Evaluating the Impact of Bilingual Lexical Resources on Cross-lingual Sentiment Projection in the Pharmaceutical Domain

Matthias Hartung, Matthias Orlikowski, Susana Veríssimo
Semalytix GmbH
Meller Straße 2, 33613 Bielefeld, Germany
{first.second}@semalytix.com

Abstract

Rolling out text analytics applications or individual components thereof to multiple input languages of interest requires scalable workflows and architectures that do not rely on manual annotation efforts or language-specific re-engineering per target language. These scalability challenges aggravate even further if specialized technical domains are targeted in multiple languages. In recent work, it has been shown that cross-lingual projection of sentiment models in deep learning frameworks based on bilingual sentiment embeddings (BLSE) is feasible without any annotated data in the target language, capitalizing on monolingual embeddings and a bilingual translation dictionary only (Barnes et al., 2018a). We use their framework and apply it to multilingual text analytics problems in the pharmaceutical domain in order to (i) investigate under which conditions the BLSE approach scales to technical domains as well, and (ii) assess the impact of different configurations of underlying lexical resources. For the language pair English/Spanish, our findings corroborate the strength of cross-lingual projection approaches such as BLSE in technical scenarios, given the availability of bilingual resources that provide broad lexical coverage, on the one hand, and complementary domain- and task-specific knowledge, on the other.

1 Introduction

Real-world text analytics in business-relevant domains usually depend on complex stacks of NLP models and components. In global markets, a main challenge faced by suppliers of text analytics services and applications arises from the complexity of business use cases, technical components needed to address them, and input data that comes from multiple languages. Obviously, approaching this challenge by attempting to build dedicated NLP stacks for each new language from scratch is not scalable, due to generally high onboarding costs for initial model development and refinement.

Against this backdrop, cross-lingual model transfer poses a promising alternative (Kozhevnikov and Titov, 2014; Zirikly and Hagiwara, 2015; Ahmad et al., 2019, inter alia). The main underlying idea is that NLP models readily existing for a source language are transfered to a new target language of interest without language-specific training data being available in this target language.

More recently, cross-lingual projection frameworks have been proposed in the literature (Barnes et al., 2018a; Barnes and Klinger, 2019) which do no longer require existing models as a source for model transfer, but induce target-language models as a result of a training procedure that jointly addresses the source and the target language based on a set of mono- and bilingual language resources available for both of them. Thus, such projection frameworks can be seen as shifting the development burden from manual annotation efforts as required for supervised learning of language-specific models from scratch to discovering suitable language resources, possibly combining them and plugging them into end-to-end learning frameworks. In light of the increased availability of such language resources as Linguistic Linked Open Data[1] (Chiarcos et al., 2012; Cimiano et al., 2020), we consider cross-lingual projection as a highly versatile methodology with a wide application potential in various multilingual NLP or text analytics problems.

Throughout this paper, we compare two prototypical model architectures for cross-lingual sentiment analysis in the pharmaceutical domain which

[1] http://linguistic-lod.org/llod-cloud

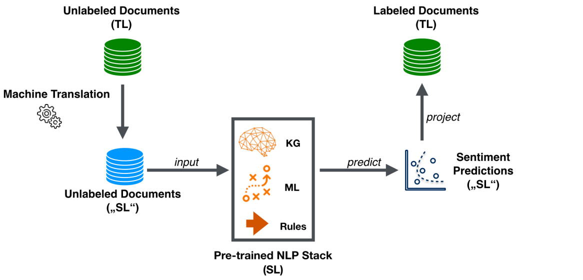

Figure 1: Translation Pipeline: High-level Overview ("SL" and "TL" referring to source and target language, respectively)

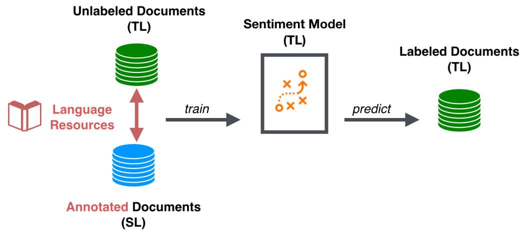

Figure 2: Cross-lingual Projection: High-level overview ("TL" referring to target language)

is our main area of interest. Henceforth, these architectures are denoted as *Translation Pipeline* and *Cross-lingual Projection*, respectively. Our main focus is on evaluating the impact of different configurations of monolingual and bilingual lexical resources on the classification performance of the induced, domain-specific sentiment model in the target language for which no supervised training signals are available.

2 Model Architectures

In this section, we describe the two prototypical model architectures underlying our experiments in cross-lingual sentiment analysis in the pharmaceutical domain: *Translation Pipeline* and *Cross-lingual Projection*. Both of them are based on the assumption that no labeled training data is available for supervised learning of a classification model in the target language.

2.1 Translation Pipeline

This architecture relies on the existence of (i) a pre-trained model for sentiment classification in the source language, and (ii) a translation engine or service covering the target/source language pair. In order to render the source-language classification model operable for sentiment analysis in the target language, target-language input texts are automatically translated into the source language before being presented to the classifier in order to yield a target prediction. A high-level depiction of this architecture is displayed in Figure 1.

Depending on the quality of the translation service, a certain proportion of noisy translations is to be expected that might impair classification performance, as the translated data may exhibit markers of sentiment that are largely different from what the model has learnt on original English data. Hence, despite the simplicity of this approach, we consider it as a robust baseline.

2.2 Cross-lingual Projection

In light of its presumed drawbacks, we aim at comparing the translation pipeline approach against a more sophisticated cross-lingual projection approach as shown in Figure 2. As a result of the projection, a language-specific classification model is induced that can be directly applied to input documents in the target language, without the need of supervised training from scratch.

Due to its parsimony in terms of language services and resources required, we use *Bilingual Sentiment Embeddings* (Barnes et al., 2018a) as an instantiation of a cross-lingual projection framework. The BLSE architecture is displayed in Figure 3. As can be seen from the figure, BLSE requires (i) monolingual word embeddings in both the source and target language, (ii) ground-truth annotations in the source language, and (iii) a bilingual dictionary that maps words from the source language to their translations in the target language. These resources provide the foundation for learning a mapping into a bilingual task-specific embedding space. The learning procedure is guided by a composite loss function based on the cross-entropy between sentiment predictions and ground truth labels in the source language and the spatial proximity of source/target pairs from the bilingual dictionary in the bilingual embedding space. The latter part enables the model to tailor target-language embeddings such that they can be used as input to a softmax classification layer that returns target-language predictions without any direct supervision in this language being available.

For formal details of the BLSE architecture, the reader is referred to Barnes et al. (2018a). For the experiments conducted in this study, we use their implementation[2] and parameter settings directly.

[2]Available from https://github.com/jbarnesspain/blse/

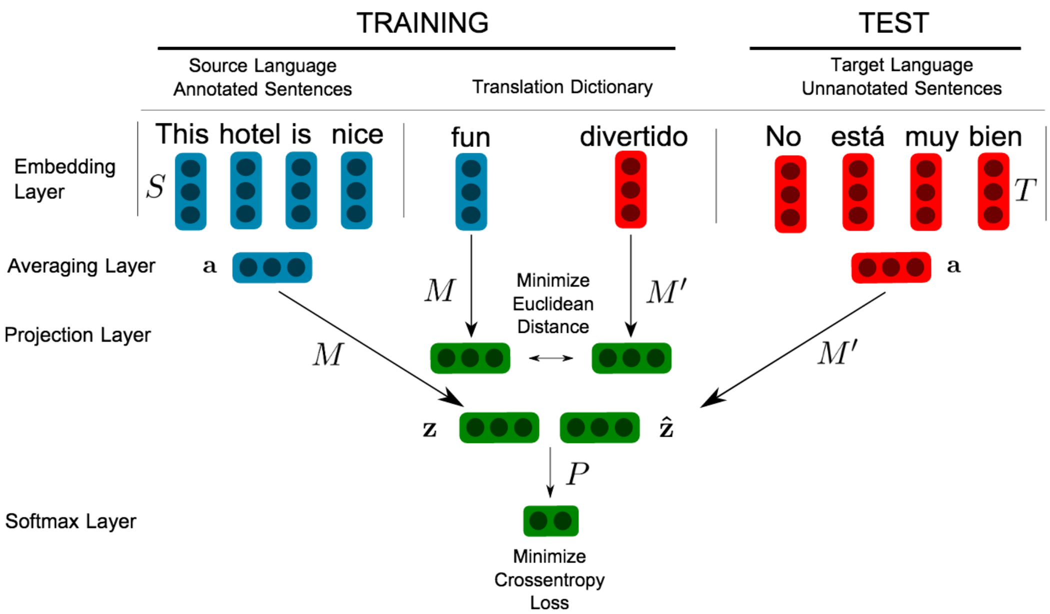

Figure 3: Overview of BLSE architecture, taken from Barnes et al. (2018).

3 Experiments and Results

In the experiments discussed below, we address the question: *What is the best configuration of lexical resources in order to achieve effective cross-lingual induction of pharma-specific sentiment models using the BLSE framework, focusing on English/Spanish as an exemplary language pair?*

3.1 Data Set and Lexical Resources

3.1.1 Data

Our corpus consists of a non-parallel sample of comparable English and Spanish transcripts of summaries of conversations between pharma representatives and medical experts. In these conversations, the medical experts are asked to state their opinions and assessments about particular medical treatments of interest (e.g., aspects of effectiveness and safety of drug products). A collection of 21,400 English summaries is manually annotated with binary sentiment labels (11,069 positives vs. 10,331 negatives) and subsequently used in order to train a classification model based on neural networks in a cross-validation setting. A set of 1,001 Spanish summaries is annotated likewise (559 positives, 442 negatives) in order to provide a gold standard for evaluation purposes only.

3.1.2 Word Embeddings

Monolingual word embeddings used in this study are selected along the two axes of *language* and *domain*:

- **google**[3]: English, open-domain, news corpus
- **PMC**[4]: English, medical domain, PubMed Central corpus
- **sg_300_es**[5]: Spanish, open-domain, Wikipedia corpus
- **scielo_wiki**[6]: Spanish, medical domain, concatenation of Scielo corpus and medical subset of Wikipedia

All embeddings have been pre-trained on the respective corpus using the word2vec algorithm (Mikolov et al., 2013).

3.1.3 Bilingual Lexicons

In order to inform the cross-lingual projection in BLSE, we apply three different lexicons that provide Spanish translations for English lexcial entries. These lexicons were selected according to the criteria given in the following.

[3] Available from https://drive.google.com/open?id=1GpyF2h0j8K5TKT7y7Aj0OyPgpFc8pMNS

[4] Available from http://bio.nlplab.org

[5] Available from https://drive.google.com/open?id=1GpyF2h0j8K5TKT7y7Aj0OyPgpFc8pMNS

[6] Available from https://zenodo.org/record/2542722#.XeUOo5NKjUK

Apertium. This lexicon was originally generated from an open-source machine translation platform (Gracia et al., 2017) and is distributed as Linguistic Linked Open Data[7]. For the purposes of this study, we use it as a *broad-coverage, open-domain* lexicon. After preprocessing and deduplication, Apertium contains 20,363 bilingual entries that are mapped at the level of word senses. Hence, in comparison to the other lexicons used here, it can be considered the linguistically most sophisticated one. In lack of a principled pharma-specific word sense disambiguation module for the source data set, however, we do not make use of the word sense information provided. Instead, all word senses pertaining to a particular lemma are used for learning the BLSE projection matrix.

Pharma. This lexicon comprises a total of 2,687 entries extracted from bilingual entity lexicalizations from the Semalytix Knowledge Graph which constitutes a large-scale repository of pharma-specific domain knowledge. Entity types contained in the graph comprise diseases and symptoms, drug products and agents, drug manufacturers, therapy areas, among various others. As such, this lexicon is designed as a source of *domain-specific* knowledge in order to render the BLSE projection matrix more sensitive to pharma-specific contents.

BingLiu. As a *task-specific* source of information, we use the sentiment lexicon originally provided by Hu and Liu (2004). In its original state, this is a monolingual English lexicon containing 5,749 entries that have been found to benefit sentiment analysis on English data in numerous previous studies. In this work, we use the bilingual extension of this resource that has been generated by Barnes et al. (2018a) via machine translation[8]. Note that we do not make use of the polarity information (i.e., positive or negative sentiment being expressed by a particular lemma) that is provided alongside each entry in this lexicon; instead, all entry pairs, irrespective of their polarity, are fed into BLSE in order to make sure that the resulting bilingual embeddings space captures open-domain sentiment-specific information to a sufficient extent.

[7]Available from https://old.datahub.io/dataset/apertium-rdf-en-es

[8]Available from https://github.com/jbarnesspain/blse/tree/master/lexicons/bingliu

Approach	Language	Accuracy
BLSE	Target (ES)	0.773
	Source (EN)	0.823
MT Pipeline	Target (ES)	0.536
	Source (EN)	0.825

Table 1: Performance of sentiment classification as obtained from BLSE and translation pipeline in terms of accuracy. Note that source and target language scores are not exactly comparable due to non-parallel evaluation data.

3.1.4 Lexicon Extensions

Domain Extension. The first extension is created by adding entries from the Pharma lexicon to Apertium, resulting in an extended version comprising 22,838 entries. Among those are 2,475 novel entries, whereas another 212 lemmas in the source language were already covered by Apertium in the first place, but either gained additional translations from the Pharma biomedical lexicon or remained unchanged due to equivalent entries in both original lexicons.

Task Extension. Analogously, entries from the BingLiu lexicon are added to Apertium in order to generate a task-specific bilingual lexicon that is better suited to sentiment classification. The resulting extension comprises 24,011 entries, 3,648 of which are novel ones.

Combined Domain and Task Extension. In order to exploit potential complementarity between domain- and task-specific knowledge when being added to Apertiums, a third variant of an extended bilingual lexicon is generated that combines the previous steps taken for domain and task extensions. Both steps are applied sequentially after one another, i.e., complementary entries from BingLiu are added to Apertium first, followed by additional entries from the Pharma lexicon. The resulting extended lexicon that comprises all three sources of information contains 29,397 unique entries.

3.2 Experiment 1: Comparison of Architectures

In this experiment, we run the translation pipeline (using the Microsoft Azure API[9] as translation service) and the cross-lingual projection framework

[9]https://azure.microsoft.com/en-us/services/cognitive-services/translator-text-api/

on the test set in order to investigate differences in their respective performance in sentiment prediction in the target language. Table 1 shows the results. For BLSE, only the best-performing settings across all word embeddings and bilingual lexicons used are reported here. For deeper analysis of individual configurations, the reader is referred to Section 3.3

As can be seen from Table 1, both approaches show a marked difference of approx. 24 points in accuracy in the target language. We conclude that, in the absence of supervision in the target language, cross-lingual projection bears a strong advantage over a simple baseline that translate target text back into the source language for which a pre-trained language-specific classification model is available. Even though cross-lingual projection cannot fully compensate for the lack of ground truth labels in the target language, the observed differences between source- and target language performance are relatively small in case of cross-lingual projection, and considerably higher in case of the machine translation baseline (ΔBLSE: -5.0; ΔMT: -28.9). While these figures are not exactly comparable due to non-parallel testing data, they still suggest that cross-lingual projection approaches offer a feasible and very robust solution towards multilingual text analytics in specialized technical domains, even in the absence of language-specific training data in the target language of interest.

Moreover, we note that the results obtained from the machine translation baseline in our experiments are in contrast to the findings by Barnes and Klinger (2019). In most of their experiments on product and hotel reviews, the machine translation approach is found to outperform BLSE. This confirms our perspective that, in addition to language transfer, domain adaptation is a key challenge in most real-world scenarios of multilingual text analytics that involve more technical domains.

3.3 Experiment 2: Impact of Lexical Resources

Based on the finding from Experiment 1 that cross-lingual projection bears the potential of outperforming sequential translation pipelines (cf. Table 1), the goal in this experiment is to gain a deeper understanding about the impact of different configurations of bilingual lexicons on this result. As a first step, we report on the performance of individual lexicons being used as the source of bilingual knowledge in cross-lingual projection (cf. Section 3.3.1 below); subsequently, a first approach towards composition of these bilingual lexicons is evaluated (cf. Section 3.3.2).

3.3.1 Performance of Individual Lexicons in Cross-lingual Sentiment Projection

Table 2 shows the results of cross-lingual projection using BLSE when each of the lexicons introduced in Section 3.1.4 above is injected into the BLSE framework as the only source of bilingual information.

We clearly observe that the best performance in the target language is due to Apertium (Acc=0.759). The considerable margin over Pharma and BingLiu confirms the status of Apertium as a linguistically rich, general-purpose source of bilingual lexical knowledge. Even though the underlying data set is highly pharma-specific, the relative individual performance of Pharma and BingLiu suggests that task-specific sentiment information benefits cross-lingual projection approaches more than task-specific technical knowledge. Still, all information sources have valuable information to contribute, which can be seen from comparing their respective performance to the machine translation baseline in Table 1: Individual gains from cross-lingual projection over the machine translation pipeline range from approx. $+10$ points (for the Pharma lexicon) to more than $+22$ points in accuracy (for Apertium).

With respect to the monolingual word embeddings involved, a clear pattern of complementarity can be observed: Apertium benefits most[10] from domain-specific embeddings in the target language, whereas the domain-specific Pharma lexicon is best complemented by open-domain embeddings in both the source and the target language. For BingLiu, the combination of sentiment-specific knowledge in the lexicon and domain knowledge from the (source) embeddings works best.

3.3.2 Performance of Extended Lexicons in Cross-lingual Sentiment Projection

Following the procedures described in Section 3.1.4, three extensions of the Apertium, Pharma and BingLiu bilingual lexicons were generated

[10]For Apertium, Pharma, and Bing Liu, Table 2 displays only the best-performing configurations of monolingual embeddings.

	Monolingual Embeddings	Target Accuracy	Source Accuracy
Apertium	google; scielo_wiki	0.759	0.814
Pharma	google; sg_es_300	0.643	0.815
Bing Liu	PMC; sg_300_es	0.676	0.815

Table 2: Accuracy scores obtained from BLSE when different individual lexicons are used. Note that source and target language scores are not directly comparable due to non-parallel evaluation data.

	Monolingual Embeddings	Target Accuracy	Source Accuracy
Apertium only	google; scielo_wiki	0.759	0.814
Pharma & Apertium	google; scielo_wiki	0.767	0.821
Bing Liu & Apertium	google; scielo_wiki	0.756	0.817
Pharma & Bing Liu & Apertium	google; scielo_wiki	0.773	0.823

Table 3: Accuracy scores obtained from BLSE when different combinations of lexicons are used. Note that source and target language scores are not directly comparable due to non-parallel evaluation data.

and evaluated with respect to their impact on cross-lingual sentiment projection using BLSE. The results of this experiment are displayed in Table 3.

As can be seen from the comparison between the settings "Apertium only" (cf. Table 2; repeated here for the sake of convenience) and the combined lexicons, lexicon composition is effective in generating richer bilingual lexical representations that result in more accurate cross-lingual projection of sentiment classifiers. Apparently, this is due to a certain degree of complementarity within the original lexicons, which can be concluded from the facts that adding pharma-specific information to Apertium is already beneficial, while adding task knowledge from BingLiu on top leads to the best results overall.

In all extensions investigated, the combination of open-domain word embeddings for English and pharma-specific ones for Spanish yields the best results. This seems to suggest that the importance of word embeddings might decrease when a sufficiently comprehensive bilingual lexicon is available that covers many aspects of domain- or task-specific lexical knowledge. This conjecture needs closer investigation in future work, though.

4 Related Work

The idea of inducing resources and components for multi-lingual text analytics across languages (instead of creating them from scratch) has been attracting considerable attention in the NLP literature over the last decades now, dating back at least to Yarowsky et al. (2001). These early works are comparatively resource-intensive themselves, however, as they assume the availability of parallel or aligned corpora, which is a requirement that is still hard to meet for many language pairs, and even more so in technical domains or specific genres of text.

In more recent work, these requirements are substantially alleviated by representation learning approaches capitalizing on bilingual word embeddings which can be induced from parallel and non-parallel corpora. Due to their genericity, bilingual embedding approaches are sufficiently versatile in order to be applied to a variety of cross-lingual text classification problems (Mogadala and Rettinger, 2016). Cross-lingual sentiment analysis, as a special case of cross-lingual text classification problems, is investigated in multiple studies from a representation learning perspective, among them Zhou et al. (2016) or Zennaki et al. (2016). In several of the approaches reviewed here, machine translation baselines are used for comparison, similarly to our experimental settings in this papers. Alternatively, machine translation services can also be used for training data generation in order to train language-specific machine learning models subsequently (Balahur and Turchi, 2012).

Most recently, Feng and Wan (2019) present a projection approach based on bilingual sentiment-specific word embeddings (UBiSE) without any cross-lingual supervision, thus reducing external resoure requirements to a minimum: Only relying on a labeled sentiment corpus in the source language, as well as monolingual embeddings for both languages, their method outperforms BLSE on the same data. In light of our results presented in this paper, it remains to be evaluated as to whether UBiSE can be scaled to more technical domains as well.

The combination of lexical resources as a means to increase the amount of lexical knowledge that informs a text analytics model or component has found to be effective in tasks or applications such as contextual synonym expansion (Sinha and Mihalcea, 2009), entity recognition and linking (Hakenberg et al., 2011), or dialogue-based information access (Uszkoreit et al., 2006).

Barnes et al. (2018b) successfully apply the BLSE architecture to the related problem of domain adaptation in sentiment classification. Similarly, Barnes and Klinger (2019) present a real-world study on sentiment detection in Twitter data with a focus on the tourism domain. In these experiments, and likewise for all approaches reviewed here, domain (or genre) adaptation is investigated independently from language transfer (i.e., keeping the text language fixed to English). In that sense, we consider our work as a first step towards addressing a real-world problem in multilingual text analytics under realistic conditions of complexity that involve cross-lingual transfer and domain adaptation simultaneously.

5 Conclusions and Perspectives

In this paper, we investigate different options in order to enable multilingual text analytics without the need of (re-)training classification models from scratch for every new target language of interest. As an example case, we consider sentiment analysis for Spanish texts from the pharmaceutical domain. Comparing a machine translation pipeline to a cross-lingual projection approach using the BLSE framework (Barnes et al., 2018a), we find that the translation baseline enables multilingual analytics with low effort, but has inherent limitations that are not easily mitigated in a robust way. In contrast, a principled framework for cross-lingual sentiment projection clearly pays off, with a comparative edge of up to $+24$ points in accuracy over said translation baseline.

Moreover, the projected model predicts the sentiment of Spanish pharmaceutical texts at levels of accuracy that are reasonably close to the ones that can be achieved using specifically trained models for comparable English texts. At the same time, the development efforts for cross-lingual model projection are easily manageable, as manual annotations are required in a source language only.

In terms of the lexical resources that are required by BLSE, we find that attention should be paid to their careful selection and combination in order to foster complementarity between task- and domain-specific information as much as possible. In our case study on sentiment projection, the bilingual LLOD resource Apertium turns out as a particularly valuable source of broad-coverage open-domain knowledge that can be seen as the backbone of effective cross-lingual projection in our experiments.

Finally, our results suggest that the full potential of lexical resource combination for the cross-lingual projection problem investigated here has not been exhausted yet, as the domain-specific Pharma lexicon is currently limited to a small set of entity types only. We plan to extend this to more varied linguistic types and surface patterns in future work.

6 Acknowledgements

For this work we have received funding from the H2020 project Prêt-à-LLOD under Grant Agreement number 825182.

7 Bibliographical References

References

Wasi Uddin Ahmad, Zhisong Zhang, Xuezhe Ma, Kai-Wei Chang, and Nanyun Peng. 2019. Cross-lingual dependency parsing with unlabeled auxiliary languages. In *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)*, pages 372–382, Hong Kong, China. Association for Computational Linguistics.

Alexandra Balahur and Marco Turchi. 2012. Multilingual sentiment analysis using machine translation? In *Proceedings of the 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis*, pages 52–60, Jeju, Korea. Association for Computational Linguistics.

Jeremy Barnes and Roman Klinger. 2019. Embedding projection for targeted cross-lingual sentiment: Model comparisons and a real-world study. *Journal of Artificial Intelligence Research*, 66:691–742.

Jeremy Barnes, Roman Klinger, and Sabine Schulte im Walde. 2018a. Bilingual sentiment embeddings: Joint projection of sentiment across languages. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2483–2493, Melbourne, Australia. Association for Computational Linguistics.

Jeremy Barnes, Roman Klinger, and Sabine Schulte im Walde. 2018b. Projecting embeddings for domain adaption: Joint modeling of sentiment analysis in diverse domains. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 818–830, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Christian Chiarcos, Sebastian Hellmann, and Sebastian Nordhoff. 2012. Linking linguistic resources. Examples from the Open Linguistics Working Group. In Christian Chiarcos, Sebastian Nordhoff, and Sebastian Hellmann, editors, *Linked Data in Linguistics. Representing Language Data and Metadata*, pages 201–216. Springer.

Philipp Cimiano, Christian Chiarcos, John McCrae, and Jorge Gracia. 2020. *Linguistic Linked Open Data*. Springer.

Yanlin Feng and Xiaojun Wan. 2019. Learning bilingual sentiment-specific word embeddings without cross-lingual supervision. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 420–429, Minneapolis, Minnesota. Association for Computational Linguistics.

Jorge Gracia, Marta Villegas, Asunción Gómez-Pérez, and Núria Bel. 2017. The apertium bilingual dictionaries on the web of data. *Semantic Web*, 9:231–240.

Jörg Hakenberg, Martin Gerner, Maximilian Haeussler, Illés Solt, Conrad Plake, Michael Schroeder, Graciela Gonzalez, Goran Nenadic, and Casey M. Bergman. 2011. The GNAT library for local and remote gene mention normalization. *Bioinformatics*, 27(19):2769–2771.

Minqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '04, pages 168–177, New York, NY, USA. ACM.

Mikhail Kozhevnikov and Ivan Titov. 2014. Cross-lingual model transfer using feature representation projection. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 579–585, Baltimore, Maryland. Association for Computational Linguistics.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representationsin vector space. https://arxiv.org/pdf/1301.3781.pdf.

Aditya Mogadala and Achim Rettinger. 2016. Bilingual word embeddings from parallel and non-parallel corpora for cross-language text classification. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 692–702, San Diego, California. Association for Computational Linguistics.

Ravi Sinha and Rada Mihalcea. 2009. Combining lexical resources for contextual synonym expansion. In *Proceedings of the International Conference RANLP-2009*, pages 404–410, Borovets, Bulgaria. Association for Computational Linguistics.

Hans Uszkoreit, Feiyu Xu, Jörg Steffen, and Ilhan Aslan. 2006. The pragmatic combination of different crosslingual resources. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*, Genoa, Italy. European Language Resources Association (ELRA).

David Yarowsky, Grace Ngai, and Richard Wicentowski. 2001. Inducing multilingual text analysis tools via robust projection across aligned corpora. In *Proceedings of the First International Conference on Human Language Technology Research*.

Othman Zennaki, Nasredine Semmar, and Laurent Besacier. 2016. Inducing multilingual text analysis tools using bidirectional recurrent neural networks. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 450–460, Osaka, Japan. The COLING 2016 Organizing Committee.

Xinjie Zhou, Xiaojun Wan, and Jianguo Xiao. 2016. Cross-lingual sentiment classification with bilingual document representation learning. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1403–1412, Berlin, Germany. Association for Computational Linguistics.

Ayah Zirikly and Masato Hagiwara. 2015. Cross-lingual transfer of named entity recognizers without parallel corpora. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 390–396, Beijing, China. Association for Computational Linguistics.