Translating Compounds by Learning Component Gloss Translation Models via Multiple Languages

Nikesh Garera and **David Yarowsky**
Department of Computer Science
Center for Language and Speech Processing
Johns Hopkins University
Baltimore, MD 21218, USA
{ngarera,yarowsky}@cs.jhu.edu

Abstract

This paper presents an approach to the translation of compound words without the need for bilingual training text, by modeling the mapping of literal component word glosses (e.g. "iron-path") into fluent English (e.g. "railway") across multiple languages. Performance is improved by adding component-sequence and learned-morphology models along with context similarity from monolingual text and optional combination with traditional bilingual-text-based translation discovery.

1 Introduction

Compound words such as *lighthouse* and *fireplace* are words that are composed of two or more component words and are often a challenge for machine translation due to their potentially complex compounding behavior and ambiguous interpretations (Rackow et al., 1992). For many languages, such words form a significant portion of the lexicon and the compounding process is further complicated by diverse morphological processes (Levi, 1978) and the properties of different compound sequences such as Noun-Noun, Adj-Adj, Adj-Noun, Verb-Verb, etc. Compounds also tend to have a high type frequency but a low token frequency which makes their translation difficult to learn using corpus-based algorithms (Tanaka and Baldwin, 2003). Furthermore, most of the literature on compound translation has been restricted to a few languages dealing with compounding phenomena specific to the language in question.

Compound	Splitting	English Gloss	Translation
Input: Distilled glosses from German-English dictionary			
Krankenhaus	Kranken-Haus	sick-house	hospital
Regenschirm	Regen-Schirm	rain-guard	umbrella
WörterBuch	Wörter-Buch	words-book	dictionary
Eisenbahn	Eisen-Bahn	**iron-path**	railroad
Input: Distilled glosses from Swedish-English dictionary			
Sjukhus	Sjhu-Khus	sick-house	hospital
Järnväg	Järn-väg	**iron-path**	railway
Ordbok	Ord-Bok	words-book	dictionary
Goal: To translate new Albanian compounds			
Hekurudhë	Hekur-Udhë	**iron-path**	**???**

Table 1: Example lexical resources used in this task and their application to translating compound words in new languages.

With these challenges in mind, the primary goal of this work is to improve the coverage of translation lexicons for compounds, as illustrated in Table 1 and Figure 1, in multiple new languages. We show how using cross-language compound evidence obtained from bilingual dictionaries can aid in compound translation. A primary motivating idea for this work is that the literal component glosses for compound words (such as "iron path" for *railway*) is often replicated in multiple languages, providing insight into the fluent translation of a similar literal gloss in a new (often resource-poor) language.

2 Resources Utilized

The only resource utilized for our compound translation lexicon algorithm is a collection of bilingual dictionaries. We used bilingual dictionary collections for 50 languages that were acquired in electronic form over the Internet or via optical character recognition (OCR) on paper dictionaries. Note that *no parallel or even monolingual corpora is required*, their use described later in the paper is optional.

3 Related Work

The compound-translation literature typically deals with these steps: 1) Compound splitting, 2) translation candidate generation and 3) translation candidate scoring. Compound splitting is generally done using translation lexicon lookup and allowing for different splitting options based on corpus frequency (Zhang et al., 2000; Koehn and Knight, 2003).

Translation candidate generation is an important phase and this is where our work differs significantly from the previous literature. Most of the previous work has been focused on generating *compositional* translation candidates, that is, the translation candidates of the compound words are lexically composed of the component word translations. This has been done by either just concatenating the translations of component words to form a candidate (Grefenstette, 1999; Cao and Li, 2002), or using syntactic templates such as "E_2 in E_1", "E_1 of E_2" to form translation candidates from the translation of the component words E_2 and E_1 (Baldwin and Tanaka, 2004), or using synsets of the component word translations to include synonyms in the compositional candidates (Navigli et al., 2003).

The above class of work in compositional-candidate generation fails to translate compounds such as *Krankenhaus (hospital)* whose component word translations are *Kranken (sick) and Haus (hospital)*, and composing *sick* and *house* in any order will not result in the correct translation (*hospital*). Another problem with using fixed syntactic templates is that they are restricted to the specific patterns occurring in the target language. We show how one can use the gloss patterns of compounds in multiple other languages to hypothesize translation candidates that are not lexically compositional.

4 Approach

Our approach to compound word translation is illustrated in Figure 1.

4.1 Splitting compound words and gloss generation with translation lexicon lookup

We first split a given source word, such as the Albanian compound *hekurudhë*, into a set of component word partitions, such as *hekur (iron)* and *udhë (path)*. Our initial approach is to consider all possible partitions based on contiguous component words found in a small dictionary for the language, as in

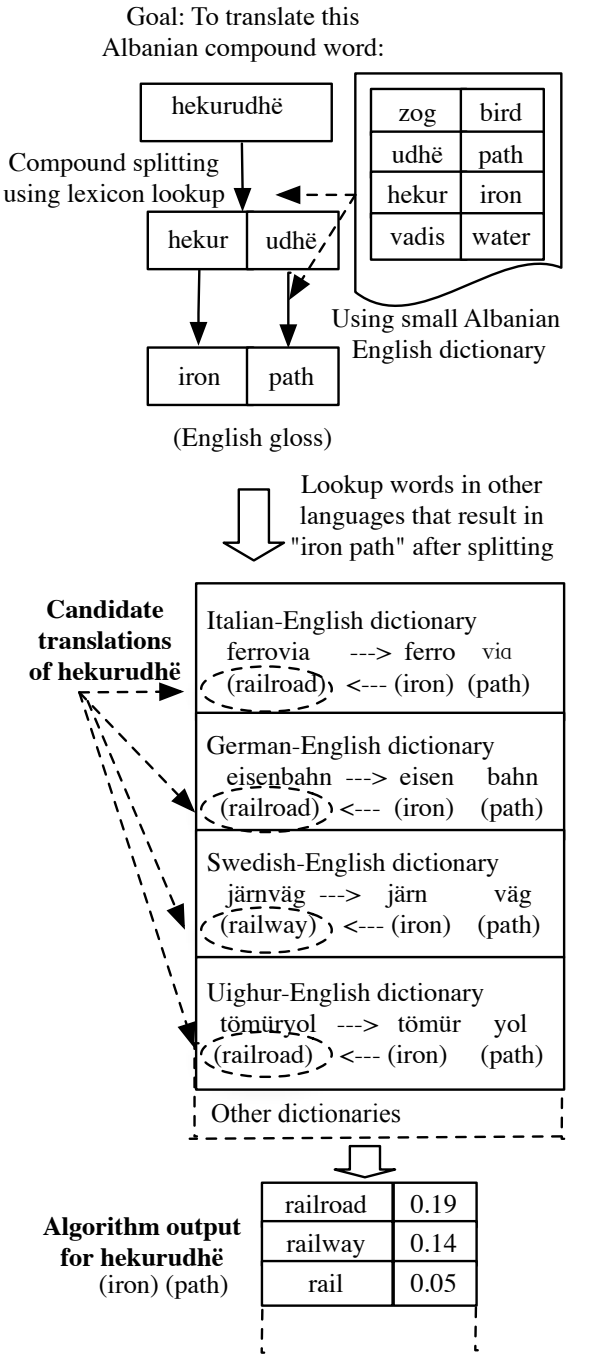

Figure 1: Illustration of using cross-language evidence using bilingual dictionaries of different languages for compound translation

Brown (2002) and Koehn and Knight (2003)[1]. For a given split, we generate its English glosses by using all possible English translations of the component words given in the dictionary of that language[2].

4.2 Using cross-language evidence from different bilingual dictionaries

For many compound words (especially for borrowings), the compounding process is identical across several languages and the literal English gloss remains the same across these languages. For example, the English word *railway* is translated as a compound word in many languages, and the English gloss of those compounds is often *"iron path"* or a similar literal meaning[3]. Thus knowing the fluent English translation of the literal gloss *"iron path"* in some relatively resource-rich language provides a vehicle for the translation from all other languages sharing that literal gloss[4]

4.3 Ranking translation candidates

The confidence in the correctness of a mapping between a literal gloss (e.g. "iron path") and fluent translation (e.g. "railroad") can be based on the number of distinct languages exhibiting this association. Thus we rank the candidate translations generated via different languages as in Figure 1 as follows: For a given target compound word, say f_c with a set of English glosses G obtained via multiple splitting options or multiple component word translations, the translation probability for a candidate translation can be computed as:

$$\begin{aligned} p(e_c|f_c) &= \sum_{g \in G} p(e_c, g|f_c) \\ &= \sum_{g \in G} p(g|f_c) \cdot p(e_c|g, f_c) \\ &= \sum_{g \in G} p(g|f_c) \cdot p(e_c|g) \end{aligned}$$

where, $p(g|f_c) = p(g_1|f_1) \cdot p(g_2|f_2)$. f_1, f_2 are the individual component-words of compound and g_1, g_2 are their translations from the existing dictionary. For human dictionaries, $p(g|f_c)$ is uniform for all $g \in G$, while variable probabilities can also be acquired from bitext or other translation discovery approaches. Also, $p(e_c|g) = \frac{freq(g,e_c)}{freq(g)}$, where $freq(g, e_c)$ is the number of times the compound word with English gloss g is translated as e_c in the bilingual dictionaries of *other* languages and $freq(g)$ is the total number of times the English gloss appears in these dictionaries.

5 Evaluation using Exact-match Translation Accuracy

For evaluation, we assess the performance of the algorithm on the following 10 languages: Albanian, Arabic, Bulgarian, Czech, Farsi, German, Hungarian, Russian, Slovak and Swedish. We detail both the average performance for these 10 languages (Avg_{10}), as well as provide individual performance details on Albanian, Bulgarian, German and Swedish. For each of the compound translation models, we report coverage (the # of compound words for which a hypothesis was generated by the algorithm) and Top1/Top10 accuracy. Top1 and Top 10 accuracy are the fraction of words for which a correct translation (listed in the evaluation dictionary) appears in the Top 1 and Top 10 translation candidates respectively, as ranked by the algorithm. Because evaluation dictionaries are often missing acceptable translations (e.g. *railroad* rather than *railway*), and any deviation from exact-match is scored as incorrect, these measures will be a lower bound on acceptable translation accuracy. Also, target language models can often select effectively among such hypothesis lists in context.

6 Comparison of different compound translation models

6.1 A simple model using literal English gloss concatenation as the translation

Our baseline model is a simple gloss concatenation model for generating compositional translation candidates on the lines of Grefenstette (1999) and Cao and Li (2002). We take the translations of the individual component-words (e.g. for the compound word *hekurudhë*, they would be *hekur (iron)* and

[1] In order to avoid inflections as component-words we limit the component-word length to at least three characters.

[2] The algorithm is allowed to generate multiple glosses "iron way," "iron road," etc. based on multiple translations of the component words. Multiple glosses only add to the number of translation candidates generated.

[3] For the gloss, "iron path", we found 10 other languages in which some compound word has the English gloss after splitting and component-word translation

[4] We do assume an existing small translation lexicon in the target language for the individual component-words, but these are often higher frequency words and present either in a basic dictionary or discoverable through corpus-based techniques.

udhë (path)) and hypothesizes three translation candidate variants: "iron path", "iron-path" and "iron-path". A test instance is scored as correct if any of these translation candidates occur in the translations of *hekurudhë* in the bilingual dictionary. This baseline performance measures how well simple literal glosses serve as translation candidates. In cases such as the German compound *Nußschale* (*nutshell*), which is a simple concatenation of the individual components *Nuß(nut)* and *Schale (shell)*, the literal gloss is correct. For this baseline, if the component-words have multiple translations, then each of the possible English gloss is ranked randomly. While Grefenstette (1999) and Cao and Li (2002) proposed re-ranking these candidates using web-data, the potential gains of this ranking are limited, as we see in Table 2 that even the *Found Acc.* is very low[5], that is for most of the cases the correct translation does not appear anywhere in the set of English glosses[6]

Language	Cmpnd wrds translated	Top1 Acc.	Top10 Acc.	Found Acc.
Albanian	4472 (10.11%)	0.001	0.010	0.020
Bulgarian	9093 (12.50%)	0.001	0.015	0.031
German	15731 (29.11%)	0.004	0.079	0.134
Swedish	18316 (31.57%)	0.005	0.068	0.111
Avg_{10}	14228 (17.84%)	0.002	0.030	0.055

Table 2: Baseline performance using unreordered literal English glosses as translations. The percentages in parentheses indicate what fraction of all the words in the test (entire) vocabulary were detected and translated as compounds.

6.2 Using bilingual dictionaries

This section describes the results from the model explained in Section 4. To recap, this model attempts to translate every test word such that there is at least one additional language whose bilingual dictionary supports an equivalent split and literal English gloss, and bases its translation hypotheses on the consensus fluent translation(s) corresponding to the literal glosses in these other languages. The performance is shown in Table 3. The substantial increase in accuracy over the baseline indicates the usefulness of such gloss-to-translation guidance from other languages. The rest of the sections detail our investigation of improvements to this model.

Language	Compound words translated	Top1 Acc.	Top10 Acc.
Albanian	3085 (6.97%)	0.185	0.332
Bulgarian	6719 (9.24%)	0.247	0.416
German	11103 (20.55%)	0.195	0.362
Swedish	12681 (21.86%)	0.188	0.346
Avg_{10}	9320.9 (11.98%)	0.184	0.326

Table 3: Coverage and accuracy for the standard model using gloss-to-fluent translation mappings learned from bilingual dictionaries in other languages (in forward order only).

6.3 Using forward and backward ordering for English gloss search

In our standard model, the literal English gloss for a source compound word (for example, *iron path*) matches glosses in other language dictionaries only in the identical order. But given that modifier/head word order often differs between languages, we test how searching for both orderings (e.g. *"iron path" and "path iron"*) can improve performance, as shown in Table 4. The percentages in parentheses show relative increase from the performance of the standard model in Section 6.2. We see a substantial improvement in both coverage and accuracy.

Language	Cmpnd wrds translated	Top1 Acc.	Top10 Acc.
Albanian	3229(+4.67%)	.217(+17.30%)	.409(+23.19%)
Bulgarian	6806(+1.29%)	.255(+3.24%)	.442(+6.25%)
German	11346(+2.19%)	.199(+2.05%)	.388(+7.18%)
Swedish	12970(+2.28%)	.189(+0.53%)	.361(+4.34%)
Avg_{10}	9603(+3.03%)	.193(+4.89%)	.362(+11.04%)

Table 4: Performance for looking up English gloss via both orderings. The percentages in parentheses are relative improvements from the performance in Table 3

6.4 Increasing coverage by automatically discovering compound morphology

For many languages, the compounding process introduces its own morphology (Figure 2). For example, in German, the word *Geschäftsführer (manager)* consists of the lexemes *Geschäft (business)* and *Führer (guide)* joined by the lexeme *-s*. For the purposes of these experiments, we will call such lexemes *fillers or middle glue characters*. Koehn and Knight (2003) used a fixed set of two known fillers *s and es* for handling German compounds. To broaden the applicability of this work to new languages without linguistic guidance, we show how such fillers

[5]Found Acc. is the fraction of examples for which the correct translation appears *anywhere* in the n-best list

[6]One explanation for this could be that for only a small percentage of compound words, their dictionary translations are formed by concatenating their English glosses. Also, Grefenstette (1999) reports much higher accuracies for German on this model because the 724 German test compounds were chosen in such a way that their correct translation is a concatenation of the possible component word translations.

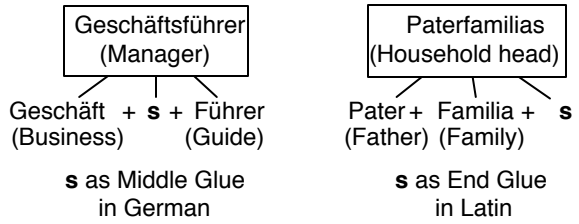

Figure 2: Illustration of compounding morphology using middle and end glue characters.

can be estimated directly from corpora in different languages. In additional to fillers, compound can also introduce morphology at the suffix or prefix of compounds, for example, in the Latin language, the lexeme *paterfamilias* contains the genitive form *familias* of the lexeme *familia (family)*, thus *s* in this case is referred to as the *"end glue"* character. To

Albanian		**Bulgarian**		**German**		**Swedish**	
Top 5 Middle Glue Character(s)							
j	0.059	О	0.129	s	0.133	s	0.132
s	0.048	И	0.046	n	0.090	l	0.051
t	0.042	Н	0.036	k	0.066	n	0.049
r	0.042	Я	0.025	h	0.042	t	0.045
i	0.038	А	0.025	f	0.037	r	0.035
Top 5 End Glue Character(s)							
m	0.146	Т	0.124	n	0.188	a	0.074
t	0.079	ЕН	0.092	t	0.167	g	0.073
s	0.059	Н	0.063	en	0.130	t	0.059
k	0.048	М	0.049	e	0.069	e	0.057
r	0.037	АМ	0.047	d	0.043	d	0.057

Table 5: Top 5 middle glues (fillers) and end glues discovered for each language along with their probability scores.

augment the splitting step outlined in Section 4.1, we allow deletion of up to two middle characters and two end characters. Then, for each glue candidate (for example *es*), we estimate its probability as the relative frequency of unique hypothesized compound words successfully using that particular glue. We rank the set of glues by their probability and take the top 10 middle and end glues for each language. A sample of glues discovered for some of the languages are shown in Table 5. The performance for the morphology step is shown in Table 6. The relative percentage improvements are with respect to the previous Section 6.3. We observe significant gain in coverage as the flexibility of glue process allows discovery of more compounds.

6.5 Re-ranking using context vector projection

We may further improve performance by re-ranking candidate translations based on the goodness of semantic "fit" between two words, as measured by

Language	**Cmpnd wrds translated**	**Top1 Acc.**	**Top10 Acc.**
Albanian	3272(+1.33%)	.214(-1.38%)	.407(-0.49%)
Bulgarian	7211(+5.95%)	.258(+1.18%)	.443(+0.23%)
German	13372(+17.86%)	.200(+0.50%)	.391(+0.77%)
Swedish	15094(+16.38%)	.190(+0.53%)	.363(+0.55%)
Avg_{10}	10273(+6.98%)	.194(+0.52%)	.363(+0.28%)

Table 6: Performance for increasing coverage by including compounding morphology. The percentages in parentheses are relative improvements from the performance in Table 4

their context similarity. This can be accomplished as in Rapp (1999) and Schafer and Yarowsky (2002) by creating bag-of-words context vectors around both the source and target language words and then projecting the source vectors into the (English) target space via the current small translation dictionary. Once in the same language space, source words and their translation hypotheses are compared via cosine similarity using their surrounding context vectors. We performed this experiment for German and Swedish and report average accuracies with and without this addition in Table 7. For monolingual corpora, we used the German and Swedish side of the Europarl corpus (Koehn, 2005) consisting of approximately 15 million and 21 million words respectively. We were able to project context vectors for an average of 4224.5 words in the two languages among all the possible compound words detected in Section 6.4. The poor Eurpoarl coverage could be due to the fact that compound words are generally technical words with low Europarl corpus frequency, especially in parliamentary proceedings. We believe that the small performance gains here are due to these limitations of the monolingual corpora.

Method	**Top1**$_{avg}$	**Top10**$_{avg}$
Original ranking	0.196	0.388
Comb. with Context Sim	0.201	0.391

Table 7: Average performance on German and Swedish with and without using context vector similarity from monolingual corpora.

6.6 Using phrase-tables if a parallel corpus is available

All previous results in this paper have been for translation lexicon discovery *without* the need for parallel bilingual text (bitext), which is often in limited supply for lower-resource languages. However, it is useful to assess how this translation lexicon dis-

covery work compares with traditional bitext-based lexicon induction (and how well the approaches can be combined). For this purpose, we used phrase tables learned by the standard statistical MT Toolkit Moses (Koehn et al., 2007). We tested the phrase-table accuracy on two languages, one for which we had a lot of parallel data available (German-English Europarl corpus with approx. 15 million words) and one for which we had relatively little parallel data (Czech-English news-commentary corpus with approx. 1 million words). This was done to see how the amount of parallel data available affects the accuracy and coverage of compound translation. Table 8 shows the performance for this experiment. For German, we see a significant improvement in accuracy and for Czech a small improvement in Top1 but a decline in Top10 accuracy. Note that these accuracies are still quite low as compared to general performance of phrase tables in an end-to-end MT system because we are measuring exact-match accuracy on a generally more challenging and often-lower-frequency lexicon subset. The third row in Table 8 for each of the languages shows that if one had a parallel corpus available, its n-best list can be combined with the n-best list of Bilingual Dictionaries algorithm to provide much higher consensus accuracy gains using weighted voting.

Method	# of words translated	Top1 Acc.	Top10 Acc.
German			
BiDict	13372	0.200	0.391
Parallel Corpus SMT	3281	0.423	0.576
Parallel + BiDict	3281	0.452	0.579
Czech			
$\text{BiDict}_{thresh=1}$	3455	0.276	0.514
Parallel Corpus SMT	309	0.285	0.404
Parallel + BiDict	309	0.359	0.599

Table 8: Performance of this paper's BiDict approach compared with and augmented with traditional statistical MT learning from bitext.

7 Quantifying the Role of Cross-languages

7.1 Coverage/Accuracy Trade off

The number of languages offering a translation hypothesis for a given literal English gloss is a useful parameter for measuring confidence in the algorithm's selection. The more distinct languages exhibiting a translation for the gloss, the higher likelihood that the majority translation will be correct rather than noise. Varying this parameter yields the coverage/accuracy trade off as shown in Figure 3.

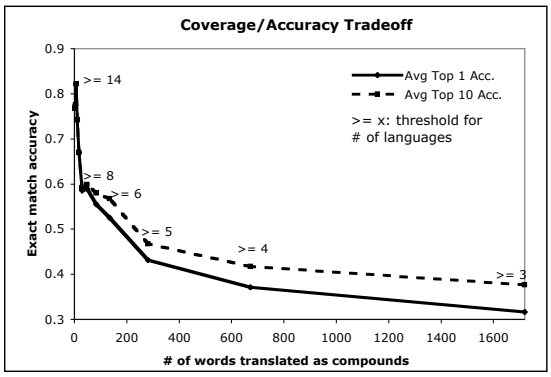

Figure 3: Coverage/Accuracy trade off curve by incrementing the minimum number of languages exhibiting a candidate translation for the source-word's literal English gloss. Accuracy here is the Top1 accuracy averaged over all 10 test languages.

7.2 Varying size of bilingual dictionaries

Figure 4 illustrates how the size of the bilingual dictionaries used for providing cross-language evidence affects translation performance. In order to take both coverage and accuracy into account, performance measure used was the F-score which is a harmonic average of Precision (the accuracy on the subset of words that could be translated) and Psuedo-recall (which is the correctly translated fraction out of total words that could be translated using 100% of the dictionary size). We can see in Figure 4 that increasing the percentage of dictionary size[7] always helps without plateauing, suggesting substantial extrapolation potential from large dictionaries.

7.3 Greedy vs Random Selection of Utilized Languages

A natural question for our compound translation algorithm is how does the choice of additional languages affect performance. We report two experiments on this question. A simple experiment is to use bilingual dictionaries of randomly selected languages and test the performance of K-randomly selected languages[8], incrementing K until it is the full set of 50 languages. The dashed lines in Figures 5

[7]Each run of choosing a percentage of dictionary size was averaged over 10 runs

[8]Each run of randomly selecting K languages was averaged over 10 runs.

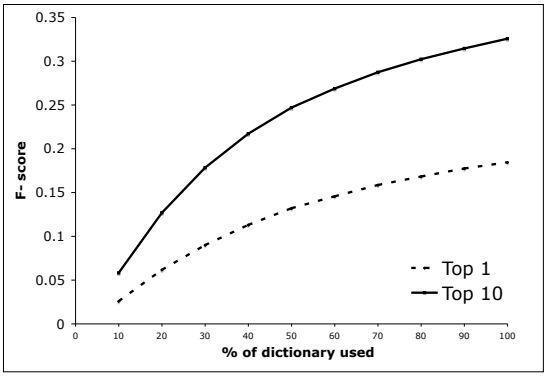

Figure 4: F-measure performance given varying sizes of the bilingual dictionaries used for cross-language evidence (as a percentage of words randomly utilized from each dictionary).

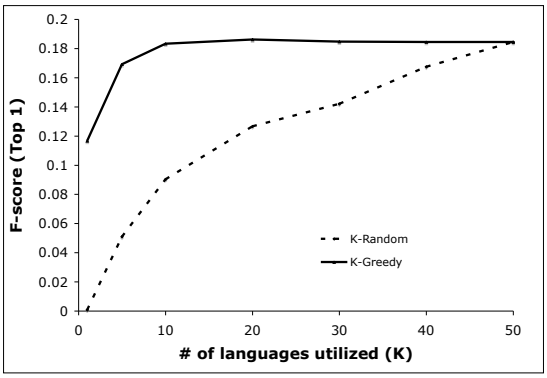

Figure 5: Top-1 match F-score performance utilizing K languages for cross-language evidence, for both a random K languages and greedy selection of the most effective K languages (typically the closest or largest dictionaries)

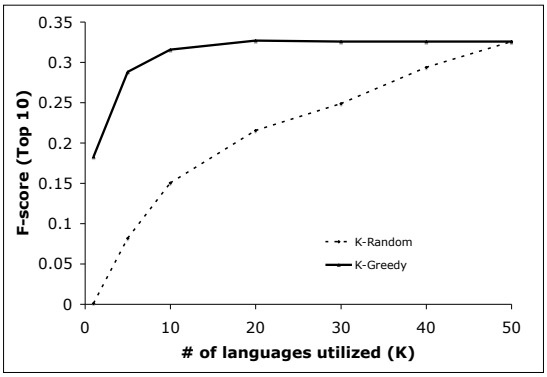

Figure 6: The performance relationship detailed in Figure 5 caption for Top-10 match F-score.

and 6 show this trend. The performance is measured by F-score as in section 7.1, where Pseudo-Recall here is the fraction of correct candidates out of the total candidates that could be translated had we used bilingual dictionaries of all the languages. We can see that adding random bilingual dictionaries helps improve the performance in a close to linear fashion. Furthermore, we observe that certain contributing languages are much more effective than others (e.g. Arabic/Farsi vs. Arabic/Czech). We use a greedy heuristic for ranking an additional cross-language, that is the number of test words for which the correct English translation can be provided by the bilingual dictionary of the respective cross-language. Figures 5 and 6 show that greedy selection of the most effective K utilized languages using this heuristic substantially accelerates performance. In fact, beyond the best 10 languages, performance plateaus and actually decreases slightly, indicating that increased noise is outweighing increased coverage.

Albanian			**Arabic**		
Russian	0.067	0.116	Farsi	0.051	0.090
+Spanish	0.100	0.169	+Spanish	0.059	0.111
+Bulgarian	0.119	0.201	+French	0.077	0.138
Bulgarian			**Czech**		
Russian	0.186	0.294	Slovak	0.177	0.289
+Hungarian	0.190	0.319	+Russian	0.222	0.368
+Swedish	0.203	0.339	+Hungarian	0.235	0.407
Farsi			**German**		
Arabic	0.031	0.047	Dutch	0.130	0.228
+Dutch	0.038	0.070	+Swedish	0.191	0.316
+Spanish	0.044	0.079	+Hungarian	0.204	0.355
Hungarian			**Russian**		
Swedish	0.073	0.108	Bulgarian	0.185	0.250
+Dutch	0.103	0.158	+Hungarian	0.199	0.292
+German	0.117	0.182	+Swedish	0.216	0.319
Slovak			**Swedish**		
Czech	0.145	0.218	German	0.120	0.188
+Russian	0.168	0.280	+Hungarian	0.152	0.264
+Hungarian	0.176	0.300	+Dutch	0.182	0.309

Table 9: Illustrating 3-best cross-languages obtained for each test language (shown in bold). Each row shows the effect of adding the respective cross-language to the set of languages in the rows above it and the corresponding F-scores (Top 1 and Top 10) achieved.

7.4 Languages found using Greedy selection

Table 9 shows the sets of the most effective three cross-languages per test language selected using the greedy heuristic explained in previous section. Unsurprisingly, related languages tend to help more than distant languages. For example, Dutch is most

effective for the test language German, and Slovak is most effective for Czech. We can also see interesting symmetries between related languages, for example: Farsi is the top language used for test language Arabic and vice-versa. Such symmetries can also be seen for other pairs of related languages such as (Czech, Slovak) and (Russian, Bulgarian). Thus, related languages are most helpful and they can be related in several ways such as etymologically, culturally and physically (such as Hungarian contact with the Germanic languages). The second point to note is that languages having large dictionaries also tend to be especially helpful, even when unrelated. This can be seen by the presence of Hungarian in top three cross-languages for most of the test languages. This is likely because Hungarian was one of the largest dictionaries and hence can provide good coverage for obtaining translation candidates of rarer or technical compounds, which may have more language universal literal glosses.

8 Conclusion

This paper has shown that successful translation of compounds can be achieved without the need for bilingual training text, by modeling the mapping of literal component-word glosses (e.g. "iron-path") into fluent English (e.g. "railway") across multiple languages. An interesting property of using such cross-language evidence is that one does need to restrict the candidate translations to compositional (or "glossy") translations, as our model allows the successful generation of more fluent non-compositional translations. We further show improved performance by adding component-sequence and learned-morphology models along with context similarity from monolingual text and optional combination with traditional bilingual-text-based translation discovery. These models show consistent performance gains across 10 diverse test languages.

9 Acknowledgments

We thank Chris Callison-Burch for providing access to phrase tables and giving valuable comments on this work as well as suggesting useful additional experiments. We also thank Markus Dreyer for helping with German examples and David Smith for giving valuable comments on initial version of the paper.

References

T. Baldwin and T. Tanaka. 2004. Translation by Machine of Complex Nominals: Getting it Right. *Proceedings of the ACL-2004 Workshop on Multiword Expressions*, pages 24–31.

R.D. Brown. 2002. Corpus-driven splitting of compound words. *Proceedings of TMI-2002*.

Y. Cao and H. Li. 2002. Base Noun Phrase translation using web data and the EM algorithm. *Proceedings of COLING-Volume 1*, pages 1–7.

G. Grefenstette. 1999. The World Wide Web as a Resource for Example-Based Machine Translation Tasks. *In ASLIB'99 Translating and the Computer 21*.

P. Koehn and K. Knight. 2003. Empirical methods for compound splitting. *Proceedings of the EACL-Volume 1*, pages 187–193.

P. Koehn, H. Hoang, A. Birch, C. Callison-Burch, M. Federico, N. Bertoldi, B. Cowan, W. Shen, C. Moran, R. Zens, et al. 2007. Moses: Open source toolkit for statistical machine translation. *Proceedings of ACL, companion volume*, pages 177–180.

P. Koehn. 2005. Europarl: A parallel corpus for statistical machine translation. *MT Summit X*.

J.N. Levi. 1978. *The Syntax and Semantics of Complex Nominals*.

R. Navigli, P. Velardi, and A. Gangemi. 2003. Ontology learning and its application to automated terminology translation. *Intelligent Systems, IEEE*, 18(1):22–31.

U. Rackow, I. Dagan, and U. Schwall. 1992. Automatic translation of noun compounds. *Proceedings of COLING-Volume 4*, pages 1249–1253.

R. Rapp. 1999. Automatic identification of word translations from unrelated English and German corpora. *Proceedings of ACL*, pages 519–526.

C. Schafer and D. Yarowsky. 2002. Inducing translation lexicons via diverse similarity measures and bridge languages. *Proceedings of COLING*, pages 1–7.

C. Schafer and D. Yarowsky. 2004. Exploiting aggregate properties of bilingual dictionaries for distinguishing senses of English words and inducing English sense clusters. *Proceedings of ACL-2004*, pages 118–121.

T. Tanaka and T. Baldwin. 2003. Noun-Noun Compound Machine Translation: A Feasibility Study on Shallow Processing. *Proceedings of the ACL-2003 Workshop on Multiword Expressions*, pages 17–24.

J. Zhang, J. Gao, and M. Zhou. 2000. Extraction of Chinese compound words: an experimental study on a very large corpus. *Proceedings of the Second Workshop on Chinese Language Processing*, pages 132–139.