Using Cross-Lingual Explicit Semantic Analysis for Improving Ontology Translation

Kartik Asooja[1] *Jorge Gracia*[1]
Nitish Aggarwal[2] *Asunción Gómez Pérez*[1]
(1) Ontology Engineering Group, UPM, Madrid, Spain
(2) DERI, NUIG, Galway, Ireland
asooja@gmail.com, jgracia@fi.upm.es, nitish.aggarwal@deri.org, asun@fi.upm.es

ABSTRACT
Semantic Web aims to allow machines to make inferences using the explicit conceptualisations contained in ontologies. By pointing to ontologies, Semantic Web-based applications are able to inter-operate and share common information easily. Nevertheless, multilingual semantic applications are still rare, owing to the fact that most online ontologies are monolingual in English. In order to solve this issue, techniques for ontology localisation and translation are needed. However, traditional machine translation is difficult to apply to ontologies, owing to the fact that ontology labels tend to be quite short in length and linguistically different from the free text paradigm. In this paper, we propose an approach to enhance machine translation of ontologies based on exploiting the well-structured concept descriptions contained in the ontology. In particular, our approach leverages the semantics contained in the ontology by using Cross Lingual Explicit Semantic Analysis (CLESA) for context-based disambiguation in phrase-based Statistical Machine Translation (SMT). The presented work is novel in the sense that application of CLESA in SMT has not been performed earlier to the best of our knowledge.

KEYWORDS: Ontology Translation, Word-Sense Disambiguation, Statistical Machine translation, Explicit Semantic Analysis, Ontology Localisation.

Second ML4HMT Workshop, pages 25–36,
COLING 2012, Mumbai, December 2012.

1 Introduction

An ontology is a formal specification of a shared conceptualization (Gruber, 1993). Since the rise of Semantic Web, many ontology-based applications have been developed, for instance in the fields of ontology-based information extraction (Buitelaar et al., 2008), semantic search (Fernandez et al., 2008) and cross lingual information extraction (Embley et al., 2011). Nevertheless, due to the fact that most of the ontologies have been documented only in English and multilingual ontologies are rare, semantic applications that exploit information across natural language barriers are uncommon. In order to cross such barriers, a critical mass of multilingual ontologies is needed, as well as methods and techniques for ontology localisation and translation. In fact, ontology localisation, or the adaptation of an ontology to a particular language and culture (Espinoza et al., 2008a) has been identified as one of the key challenges of the multilingual Semantic Web (Gracia et al., 2012).

Translation of an ontology documented in a source language into target language is one of the most important steps in ontology localisation. Translating the ontology affects the lexical layer of an ontology. This layer includes all the natural language description including labels, comments, definitions, and associated documentation to make that ontology understandable for humans (Cimiano et al., 2010).

Ideally, ontology translation has to be supported by automatic methods, as finding domain experts knowing many languages is very difficult and expensive. It can be achieved by using machine translation (MT) techniques. Unfortunately, the labels in the ontologies pose extra challenges for standard practices in MT because of the different linguistic structure and short text length of the ontology labels compared to the free text paradigm (McCrae et al., 2011). In fact, ontology labels need not to be fully grammatically-correct sentences. Thus, a single ontology label typically constitutes a poor context to disambiguate the candidate translations of a lexical entry contained in that label.

It has been shown that performing word sense disambiguation (WSD) using the surrounding words for disambiguating the possible translations improve machine translation (Carpuat and Wu, 2007) (Chan et al., 2007). Following a similar intuition, such context disambiguation can also be applied to the translation of ontologies (Espinoza et al., 2008a). In that case, the ontology concepts are precisely defined and related to other concepts. Thus, the context of a concept can be enriched with the labels and textual descriptions of its connected concepts, and such context can be exploited for semantic disambiguation.

Therefore, we want to leverage the context from the ontology for improving the translation of labels. In this work, we use Cross Lingual Explicit Semantic Analysis (CLESA) based context disambiguation between the ontology context and the translation candidates, to rank the candidates, in the phrase-based Statistical Machine Translation (SMT) architecture. In our experiments, we use the labels of the connected entities of the source label in the ontology as the ontological context for any lexical entry, which comes from the source label.

This paper describes an approach that exploits ontological context from the ontology for improving automatic translation of the ontology labels. In particular, we have investigated the use of CLESA in SMT for this purpose. The remainder of this paper

is structured as follows: Section 2 discusses some background required for better understanding of the rest of the paper. Section 3 describes the approach for using CLESA based WSD in SMT for ontology translation. Section 4 explains the evaluation setup and reports the experimental results. Section 5 describes some related work about the translation of ontologies. Finally, conclusions and future work are reported in the final section of the paper.

2 Background

In order to allow a better understanding of the rest of the paper, we briefly introduce here some basic notions of the techniques used in our approach.

2.1 Statistical Machine Translation

The statistical machine translation model utilizes the standard source-channel approach for statistically modeling the translation problem (Koehn et al., 2003) as follows:

$$argmax_{tgt}P(tgt|src) = argmax_{tgt}P(src|tgt)\ \ P_{LangModel}(tgt) \quad (1)$$

In equation 1, src and tgt refer to the source phrase and translated phrase respectively. The heuristic-based search is performed by the machine translation decoder to deduce the translation candidate with the maximum probability given the source phrase.

Phrase-based models generally perform better than word-based models as the phrase-based model tries to learn more of the local context and reduces the restrictions of word-based translation by translating whole sequences of words (Koehn et al., 2003). The phrases here are a sequence of words with all possible n-grams rather than only the linguistically correct phrases.

2.2 Cross Lingual Explicit Semantic Analysis

Explicit Semantic Analysis (ESA) was introduced by (Gabrilovich and Markovitch, 2007), and allows the semantic comparison of two texts with the help of explicitly defined concepts. In contrast, other techniques such as Latent Semantic Analysis (Landauer and Foltz, 1998) and Latent Dirichlet Allocation (Blei et al., 2003) build unsupervised concepts considering the correlations of the terms in the data. ESA is an algebraic model in which the text is represented with a vector of the explicit concepts as dimensions. The magnitude of each dimension in the vector is the associativity weight of the text to that explicit concept/dimension. To quantify this associativity, the textual content related to the explicit concept/dimension is utilized. This weight can be calculated by considering different methods, for instance, tf-idf score. A possible way of defining concepts in ESA is by means of using the Wikipedia [1] titles as dimensions of the model and the corresponding articles for calculating the associativity weight (Gabrilovich and Markovitch, 2007), thus taking advantage of the vast coverage of the community-developed Wikipedia. A compelling characteristic of Wikipedia is the large collective knowledge available in multiple languages, which facilitates an extension of existing ESA for multiple languages called Cross-lingual

[1] http://www.wikipedia.org/

Explicit Semantic Analysis (CLESA) (Sorg and Cimiano, 2008). The articles in Wikipedia are linked together across language, and this cross-lingual linked structure can provide a mapping of a vector in one language to the other. Thus, Wikipedia provides the comparable corpus in different languages, which is required by CLESA.

To understand CLESA, lets take two terms t_s in source language and t_t in the target language. As a first step, a concept vector for t_s is created using the Wikipedia corpus in the source language. Similarly, the concept vector for t_t is created in the target language. Then, one of the concept vectors can be converted to the other language by using the cross-lingual mappings provided by Wikipedia. After obtaining both of the concept vectors in one language, the relatedness of the terms t_s and t_t can be calculated by using cosine product, similar to ESA. For better efficiency, we chose to make a multilingual index by composing poly-lingual Wikipedia articles using the cross-lingual mappings. In such a case, no conversion of the concept vector in one language to the other is required. It is possible by representing the Wikipedia concept with some unique name common to all languages such as, for instance, the Uniform Resource Identifier (URI) of the English Wikipedia.

3 CLESA with SMT for Translating Ontologies

SMT systems implicitly use the local context for a better lexical choice during the translation (Carpuat and Wu, 2005). Accordingly, it is natural to assume that a focused WSD system integrated with SMT system might produce better translations. We follow the direct incorporation of WSD into SMT system as a multi-word phrasal lexical disambiguation system (Carpuat and Wu, 2007).

The WSD probability score calculated by using CLESA is added as an additional feature in the log-linear translation model. The CLESA based score would depend on the ontology in which the source label lies and ergo, the context of the ontology would be used to disambiguate the translation candidates. Equation 2 shows the integration of WSD in the standard phrase-based MT.

$$argmax_{tgt} P(tgt|src, O) = argmax_{tgt} P_{Translation}(src|tgt) P_{LangModel}(tgt) P_{Semantic}(tgt|O) \quad (2)$$

Here, the computation of equation 2 requires a heuristic search by the decoder to seek the best translation given the ontology O and the source phrase. The factor $P_{Semantic}(tgt|O)$ provides the probability score for a translation candidate given the ontology. This score is found by calculating the CLESA based semantic relatedness between the ontological context and the translation candidates. There can be several possibilities for selecting the context from the ontology, including the option to use the structure of the ontology for disambiguation (Espinoza et al., 2008a). For our experiments, the ontological context consists of labels of the connected entities to the source label in the ontology. Thereupon, we take a bag of words used in all the labels of the ontology and build the concept vector for the ontology to compare it with the concept vector of the translation candidates. We have employed Stanford Phrasal library (Cer et al., 2010), which is a phrase-based SMT system, in our architecture. It easily allows the integration of new features into the decoding model along with the already available features in the library (Cer et al., 2010).

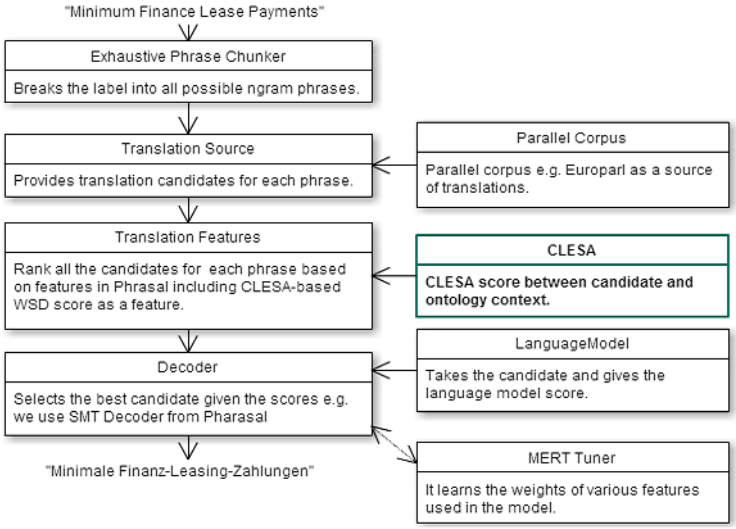

Figure 1: Phrase-based SMT architecture with CLESA integration.

Fig.1 shows the architecture applied for translating the ontologies. As an example, suppose that the SMT system has already been trained with some parallel corpus, for instance, Europarl corpus (Koehn, 2005). It receives an example English label "Minimum Finance Lease Payments" from a source ontology to translate it into German. The label is broken into a phrase chunk list containing all the possible phrases by the *Exhaustive Phrase Chunker*. As a next step, the *Translation Source* provides all the possible translation candidates for each phrase chunk in the chunk list. *Translation Source* can be a phrase table made from some parallel corpus like Europarl. Then, scores are assigned to all the translation candidates based on several standard *Translation Features* exisiting in the Phrasal library. As an additional feature, we introduce one more score based on the *CLESA* based semantic relevance of the candidate against the source ontology context, which includes all its labels. All these feature scores are combined by a log-linear model. The *MERT Tuner* (Och, 2003) is just used once to learn the weights of various features used in the model for a particular language pair. In the final step, the decoder performs search over all the translation candidates given the scores from *Translation Features* and the *Language Model*, and makes the German translation "Minimale Finanz-Leasing-Zahlungen".

For implementing CLESA, we followed an information retrieval based approach by creating a Lucene [2] inverted index of a Wikipedia dump from Jan, 2012. As a preprocessing step, all the Wikipedia namespace type articles such as mediaWiki, talk, help etc. were removed. For creating the weighted vector of concepts for a translation candidate in the

[2]http://lucene.apache.org/core/

target language, the term is searched over the Wikipedia index of the respective language to retrieve the top associated Wikipedia concepts and the Lucene ranking scores are taken as the associativity weights of the concepts to the term. We took the top 2000 Wikipedia concepts for our experiment as we found that increasing this number did not have any major effect on the translation metrics, but it significantly increases the computational time. As the ontological context for any phrase chunk, we use the source label along with the labels from the connected entities to the source label in the ontology. Thus, the concept vector for the ontological context is created by searching the ontological context in the Wikipedia index of the source language.

4 Evaluation

To evaluate the integration of CLESA in the SMT architecture, we perform the translation of several ontologies and compare the results, against reference translations, with the translations performed by a baseline SMT system. We used widely accepted machine translation metrics in our evaluation: WER (Popović and Ney, 2007), BLEU (Papineni et al., 2002), NIST(Doddington, 2002), METEOR (Banerjee and Lavie, 2005). All the translations were performed for English to Spanish, English to German and English to Dutch language pairs.

4.1 Experimental Setup

To build a baseline SMT system, we used the Stanford Phrasal library trained on EuroParl corpus (Koehn, 2005). In order to define our benchmark, we have used some multilingual ontologies available online (See table 1). For tuning the SMT system using MERT tuner, IFRS ontology was used as it contains 2757 labels (McCrae et al., 2011) for each language in the consideration, which is quite large against the number of labels present in the ontologies used for evaluation. We used a monolingual version of each ontology as input to the evaluation process. Then, we used the labels in the target language as reference translations and compared them to the translations obtained in the process. Finally, the evaluation metrics were computed. To test the effect of CLESA-based disambiguation in SMT, we run the experiments both with our SMT baseline system and with the CLESA integrated in the baseline system.

4.2 Results and Discussions

Tables 2, 3 and 4 show the results in our experiments for the English to German, English to Spanish and English to Dutch language pairs respectively.

We can see that the metric scores are low, which could be mainly because of lower word/phrase coverage. Although, the results show an improvement in BLEU-2, METEOR, NIST and WER (WER is better if the score is low) but not in BLEU-4. This is the result of the linguistic differences between free-text and ontology labels. Labels of an ontology generally tend to be shorter in length, therefore BLEU-2 (BLEU with 2-grams) gives better correlation with the reference translations than BLEU-4 (BLEU with 4-grams). It is probably because the average number of tokens is less than 4 in the evaluated ontologies. These metrics

Ontology	English	Spanish	German	Dutch
GeoSkills	211	46	238	360
Crop-Wild Relatives Ontology	1030	1025	0	0
FOAF	88	79	0	0
Housing Benefits	841	0	0	841
Open EHR Reference	36	36	0	0
Registratie Bedrijven	854	0	0	854
DOAP	47	36	35	0
ITCC CI 2011	417	0	417	0
Open EHR Demographics	24	24	0	0

Table 1: Multilingual Ontologies with the number of labels in the respective languages

Ontology		BLEU-4	BLEU-2	METEOR	NIST	WER
DOAP	Baseline	0.0	0.0	0.014	0.101	1.176
	CLESA	0.0	0.0	0.014	0.101	1.176
ITCC CI 2011	Baseline	0.0	0.022	0.043	0.791	1.070
	CLESA	0.0	0.022	0.044	0.802	1.068
GeoSkills	Baseline	0.0	0.0	0.032	0.509	1.214
	CLESA	0.0	0.0	0.034	0.523	1.209
Summary	Baseline	**0.0**	**0.014**	**0.038**	**0.669**	**1.118**
	CLESA	**0.0**	**0.014**	**0.039**	**0.680**	**1.117**

Table 2: Baseline and Baseline+CLESA scores for English to German

do not suit well to the task of ontology translation as they do in the free text paradigm (McCrae et al., 2011). Therefore, there is a need for the development of new metrics for evaluating the translation of ontologies.

From the result tables, we can see that the use of the CLESA ranker slightly improves the baseline results in most of the cases. The improvement is little because the integration of CLESA does not provide new translation candidates to the system, it just gives more weightage to the ones which are semantically more related to the ontological context.

5 Related Work

Label-Translator, developed as a NEON plug-in (Espinoza et al., 2008b), is one of the initial initiatives to automatically localize the ontology. It does not follow SMT-centered approach (Espinoza et al., 2008a). As a first step, it collects the candidate translations for a label by consulting different bilingual linguistic resources and translation services such as Google Translate. Then, it performs WSD by using the ontological context of the label against the candidates for selecting the best one. The context in which those candidates appear in different domains is taken from various multilingual ontologies and linguistic resources such as EuroWordnet (Vossen, 1998). One of the pre-requisites of Label-Translator is

Ontology		BLEU-4	BLEU-2	METEOR	NIST	WER
DOAP	Baseline	0.0	0.145	0.204	1.891	0.853
	CLESA	0.0	0.149	0.211	1.985	0.853
Open EHR Demographics	Baseline	0.0	0.0	0.095	0.736	1.028
	CLESA	0.0	0.0	0.095	0.736	1.028
CWR	Baseline	0.075	0.180	0.170	3.072	0.983
	CLESA	0.074	0.180	0.175	3.152	0.986
Open EHR Reference	Baseline	0.0	0.152	0.206	1.516	0.933
	CLESA	0.0	0.155	0.220	1.600	0.920
GeoSkills	Baseline	0.256	0.254	0.246	2.289	0.938
	CLESA	0.0	0.230	0.240	2.202	0.954
FOAF	Baseline	0.0	0.187	0.204	2.487	0.874
	CLESA	0.0	0.187	0.204	2.487	0.874
Summary	Baseline	**0.069**	**0.177**	**0.175**	**2.888**	**0.971**
	CLESA	**0.061**	**0.177**	**0.179**	**2.958**	**0.973**

Table 3: Baseline and Baseline+CLESA scores for English to Spanish

Ontology		BLEU-4	BLEU-2	METEOR	NIST	WER
Registratie Bedrijven	Baseline	0.0	0.113	0.112	1.540	0.955
	CLESA	0.0	0.113	0.113	1.550	0.954
Housing Benefits	Baseline	0.0	0.128	0.120	1.530	0.908
	CLESA	0.0	0.127	0.120	1.530	0.910
GeoSkills	Baseline	0.0	0.099	0.076	1.181	1.113
	CLESA	0.0	0.100	0.079	1.230	1.108
Summary	Baseline	**0.0**	**0.117**	**0.113**	**1.520**	**0.945**
	CLESA	**0.0**	**0.117**	**0.114**	**1.528**	**0.944**

Table 4: Baseline and Baseline+CLESA scores for English to Dutch

that it relies on the existence of the candidate translations in EuroWordNet (or similar resources) in order to operate. On the contrary, the CLESA-based approach does not suffer such limitation. Our approach does not, therefore, depend on the availability of external translation services. Furthermore, thanks to the wide language coverage of Wikipedia, the extension of the CLESA-based approach to other language pairs is straightforward.

The problem of translating ontologies has already been discussed in the context of SMT (McCrae et al., 2011), although, not much work has been done in actually experimenting with WSD in a SMT system for translating ontologies.

Therefore, we integrated CLESA into a phrase-based SMT architecture for translating labels of the ontologies. CLESA is shown to perform better than the latent concept models in the context of cross lingual information retrieval task (Cimiano et al., 2009), which motivated us to use it in SMT also.

Conclusion

We have presented an approach for ontology translation that uses CLESA for leveraging the ontological context in a Statistical Machine Translation process. Integration of CLESA based disambiguation using all the ontology labels in SMT architecture, provides the selection of the translation candidates given the ontological context, in contrast to the standard phrase-based model, which considers only the local context in the label. The results show little improvements over the baseline scores for most of the evaluation metrics, thus proving that exploring the ontology context based disambiguation may be beneficial in the process of translating the ontologies. Nevertheless, more research is needed in that direction in order to attain better results. As future work, we plan to investigate better ways of exploiting the ontological context for machine translation of labels and to compare our system against the Label-Translator.

Acknowledgements

This work is supported in part by the European Union under Grant No. 248458 for the Monnet project.

References

Banerjee, S. and Lavie, A. (2005). Meteor: An automatic metric for mt evaluation with improved correlation with human judgments. pages 65–72.

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3:993–1022.

Buitelaar, P., Cimiano, P., Frank, A., Hartung, M., and Racioppa, S. (2008). Ontology-based information extraction and integration from heterogeneous data sources. *International Journal of Human Computer Studies (JHCS)*, 66:759–788.

Carpuat, M. and Wu, D. (2005). Evaluating the word sense disambiguation performance of statistical machine translation.

Carpuat, M. and Wu, D. (2007). Improving statistical machine translation using word sense disambiguation. In *In The 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL 2007*, pages 61–72.

Cer, D., Galley, M., Jurafsky, D., and Manning, C. D. (2010). Phrasal: a toolkit for statistical machine translation with facilities for extraction and incorporation of arbitrary model features. In *Proceedings of the NAACL HLT 2010 Demonstration Session*, HLT-DEMO '10, pages 9–12, Stroudsburg, PA, USA. Association for Computational Linguistics.

Chan, Y. S., Ng, H. T., and Chiang, D. (2007). Word sense disambiguation improves statistical machine translation. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 33–40, Prague, Czech Republic. Association for Computational Linguistics.

Cimiano, P., Montiel-Ponsoda, E., Buitelaar, P., Espinoza, M., and Gómez-Pérez, A. (2010). A note on ontology localization. *Appl. Ontol.*, 5(2):127–137.

Cimiano, P., Schultz, A., Sizov, S., Sorg, P., and Staab, S. (2009). Explicit versus latent concept models for cross-language information retrieval. In *Proceedings of the 21st international jont conference on Artifical intelligence*, IJCAI'09, pages 1513–1518, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Doddington, G. (2002). Automatic evaluation of machine translation quality using n-gram co-occurrence statistics. In *Proceedings of the second international conference on Human Language Technology Research*, HLT '02, pages 138–145, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Embley, D. W., Liddle, S. W., Lonsdale, D. W., and Tijerino, Y. (2011). Multilingual ontologies for cross-language information extraction and semantic search. In *Proceedings of the 30th international conference on Conceptual modeling*, ER'11, pages 147–160, Berlin, Heidelberg. Springer-Verlag.

Espinoza, M., Gómez-Pérez, A., and Mena, E. (2008a). Enriching an ontology with multilingual information. In *Proceedings of the 5th European semantic web conference on The semantic web:research and applications*, ESWC'08, pages 333–347, Berlin, Heidelberg. Springer-Verlag.

Espinoza, M., Gómez-Pérez, A., and Mena, E. (2008b). Labeltranslator - a tool to automatically localize an ontology. In *Proceedings of the 5th European semantic web conference on The semantic web: research and applications*, ESWC 08, pages 792–796, Berlin, Heidelberg. Springer-Verlag.

Fernandez, M., Lopez, V., Sabou, M., Uren, V., Vallet, D., Motta, E., and Castells, P. (2008). Semantic search meets the web. In *Proceedings of the 2008 IEEE International Conference on Semantic Computing*, ICSC '08, pages 253–260, Washington, DC, USA. IEEE Computer Society.

Gabrilovich, E. and Markovitch, S. (2007). Computing semantic relatedness using wikipedia-based explicit semantic analysis. In *Proceedings of the 20th international joint conference on Artifical intelligence*, IJCAI'07, pages 1606–1611, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Gracia, J., Montiel-Ponsoda, E., Cimiano, P., Gómez-Pérez, A., Buitelaar, P., and McCrae, J. (2012). Challenges for the multilingual web of data. *Web Semant.*, 11:63–71.

Gruber, T. R. (1993). A translation approach to portable ontology specifications. *KNOWLEDGE ACQUISITION*, 5:199–220.

Koehn, P. (2005). Europarl: A Parallel Corpus for Statistical Machine Translation. In *Conference Proceedings: the tenth Machine Translation Summit*, pages 79–86, Phuket, Thailand. AAMT, AAMT.

Koehn, P., Och, F. J., and Marcu, D. (2003). 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 - Volume 1*, NAACL '03, pages 48–54, Stroudsburg, PA, USA. Association for Computational Linguistics.

Landauer, T. K. and Foltz, P. W. (1998). An Introduction To Latent Semantic Analysis.

McCrae, J., Espinoza, M., Montiel-Ponsoda, E., Aguado-de Cea, G., and Cimiano, P. (2011). Combining statistical and semantic approaches to the translation of ontologies and taxonomies. In *Proceedings of the Fifth Workshop on Syntax, Semantics and Structure in Statistical Translation*, SSST-5, pages 116–125, Stroudsburg, PA, USA. Association for Computational Linguistics.

Och, F. J. (2003). Minimum error rate training in statistical machine translation. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics - Volume 1*, ACL '03, pages 160–167, Stroudsburg, PA, USA. Association for Computational Linguistics.

Papineni, K., Roukos, S., Ward, T., and jing Zhu, W. (2002). Bleu: a method for automatic evaluation of machine translation. pages 311–318.

Popović, M. and Ney, H. (2007). Word error rates: decomposition over pos classes and applications for error analysis. In *Proceedings of the Second Workshop on Statistical Machine Translation*, StatMT '07, pages 48–55, Stroudsburg, PA, USA. Association for Computational Linguistics.

Sorg, P. and Cimiano, P. (2008). Cross-lingual information retrieval with explicit semantic analysis. In *Working Notes for the CLEF 2008 Workshop*.

Vossen, P., editor (1998). *EuroWordNet: a multilingual database with lexical semantic networks*. Kluwer Academic Publishers, Norwell, MA, USA.