A Hybrid Approach to Compiling Bilingual Dictionaries of Medical Terms from Parallel Corpora

Georgios Kontonatsios(✉), Claudiu Mihăilă, Ioannis Korkontzelos, Paul Thompson, and Sophia Ananiadou

The National Centre for Text Mining, The University of Manchester,
131 Princess Street, Manchester M1 7DN, UK
{kontonag,mihailac,korkonti,thompsop,ananiads}@cs.man.ac.uk

Abstract. Existing bilingual dictionaries of technical terms suffer from limited coverage and are only available for a small number of language pairs. In response to these problems, we present a method for automatically constructing and updating bilingual dictionaries of medical terms by exploiting parallel corpora. We focus on the extraction of multi-word terms, which constitute a challenging problem for term alignment algorithms. We apply our method to two low resourced language pairs, namely English-Greek and English-Romanian, for which such resources did not previously exist in the medical domain. Our approach combines two term alignment models to improve the accuracy of the extracted medical term translations. Evaluation results show that the precision of our method is 86 % and 81 % for English-Greek and English-Romanian respectively, considering only the highest ranked candidate translation.

1 Introduction

Bilingual dictionaries of technical terms are important resources for many *Natural Language Processing* (NLP) tasks. In *Statistical Machine Translation* (SMT), an up-to-date, bilingual dictionary can be used to dynamically update the SMT system in order to provide more accurate translations of unseen terms [12,15], i.e., current SMT approaches fail to translate terms that do not occur in the training data. In *Cross-Language Information Retrieval*, the use of a bilingual dictionary to expand queries is reported to enhance the overall performance [2]. Bilingual dictionaries can also be particularly useful for human translators who are not familiar with the domain specific terminology [10].

However, manually creating and updating bilingual dictionaries of technical terms is a labourious process. Especially in the biomedical domain, where there is a proliferation of newly coined terms [23], keeping such resources up to date can be extremely costly. For this reason, many existing bilingual dictionaries of technical terms remain incomplete and only cover a limited number of language pairs.

L. Besacier et al. (Eds.): SLSP 2014, LNAI 8791, pp. 57–69, 2014.
DOI: 10.1007/978-3-319-11397-5_4

The UMLS metathesaurus[1] is a popular, multilingual terminological resource for the biomedical domain. UMLS contains terms in more than 20 languages, which are linked together using common *concept* ids. However, UMLS does not index terms in Greek and in Romanian. Hence, our goals is it expand UMLS for these two low resourced languages. In the English part of UMLS, MWTs correspond to phrases that can have various syntactic forms, e.g., noun phrases (*bipolar disorder*) or adjective phrases (*abnormally high*), and which cover a wide range of biomedical concepts, e.g., qualitative concepts (*moderate to severe*), diseases or syndromes *liver cirrhosis*, pharmacological substances (*lactose monohydrate*), etc. In contrast to previous approaches, which have been concerned exclusively with identifying translations of *noun phrases* [9,19,28], we do not restrict the term alignment problem to specific syntactic categories. Without this restriction, our method is general enough to extract candidate term translations for all types of MWTs appearing in UMLS, thus providing the potential to significantly increase coverage for other languages.

Most existing term alignment algorithms suggest N ranked candidate translations for each source term (usually in the range of [1, 20]). Translation accuracy is evaluated by determining the percentage of source terms whose top N candidates contain a correct translation. Naturally, as N increases, the system performance improves, because a greater number of candidate translations is being considered. Thus, the evaluation results only tell us, if a correct translation exists, it will be somewhere amongst the N possible translation candidates. Dictionaries created using such methods (i.e., with N candidate translations for each word) are noisy, and can only be useful for a limited number of applications, e.g., in SMT systems, which use a language model to select the most probable translation out of N possible candidates. However, [8] showed that such dictionaries decreased the translation accuracy of human translators. Furthermore, automatically compiled dictionaries cannot be used to update high-quality terminological resources before human translators have removed the noisy candidates. In this paper, we aim to improve the translation accuracy on the first ranked candidate.

Our novel method firstly analyses a parallel corpus to identify terms in the source language which have a corresponding entry in UMLS. Additionally, each source term is annotated with: (a) concept id and (b) semantic category that are derived from UMLS. As a second step, we apply a term alignment method to obtain the translation equivalence in the target language. Finally, we propagate the concept id and the semantic category from the source term to the corresponding target translation.

To obtain the translation of a source term, we investigate three term alignment methods, i.e.: (a) a phrase alignment module which is part of an SMT toolkit (SPA) (b) a supervised machine learning approach that uses character n-grams, i.e., a Random Forest (RF) classifier and (c) the intersection of the above two methods, which we call the *voting* system.

For evaluation purposes, we sampled 1000 terms with their corresponding translations, and we asked bilingual speakers of both English-Greek and

[1] nlm.nih.gov/research/umls

English-Romanian to judge the translations. The results that we obtained showed that the voting method achieved the best translation accuracy for the *top 1* candidate translation by a significant margin. Furthermore, the same voting method exhibited competitive performance in the translation of both frequent and rare terms, in contrast to SPA, which was shown to be largely dependent on the frequency of terms in the corpus.

2 Related Work

Early works on term alignment immediately recognised the potential benefits of automatically constructed bilingual terminological resources. At AT&T labs, [6] presented *Termight*, a term alignment tool that was shown to be useful for human translators of technical manuals, despite of its fairly low accuracy (40 % translation accuracy on the top 1 candidate for English-German). *Termight* firstly performed word alignment [7] to compile a bilingual list of single words from the parallel corpus. A simple heuristic was then used to extract MWT translations, i.e., for each source MWT, candidate translations were defined as sequences whose first and last words were aligned with any of the words contained in the source MWT.

Several approaches have explored statistical methods to align MWTs in parallel corpora. Reference [25] investigated the alignment of collocations which are frequently occurring word pairs (e.g., *powerful computer*). *Champollion*, their proposed system, was iteratively building a translation by selecting words in the target language that are highly correlated (Dice coefficient) with the input source collocation. Their method achieved competitive performance of approximately 70 % translation accuracy on an English-French parallel corpus. Other approaches have investigated the use of co-occurrence frequency [9] and mutual information [5].

While statistical approaches for term alignment are frequently reported in the literature, several other techniques have been explored, including machine learning, distributional methods and hybrid approaches. Reference [19] introduced a machine learning method, an Expectation-Maximisation algorithm, for extracting translations of noun-phrases from an English-French parallel corpus. The authors reported an accuracy of 90 %, but only when considering the 100 highest ranking correspondences. Reference [3] used a simple distributional semantics approach, namely a *Vector Space model*. They constructed boolean term vectors in a source and target language of size N, where N is the number of sentences in the corpus. Each dimension of a vector corresponded to a sentence and its value indicated whether or not the term appeared in the sentence. For ranking candidate translations they used the *Jaccard Index* between the vector of a source and target term. For evaluation, they did not compute the standard precision and recall. They rather measured the effect of using the extracted bilingual dictionary within an SMT system. The results obtained showed a small improvement of +0.30 BLEU points [22] over the baseline SMT system. Reference [29] introduced a hybrid collocation alignment system that combines statistical (log likelihood ratio) and syntactic information (part-of-speech patterns). Their system achieved a competitive performance when applied on a distant language pair, namely English-Chinese.

In contrast to previous works that applied their methods to well-resourced language pairs, e.g., English-French, few approaches have focussed on low-resourced language pairs. Reference [28] applied term alignment to extend the Slovenian WordNet. Their method relied on word alignments and on lexico-syntactic patterns. In total, they were able to extract 5,597 new multi-word terms for the Slovenian WordNet, of which 2,881 were correct.

In this work, we aim at enriching UMLS with low-resourced languages that previously were absent from the thesaurus. Overall, our system retrieved 5,926 and 5,446 multi-word terms with a translation accuracy of 86 % and 81 % on the top 1 ranked candidate for English-Greek and English-Romanian respectively.

3 Methodology

We use *EMEA*, a biomedical parallel corpus from the European Medicines Agency [27]. The corpus contains approximately 1,500 sentence-aligned documents in 22 European languages. As a first step, English MWTs are identified in the corresponding part of EMEA, using a monolingual term extraction tool. For this, we use MetaMap [1], which automatically recognises biomedical terms in an input English text. Furthermore, MetaMap assigns a UMLS concept id and a semantic category to each term. In this way, the extracted English terms are mapped to the UMLS metathesaurus. This step identified 17,907 unique English MWTs in the English-Greek part of the corpus and 16,625 MWTs in the English-Romanian part of the corpus.

For the target part of the parallel corpus (Greek/Romanian), we extract candidate translations by simply considering all contiguous sequences containing up to four words. These candidates are used in both SPA and RF. Additionally, the system computes the intersection of the alignment methods (voting system) in order to improve the quality of the extracted dictionaries.

3.1 Statistical Phrase Alignment

We adopt a standard approach used in Statistical Machine Translation to align phrases from parallel corpora [16]. The *Statistical Phrase Alignment* (SPA) method builds phrase alignments using a single-word bilingual dictionary that was previously extracted from the parallel corpus. The performance of SPA is heavily dependent on the quality of the word alignment module. Word alignments are established using GIZA++ [21], an open source implementation of the 5 IBM-models [4]. GIZA++ is trained on both translation directions and extracts two word alignment tables $L_s \rightarrow L_t$ and $L_t \rightarrow L_s$ between a source L_s and target L_t language. Then we combine the two tables using the *grow-diag-final* heuristic which starts with all alignments found in the intersection of the tables and then adds neighbouring alignments of the union set. The merged translational table yields a better recall and precision of word-to-word correspondences than the intersection or union of the two tables.

Using the word alignments that are established as described above, we extract any pair of phrases (s^n, t^m), where s^n is a source phrase containing of n words and t^m a target phrase of m words, with the condition that the words in s^n are aligned to the words in t^m and not to any other words outside t^m. Finally, the candidate phrase pairs are ranked using the lexical translation probability.

3.2 Random Forest Aligner

References [17,18] introduced an Random Forest (RF) aligner that is able to automatically learn association rules of textual units between a source and target language when trained on a corpus of positive and negative examples. The method is based on the hypothesis that terms across languages are constructed using semantically equivalent textual units. Hence, if we know the translations of the basic building blocks of a term, e.g., morphemes, prefixes, suffices, we can predict the term in a target language. Table 1 illustrates an example of the training and prediction process of the RF aligner. In this toy example, the RF aligner is trained on two English-Greek and English-Romanian instances and learns how to translate the morphemes *cardio* and *vascular*. Once trained, the model uses the previously learned associations of textual units to extract new term translations, e.g., <*cardio-vascular, καρδι-αγγειακό*, cardio-vascular>.

Table 1. Example of training and prediction process of RF Aligner

	English	**Greek**	**Romanian**
training	**cardio**-*myopathy*	*μυο*-**καρδιο**-*πάθεια*	**cardio**-*miopatie*
	extra-**vascular**	*εξω*-**αγγ ειακό**	*extra*-**vascular**
prediction	**cardio-vascular**	**καρδι-αγγειακό**	**cardio-vascular**

The RF aligner is a supervised, machine learning method. It uses n-grams (size of $[2, 5]$ characters) to represent a pair of terms in a source and in a target language. The feature vectors have a fixed size of $2q$. The first q n-grams are extracted from the source term while the last q n-grams from the target term. For dimensionality reduction, we considered only the $1,000$ (500 source and 500 target) most frequent character n-grams. Given a term in a source language, the model outputs a list of the top N ranked candidate translations. For ranking candidate translations, we use the prediction confidence, i.e., *classification margin*.

To train the RF aligner, we use *BabelNet* [20]. *BabelNet* is a multilingual, multi-domain, encyclopedic dictionary containing terms in 50 languages including English, Greek and Romanian. For both language pairs (English-Greek and English-Romanian) we select $10K$ positive and $10K$ negative instances to learn positive and negative associations of character grams between the source and target language. *Pseudo-negative* instances are created by randomly matching non-translation terms.

In addition to the two alignment models, we use a voted system that considers the intersection of the two models. The combined system is expected to increase the accuracy of the automatically extracted dictionaries.

4 Experiments

In this section, we evaluate the three dictionary extraction methods, namely SPA, RF and the voted system on two biomedical parallel corpora for English-Greek and English-Romanian.

We follow a standard evaluation methodology reported in the literature. We select the top N ranked translations for each source term as given by the term alignment methods and we mark a correct answer when a true translation is found among the top N candidates. In some cases both the RF and SPA failed to propose any candidate translations for a given source term (the list of ranked candidates was empty). Based upon, we use two evaluation metrics in our experiments, namely the top-N precision and top-N recall. The top-N recall is defined as the percentage of source terms whose top N candidates contain a correct translation. The top-N precision is the percentage of source terms whose list of top N candidates: (a) is not empty and (b) contains a correct translation.

4.1 Results

For evaluation purposes, we randomly sampled $1,000$ English MWTs. For each English MWT, we selected the top 20 translation candidates. We asked two people whose native languages are Greek and Romanian, respectively, and who are fluent English speakers, to manually judge the translations of the English MWTs. The curators marked only exact translations as being correct.

Figures 1 and 2 show the top 20 precision of the dictionary extraction methods. We note that the performance of SPA is consistently better than RF in both datasets. SPA achieves a precision of 93.2 % and 89 % on the top 20 candidates, while RF achieves a precision of 67.1 % and 67.7 % for English-Greek and English-Romanian, respectively. A possible explanation for the poor performance of RF is that we trained the model on an *out-of-domain* lexicon, namely *BabelNet*, due to the lack of existing bilingual biomedical dictionaries for the two language pairs. Hence, the model does not explicitly learn character n-gram mappings of biomedical terms, which leads to noisy translations when we apply the model to *in-domain* datasets.

SPA achieves robust performance when considering the top 20 candidate translations. However, the performance of SPA declines steadily as fewer translation candidates are considered (49 % for English-Greek and 48 % for English-Romanian on the top 1 candidate). Thus, although SPA is able to determine

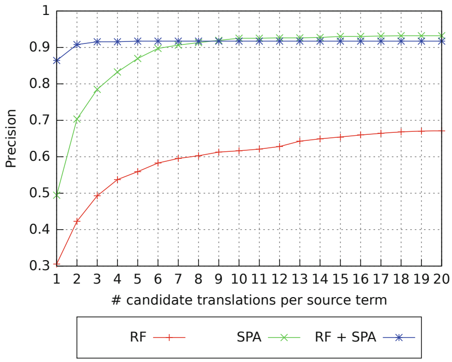

Fig. 1. Precision of top N candidates of SPA, RF and the voted method (SPA + RF) for English-Greek

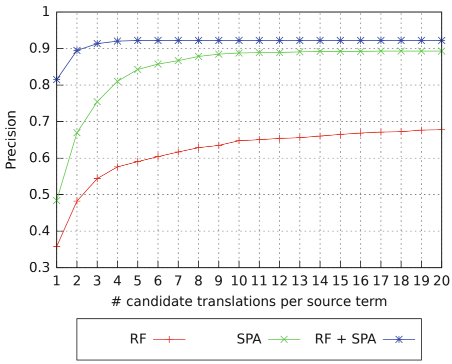

Fig. 2. Precision of top N candidates of SPA, RF and the voted method (SPA + RF) for English-Romanian

an exact translation in most cases, this could lie anywhere within the top 20 candidates.

In an attempt to improve the precision of higher ranking candidates, we considered the intersection between RF and SPA. This "hybrid" method achieved significantly better performance than either RF or SPA when considering only the highest ranked candidate (86 % and 81 % for English-Greek and English-Romanian, respectively). For the top 20 candidates, the performance of the voted system is approximately the same with SPA.

The recall which determines the coverage of the extracted dictionary is a further important feature of term alignment methods. Figures 3 and 4 illustrate the recall of SPA, RF and the voted system on an increasing number of ranked candidate translations. We observe that the recall of RF is significantly better than the recall of SPA and the voted system. For the top 20 candidate translations, RF achieves the best observed recall of approximately 67 % for both

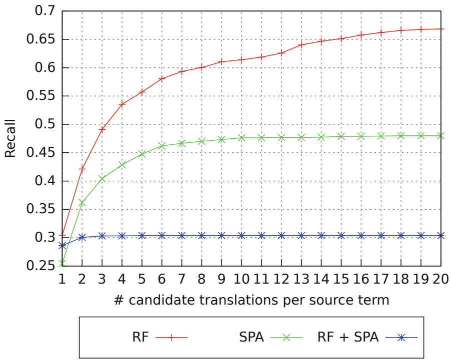

Fig. 3. Recall of top N candidates of SPA, RF and the voted method (SPA + RF) for English-Greek

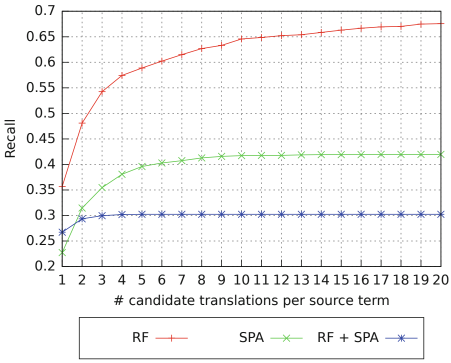

Fig. 4. Recall of top N candidates of SPA, RF and the voted method (SPA + RF) for English-Romanian

English-Greek and English-Romanian. In contrast, the voted system performs poorly, with a total coverage of 30 % on the top 20 candidates. The low recall of the voted system is caused by the significantly decreased coverage of SPA compared to RF. This impacts negatively on the voted system, since its results constitute the intersection of the two alignment methods. SPA extracted correct translations for only 48 % and 40 % of the English terms from the English-Greek and English-Romanian corpus, respectively.

4.2 Error Analysis

In this subsection, we discuss the results of an error analysis that was performed to reveal common noisy translations extracted by RF and SPA.

In the case of the RF, we identified two types of erroneous translations: (a) partial matches and (b) discontinuous translations. Partial matches refer to the cases where RF translated part of the English term but failed to retrieve an exact match. For example, the English term "urea cycle disorder" was partially translated into Romanian as "tulburărilor ciclului" (missing translation for "urea"). Additionally, in several cases, a partial translation was ranked higher than the exact match, which led to a decrease in the precision for top 1 candidate. The top ranked Greek translation for "urea cycle disorders" was "διαταραχών" (urea cycle) while the exact matched (διαταραχών του κύκλου της) was ranked fifth. In fewer cases, the translation of an English term occurred as discontinuous sequence in the target corpus. For example, the term "metabolic diseases" (*boli de metabolism*) occurred in the Romanian corpus as a discontinuous sequence with the span "boli ereditare de metabolism" (*hereditary metabolic diseases*). However, the current implementation of RF searches for candidate translations only in contiguous sequences, in order to minimise the number of classification instances. Hence, these type of translations were not captured by RF.

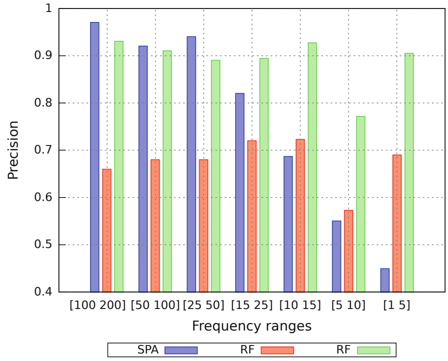

Fig. 5. Precision of top 20 candidates on different frequency ranges of SPA, RF and the voted method (SPA + RF), English-Greek dataset

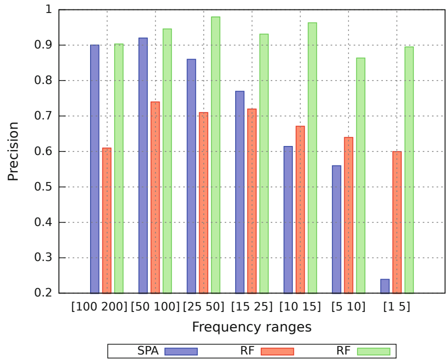

Fig. 6. Precision of top 20 candidates on different frequency ranges of SPA, RF and the voted method (SPA + RF), English-Romanian dataset

SPA is a statistical based alignment tool whose performance is largely affected by the frequency of the terms in the corpus. For high frequency terms, SPA has stronger evidence of term alignment and extracts correct translations with higher confidence. However, we predicted that the alignment quality would decrease for rarely occurring terms.

To further investigate this intuition, we evaluated the performance of SPA on terms having varying frequencies in the corpus. We segmented the $1,000$ test terms for both English-Greek and English-Romanian, that were previously evaluated by the human curators, into 7 frequency ranges, from high-frequency to rare terms. Then we computed the precision for the top 20 candidates for RF, SPA and the voted system on those frequency ranges. Figures 5 and 6 show the performance of the term alignment methods. We observe that for frequent terms (i.e., those terms occurring 25 times or more in the corpus), SPA shows a robust performance for both language pairs. However, for less frequent terms, the precision of SPA steadily decreases. In contrast to SPA, RF does not exploit any corpus statistics to align terms between a source and a target language and as a result, its performance is not dependent on the frequency of occurrence of a term. Accordingly, the precision of RF fluctuates only slightly over different term frequency ranges. Furthermore, the voted system presents a stable precision over different frequency ranges, since it is the intersection of RF and SPA. Hence, we can conclude that the dictionary extracted by the voted system is robust to differences in frequency ranges of terms.

5 Conclusion

In this paper, we have presented a hybrid approach to the automatic compilation of bilingual dictionaries of biomedical terms from parallel corpora. Our novel voted system combines a supervised machine learning approach, i.e., a Random Forest (RF) aligner, with a Statistical Phrase Alignment (SPA) alignment method, to improve the accuracy of extracted term translations. We have applied our method to two low-resourced language pairs, namely English-Greek and English-Romanian, and candidate translations have been evaluated by bilingual speakers of these two language pairs. The voted system exhibits significantly better translation accuracy for the highest ranked translation candidate than either the RF or SPA methods, when they are used in isolation. In addition, the voted system achieved a considerably better performance in translating rarely occurring terms than SPA.

As future work, we plan to exploit other sources of information in order to increase the size of the automatically extracted bilingual dictionaries. Parallel corpora are useful resources for SMT and for compiling high-quality bilingual dictionaries. However, such corpora are expensive to construct because human translators need to provide the translations of the source documents. Consequently, parallel corpora are of limited size and they quickly become out-of-date (and thus are unlikely to contain neologisms). Additionally, they are not available for every domain or language pair. Due to the sparsity of parallel document collections, researchers have started to explore comparable corpora, since they are more readily available, more up-to-date and they are easier and cheaper to construct. In common with a parallel corpus, a comparable corpus is a collection of documents in a source and target language. However, in contrast to a parallel corpus, the documents in a comparable corpus are not direct translations of each other. Rather, they share common features, such as covering the same topic, domain, time period, etc. Large comparable collections can be readily constructed using freely available multilingual resources, e.g., Wikipedia [13,26,30]. This means that comparable corpora constitute a promising resource to aid in the construction and maintenance of bilingual dictionaries, especially when parallel corpora are limited or unavailable for a given language pair.

Whilst the hybrid system described in this paper cannot be applied directly to comparable corpora, since the SPA module can only process parallel data, we are planning to incorporate a context-based method into our system. This method is widely used to facilitate term alignment approaches involving comparable corpora. Context-based methods approaches (context vectors) [11,24] adapt the *distributional hypothesis* [14] to extract term translations from comparable corpora. They hypothesise that a term and its translation tend to appear in similar lexical contexts. Intuitively, the RF aligner and the context-based approach are complementary, since RF exploits the internal structure of terms, while context vectors use the surrounding lexical context. Therefore, it will be interesting to investigate how these two methods can be combined within a hybrid system.

Acknowledgements. This work was funded by the European Community's Seventh Framework Program (FP7/2007–2013) [grant number 318736 (OSSMETER)].

References

1. Aronson, A.R.: Effective mapping of biomedical text to the umls metathesaurus: the metamap program. In: Proceedings of the AMIA Symposium, p. 17. American Medical Informatics Association (2001)
2. Ballesteros, L., Croft, W.: Phrasal translation and query expansion techniques for cross-language information retrieval. In: ACM SIGIR Forum, vol. 31, pp. 84–91. ACM (1997)
3. Bouamor, D., Semmar, N., Zweigenbaum, P.: Identifying bilingual multi-word expressions for statistical machine translation. In: LREC, pp. 674–679 (2012)
4. Brown, P., Pietra, V., Pietra, S., Mercer, R.: The mathematics of statistical machine translation: parameter estimation. Comput. linguist. **19**(2), 263–311 (1993)
5. Church, K.W., Hanks, P.: Word association norms, mutual information, and lexicography. Comput. linguist. **16**(1), 22–29 (1990)
6. Dagan, I., Church, K.: Termight: identifying and translating technical terminology. In: Proceedings of the Fourth Conference on Applied Natural Language Processing, pp. 34–40. Association for Computational Linguistics (1994)

7. Dagan, I., Church, K.W., Gale, W.A.: Robust bilingual word alignment for machine aided translation. In: Proceedings of the Workshop on Very Large Corpora, pp. 1–8 (1993)
8. Delpech, E.: Evaluation of terminologies acquired from comparable corpora: an application perspective. In: Proceedings of the 18th International Nordic Conference of Computational Linguistics (NODALIDA 2011), pp. 66–73 (2011)
9. Van der Eijk, P.: Automating the acquisition of bilingual terminology. In: Proceedings of the Sixth Conference on European Chapter of the Association for Computational Linguistics, pp. 113–119. Association for Computational Linguistics (1993)
10. Fung, P., McKeown, K.: A technical word-and term-translation aid using noisy parallel corpora across language groups. Mach. Transl. **12**(1), 53–87 (1997)
11. Fung, P., Yee, L.Y.: An ir approach for translating new words from nonparallel, comparable texts. In: Proceedings of the 17th International Conference on Computational linguistics, vol. 1, pp. 414–420. Association for Computational Linguistics (1998)
12. Habash, N.: Four techniques for online handling of out-of-vocabulary words in arabic-english statistical machine translation. In: Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Short Papers, pp. 57–60. Association for Computational Linguistics (2008)
13. Haghighi, A., Liang, P., Berg-Kirkpatrick, T., Klein, D.: Learning bilingual lexicons from monolingual corpora. In: ACL, vol. 2008, pp. 771–779 (2008)
14. Harris, Z.: Distributional structure. Word (1954)
15. Irvine, A., Callison-Burch, C.: Combining bilingual and comparable corpora for low resource machine translation. In: Proceedings of the Eighth Workshop on Statistical Machine Translation. Association for Computational Linguistics, August 2013
16. Koehn, P., Och, F.J., Marcu, D.: Statistical phrase-based translation. In: Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology, vol. 1, pp. 48–54. Association for Computational Linguistics (2003)
17. Kontonatsios, G., Korkontzelos, I., Tsujii, J., Ananiadou, S.: Using a random forest classifier to compile bilingual dictionaries of technical terms from comparable corpora. In: Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics: Short Papers, vol. 2, pp. 111–116. Association for Computational Linguistics, April 2014, http://www.aclweb.org/anthology/E14-4022
18. Kontonatsios, G., Korkontzelos, I., Tsujii, J., Ananiadou, S.: Using random forest to recognise translation equivalents of biomedical terms across languages. In: Proceedings of the Sixth Workshop on Building and Using Comparable Corpora, pp. 95–104. Association for Computational Linguistics, August 2013, http://www.aclweb.org/anthology/W13-2512
19. Kupiec, J.: An algorithm for finding noun phrase correspondences in bilingual corpora. In: Proceedings of the 31st Annual Meeting on Association for Computational Linguistics, pp. 17–22. Association for Computational Linguistics (1993)
20. Navigli, R., Ponzetto, S.P.: BabelNet: the automatic construction, evaluation and application of a wide-coverage multilingual semantic network. Artif. Intell. **193**, 217–250 (2012)
21. Och, F.J., Ney, H.: A systematic comparison of various statistical alignment models. Comput. linguist. **29**(1), 19–51 (2003)

22. Papineni, K., Roukos, S., Ward, T., Zhu, W.J.: Bleu: a method for automatic evaluation of machine translation. In: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, pp. 311–318. Association for Computational Linguistics (2002)
23. Pustejovsky, J., Castano, J., Cochran, B., Kotecki, M., Morrell, M.: Automatic extraction of acronym-meaning pairs from medline databases. Studies in health technology and informatics, pp. 371–375 (2001)
24. Rapp, R.: Automatic identification of word translations from unrelated english and german corpora. In: Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics on Computational Linguistics, pp. 519–526. Association for Computational Linguistics (1999)
25. Smadja, F., McKeown, K.R., Hatzivassiloglou, V.: Translating collocations for bilingual lexicons: a statistical approach. Comput. linguist. **22**(1), 1–38 (1996)
26. Tamura, A., Watanabe, T., Sumita, E.: Bilingual lexicon extraction from comparable corpora using label propagation. In: Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pp. 24–36. Association for Computational Linguistics (2012)
27. Tiedemann, J.: News from opus-a collection of multilingual parallel corpora with tools and interfaces. In: Recent Advances in Natural Language Processing, vol. 5, pp. 237–248 (2009)
28. Vintar, S., Fiser, D.: Harvesting multi-word expressions from parallel corpora. In: LREC (2008)
29. Wu, C.C., Chang, J.S.: Bilingual collocation extraction based on syntactic and statistical analyses. In: ROCLING (2003)
30. Yu, K., Tsujii, J.: Bilingual dictionary extraction from wikipedia. In: Proceedings of Machine Translation Summit XII, pp. 379–386 (2009)