Examining the Validity of Cross-Lingual Word Sense Disambiguation

Els Lefever and Veronique Hoste

Abstract—This paper describes a set of experiments in which the viability of a classification-based Word Sense Disambiguation system that uses evidence from multiple languages is investigated. Instead of using a predefined monolingual sense-inventory such as WordNet, we use a language-independent framework and start from a manually constructed gold standard in which the word senses are made up by the translations that result from word alignments on a parallel corpus. To train and test the classifier, we used English as an input language and we incorporated the translations of our target words in five languages (viz. Spanish, Italian, French, Dutch and German) as features in the feature vectors. Our results show that the multilingual approach outperforms the classification experiments where no additional evidence from other languages is used. These results confirm our initial hypothesis that each language adds evidence to further refine the senses of a given word. This allows us to develop a proof of concept for a multilingual approach to Word Sense Disambiguation.

Index Terms—Word Sense Disambiguation, multilingual, cross-lingual.

I. INTRODUCTION

WORD Sense Disambiguation (WSD) is the NLP task that consists in selecting the correct sense of a polysemous word in a given context. For a detailed overview of the main WSD approaches we refer to Agirre and Edmonds [1] and Navigli [2]. State-of-the-art WSD systems are mainly supervised systems, trained on large sense-tagged corpora, where human annotators have labeled each instance of the target word with a label from a predefined sense inventory such as WordNet [3]. Two important problems arise with this approach. Firstly, large sense-tagged corpora and sense inventories are very time-consuming and expensive to build. As a result they are extremely scarce for languages other than English. In addition, there is a growing conviction within the WSD community that WSD should not be tested as a stand-alone NLP task, but should be integrated in real applications such as Machine Translation and cross-lingual information retrieval [4].

In this paper, we describe the construction of a multilingual WSD system that takes an English ambiguous word and its context as input, and outputs correct translations for this ambiguous word in a given focus language. For our experiments we trained a classifier for five focus languages (viz. Italian, German, Dutch, Spanish and French). In addition to a set of local context features, we included the translations in the four other languages (depending on the focus language of the classifier) in the feature vector. All translations are retrieved from the parallel corpus Europarl [5].

Using a parallel corpus, such as for example Europarl, instead of human defined sense-labels offers some advantages: (1) for most languages we do not have large sense-annotated corpora or sense inventories, (2) using corpus translations should make it easier to integrate the WSD module into real multilingual applications and (3) this approach implicitly deals with the granularity problem, as fine sense distinctions (that are often listed in electronic sense inventories) are only relevant in case they get lexicalized in the target translations. The idea to use translations from parallel corpora to distinguish between word senses is based on the hypothesis that different meanings of a polysemous word are lexicalized across languages [6], [7]. Many WSD studies have already shown the validity of this cross-lingual evidence idea. Most of these studies have focused on bilingual WSD (E.g.[8], [9], [10]) or on the combination of existing WordNets with multilingual evidence (E.g. [11]).

In order to use the parallel texts to train a WSD classifier, most systems lump different senses of the ambiguous target word together if they are translated in the same way (E.g. Chan and Ng [12]), which reflects the problem of assigning unique translations to each sense of a noun. If we take for instance the English word *mouse*, this is translated in French as *souris*, both for the animal and the computer sense of the word. In order to construct and refine a multilingual sense inventory reflecting the different senses of a given word, more translations are required to increase the chance that the different word senses are lexicalized differently across the different languages. To our knowledge, however, it has not been shown experimentally if and how much multilingual evidence from a parallel corpus indeed helps to perform classification-based WSD for a given target language. In the experiments reported in this paper, we included evidence from up to 4 languages into the feature vectors of a multilingual lexical sample WSD classifier.

The remainder of this paper is organized as follows. Section II describes the data set we used for the experiments. Section III presents the construction of the feature vectors, and gives more insights in the classifier that was built. Section IV gives an overview of the experiments and we finally draw conclusions and present some future research in Section V.

The authors are with the LT3, University College Ghent, Groot-Brittanniëlaan 45, Ghent, Belgium and Dpt. of Applied Mathematics and Computer Science, Ghent University, Krijgslaan 281(S9), Ghent, Belgium (e-mail: {Els.Lefever,Veronique.Hoste}@hogent.be).

II. Data

In order to construct our sense inventory, we extracted the translations of our ambiguous target words from the parallel corpus Europarl [5]. We selected 6 languages from the 11 European languages represented in the corpus, viz. English (our target language), Dutch, French, German, Italian and Spanish. As our approach is both language- and corpus-independent, and all steps can be run in an automatic way, we can easily add other languages and extend or replace the corpus that was used.

All Europarl data were already sentence-aligned using a tool based on the Gale and Church algorithm [13], which was part of the corpus. We only considered the intersected 1-1 sentence alignments between English and the five other languages (see also [11] for a similar strategy). The experiments were performed on a lexical sample of five ambiguous words, being *bank*, *plant*, *movement*, *occupation* and *passage*, which were collected in the framework of the SemEval-2 Cross-Lingual Word Sense Disambiguation task. The six-language sentence aligned corpus, as well as the test set and corresponding gold standard, can be downloaded from the task website[1].

After the selection of all English sentences containing these target nouns and the aligned sentences in the five target languages, we used GIZA++ [14] word alignment on the selected sentences to retrieve the set of possible translations for our ambiguous target words. All alignments were manually checked afterwards. In cases where one single target word (E.g. **occupation**) led to a multiword translation (e.g *actividad profesional* in Spanish) or to a compound (e.g *beroepsbezigheden* in Dutch and *Berufstätigkeit* in German), we kept the multi-part translation as a valid translation suggestion.

All sentences containing the target words were preprocessed by means of a memory-based shallow parser (MBSP) [15], that performs tokenization, Part-of-Speech tagging and text chunking. On the basis of these preprocessed data, we built a feature vector which contains information related to the target word itself as well as local patterns around the target word. Table I shows the size of the instance base for each of the ambiguous words, whereas Figure 1 lists the number of classes per ambiguous target word in the five focus languages.

TABLE I
Size of the instance base per ambiguous target word

	Number of instances
bank	4029
movement	4222
occupation	634
passage	238
plant	1631

Figure 1 also suggests that due to the high number of unique translations in Dutch and German, mainly due to

their compounding strategies, the classification task will be especially challenging for these two languages.

As Figure 1 shows, the polysemy of the target words is considerably high in all five target languages. Even for the romance languages, where the number of compound translations is rather low, the classifier has to choose from a substantial number of possible classes. Example 1 illustrates this by listing the French translations that were retrieved for the English word *plant* (*NULL* refers to a null link from the word alignment):

(1) centrale, installation, plante, usine, végétal, NULL, phyto-sanitaire, entreprise, incinérateur, station, pesticide, site, flore, unité, atelier, plant, phytopharmaceutique, établissement, culture, réacteur, protéagineux, centre, implantation, oléoprotéagineux, équipement, horticulture, phytogénétique, exploitation, végétation, outil, plantation, sucrerie, société, fabrique, four, immobilisation, céréale, espèce, séchoir, production, claque, arsenal, ceps, poêle, récolte, plateforme, artémisinine, fabrication, phytogénéticien, oléagineux, glacière, espèce végétale, chou, tranche, Plante, installation incinérateur.

III. Experimental Set-up

We consider the WSD task as a classification task: given a feature vector containing the ambiguous word and the context as features, a classifier predicts the correct sense (or translation in our case) for this specific instance.

A. Feature Vectors

For our initial feature set we started off with the traditional features that have shown to be useful for WSD [1]:

– features related to the **target word itself** being the word form of the target word, the lemma, Part-of-Speech and chunk information
– **local context features** related to a window of three words preceding and following the target word containing for each of these words their full form, lemma, Part-of-Speech and syntactic dependencies.

In addition to these well known WSD features, we integrated the translations of the target word in the other languages (Spanish, German, Italian, Dutch and French depending on the desired classification output) as separate features into the feature vector. Example 2 lists the feature vector for one of the instances in the training base of the Dutch classifier. The first features contain the word form, PoS-tag and chunk information for the three words preceding the target word, the target word itself and for the three words following the target word. In addition we added the aligned translations for the target word in the four additional languages (being German, Spanish, Italian and French for the Dutch classifier). The last field contains the classification label, which is the aligned Dutch translation in this case.
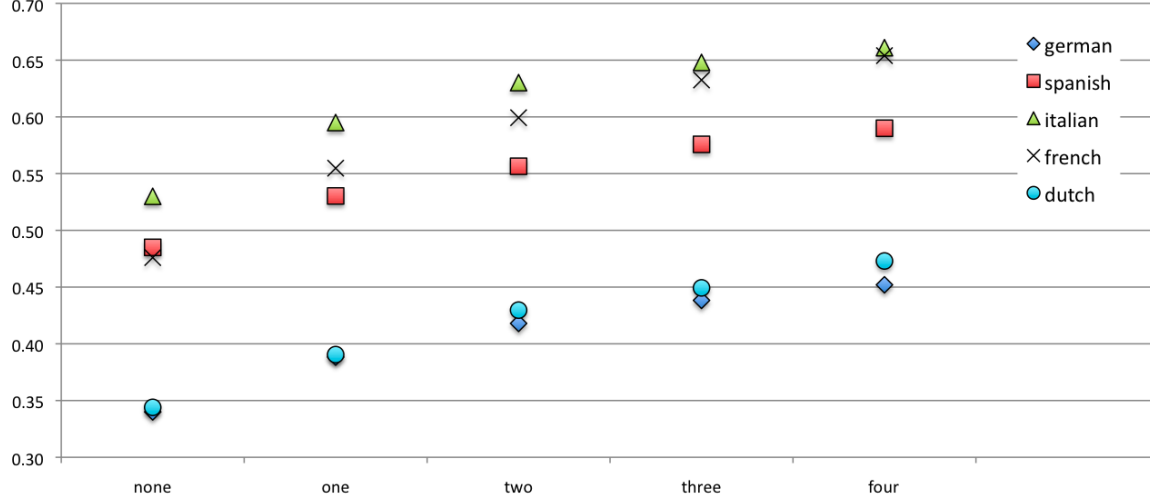
Fig. 1. Number of unique translations per language and per ambiguous word.

(2) English input sentence for the word **bank**:

This is why the Commission resolved on raising a complaint against these two banks at its last meeting, and I hope that Parliament approves this step.

Feature vector:

against these two against these two IN DT CD I-PP I-NP I-NP banks bank NNS I-NP at its last at its last IN PRP JJ I-PP I-NP I-NP Bank banco banca banque bank

Incorporating the translations in our feature vector allows us to develop a proof of concept for a multilingual approach to Word Sense Disambiguation. This multilingual approach will consist of two steps: (1) we first examine whether evidence from different languages can lead to better sense discrimination (which is the scope of this paper) and (2) in a following step we will then introduce additional cross-lingual evidence (bag-of-words features containing all content words from the aligned translations) in the feature vectors for our WSD classifier. An automatic sense discrimination step can then be applied on the training feature base.

Unsupervised approaches to sense discrimination know a long research history. The idea to use distributional methods to cluster words that appear in similar contexts corpora has been succesfully applied on monolingual corpora (E.g. [16], [17]), as well as on parallel corpora. Previous research on parallel corpora [18], [7] confirmed the use of cross-lingual lexicalization as a criterion for performing sense discrimination. Whereas in previous research on cross-lingual WSD the evidence from the aligned sentences was mainly used to enrich WordNet information, our approach does not require any external resources. With our experiments we want to examine to which extent evidence from other languages, without additional information from external lexical resources, helps to detect correct sense distinctions that result in a better WSD classification output (or translation in our case).

B. Classification

To train our WSD classifier, we used the memory-based learning (MBL) algorithms implemented in TIMBL [19], which have been shown to perform well on WSD [20]. We performed heuristic experiments to define the parameter settings for the classifier, leading to the selection of the Jeffrey Divergence distance metric, Gain Ratio [21] feature weighting and $k = 7$ as number of nearest distances. In future work, we plan to use a genetic algorithm to perform joint feature selection and parameter optimization per ambiguous word [22].

IV. EVALUATION

For the evaluation, we performed 10-fold cross-validation on the instance bases. As a baseline, we selected the most frequent translation that was given by the automatic word alignment. We added the translations in the other languages that resulted from the word alignment as features to our feature vector and built classifiers for each target word for all five supported languages. Since we aim to investigate the impact of cross-lingual evidence on WSD, we deliberately chose to use the manually verified gold standard word alignments. Our classification results can thus be considered as an upper bound for this task, as the automatic word alignments will presumably lead to lower performance figures.

An overview of the classification results for the romance languages (French, Italian, Spanish) can be found in Table II, whereas the classification results for Dutch and German are in Table III. Figure 2 illustrates the classification results per language for 2 ambiguous words, viz "bank" and "plant" when averaging over the translations in the feature vector.

The results show that even the simple classifier which does not incorporate translation features, beats the most frequent translation baseline for all languages (except for *occupation* in Spanish and Italian), although we can improve a lot on the

TABLE II
FRENCH (TOP LEFT), ITALIAN (TOP RIGHT) AND SPANISH (BOTTOM LEFT) RESULTS FOR A VARYING NUMBER OF TRANSLATION FEATURES INCLUDING THE OTHER FOUR LANGUAGES VIZ. ITALIAN (I), SPANISH (E), GERMAN (D), DUTCH (N) AND FRENCH (F)

French	bank	move-ment	occu-pation	passage	plant
Baseline	55.8	44.7	75.5	50.0	20.7
all four translation features					
IEDN	**84.9**	**71.7**	**82.8**	**60.3**	**65.4**
Three translation features					
I,E,D	84.5	70.9	80.8	59.5	63.7
E,D,N	84.0	70.7	81.6	59.1	63.7
I,D,N	83.9	70.7	82.0	59.1	61.3
I,E,N	84.6	71.3	81.2	57.4	64.3
Two translation features					
E, D	83.2	69.2	80.0	59.9	60.8
I, D	83.1	69.8	80.1	58.7	58.8
D, N	82.8	69.1	80.9	57.4	58.6
I, E	84.3	69.8	80.0	57.8	61.0
E, N	83.2	69.8	80.5	57.4	61.0
I, N	83.2	70.1	81.1	57.8	59.4
One translation feature					
D	81.4	67.5	78.9	58.7	54.0
E	83.0	67.7	79.2	56.5	56.4
I	82.4	68.4	79.5	57.4	56.1
N	82.0	68.0	80.5	57.4	55.4
No translation features					
none	83.5	65.6	76.5	55.3	47.6
Only translation features					
only	85.8	73.3	82.8	62.9	69.0

Italian	bank	move-ment	occu-pation	passage	plant
Baseline	54.6	51.9	78.7	37.1	32.8
all four translation features					
EFDN	**83.1**	**80.2**	**81.1**	**40.1**	**66.1**
Three translation features					
E,F,D	82.7	79.6	81.1	40.1	65.1
F,D,N	82.8	79.7	79.2	40.9	64.2
E,D,N	82.6	79.2	81.0	40.5	64.6
E,F,N	82.8	80.0	81.0	40.5	65.3
Two translation features					
F, D	82.0	78.6	79.3	40.5	63.4
E, D	81.8	78.5	80.9	40.5	62.1
D, N	81.4	77.8	78.5	40.9	62.4
E, F	82.3	79.5	80.9	40.1	64.3
F, N	82.4	79.0	79.2	41.4	63.2
E, N	82.1	78.7	80.1	40.1	62.7
One translation feature					
D	80.0	76.8	77.9	40.5	59.4
F	81.4	78.0	79.2	40.9	61.1
E	81.4	77.5	80.6	38.4	58.1
N	80.9	77.2	78.1	39.7	59.4
No translation features					
none	79.5	75.2	78.1	38.0	53.0
Only translation features					
only	83.9	81.4	81.6	42.6	67.3

Spanish	bank	move-ment	occu-pation	passage	plant
Baseline	58.8	51.0	81.6	24.1	30.1
all four translation features					
IFDN	**90.0**	**80.8**	**83.0**	**38.0**	**59.0**
Three translation features					
I,F,D	89.6	80.6	82.8	35.9	58.6
F,D,N	89.1	79.6	82.7	37.6	57.1
I,D,N	89.4	79.4	82.4	37.6	55.9
I,F,N	89.8	80.3	82.7	35.4	58.7
Two translation features					
F, D	88.9	79.1	82.7	35.9	55.9
I, D	88.7	79.0	82.4	36.3	54.3
D, N	88.0	78.0	82.0	38.0	53.7
I, F	89.4	79.9	82.5	34.2	57.8
F, N	89.0	79.2	82.2	35.4	57.3
I, N	89.3	78.6	82.4	34.2	54.9
One translation feature					
D	87.2	77.3	82.2	37.1	50.8
F	88.7	78.3	82.7	34.2	55.1
I	88.7	78.3	81.6	32.5	53.6
N	87.7	77.1	81.9	34.6	52.6
No translation features					
none	86.5	75.8	80.6	32.9	48.5
Only translation features					
only	89.9	82.0	83.0	40.9	63.4

TABLE III
DUTCH (LEFT) AND GERMAN (RIGHT) RESULTS FOR A VARYING NUMBER OF TRANSLATION FEATURES INCLUDING THE OTHER FOUR LANGUAGES VIZ. ITALIAN (I), SPANISH (E), GERMAN (D), DUTCH (N) AND FRENCH (F)

Dutch	bank	move-ment	occu-pation	passage	plant
Baseline	33.4	46.7	60.6	26.7	12.0
all four translation features					
IEDF	**80.3**	**65.8**	**69.3**	**36.3**	**47.3**
Three translation features					
I,E,D	80.0	65.1	68.9	35.0	44.2
E,D,F	79.4	65.2	69.0	34.6	45.8
I,D,F	79.4	65.5	69.2	36.3	45.2
I,E,F	79.1	63.7	68.2	35.4	44.5
Two translation features					
E, D	79.2	64.4	67.6	35.0	45.2
I, D	79.0	64.3	68.5	34.6	42.7
D, F	78.8	64.9	68.8	35.0	43.8
I, E	79.0	62.9	66.3	34.6	41.2
E, F	78.4	63.3	67.7	34.6	42.7
I, F	78.0	63.1	68.2	35.0	42.2
One translation feature					
D	77.8	63.5	67.6	35.0	40.4
E	78.1	62.1	65.3	33.3	37.1
I	77.7	62.1	66.3	33.8	38.9
F	77.3	62.1	67.6	33.8	39.8
No translation features					
none	76.6	60.8	65.2	31.7	34.4
Only translation features					
only	80.0	64.1	69.6	34.6	47.3

| German	bank	move-ment	occu-pation	passage	plant
Baseline	36.7	32.3	39.0	20.3	14.0
all four translation features					
IEFN	**82.8**	**57.1**	**48.3**	**32.9**	**45.2**
Three translation features					
I,E,N	82.5	57.0	47.9	31.2	44.0
E,F,N	82.5	57.2	47.7	32.1	43.9
I,E,F	81.7	55.8	47.5	31.6	42.9
F,I,N	82.6	57.2	48.3	31.6	44.5
Two translation features					
E, F	81.6	55.6	45.5	31.2	41.1
I, F	81.6	55.5	46.9	31.2	41.6
F, N	82.3	56.9	47.2	30.4	42.9
I, E	81.6	55.3	46.4	29.5	41.1
E, N	82.2	56.6	46.7	30.0	41.6
I, N	82.2	57.1	48.0	30.0	42.5
One translation feature					
F	81.1	54.8	45.5	30.0	39.2
E	81.1	54.7	43.6	28.7	36.6
I	81.3	55.1	45.0	29.5	39.1
N	81.9	56.1	46.7	28.3	40.4
No translation features					
none	80.5	53.5	42.1	27.8	34.0
Only translation features					
only	73.1	51.1	50.4	32.5	43.8
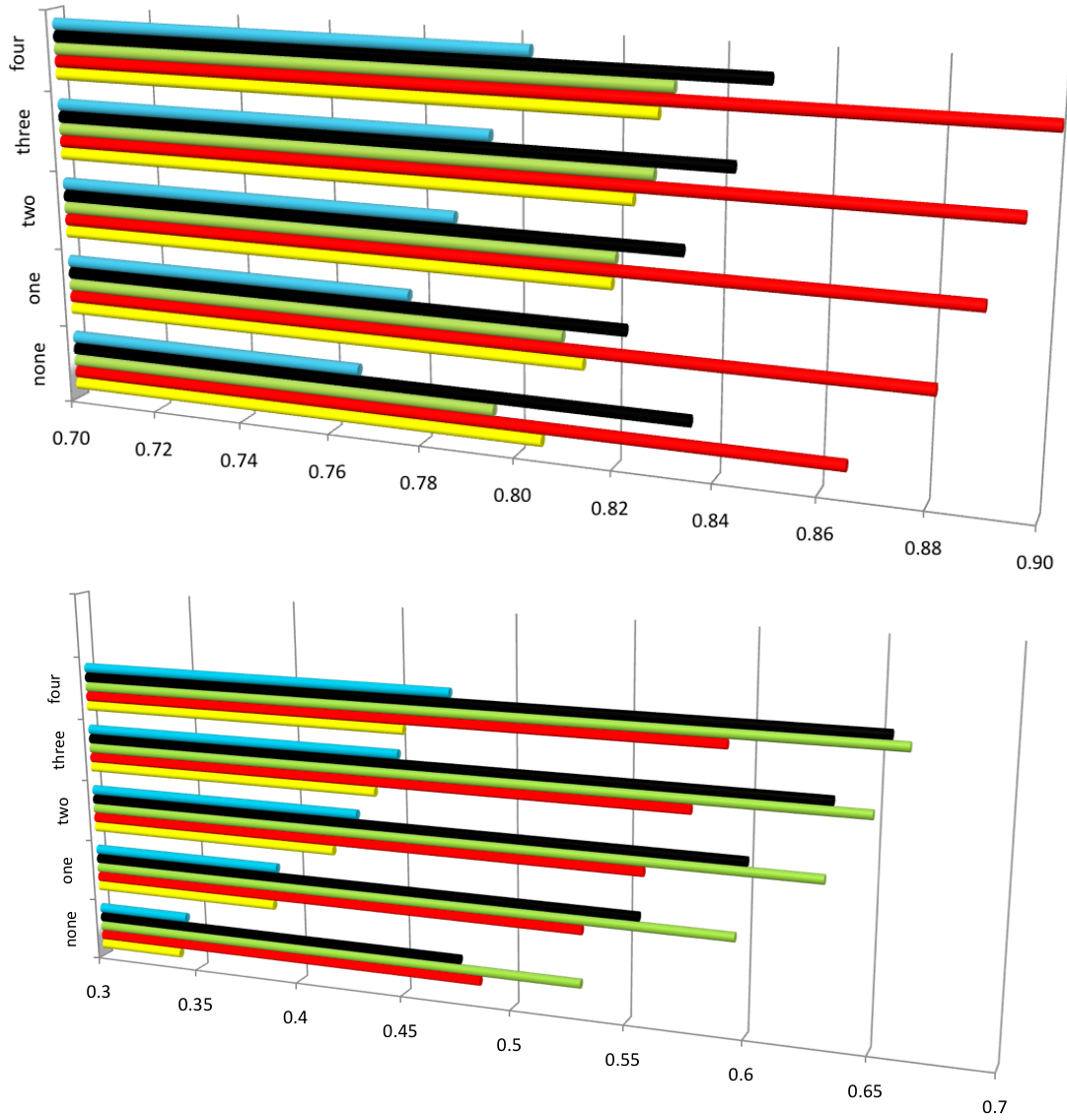
Fig. 2. Classification results for "bank" and "plant" for each of the target languages. The languages are resp. from top to bottom: Dutch, French, Italian, Spanish and German.

feature base level (e.g. by adding bag of word features for a broader context, etc.).

The scores clearly confirm the validity of our hypothesis: the experiments using all different translations as features are constantly better than the ones using less or no multilingual evidence. This conclusion holds for all five classification results. In addition, the scores also degrade relatively to the number of translation features that is used. This allows us to conclude that incorporating multilingual information in the feature vectors helps the classifier to choose more reliable and finer sense distinctions, which results in better translations in our case. Moreover, the more translations (in different languages) are incorporated in the feature vector, the better the classification results get. Another striking observation is that the classifier that solely relies on translation features

(*Only translation features*) often beats the classifier that incorporates all context and translation features. There are, however, two limitations to our experimental framework. We have not experimented with a higher number of languages, and as a consequence we can not estimate from which number of languages the performance would start to degrade. In addition, another interesting line of research would be to include languages belonging to more distant language families.

The experimental results also reveal remarkable differences between the different languages. This can probably be explained by the difference in morphological structure between the two language families. As Dutch and German tend to concatenate the parts of compounds in one orthographic unit, whereas the romance languages (French, Italian, Spanish) keep these parts separated by spaces, this often results in compound

translations in German and Dutch. As a result, the number of different classes this classifier has to choose from, is much larger (as already shown in Figure 1). This difference is also reflected in the baselines, where the French, Italian and Spanish baseline is clearly higher than the Dutch or German one for most words.

Another interesting observation to make is that languages from the same language branch seem to contribute more to a correct classification result. The results show for instance that for the Spanish classifier, the use of Italian and French translations in the feature vector results in better classification scores, whereas for German, the incorporation of the Dutch translations in the feature vector seems to contribute most for choosing a correct translation. More experiments with other words and languages will allow us to examine whether this trend can be confirmed. Previous research on this topic has ended in contradictory results: Ide [18] showed that there was no relationship between sense discrimination and language distance, whereas Resnik and Yarowsky [6] found that languages from other language families tend to lexicalize more sense distinctions.

Our results clearly show that adding more multilingual evidence to the feature vector helps the WSD classifier to predict more accurate translations. The logical next step is to integrate this multilingual information into a real WSD application. In order to do so we will use the multilingual evidence from the parallel corpus to enrich our training vectors. Instead of only incorporating the aligned translations from the other languages, we will add all content words from the aligned translations as bag-of-word features to the feature vector. We will also develop a strategy to generate the corresponding translation features for the test instances. Both the local context features of the English target word and the cross-lingual evidence will be taken into account for computing the similarity scores between the test input and the training instance base. The expected outcome, based on the results we showed in this paper, is that each language can contribute to make finer sense distinctions and thus to provide more contextually accurate translations for the ambiguous target words.

V. CONCLUSION AND FUTURE WORK

We presented preliminary results for a multilingual Word Sense Disambiguation system, which does not use labels from a predefined sense inventory, but translations that are retrieved by running word alignment on a parallel corpus. Although there is still a lot of room for improvement on the feature base, the scores of all five WSD systems constantly beat the most frequent translation baseline. The results allow us to develop a proof of concept that multilingual evidence in the feature vector, helps the classifier to make more reliable and finer sense distinctions, which result in better translations. We also observed that adding translations from the same language branch seems to help the classifier best to predict a correct translation in the focus language.

In future work, we want to run additional experiments with different classifiers on a larger sample of ambiguous words. We also wish to improve the classification results by performing joint feature selection and parameter optimization per ambiguous target word (E.g. by using a genetic algorithm approach). In addition, we also plan to include more multi-lingual evidence in a real WSD set-up. Therefore we will store the bag-of-words translation features resulting from the aligned translations in the training feature vectors, and add the automatically generated corresponding translation features for the test sentences to the test feature vectors.

REFERENCES

[1] E. Agirre and P. Edmonds, Eds., *Word Sense Disambiguation*, ser. Text, Speech and Language Technology. Dordrecht: Springer, 2006.

[2] R. Navigli, "Word sense disambiguation: a survey," in *ACM Computing Surveys*, 2009, vol. 41, no. 2, pp. 1–69.

[3] C. Fellbaum, *WordNet: An Electronic Lexical Database*. MIT Press, 1998.

[4] A. Otegi, E. Agirre, and G. Rigau, "Ixa at clef 2008 robust-wsd task: Using word sense disambiguation for (cross lingual) information retrieval," in *Evaluating Systems for Multilingual and Multimodal Information Access 9th Workshop of the Cross-Language Evaluation Forum, CLEF 2008, Aarhus, Denmark, September 17-19, 2008.*, 2009.

[5] P. Koehn, "Europarl: A parallel corpus for statistical machine translation," in *Proceedings of the MT Summit*, 2005.

[6] P. Resnik and D. Yarowsky, "Distinguishing systems and distinguishing senses: New evaluation methods for word sense disambiguation," *Natural Language Engineering*, vol. 5, no. 3, pp. 113–133, 2000.

[7] N. Ide, T. Erjavec, and D. Tufis, "Sense discrimination with parallel corpora," in *Proceedings of ACL Workshop on Word Sense Disambiguation: Recent Successes and Future Directions*, Philadelphia, PA, 2002, pp. 54–60.

[8] W. Gale, K. Church, and D. Yarowsky, "A method for disambiguating word senses in a large corpus," in *Computers and the Humanities*, 1993, vol. 26, pp. 415–439.

[9] H. Ng, B. Wang, and Y. Chan, "Exploiting parallel texts for word sense disambiguation: An empirical study," in *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, Santa Cruz, 2003, pp. 455–462.

[10] M. Diab and P. Resnik, "An unsupervised method for word sense tagging using parallel corpora," in *Proceedings of ACL*, 2002, pp. 255–262.

[11] D. Tufiş, R. Ion, and N. Ide, "Fine-Grained Word Sense Disambiguation Based on Parallel Corpora, Word Alignment, Word Clustering and Aligned Wordnets," in *Proceedings of the 20th International Conference on Computational Linguistics (COLING 2004)*. Geneva, Switzerland: Association for Computational Linguistics, Aug. 2004, pp. 1312–1318.

[12] Y. Chan and H. Ng, "Scaling up word sense disambiguation via parallel texts," in *AAAI'05: Proceedings of the 20th national conference on Artificial intelligence*. AAAI Press, 2005, pp. 1037–1042.

[13] W. Gale and K. Church, "A program for aligning sentences in bilingual corpora," in *Computational Linguistics*, 1991, pp. 177–184.

[14] F. Och and H. Ney, "A systematic comparison of various statistical alignment models," *Computational Linguistics*, vol. 29, no. 1, pp. 19–51, 2003.

[15] W. Daelemans and A. van den Bosch, *Memory-Based Language Processing*. Cambridge University Press, 2005.

[16] H. Schütze, "Automatic word sense discrimination," *Computational Linguistics*, vol. 24, no. 1, pp. 97–123, 1998.

[17] A. Purandare and T. Pedersen, "Word sense discrimination by clustering contexts in vector and similarity spaces," in *Proceedings of the Conference on Computational Natural Language Learning*, 2004, pp. 41–48.

[18] N. Ide, "Parallel translations as sense discriminators," in *SIGLEX Workshop On Standardizing Lexical Resources*, 1999.

[19] W. Daelemans, J. Zavrel, and K. v. d. B. van der Sloot, "Timbl: Tilburg memory-based learner, version 4.3, reference guide," Tilburg University, Tech. Rep. ILK Technical Report - ILK 02-10, 2002.

[20] V. Hoste, I. Hendrickx, W. Daelemans, and A. van den Bosch, "Parameter optimization for machine-learning of word sense disambiguation," *Natural Language Engineering, Special Issue on Word Sense Disambiguation Systems*, vol. 8, pp. 311–325, 2002.

[21] J. Quinlan, *C4.5: Programs for machine learning*. Morgan Kaufmann, San Mateo, CA, 1993.

[22] W. Daelemans, V. Hoste, F. De Meulder, and B. Naudts, "Combined optimization of feature selection and algorithm parameter interaction in machine learning of language," in *Proceedings of the 14th European Conference on Machine Learning (ECML-2003)*, 2003, pp. 84–95.