Building a Multilingual Named Entity-Annotated Corpus Using Annotation Projection

Maud Ehrmann, Marco Turchi, Ralf Steinberger
European Commission - Joint Research Centre (JRC), IPSC - GlobeSec
Via Fermi 2749, 21020 Ispra (VA) - Italy
name.surname@jrc.ec.europa.eu

Abstract

As developers of a highly multilingual named entity recognition (NER) system, we face an evaluation resource bottleneck problem: we need evaluation data in many languages, the annotation should not be too time-consuming, and the evaluation results across languages should be comparable. We solve the problem by automatically annotating the English version of a multi-parallel corpus and by projecting the annotations into all the other language versions. For the translation of English entities, we use a phrase-based statistical machine translation system as well as a lookup of known names from a multilingual name database. For the projection, we incrementally apply different methods: perfect string matching, perfect consonant signature matching and edit distance similarity. The resulting annotated parallel corpus will be made available for reuse.

1 Introduction

Named Entity recognition is a well-established task, acknowledged as fundamental to a wide variety of natural language processing (NLP) applications (Nadeau and Sekine, 2007). As for other text mining applications, annotated corpora constitute a crucial and constant need for named entity recognition (NER). Within a development or training framework, annotated corpora are used as models from which machine learning systems, or computational linguists, can infer rules and decision criteria; within an evaluation framework, they are used as a gold standard to assess systems' performances and help to guide their quality improvement, *e.g.* via non-regression tests.

During the last decade, several named entity (NE) annotated corpora were built, thanks to a large series of evaluation campaigns (Fort et al., 2009). However, most of these gold-standard data are available only for English or for a few languages. Even if unsupervised methods tried to overcome this difficulty, the shortage of annotated data for the large majority of the world's languages remains a problem. An obvious solution is to manually produce annotated corpora, but it is a complex and time-consuming task and it may be difficult to find experts in each specific language.

Beyond the scarcity of annotated corpora, another issue lies in the fact that annotation schemas or guidelines usually differ from one annotated corpus to another: named entity extents can be different (*e.g.* inclusion or not of the function in a person name), as well as entity types and granularity (*e.g.* some corpora may consider product names, whereas others will differentiate, within this category, vehicles, awards and documents, and others will not even consider product names). Such divergences should be expected, as annotated corpora are built according to different applications. However, they constitute a real issue, particularly when developing or evaluating multilingual NE recognition systems and the effort to reuse existing annotated data collections is big.

Our goal is to automatically build a set of multilingual named entity-annotated corpora, taking advantage of the existence of parallel corpora (bilingual or multiparallel). Traditionally used in the field of Machine Translation, parallel corpora have been exploited in recent years in various NLP tasks, including linguistic annotation, with the creation of annotated corpora. The underlining principle is *annotation projection*, where annotations available for a text in one language can be projected, thanks to the alignment, to the corresponding text in another language, creating herewith a newly annotated corpus for a new language.

This paper presents how we applied this method to named entity annotations, projecting automatically annotated English entities to, firstly, French, Spanish, German and Czech multiparallel corpora and, secondly, Russian parallel corpora. We experimented with several annotation projection techniques: starting from the baseline of simply searching for the English name string in foreign text, results improve gradually by adding new information and varying projection methods. Our objective is to make freely available named entity annotated-corpora in a large set of languages, with a quality similar to that of manually annotated data.

This method shows several advantages. Firstly it could be a way of overcoming the NE-annotated data shortage problem. Then, it could solve the non-harmonized annotation issue: if the projected annotations (on the target side) always come from the same automatic recognition system (on the source side), then we obtain annotated corpora in different languages, but with a common annotation schema. The use of multiparallel corpora also presents the benefit of ensuring the comparability of NER system results across languages; morever, as named entity recognition systems are domain-sensitive, it could be relevant to evaluate multilingual NER systems on equivalent tasks.

The remainder of the paper is organized as follows. We introduce related work (section 2), then present our NE projection method (section 3), report the results (section 4) and finally conclude and propose some elements for future work (section 5).

2 Related Work

Regarding the automatic acquisition of NE annotated corpora, some work investigates how to constitute monolingual annotated data (An et al., 2003; Nothman et al., 2008).

With respect to parallel corpora, their exploitation has been growing in recent years, showing their usefulness in various NLP tasks like word sense disambiguation or cross-lingual tagging (refer to the state of art presented by Bentivolgi and Pianta (2005)). With respect to cross-lingual knowledge induction, multiple work addressed the challenge of automatic parallel treebank building (Lavie et al., 2008; Hwa et al., 2005), whereas (Padó and Lapata, 2009; Bentivogli and Pianta, 2005) explored semantic information projection.

Several researchers investigated named entity annotation and parallel corpora exploitation. Yarowsky *et al.* (2001) carried out some pioneer experiments, investigating the feasibility of annotation projection over four NLP tasks, including named entity recognition. The goal was to automatically induce stand-alone text analysis tools via robust (and noisy) annotation projection. More recently, Ma (2010) applied a co-training algorithm on unlabelled bilingual data (English-Chinese), showing that NE taggers can complement and improve each other while working together on parallel corpora. Samy *et al.* (2005) developed a named entity recognizer for Arabic, leveraging an Arabic-Spanish parallel corpus aligned at sentence level and POS tagged. With a slightly different goal, Klementiev and Roth (2008) proposed an algorithm for cross-lingual multiword NE discovery in a bilingual weakly temporally aligned corpus. The work of Volk *et al.* (2010) on combining parallel treebanks and geo-tagging showed similar results to what we offer, with the difference that they focused on the location type only and worked with a bilingual French-German corpus. Finally, Shah *et al.* (2010) designed a Machine Translation-based approach to NER which includes a NE annotation projection phase based on word alignment.

These approaches aimed at developing/improving NER systems and parallel annotated corpora seemed to be a positive side-effect of these experiments. In comparison, our work differs from that mentioned here in that we aim at developing an annotated multilingual parallel corpus for evaluation purposes. Using a multilingual parallel corpus is beneficial over using a bilingual corpus in that we save more annotation time. More importantly, text type, entity type distribution, and entity annotation specifications are the same across all languages, resulting in a more useful evaluation resource. We will make this multi-parallel corpus freely available to other system developers.

3 Named Entity Annotation Projection

Given a multiparallel corpus and a monolingual NER system, our objective is to automatically provide NE annotations for each text of the aligned corpora. A possible solution to project a named entity between two aligned texts is to translate this entity; accordingly, our multilingual NE annotation projection method relies, for the most part, on the use of a statistical machine translation system.

We used a multiparallel corpus in English, French, Spanish, German and Czech (news texts coming from the WMT shared tasks (Callison-Burch et al., 2009)), hereafter *En-4*, and an English-Russian one (union of two news data sets (Klyueva and Bojar, 2008; Rafalovitch and Dale, 2009)), hereafter *En-Ru*. For each language, *En-4* has a training set of roughly 70,000 sentence pairs and a test set of 2,490 sentence pairs, against 160,000 and 2,700 respectively for *En-Ru*. We used the test sets for the annotation projection. The next sections detail each step of the NE annotation projection process.

3.1 Automatic annotation of Source Named Entities

The first step is to annotate NEs in one corpus in a given language. We chose to annotate English entities of type *Person* (including titles), *Location* and *Organisation* and tried to project them in the corresponding texts in other languages. As a matter of fact, English is a resource-rich language with already existing efficient tools, but one may choose another source language, according to his/her goals and constraints. We used an in-house NER system (Steinberger and Pouliquen, 2007; Crawley and Wagner, 2010) to process the English source side text (any NER system or even manual annotation could have been used at this stage). Obviously, the NER system's quality is a crucial element that determines the projection quality. In the English texts of the *En-4* and *En-Ru* corpora, the NER system annotated 826 unique entities (corresponding to 1,395 occurrences) and 674 (1,312 occurrences) respectively.

3.2 Source Named Entity Translation

The second step corresponds to the translation of the previously extracted entities into the target languages. We make use of two different NE translation sources: translations resulting from the application of a Phrase-Based Statistical Machine Translation system (PBSMT), and translations resulting from the exploitation of a multilingual Named Entity database.

3.2.1 PBSMT System

One of the most popular classes of statistical machine translation (SMT) systems is the Phrase-Based Model (Koehn, 2010). It is an extension of the noisy channel model, introduced by (Brown et al., 1994), using phrases rather than words. A source sentence f is segmented into a sequence of I phrases and the same is done for the target sentence e, where the notion of phrase is not related to any grammatical assumption: a phrase is an n-gram. The best translation e of f is obtained by maximizing the PBSMT model probability $p(e|f)$, relying on three components: the probability of translating a phrase e_i into a phrase f_i, the distance-based reordering model and the language model probability.

Phrases and probabilities are estimated processing the parallel data. Word to word alignment is firstly extracted running the IBM models (Brown et al., 1994), and then proximity rules are applied to obtain phrases, see (Koehn, 2010). Probabilities are estimated counting the frequency of the phrases in the parallel corpus. In this work, we used the open source PBSMT system Moses (Koehn et al., 2007).

Since Named Entities correspond most of the time to small sets of contiguous words (phrases), the phrase-based model appeared to be well-suited to translate this kind of units. Instead of running a whole SMT system, we could have used the word alignment only or done a simple phrase-table lookup. By choosing the first option, we would have been dependent on the NE alignment quality and, by choosing the second one, we would have lost another advantage of the PBSMT system: its decoder's capacities, which allows the reconstruction of the good target phrase even if spread over different phrases (a NE could be cut into different phrases during the phrase table extraction). These choices were confirmed by preliminary experiments.

Experimental framework We translate Named Entities in isolation and not the full sentences where they occur. Translating the full sentences would have implied finding again the entities in the output, which seemed quite complicated and time-consuming. Regarding the training phase, we chose a specific configuration that does not correspond to the classical idea of translation: we trained the PBSMT system using the training sets of the corpora *plus* the parallel sentences that we want to annotate, *i.e.* the test sets. It means that the translation system should know how to translate a source entity because it has seen it in the training data; this reduces the number of completely untranslated entities. Finally, with respect to the SMT output, we did not only consider the most probable translation but took into account the top

15 ranked translations according to $p(e|f)$.

Correction Phase Entity translations are not always correct because the PBSMT system tries to reproduce the most readable sentence driven by the language model; in this way, the translation system may add articles, prepositions or in some cases groups of words before or after the entity name. For example, the french translation of *Afghanistan* is *en Afghanistan* and the translation of *Germany* is *l' Allemagne*. In these cases, only *Afghanistan* and *Allemagne* should be projected, as prepositions and articles cannot be part of proper names in French. We could observe similar phenomena in other languages. To address this problem, we post-processed the translations in a simple way: applying stopword lists. This allowed us to correct a certain number of entities for each language, even if some wrong entities could remain in the translation list. Before projecting these "corrected" translated entities in the aligned corpora, we asked bilingual annotators to check the correctness of the translated entities, according to a set of evaluation categories that identifies possible translation errors. In all languages, the main problems seem to be the addition and subtraction of word(s) during the translation phase (En: *tariq ramadan* Fr: *peut-être tariq ramadan*). More details about this evaluation are reported in (Ehrmann and Turchi, 2010).

3.2.2 External Named Entity Resource

In addition to the SMT approach, we benefit from an external multilingual named entity database; it contains, among others, translations and transliterations of entity names in several languages. By querying this database, we retrieved, for each English entity, a list of translated entities (that may have different spellings) in a given language.[1]

The information coming from the external resource is quite reliable, because part of the entity names has been manually checked. However, it is not exhaustive. On the contrary, the SMT system provides translations almost every time, but they may be incorrect. In other words, information coming from the external resource and the SMT system can complement each other, the former boosting precision and the latter ensuring recall. For example, *Sakharov Prize for Freedom of Thought* is correctly translated by the SMT system for each language while the database does not contain this name.

3.3 Annotation Projection Methods

Once we have a list of possible translations (or candidates) for a given NE in an English sentence, we try to project it into the corresponding sentences of the aligned corpora. We incrementally apply different projection strategies.

String matching The first projection method we use is a strict string matching: the candidate is present or not in the translated sentence. With this method, we are able to project the entity *european parliament*[2] from the English sentence to the corresponding Spanish one in the following example:

English: recipients of the 2005 sakharov prize from the <organization>european parliament</organization>...
Candidate list: parlamento europeo, presidente del parlamento europeo, parlamento de europa, parlamento europea
Spanish: las "Damas de Blanco", galardonadas con el premio sajarov 2005 otorgado por el <organization> parlamento europeo </organization>, ...

This method is rigorous and does not allow to catch named entities showing different spellings or morphological variants that are not present in the candidate list. The following is an example where the entity (*tariq ramadan*) cannot be projected in the target Czech sentence:

English: with the possible exception of <person> tariq ramadan </person>...
Candidate list: tariq ramadan, ramadan tariq
Czech: nevyjímaje tarika ramadana

Consonant Signature matching If the string matching method does not retrieve any result, then we try to match candidates and potential NE over consonant signatures. The consonant signature of a token is obtained by first producing a "normalised" form and then by removing the vowels, as described in (Steinberger and Pouliquen, 2007). The normalised form is produced through the application of a small set of transformation rules based on empirically observed regularities between name variants (double to single consonant, *ck* to *k*, *ou* to *u* etc.). We compare the first candidate token to each sentence token and if there

[1] The database contains 134,046 en-fr NE translations, 157,442 en-es, 156,363 en-de, 2,807 en-cs and 65,916 en-ru.

[2] During the projection step we work on lower-case texts.

is an exact match between their consonant signatures, we continue the comparison with the next tokens until the end of the candidate unit. Considering again the *tarik ramadan* example and its consonant signature [trk - rmdn], this method allows to project its person tag onto the string *tarika ramadana* which, even if not present in the candidate list, has the same consonant signature.

Similarity Distance Finally, for cases where the consonant signature matching method fails, we attempt to project the NE by computing a similarity measure between the consonant groups. Reproducing the work done by (Pouliquen, 2008), we applied a cost-based Levenshtein edit distance, "*where the difference between two letters is not binary but depends on the distance between two letters*". This distance is learned from a set of existing named entity variants. By looking at several examples, we empirically determined the threshold of 0.7, above which the similarity shows good candidates for matching. With this third method, we succeed to project some more candidates, as illustrated by this example: the name *samantha geimer* can be projected from English to Czech, thanks to the calculation of the string similarity distance between the two groups [smnth - gmr] and [smnth - gmrv]:

English: the lawyer of samantha geimer, the victim...
Candidate list: samantha geimer, geimer samantha
Czech: právní zástupkyne obeti, samanthy geimerové

4 Results

4.1 Experimental settings

We ran several experiments according to various set-ups. First of all, we started from the baseline of simply searching for the English named entities in the foreign texts. Then, during the source NE annotation step, we noticed the presence of wrong English entities. We are not interested in evaluating the quality of the NER system that we used but we wondered how it can affect our projection performance. Therefore, we manually corrected the English entities of the *En-4* corpus; performance results are reported according to corrected and non-corrected source entities. Finally, we evaluate the performance of the projection combining different translation approaches. English entities are translated using: (1) external information: for each language pair, a list of English-Foreign entity associations is used as a look-up table (*DB* in Table 1), (2) machine translation system (*SMT*) and (3) external information and machine translation system together: a list of all possible translations is associated to each English entity[3](*ALL*). Moreover, with respect to the SMT approach (case 2), we consider two different SMT outputs: (2a) highest-ranking translation (*SMT-1*) and (2b) top 15 ranked translations (*SMT-15*). By considering the less probable translations up to 15, we expect to cover as much as possible morphological variations in inflected languages.

4.2 Results

As we do not have a reference corpus, we only compute projection Recall. In the future, we plan to manually annotate a part of the multilingual set to evaluate Precision.

Recall results are presented in Table 1. As said above, we combined several translation approaches and projection methods. First it should be noted that the baseline gives quite good results for target languages of the same alphabet (from 0.3 to 0.5 in the *En-4* corpus). Most of the successful English projections are for person names, but performance decreases with inflected languages. Adding external information (*DB*) brings some improvements but it all depends on the amount of translations available in the database, as shown by the difference in gains between French (+12 pts) and Czech (+5 pts). By taking into account the highest-ranking translation (SMT-1), recall improves quite significantly for each target language, although Czech and Russian show lower results. Merging of external and SMT-1 translations (*ALL with SMT-1*) produces small improvements.

Overall results improve even more considering more SMT translations (SMT-15) and varying projection methods. As evidenced by Figure 1, taking into account less probable translations emitted by the SMT system yields significant improvements, especially for inflected languages (+8pts for French, +24 for Czech and +37 for Russian). Then, applying different projection methods for the remaining non-projected entities increases again the results (+0.4 pts on average for all languages), consonant signature and similarity measure giving more or less the same contribution. Adding external information on top of this

[3] If more than one translation matches the target sentence, it is counted only one time.

Translation configurations	French	Spanish	German	Czech	Russian
Baseline	0.493	0.415	0.494	0.312	0.041
Baseline (corrNE)	0.508	0.431	0.516	0.323	0.041
DB	0.628	0.59	0.631	0.375	0.201
SMT-1	0.840	0.846	0.836	0.604	0.433
ALL (with SMT1)	0.869	0.852	0.857	0.594	-
SMT-15	0.929	0.917	0.921	0.837	0.803
SMT-15 + csnt	0.940	0.933	0.933	0.879	0.842
SMT-15 + cnst + sim	0.953	0.942	0.947	0.919	0.867
ALL (with SMT-15)	0.93	0.916	0.924	0.831	0.803
ALL (with SMT-15) + cnst + sim	0.954	0.943	0.95	0.918	0.867

Table 1: Projection Recall performance according to various translation configurations and projection methods. Recall is computed relative to the total number of English annotated entities in each corpus. *CorrNE* = corrected English Named Entities; *csnt* = consonant signature and *sim* = similarity measure. Apart from Baseline, all results are computed with corrected English entities.

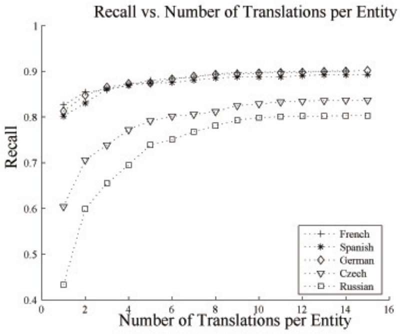

Figure 1: Recall projection performance according to the number of SMT translations (with string matching projection method).

configuration (SMT15 + csnt + sim + DB) brings only small improvements. At the end, best results range from 0.86 to 0.95, showing significant improvements over the baseline. Among the different approaches we tried out, the most beneficial ones are SMT-15 for the translation and the combination of three methods for the projection, particularly in the case of highly inflected languages.

4.3 Error Analysis

Non-projected entities are approximately the same across languages. We identified four main reasons of non-projection. First, as already pointed out, it happens that some English NEs are wrongly annotated, even when manually corrected. This can be illustrated with the following case: in the English entity *iraqi prime minister nouri al-maliki* only *prime minister nouri al-maliki* is annotated and, in consequence, it is not possible to project the Spanish translation *primer ministro nouri al-maliki* on the target *primer ministro iraquí nouri al-maliki*. Then, entity translations can be incorrect. We can report this example: the English entity *state secretary peter wichert* is wrongly translated by *secretario de estado peter wichert habría solicitado*, which make the projection impossible. Furthermore, human sentence translations across parallel texts are not always equivalent, which sometimes block the projection, even with correctly translated entities: *European Court of Justice* appears as *corte europeo* in the Spanish sentence, whereas we try to project *corte europea de justicia*. Finally, there are some hopeless cases combining all sorts of mistakes.

5 Conclusion and Future Work

This work showed how parallel corpora can support the automatic creation of multilingual NE annotated-corpora. By projecting NE annotations across aligned texts in different languages, we solved the evaluation resource bottleneck problem, saving annotation time and providing comparable annotated data. The resource will be made available http://langtech.jrc.it/. Our approach can be improved in several ways. In order to make the source language annotation step more "objective" and reliable, we intend to combine different NE recognition systems through a voting system. Then, we plan to evaluate the precision of the projection. In addition, it could be interesting to project more fine-grained information, considering NE sub-parts like functions, titles, etc. At last, we are currently working on Italian and Hungarian and we intend to reproduce this work on other parallel corpora, including for resource-poor languages.

References

An, J., Lee, S. and Lee, G. (2003) Automatic acquisition of named entity tagged corpus from world wide web. In *Proceedings of ACL (ACL'03)*, Sapporo.

Bentivogli, L. and Pianta, E. (2005) Exploiting parallel texts in the creation of multilingual semantically annotated resources: the MultiSemCor Corpus. In *Natural Language Engineering* pp. 247–261, Cambridge University Press.

Bering, C., Drozdzynski, W., Erbach, G., Guasch, C., Homola and others. (2003) Corpora and evaluation tools for multilingual NE grammar development. In *Proceedings of Multilingual Corpora - Linguistic Requirements and Technical Perspectives*, Lancaster.

Brown, P.F., Della Pietra, S., Della Pietra, V.J. and Mercer R.L.(1994). The Mathematic of Statistical Machine Translation: Parameter Estimation. In *Computational Linguistics*, 19(2):263–311.

Callison-Burch, C., Koehn, P., Monz, C. and Schroeder, J. (2009) Findings of the 2009 Workshop on Statistical Machine Translation. In *Proceedings of the Fourth WMT'09*, Athens.

Crawley, J. B. and Wagner, G. (2010). Desktop text mining for law enforcement. In *Proceedings of ISI'10*, Vancouver.

Ehrmann, M. and Turchi, M. (2010). Building Multilingual Named Entity Annotated Corpora Exploiting Parallel Corpora. In *Proceedings of AEPC*, Tartu, Estonia.

Fort, K., Ehrmann M. and Nazarenko, A. (2009) Towards a Methodology for Named Entities Annotation. In *Proceedings of LAWIII*, Singapore.

Hwa, R., Resnik, P., Weinberg, A., Cabezas, C. and Kolak, O. (2005). Bootstrapping parsers via syntactic projection across parallel texts. In *Natural Language Engineering*, 11(3).

Klementiev, A. and Roth, D. NE Transliteration and Discovery from Multilingual Corpora. (2008) In *Learning Machine Translation.* MIT Press.

Klyueva N. and Bojar O. UMC 0.1: Czech-Russian-English Multilingual Corpus. (2008) In *Proceedings of International Conference Corpus Linguistics*.

Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N. and others (2007). Moses: Open source toolkit for statistical machine translation. ACL, 45(2), Columbus, Oh, USA.

Koehn, P. (2010). Statistical Machine Translation. Cambridge Univ. Press.

Lavie, A., Parlikar, A. and Ambati, V. (2008). Syntax-driven learning of sub-sentential translation equivalents and translation rules from parsed parallel corpora. In *Proceedings of theHLT-SSST-2 workshop*, Columbus, Ohio.

Paroubek, P., Chaudiron, S. and Hirschman, V. (2007). Principles of Evaluation in Natural Language Processing. In *TAL*, 48-1

Ma, X. (2010) Toward a Named Entity Aligned Bilingual Corpus. In *Proceedings of the Seventh LREC Conference*, Malta.

Nadeau, D., and Sekine, S. (2007) A survey of named entity recognition and classification. In *Linguisticae Investigaciones*, 30-1, pp. 3-26.

Nothman, J., Curran, J., and Murphy, T. (2008) Transforming Wikipedia into named entity training data. In *Proceedings of the ALTA Workshop*, Hobart.

Padó, S. and Lapata, M. (2009) Cross-linguistic projection of role-semantic information. In *Journal of Artificial Intelligence Research*, 36.

Pouliquen, B. (2008) Similarity of names across scripts: Edit distance using learned costs of n-grams In *Advances in Natural Language Processing*, Sringer.

Rafalovitch, A. and Dale, R.(2009) United Nations General Assembly Resolutions: A Six-Language Parallel Corpus. In *Proceedings of the MT Summit*

Samy, D., Moreno-Sandoval, A. and Guirao, J.M. (2005). A Proposal for an Arabic Named Entity Tagger Leveraging a Parallel Corpus (Spanish-Arabic). In *Proceedings of RANLP Conference*, Borovets, Bulgaria.

Shah R., Lin B., Gershman A. and Frederking R. (2010). SYNERGY: A Named Entity Recognition System for Resource-scarce Languages such as Swahili using Online Machine Translation. In *Proceedings of (AfLaT), LREC*, Valleta, Malta.

Steinberger, R. and Pouliquen B. (2007). Cross-lingual Named Entity Recognition. In *Named Entities - Recognition, Classification and Use*, Benjamins Current Topics, Vol. 19, pp. 137-164.

Turchi, M., DeBie, T. and Cristianini N. (2008). Learning Performance of a Machine Translation System: a Statistical and Computational Analysis. In *Proceedings of the Third WMT'08*, Columbus, Oh, USA.

Volk, M., Goehring, A. and Marek, T. (2010) Combining Parallel Treebanks and Geo-Tagging. In *Proceedings of The Fourth LAW Workshop*, Uppsala.

Yarowsky, D., Ngai, G. and Wicentowski, R. (2001) Inducing Multilingual Text Analysis Tools via Robust Projection across Aligned Corpora. In *Proceedings of HLT'01*, San Diego.