three different strategies to use BabelNet to handle OOVs. Section 4 describes our experiments and analysis. In Section 5, we conclude and provide some avenues for future research.

2. Related Work

There has been a long line of research on handling OOVs in SMT. In this section, we briefly introduce some representative work in terms of the methods of processing OOVs.

Lexicon-induction-based and morpho-syntactic methods are commonly used for handling unknown words by creating a bilingual lexicon for OOVs. By extending this work,

---

[1]http://babelnet.org/about

Habash (2008) presents techniques for online treatment of OOVs for Arabic-to-English such as spelling expansion and morphological expansion. Huang et al. (2011) propose to combine sublexical/constituent translations of an OOV word or phrase to generate its translations.

Pivot language techniques take advantage of available parallel data between the source language and a third language to handle the problem of OOVs. Using a pivot language, OOVs are translated into a third language and back into the source language and thereby paraphrases for those OOVs are extracted (Callison-Burch et al., 2006).

Semantic methods are based on the distributional hypothesis that words appearing in the same contexts tend to have similar meanings. Paraphrases can express similar meaning of different words or phrases that are useful to alleviate the OOV problem. Marton et al. (2009) use a monolingual text on the source side to find paraphrases for OOVs for which translations are available. The translations of these paraphrases are then used as the translations of the OOV word. Du et al. (2010) constructed a source-side paraphrase lattice to handle OOVs and allow the decoder to decide which paraphrase candidate is the best option for the translation.

Instead of replacing OOVs, Zhang et al. (2012) propose a different way of using semantics for handling OOVs. They focus on keeping the untranslated words in the correct position in the translation, i.e. employing the distributional semantic model and the bidirectional language model to determine the semantic function which the unknown words serve in the test sentence, and keeping the semantic function unchanged in the translation process. In this way, unknown words will help the phrase reordering and lexical selection of their surrounding words even though they themselves still remain untranslated.

For OOVs that are transliterations, a grapheme-based model maps directly from source graphemes to target graphemes. In this model, phonetic information or pronunciation is used, and thus an additional processing step of converting source grapheme to source phoneme is required. For example, Lehal and Saini (2012) propose a hybrid transliteration approach using both the grapheme-based transliteration model and the phoneme-based model. Different from the methods above, we utilize an extra semantic resource to handle the problem of OOVs. Specifically, we use BabelNet (i) as direct parallel data, or (ii) to retrieve the translations of OOVs via an API call. Experimental results on different language pairs show that BabelNet is helpful in improving translation performance of an SMT system.

3. Strategies for Using BabelNet in SMT

In this section, we introduce three different strategies to utilize BabelNet resources in SMT.

3.1. Direct Training Data

This strategy is straightforward. We first retrieve the whole source and target terms in BabelNet and then perform set of word alignment to obtain a bilingual dictionary. Then clean up the dictionary and add it to the initial training data. The entries in the dictionary are not only single words, but also multi-word expressions.

3.2. Domain Adaptation

As mentioned, BabelNet integrates many knowledge resources, so its entries come from many different domains. However, for practical use, we often use domain-specific data sets to build SMT systems, e.g. Newswire, Europarl (Koehn, 2005), DGT etc. In order to make good use of the knowledge in the BabelNet dictionary rather than simply adding it to the training data, we propose to use different domain adaptation methods, namely the entry-match strategy and Cross Entropy Difference (CED) (Moore and Lewis, 2010; Axelrod et al., 2011; Haque et al., 2014), to select in-domain entries.

The entry match strategy uses a straightforward way to select in-domain entries, i.e. if either the source side or the target side of one entry-pair occurs in the initial training data, then this entry-pair will be selected. The advantage of this method is that it can increase the probability estimation of existing words. However, the disadvantage is that it cannot handle OOVs.

We also use a domain-adaptation method which has a good performance record – CED – to facilitate entry selection from the bilingual dictionary. In this method, given an in-domain (ID) corpus I and a general corpus O, language models are built from both, and each sentence s in O is scored according to the entropy difference, as in (1):

$$score(s) = H_I(s) - H_O(s) \qquad (1)$$

where $H_I(s)$ is the entropy of s in the ID corpus I, and $H_O(s)$ is the entropy of s in the out-of-domain (OOD) corpus O. The advantage of this method is that it can handle OOVs to some extent. However, the potential problem is that the selection accuracy might be affected by short entries, i.e. there are many entries that only contain either a single word or a pair of words, which may cause some context information to be lost for domain adaptation. Therefore, in order to minimize this influence, we use bigrams to build the language models.

3.3. OOV Post-processing: BabelNet API

There is a lot of noise in the extracted dictionary. For example, on the Chinese side of the English–Chinese dictionary, an entry might occur in Simplified or Traditional Chinese, or an entry might be segmented into words or characters. More importantly, there are many possible target terms for a given source term which come from different domains. In our experiments, using BabelNet bilingual entries as the training data does not perform well, so we propose to directly call the BabelNet API (Navigli and Ponzetto, 2012) to post-process OOVs in the translation of a source sentence.

We use the BabelNet API to call the precompiled index bundle (version: 2.5.1)[2] to retrieve the translation for an OOV. Specifically, an OOV in the source sentence is automatically marked in the output of the SMT decoder, and then we recognize this OOV and retrieve its 1-best candidate by calling the BabelNet API. Finally, we replace the OOV in the target side by the candidate translation.

---

[2]http://babelnet.org/download

English – Training Data			Polish – Training Data		
#sen	#word	#entry	#sen	#word	#entry
518,155	11,270,214	52,247	518,155	9,743,192	144,146

Table 1: Statistics of Europarl EN–PL data for the model training

English – Test Set			Polish – Test Set		
#sen	#word	#entry	#sen	#word	#entry
2,000	47,194	4,063	2,000	39,956	7,451

Table 2: Statistics of Europarl EN–PL data for the test set

Chinese – Training Data			English – Training Data		
#sen	#word	#entry	#sen	#word	#entry
270,794	9,582,189	102,035	270,794	10,319,019	81,036

Table 3: Statistics of FBIS ZH–EN data for model training

Chinese – Test Set			English – Test Set (4 Refs)		
#sen	#word	#entry	#sen	#word	#entry
1,082	30,489	5,684	1,082	142,794	7,552

Table 4: Statistics of FBIS ZH–EN data for the test set

4. Experiments and Analysis

4.1. Experimental Settings

We use Moses (Koehn et al., 2007) as the SMT system and configure the argument '–mark-unknown' to mark up the OOVs in the translation. Experiments are conducted on English–Polish (EN–PL), English–Chinese (EN–ZH) and Chinese–English (ZH–EN) translation tasks.

4.2. Experiments on Strategy 1: Direct Training Data (DTD)

In this section, we verify the impact of Strategy 1 on system performance, i.e. directly adding the bilingual dictionary to the training data to build the SMT system.

4.2.1. Data Statistics

Regarding the EN–PL task, the initial training data comes from Europarl which contains 518,155 sentence pairs, and the devset and test set each contain 2,000 sentence pairs which are randomly extracted from the Europarl data set. The statistics of the data are shown in Table 1 and Table 2. Regarding the ZH–EN and EN–ZH tasks, the training data comes from NIST FBIS that contains 270,794 sentence pairs, the devset is the NIST 2006 current set that includes 1,664 sentences with 4 references for each, and the test set is the NIST 2005 current set that contains 1,082 sentences with 4 references for each. The statistics of the data are shown in Table 3 and Table 4.

4.2.2. BabelNet Bilingual Dictionary and Denoising

The EN–PL BabelNet dictionary contains 6,199,888 bilingual entries. However, the raw data contains a lot of noise, so we performed some pre-processing of the dictionary, including:

- if East_Asian characters are included in either the English or Polish side, we remove this pair;

- if the English side contains symbols which are neither letters nor digits, then we remove this pair;

- if either side contains punctuation, then we remove this pair;

- if the English side is the same as the Polish side, then we remove this pair;

- if the ratio of the word-level entry lengths between the English side and the Polish side is less than 0.5 or greater than 2, then we remove it. This rule is based on the fact that 99% of pairs of EN–PL sentences in Europarl training data fall within this range.

After this clean-up, we are left with $2,215,248$ pairs. Note that almost 4 million sentence-pairs – or 64.3% of the original date – are filtered out, indicating that the EN–PL data is overall not of high quality. Nonetheless, 2.2 million sentence-pairs is still more than a reasonable amount of additional data for MT engine training.

The ZH–EN BabelNet dictionary contains 5,975,619 bilingual entries. As we did for the EN–PL dictionary, we also filtered the ZH–EN dictionary as follows:

- many Chinese characters are encoded as UTF-8 Traditional format (BIG5), so we convert them to UTF-8 Simplified format (GBK);

- if the English side contains symbols which are neither letters nor digits, then we remove this pair;

system	EN–PL		EN–ZH		ZH–EN	
	BLEU4(%)	TER(%)	BLEU4(%)	TER(%)	BLEU4(%)	TER(%)
Baseline	**24.57**	**58.47**	11.62	72.03	27.08	67.90
DTD-CLEAN-1	24.15	58.91	**12.53**	**71.25**	**27.53**	**66.20**
DTD-CLEAN-10	24.12	59.09	12.36	71.34	26.60	67.82

Table 5: Results of Strategy 1 on two language pairs

	EN–PL			EN–ZH		
	CEM	UEM	CED	CEM	UEM	CED
#entry	134,057	715,404	868,423	284,990	1,360,244	879,257

Table 6: Statistics of the entry match methods on different language pairs

system	EN–PL		EN–ZH		ZH–EN	
	BLEU4(%)	TER(%)	BLEU4(%)	TER(%)	BLEU4(%)	TER(%)
Baseline	24.57	58.47	11.62	72.03	27.08	67.90
DTD-CLEAN-1	24.15	58.91	12.53	71.25	27.53	66.20
CEM	24.71	58.29	12.23	71.50	27.91	66.05
UEM	24.59	58.31	**12.76**	**70.81**	**28.47**	**65.78**
CED	**24.72**	**58.29**	12.73	71.24	27.23	66.97

Table 7: Results of domain adaptation methods on different settings

- if the Chinese side does not contain Chinese characters, then we remove this pair;

- if the English side is the same as the Chinese side, then we remove this pair;

- we first remove the spaces between any characters on the Chinese side, and then re-segment into words to remain consistent with the training data.

Finally, we obtain $5,493,323$ Chinese–English pairs from the original $5,975,619$ pairs. That is, far less data is filtered out than for EN–PL; for ZH–EN, just 8% of the original data is discarded, indicating that on the whole, the ZH–EN data is of superior quality than the EN–PL entries.

4.2.3. Experimental Results

The results for EN–PL and ZH–EN language pairs on different data sets are shown in Table 5, where 'Baseline' indicates the system does not contain any BabelNet entries; 'DTD-CLEAN-1' indicates that the cleaned BabelNet entries are only repeated once in the training data; and 'DTD-CLEAN-10' indicates that the cleaned BabelNet entries are repeated 10 times in the training data.

It can be seen that (i) using the BabelNet dictionary as the training data on the EN–PL task does not result in better performance; (ii) however, it achieves better performance on ZH–EN and EN–ZH tasks compared to the baselines; (iii) repeating the occurrences of the BabelNet entries does not improve the quality, but actually results in worse performance than when using DTD-CLEAN-1; (iv) for different languages, BabelNet has an unstable impact on system performance.

As far as the above results are concerned, we infer that (i) the domain of the BabelNet dictionary might have a sig-

nificant difference compared to that of the initial training data; and (ii) the entries need to be further cleaned up to obtain more domain-related data. Therefore, we propose to use domain-adaptation strategies to verify the contribution of BabelNet as described in the next section.

4.3. Experiments on Strategy 2: Domain Adaptation

In this section, we use two different domain-adaptation methods to select in-domain entries from the cleaned BabelNet dictionary as described in Section 3.2.

4.3.1. Entry Match

We utilize two entry-match methods:

- Co-occurrence Entry Match (CEM): if both the source side and the target side of an entry occur in the initial training data, then we select it;

- Unilateral Entry Match (UEM): if either the source side or the target side of an entry occurs in the initial training data, we select it.

The statistics regarding the selected entries are shown in Table 6.

4.3.2. CED

The numbers of selected entries using the CED method are shown in Table 6. Compared to the other data selection methods, quite similar amounts of data are selected for both EN–PL and EN–ZH using this approach, with around 870K BabelNet entries.

4.3.3. Experimental Results

The results for EN–PL and ZH–EN language pairs on these different domain-adaptation methods are shown in Table 7.

Task	#OOVs	#Translated	Ratio (%)
EN–PL	246	70	28.46
EN–ZH	940	325	34.57
ZH–EN	1011	211	20.87

Table 8: Statistics of OOVs in experiments of Strategy 3

system	EN–PL		EN–ZH		ZH–EN	
	BLEU4(%)	TER(%)	BLEU4(%)	TER(%)	BLEU4(%)	TER(%)
Baseline	24.57	58.47	11.62	72.03	27.08	67.90
DTD-CLEAN-1	24.15	58.91	12.53	71.25	27.53	66.20
Best-DoAdpt	**24.72**	**58.29**	**12.76**	**70.81**	**28.47**	**65.78**
API	24.71	58.36	11.71	71.84	27.23	67.77

Table 9: Comparison between BabelNet API method and others

In terms of BLEU (Papineni et al., 2002) and TER (Snover et al., 2006), we can make the following observations: (i) all three domain adaptation methods (CEM, UEM and CED) outperform the baselines on both EN–ZH and ZH–EN tasks; (ii) the UEM method performs best on EN–ZH and ZH–EN tasks, while CED performs best on EN–PL; (iii) for the EN–PL task, all domain-adaptation methods perform better than DTD-CLEAN-1 which shows that the data-selection methods are better at removing out-of-domain and noisy data.

Based on these results, we can say that the UEM and CED methods are effective and feasible in selecting useful or in-domain data from the noisy, out-of-domain dictionary.

4.4. Experiments on Strategy 3: BabelNet API

The statistics of OOVs occurring in the test sets in terms of EN–PL, EN–ZH and ZH–EN tasks are shown in Table 8. Here '#Translated' indicates the number of translations of OOVs that can be retrieved in BabelNet; the last column indicates the ratio of how many OOVs are translated. We can see that only a small proportion of OOVs can be translated by BabelNet.

The results of calling the BabelNet API to process OOVs are shown in Table 9. 'Best-DoAdpt' gives the best result of all domain-adaption methods. 'API' refers to the offline BabelNet API-call method. We can see that the BabelNet API method did not beat the best domain-adaptation method on all tasks in terms of BLEU and TER. However, it does improve system performance compared to the baselines, which shows that using BabelNet can alleviate the issue of unknown words to some extent.

However, the improvements are not significant (Koehn, 2004). The possible reasons for this include:

- Most OOVs cannot be retrieved by BabelNet;

- Due to lack of context, the retrieved translation for an OOV might not be correct;

- Some retrieved translations need to be processed further, e.g. using BIG5 and simplified encoding for Chinese, tokenization or segmentation etc.

From these results, we can see that the domain-adaptation methods are more effective in utilizing BabelNet resources.

5. Conclusions and Future Work

In this paper, we proposed three strategies for using BabelNet to augment SMT, namely using direct training data, domain adaptation and BabelNet API methods. Experiments conducted on EN–PL, EN–ZH and ZH–EN tasks show that the domain-adaptation strategy is the most effective out of the three strategies in using BabelNet resources to improve system performance.

In future work, we intend to carry out further studies on the use of BabelNet for SMT regarding (i) using the latest version of BabelNet and online Web service-based API to process OOVs; (ii) examining BabelNet on more language pairs; (iii) studying different components of using BabelNet resources to augment SMT, such as supervising word alignment, phrase extraction and decoding.

6. Bibliographical References

Axelrod, A., He, X., and Gao, J. (2011). Domain Adaptation via Pseudo InDomain Data Selection. In *Proceedings of the 25th Pacific Asia Conference on Language, Information and Computation*, pages 20–30, Singapore.

Callison-Burch, C., Koehn, P., and Osborne, M. (2006). Improved statistical machine translation using paraphrases. In *Proceedings of the NAACL*, pages 17–24, New York City, USA.

Daume-III, H. and Jagarlamudi, J. (2011). Domain adaptation for machine translation by mining unseen words. In *ACL-HLT 2011: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 407–412, Portland, Oregon, USA.

Du, J., Jiang, J., and Way, A. (2010). Facilitating translation using source language paraphrase lattices. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 420–429, Cambridge, MA.

Ehrmann, M., Cecconi, F., Vannella, D., Mccrae, J. P., Cimiano, P., and Navigli, R. (2014). Representing multilingual data as linked data: the case of babelnet 2.0. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation)*, pages 401–408, Reykjavik, Iceland, May.

Elbedweihy, K., Wrigley, S., and Ciravegna, F. (2013). Using babelnet in bridging the gap between natural language queries and linked data concepts. In *Proceedings of the 12th International Semantic Web Conference*.

Garera, N., Callison-Burch, C., and Yarowsky, D. (2009). Improving translation lexicon induction from monolingual corpora via dependency contexts and part-of-speech equivalences. In *Proceedings of the 13th Conference on Computational Natural Language Learning*, pages 129–137, Boulder, Colorado, June.

Habash, N. (2008). Four techniques for online handling of out-of-vocabularywords in arabic-english statistical machine translation. In *Proceedings of ACL-08*, pages 57–60, Columbus, Ohio, June. Association for Computational Linguistics.

Haghighi, A., Liang, P., Berg-Kirkpatrick, T., and Klein, D. (2008). Learning bilingual lexicons from monolingual corpora. In *Proceedings of ACL-08: HLT*, pages 771–779, Columbus, Ohio, June.

Haque, R., Penkale, S., and Way, A. (2014). Bilingual Termbank Creation via Log-Likelihood Comparison and Phrase-Based Statistical Machine Translation. In *Proceedings of the 4th International Workshop on Computational Terminology (COLING2014)*, pages 42–51, Dublin, Ireland, August.

Huang, C.-C., Yen, H.-C., Yang, P.-C., Huang, S.-T., , and Chang, J. S. (2011). Using sublexical translations to handle the oov problem in machine translation. *ACM Transactions on Asian Language Information Processing*, 10(3):16.

Jadidinejad, A. H. (2013). Unsupervised information extraction using babelnet and dbpedia. In *Proceedings of the Making Sense of Microposts (MSM2013) Concept Extraction Challenge*, pages 54–56, Rio de Janeiro, Brazil, May.

Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., Cowan, B., Shen, W., Moran, C., Zens, R., Dyer, C., Bojar, O., Constantin, A., and Herbst, E. (2007). Moses: open source toolkit for statistical machine translation. In *Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions, ACL 2007*, pages 177–180, Prague, Czech Republic, June.

Koehn, P. (2004). Statistical significance tests for machine translation evaluation. In *Proceedings of EMNLP 2004*, pages 388–395, Barcelona, Spain, July.

Koehn, P. (2005). Europarl: a parallel corpus for statistical machine translation. In *MT Summit X: Conference Proceedings: the tenth Machine Translation Summit*, pages 79–86, Phuket, Thailand.

Lehal, G. S. and Saini, T. S. (2012). Conversion between scripts of punjabi: Beyond simple transliteration. In *Proceedings of the 24th International Conference on Computational Linguistics)*, pages 633–642, Mumbai, India, December.

Luo, J., Tinsley, J., and Lepage, Y. (2013). Exploiting parallel corpus for handling out-of-vocabularywords. In *Proceedings of the 27th PACLIC*, pages 399–408, Taipei, Taiwan, November.

Marton, Y., Callison-Burch, C., and Resnik, P. (2009). Improved statistical machine translation using monolingually-derived paraphrases. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 381–390, Singapore, Singapore, August.

Moore, R. C. and Lewis, W. (2010). Intelligent Selection of Language Model Training Data. In *Proceedings of the Association for Computational Linguistics (ACL) 2010*, pages 220–224, Uppsala, Sweden, July.

Moro, A., Navigli, R., Tucci, F. M., and Passonneau, R. J. (2014). Annotating the masc corpus with babelnet. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation*, pages 4214–4219, Reykjavik, Iceland, May.

Navigli, R. and Ponzetto, S. P. (2010). BabelNet: Building a Very Large Multilingual Semantic Network. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 216–225, Uppsala, Sweden, July.

Navigli, R. and Ponzetto, S. P. (2012). Multilingual WSD with Just a Few Lines of Code: the BabelNet API. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 67–72, Jeju, Republic of Korea, July.

Navigli, R., Jurgens, D., and Vannella, D. (2013). Semeval-2013 task 12: Multilingual word sense disambiguation. In *Proceedings of the Seventh International Workshop on Semantic Evaluation*, pages 222–231, Atlanta, Georgia, June.

Navigli, R. (2012). Babelnet goes to the (multilingual) semantic web. In *Proceedings of the ISWC 2012 Workshop on Multilingual Semantic Web*, Boston, USA, November.

Papineni, K., Roukos, S., Ward, T., and Zhu, W.-J. (2002). BLEU: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pages 311–318, Philadelphia, USA, July.

Popovic, M. and Ney, H. (2004). Towards the use of word stems and suffixes for statistical machine translation. In *Proceedings of Language Resources and Evaluation*, pages 1585–1588, Lisbon, Portugal, May.

Roberto Navigli, S. P. P. (2012). BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence*, 193:217–250.

Sadat, F. and Habash, N. (2006). Combination of arabic preprocessing schemes for statistical machine transla-

tion. In *Proceedings of the 21st COLING and 44th ACL*, pages 1–8, Sydney, Australia, July.

Snover, M., Dorr, B., Schwartz, R., Micciulla, L., and Makhoul, J. (2006). A study of translation edit rate with targeted human annotation. In *Proceedings of the Association for Machine Translation in the Americas*, pages 223–231, Cambridge, USA, August.

Zhang, J., Zhai, F., and Zong, C. (2012). Handling unknown words in statistical machine translation from a new perspective. In *Proceedings of the 2nd Natural Language Processing and Chinese Computing*, pages 176–187, Beijing, China, October.