# Unsupervised Induction of Cross-lingual Semantic Relations

**Mike Lewis**
School of Informatics
University of Edinburgh
Edinburgh, EH8 9AB, UK
`mike.lewis@ed.ac.uk`

**Mark Steedman**
School of Informatics
University of Edinburgh
Edinburgh, EH8 9AB, UK
`steedman@inf.ed.ac.uk`

## Abstract

Creating a language-independent meaning representation would benefit many cross-lingual NLP tasks. We introduce the first unsupervised approach to this problem, learning clusters of semantically equivalent English and French relations between referring expressions, based on their named-entity arguments in large monolingual corpora. The clusters can be used as language-independent semantic relations, by mapping clustered expressions in different languages onto the same relation. Our approach needs no parallel text for training, but outperforms a baseline that uses machine translation on a cross-lingual question answering task. We also show how to use the semantics to improve the accuracy of machine translation, by using it in a simple reranker.

## 1 Introduction

Identifying a language-independent semantics is a major long term goal of computational linguistics, and is interesting both theoretically and for practical applications. It assumes that semantically equivalent sentences in any language can be mapped onto a common meaning representation. Such a representation would be of great utility for tasks such as translation, relation extraction, summarization, question answering, and information retrieval. Regardless of whether it is even possible to create such a semantics, we show that an incomplete version can be useful for downstream tasks.

Semantic machine translation aims to map a source language to a language-independent meaning representation, and then generate the target language translation from this. It is hoped this would alleviate the difficulties of simpler models when translating between languages with very different word ordering and syntax (Vauquois, 1968). Despite many attempts to define interlingual representations (Mitamura et al., 1991; Beale et al., 1995; Banarescu et al., 2013), state-of-the-art machine translation still uses phrase-based models (Koehn et al., 2007). The major obstacle to defining interlinguas has been devising a meaning representation that is language-independent, but capable of expressing the limitless number of meanings that natural languages can express (Dorr et al., 2004).

Our approach avoids this problem by utilizing the methods of distributional semantics. Recent work has shown that paraphrases of expressions can be learned by clustering those with similar arguments (Poon and Domingos, 2009; Yao et al., 2011; Lewis and Steedman, 2013)—for example learning that *X wrote Y* and *X is the author of Y* are equivalent if they appear in a corpus with similar *(X, Y)* argument-pairs such as $\{$*(Shakespeare, Macbeth), (Dickens, Oliver Twist)*$\}$. We extend this to the multilingual case, aiming to also map the French equivalents *X a écrit Y* and *Y est un roman de X* on to the same cluster as the English paraphrases. Conceptually, we treat a foreign expression as a paraphrase of an English expression. The cluster identifier can be used as a predicate in a logical form, suggesting that the fundamental predicates of an interlingua can be learnt in an unsupervised manner via clustering.

In this paper we focus on learning binary relations between named entities. This problem is much simpler than attempting complete interlingual semantic

interpretation, but the approach could be generalized. This class of expressions has proved extremely useful in the monolingual case, with direct applications for question answering and relation extraction (Poon and Domingos, 2009; Mintz et al., 2009), and we demonstrate how to use them to improve machine translation. It is important to be able to extract knowledge across languages, as many facts will not be expressed in all languages—either due to less-complete encyclopedias being available in some languages, or facts being most relevant to a single country.

In contrast to most previous work on machine translation and cross-lingual clustering, our method requires no parallel text (see Section 8 for discussion of some exceptions). It instead exploits an alignment between named-entities in different languages. The limited size of parallel corpora is a significant bottleneck for machine translation (Resnik and Smith, 2003), whereas our approach can be used on much larger monolingual corpora. This means it is potentially useful for language-pairs where little parallel text is available, for domain adaptation, or for semi-supervised approaches.

## 2  Basic Approach

Our work builds on clustering-based approaches to monolingual distributional semantics, aiming to create clusters of semantically equivalent predicates, based on their arguments in a corpus. In each language, we first map each sentence in a large monolingual corpus onto a simple logical form, by extracting binary predicates between named entities. Then, we cluster predicates both within and between languages into those with similar arguments.

When parsing a new sentence, instead of using the monolingual predicate, we use the cluster identifier as a language-independent semantic relation, as shown in Figure 1. The resulting logical form can be used for inference in question answering.

Unlike traditional approaches to translation, this does not require parallel text—but it does impose some additional constraints on language resources. Our approach requires:

- A large amount of factual text, as we rely on the same facts being expressed in different languages. We use Wikipedia, which contains ar-

ticles in 250 languages, including 121 with at least 10,000 articles.[1] Other domains, such as Newswire, may also be effective.

- A method for extracting binary relations from sentences. This is straightforward from dependency parses, which are available for many languages. It is also possible without a parser, with some language-specific work (Fader et al., 2011). We describe our approach in Section 3.

- A method for linking entities in the training data to some canonical representation. McNamee et al. (2011) report good results on this task in 21 languages. We describe our method for this in Section 4.1.

## 3  Predicate Extraction

Our method relies on extracting binary predicates between entities from sentences. Various representations have been suggested for binary predicates, such as Reverb patterns (Fader et al., 2011), dependency paths (Lin and Pantel, 2001; Yao et al., 2011), and binarized predicate-argument relations derived from a CCG-parse (Lewis and Steedman, 2013). Our approach is formalism-independent, and is compatible with any method of expressing binary predicates.

We choose the CCG-based parser of Lewis and Steedman (2013) for several reasons. It outputs a logical form derived automatically from the CCG-parse, containing predicates such as: $write_{arg0,arg1}(shakespeare,macbeth)$. By using the close relationship between the CCG syntax and semantics, it is able to generalize over many semantically equivalent syntactic constructions (such as passives, conjunctions and relative clauses), meaning we can map both *Shakespeare wrote Macbeth* and *Macbeth was written by Shakespeare* to the same logical form. Using a dependency-based representation, these would have different predicates, which would need to be clustered later. CCG also has a well developed theory of operator semantics (Steedman, 2012), so is able to represent semantic operators such as quantifiers, negation and tense—understanding these is crucial to high performance on question answering or translation tasks. As in
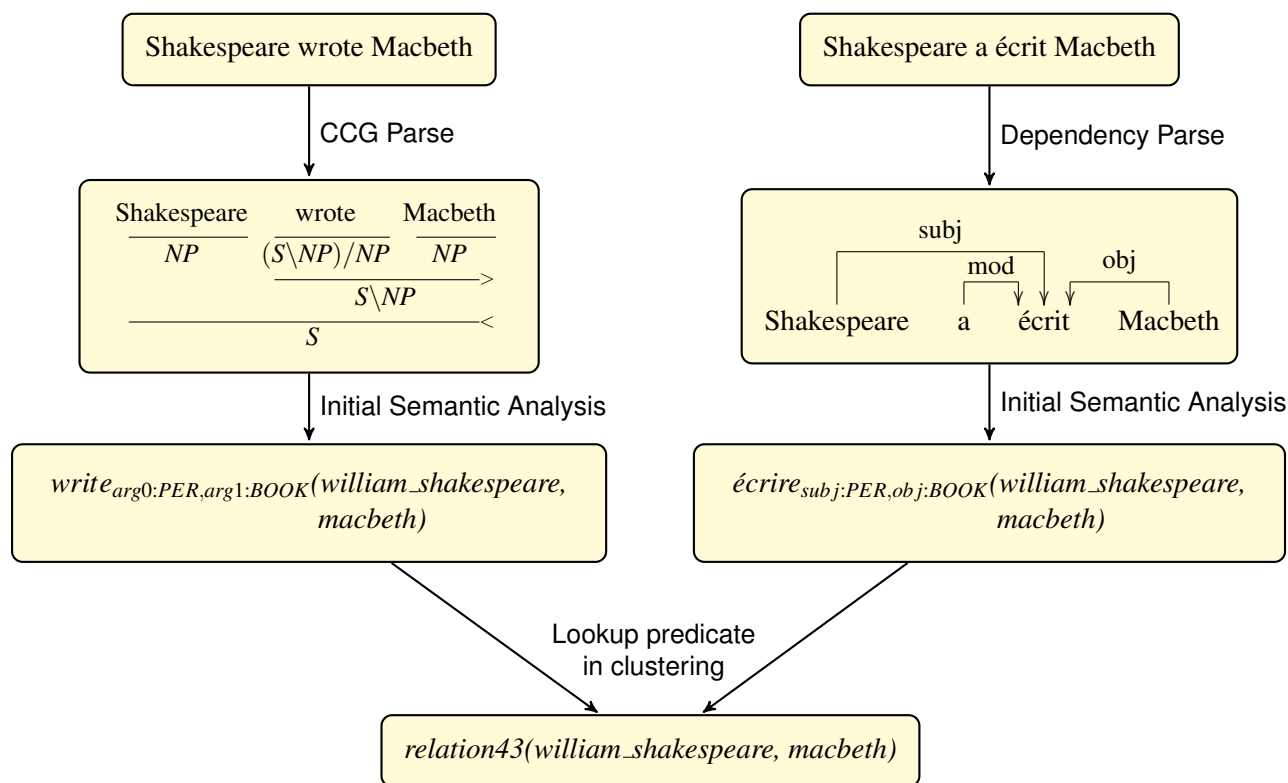
---

[1]As of June 2013.

Figure 1: Example showing how our system can map sentences in different languages to the same meaning representation, assuming we have clustered the equivalent predicates $write_{arg0:PER,arg1:BOOK}$ and $écrire_{subj:PER,obj:BOOK}$

Lewis and Steedman (2013), clusters derived from the output from the parser can be integrated into the lexicon, allowing us to build logical forms which capture both operator and lexical semantics.

Accurate CCG syntactic parsers are currently only available for English, whereas dependency treebanks and parsers exist for many languages (Buchholz and Marsi, 2006). Consequently, for French we use the dependency path representation, which captures the nodes and edges connecting two named entities in a dependency parse. The extraction of these paths is language-independent, and does not depend on the dependency grammar used, which means our approach could be adapted to new languages with minimal work.

## 4 Entity Semantics

### 4.1 Entity Linking

As discussed, our approach assumes that semantically similar predicates will have similar argument entities. This requires us to be able to identify coreferring entities across languages during training. In the monolingual case, it suffices to represent entities by the string used in the sentence. This is inadequate in the multilingual case, as many entities may be referred to by different names in different languages—for example the *United States* translates as *les États-Unis* in French and *die Vereinigte Staaten* in German. This problem is worsened by the ambiguity of named-entity strings—for example, in the context of a sports article, *United States* may refer specifically to a team, rather than a country.

Recent work on multilingual named-entity linking (McNamee et al., 2011) shows how to link named entities in multiple languages onto English Wikipedia articles, which can be used as unique identifiers for entities. This means that we could gain the information we need from unrestricted text. However, as we use Wikipedia itself for our training corpora, we can bootstrap entity information directly from its markup. Wikipedia contains cross-language links, e.g. between the United States articles in different languages, allowing us to determine the equivalence of entities in differ-

ent languages. Wikipedia links also help us automatically disambiguate entities to a given article. For unlinked named-entity mentions, we perform some simple heuristic co-reference—based on word-overlap with previously mentioned entities in the document, whether the mention name is the title of a Wikipedia article, or whether the mention name is a Freebase (Bollacker et al., 2008) alias of an entity. We emphasise that this does not mean our approach is only applicable to the Wikipedia corpus.

## 4.2 Entity Typing

It has become standard in clustering approaches to distributional semantics to assign types to predicates before clustering, and only cluster predicates with the same type (Schoenmackers et al., 2010; Berant et al., 2011; Yao et al., 2012). This is useful for resolving ambiguity—for example the phrase *born in* may express a place-of-birth or date-of-birth relation depending on whether its second argument has a *LOC* or *DAT* type. Ambiguous expressions may translate differently in other languages—for example, the two interpretations of *was born in* translate in French as *est né à* and *est né en* respectively. The type of a predicate is determined by the type of its arguments, and predicates with different types are treated as distinct.

Lewis and Steedman (2013) induce an unsupervised model of entity types using Latent Dirichlet Allocation (Blei et al., 2003), based on selectional preferences of verbs and argument-taking nouns. When applied cross-linguistically, we found this technique tended to create language-specific topics. Instead, we exploit the fact that many Wikipedia entities are linked to the Freebase database, which has a detailed manually-built type-schema. This means for a Wikipedia entity, we can look up its set of types in Freebase.[2] We use the simplified type-set of 112 types created by Ling and Weld (2012). Where entities have multiple types (for example, Shakespeare is both an *author* and a *person*), we create a separate relation for each type.

---

[2]Named entities not present in Freebase are ignored during training.

## 5 Relation Clustering

Predicates are clustered into those which are semantically equivalent, based on their argument-pairs in a corpus. The initial semantic analysis is run over the corpora, and for each predicate we build a vector containing counts for each of its argument-pairs (we divide these counts by the overall frequency of an argument-pair in the corpus, so that rarer argument-pairs are more significant). These vectors are used to compute similarity between predicates.

First, we run the clustering algorithm on each language independently, and then we attempt to find an alignment between the clusters. Duc et al. (2011) and Täckström et al. (2012) use similar two-step approaches. Running the clustering on both languages simultaneously was found to produce many clusters only containing predicates from a single language. This appears to be because even if predicates in two different languages are truth-conditionally equivalent, the language biases the sample of entity-pairs found in a corpus. For example, the French verb *écrire* may contain more French author/book pairs than the English equivalent *write*. This difference can make the verbs appear to represent different predicates to the clustering algorithm. Our two-step approach also means that advances in monolingual clustering should directly lead to improved cross-lingual clusters.

## 5.1 Monolingual Clustering

Following Lewis and Steedman (2013), we use the Chinese Whispers algorithm (Biemann, 2006) for monolingual clustering—summarized in Algorithm 1. The algorithm is non-parametric, meaning that the number of relation clusters is induced from the data, and highly scalable. We create a separate graph for each type of predicate in each language—for example, predicates between types *AUTHOR* and *BOOK* in French (so only predicates with the same type will be clustered). We create one node per predicate in the graph, and edges represent the distributional similarity between the predicates.

The distributional similarity between a pair of predicates is calculated as the cosine-similarity of their argument pair vectors in the corpus. Many more sophisticated approaches to determining similarity have been proposed (Kotlerman et al., 2010;

684

Weisman et al., 2012), and future work should explore these. We prune nodes with less than 25 occurrences, edges of weight less than 0.05, and a short list of stop predicates. We find many of our French dependency paths do not have a clear semantic interpretation, so add the requirement that dependency paths contain at least one content word, contain at most 5 edges, and that one of the dependencies connected to the root is subject, object or the French preposition *de*.

---

**Data**: Set of predicates $P$
**Result**: A cluster assignment $r_p$ for all $p \in P$
$\forall p \in P : r_p \longleftarrow$ unique cluster identifier;
**while** *not converged* **do**
    randomize order of $P$
    **for** $p \in P$ **do**
        $r_p \longleftarrow \arg\max_r \sum_{p'} \mathbb{1}_{r=r_{p'}} sim(p,p')$
    **end**
**end**

**Algorithm 1:** Chinese Whispers algorithm, used for monolingual predicate clustering. $sim(p,p')$ is the distributional similarity between $p$ and $p'$, and $\mathbb{1}_{r=r'}$ is 1 iff r=r' and 0 otherwise

---

## 5.2 Cross-lingual Cluster Alignment

We use a simple greedy procedure to find an alignment between the monolingual clusters in different languages. First, the entity-pair vectors for each predicate in a relation cluster are merged. Then, the cosine similarity between entity-pair vectors for clusters in different languages is calculated—we base this only on argument-pairs that occur in both languages, to reduce the potential bias of some entities being more relevant to one language. Clusters are then greedily aligned, in order of their similarity, as in Algorithm 2 (pruning similarities less than 0.01). This means that clusters are aligned with their most similar foreign cluster. We only attempt to align clusters with the same argument types.

## 6 Cross Lingual Question Answering Experiments

We evaluate our system on English and French, using Wikipedia for corpora. The English corpus is POS-tagged and CCG-parsed with the C&C tools

---

**Data**: Sets of monolingual relation clusters $R_{L1}$ and $R_{L2}$
**Result**: An alignment between the monolingual clusters $A$
$A \longleftarrow \{\}$;
**while** $R_{L1} \neq \{\} \wedge R_{L2} \neq \{\}$ **do**
    $(r1, r2) \longleftarrow \underset{(r_1,r_2)\in R_{L1}\times R_{L2}}{\arg\max} \ sim(r1, r2)$;
    $A \longleftarrow A \cup \{(r1, r2)\}$;
    $R_{L1} \longleftarrow R_{L1}/\{r1\}$;
    $R_{L2} \longleftarrow R_{L2}/\{r2\}$;
**end**

**Algorithm 2:** Cluster alignment algorithm

| English | French |
|---|---|
| X invades Y | X envahit Y |
|  | invasion de Y par X |
| X orbits Y | X est un satellite de Y |
|  | X est une lune de Y |
| X is a skyscraper in Y | X est un gratte-ciel de Y |
| X is a novel by Y | X est un roman de Y |
| X joins Y | X adhère à Y |
| X is a member of Y | X entre dans Y |
|  | X rejoint Y |

Table 1: Some example cross-lingual clusters. Predicates are given in a human-readable form, and predicate types are suppressed.

(Clark and Curran, 2004). The French corpus is tagged with MElt (Denis et al., 2009) and parsed with MaltParser (Nivre et al., 2007), trained on the French Treebank (Candito et al., 2010). Wikipedia markup is filtered using Wikiprep (Gabrilovich and Markovitch, 2007)—replacing internal links with the name of their target article, to help entity linking. Some example clusters learnt by our model are shown in Table 1. We find that the cross-lingual clusters typically contain more French expressions than English, possibly due to the differing sizes of the corpora—adjusting the parameters in Section 5 results in larger clusters, but introduces noise.

## 6.1 Experimental Setup

We evaluate our system on a cross-lingual question answering task, similar to monolingual QA evaluations by Poon and Domingos (2009) and Lewis and

Steedman (2013). A question is asked in language *L*, and is answered by the system from a corpus of language *L'*. Human annotators are shown the question, answer entity, and the sentence that provided the answer, and are then asked whether the answer is a reasonable conclusion based on the sentence. Whilst this task is much easier than full translation, it is both a practical application for our approach, and a reasonably direct extrinsic evaluation for our cross-lingual clusters.

Following Poon and Domingos (2009) and Lewis and Steedman (2013), the question dataset is automatically generated from the corpus. This approach has the advantage of evaluating on expressions in proportion to their corpus frequency, so understanding frequent expressions is more important than rare ones. We then sample 1000 questions for each language, by extracting binary relations matching certain patterns ($X \overset{nsubj}{\leftarrow} verb \overset{dobj}{\rightarrow} Y$, $X \overset{nsubj}{\leftarrow} verb \overset{pobj}{\rightarrow} Y$ or $X \overset{nsubj}{\leftarrow} be \overset{dobj}{\rightarrow} noun \overset{pobj}{\rightarrow} Y$), and removing one of the arguments. For example, from the sentence *Obama lives in Washington* we create the questions *X lives in Washington?*, and *Obama lives in X?*.[3] Answers are judged by fluent bilingual humans, and do not have to match the entity that originally instantiated *X*. Multiple answers can be returned for the same question.

Our system attempts this task by mapping both the question and candidate answer sentences (which will be in a different language to the question) on to a logical form using the clusters, and determining whether they express the same relation. This tests the ability of our approach to cluster expressions into those which are semantically equivalent between languages. It is possible for entities to have multiple types (see Section 4.2), and answers are ranked by the number of types in which the entailment relation is predicted to hold.

---

[3]Questions are given in a declarative form, to make the tasks simpler for the machine translation baseline. We found the machine translation performed poorly on questions such as *What is Obama the president of?*, as inverted word-orders and long-range dependencies are difficult to handle with re-ordering models and language models (though are straightforward to handle for a CCG system (Clark et al., 2004)). We find that machine translation performs much better on declarative equivalents, such as: *Obama is the president of X*.

## 6.2 Baseline

Our baseline makes use of the Moses machine translation system (Koehn et al., 2007), and is similar to previous approaches to cross-lingual question answering such as Ahn et al. (2004). We train a Moses model on the Europarl corpus (Koehn, 2005). First, the question is translated from language *L* to *L'*, taking the 50-best translations. As the questions are typically shorter than corpus sentences, this is substantially easier for the machine-translation than translating the corpus. These are then parsed, and patterns are extracted (as in Section 3). We also manually supply a translation of the named-entity in the question (based on the Freebase entity name translation), to avoid penalizing the translation system for failing to translate named-entities that have not been seen in its training data. These patterns are then used to find answers to the questions. Answers are ranked by the score of the best translation that produced the pattern. Figure 2 illustrates this pipeline.

The choice of languages is very favourable to the machine-translation system, English and French have similar word-order, and there is a large amount of parallel text available (Koehn and Monz, 2006). Our system works with any word-order, and does not require parallel text for training, so we would expect better performance relative to machine-translation on other language pairs. Future work will experiment with more diverse languages. The sentences to be translated are also very short, reducing the potential for error.

## 6.3 Results

Results are shown in Table 3, based on a sample of 100 answers from the output of each of the systems. Unsurprisingly, the machine-translation has high accuracy on this task, given the choice of languages and the short queries. Pleasingly, our clusters achieve similar accuracy, with much greater recall, with no usage of parallel text.

Examining the results, we see that the distribution of answers is highly skewed for all systems, with many answers to a smaller number of questions (multiple answers can be returned to the same question). This is due to the Zipfian nature of language, the difficulty of the task (which is far from

| Question | Answer |
|---|---|
| X dies in Moscow | **Sergueï Guerassimov** meurt d'une crise cardiaque le mardi 26 novembre 1985 à Moscou |
| Germany invades X | . . . depuis l'invasion de la **Pologne** par l'Allemagne et l'URSS |
| X wins the FA Cup | **Portsmouth FC** remporte la FA Challenge Cup en s'imposant en finale face à Wolverhampton Wanderers FC |
| X is a band from Finland | **Yearning** est un groupe Finlande de doom metal atmosphérique |
| X vit en France | **Dewi Sukarno** . . . has lived in different countries including Switzerland, France and the United States |
| X bat Kurt Angle | **Anderson** defeated Kurt Angle and Abyss to advance to the finals |
| X est une ville de Kirghizistan | **Il'chibay** is a village in the Issyk Kul Province of Kyrgyzstan |

Table 2: Example questions correctly answered using our clusters, with the answer entity highlighted in bold.
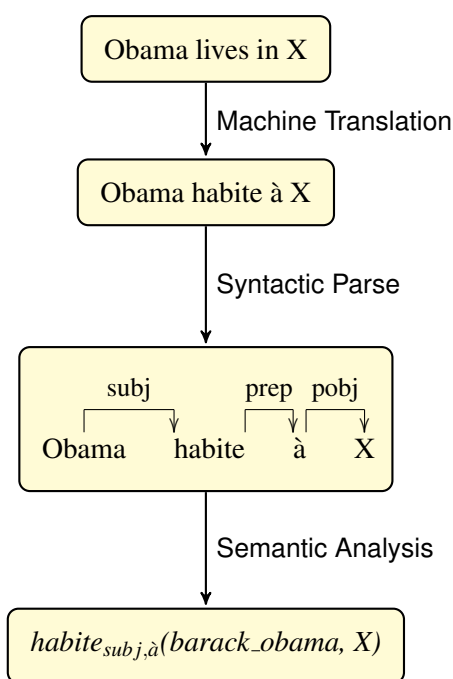


Figure 2: Pipeline used by baseline system for answering French questions. The pattern extracted from the translated sentence is used to search for answers in an English corpus.

| English→ French | Answers | Correct |
|---|---|---|
| Baseline | 269 | 86% |
| Clusters (best 270) | 270 | 100% |
| Clusters (all) | 1032 | 72% |
| **French→ English** | **Answers** | **Correct** |
| Baseline | 274 | 85% |
| Clusters (all) | 401 | 93% |

Table 3: Results on wide-coverage Question Answering task. Best-N results are shown to illustrate the accuracy of our cluster-based system at the same rank as the baseline. It is not possible to give a recall figure, as the total number of correct answers in the corpus is unknown. English→ French results are from the full French Wikipedia corpus, whereas French→ English results are from a 10% sample.

solved in the monolingual case), and the possibility that questions may have no answers in the foreign corpus. This is particularly true for the clustering approach—although the clustering system finds more answers with the English corpus, the baseline system answers slightly more unique questions (57 vs 66). The 1032 answers found by the clusters in the French corpus came from just 56 questions (compared to 29 unique questions answered by the baseline). This suggests that the translations found by the clustering can be more useful than those of Moses on this task—for example, it may find an equivalence between a rare French term and a common related English term, where machine translation may only find a more literal translation.

Despite this, we see the clusters have learnt to

687

paraphrase a variety of relations between languages with high accuracy, suggesting that there is much potential for the use of unsupervised clusters in cross-lingual semantic applications. Some examples answers are given in Table 2. Most of the errors are caused by a small number of questions.

# 7 Translation Reranking Experiments

Ultimately, we would like to be able to translate using semantic parsing with cross-lingual clusters. As a step towards this, we investigated whether we could rerank the output of a machine translation system, on the basis of whether the semantic parse of the source sentence is consistent with that of candidate translations.

We sample French sentences where we can produce a semantic parse (i.e. we can extract a predicate between named entities that maps to a cross-lingual cluster). These sentences are translated to English using Moses, taking the 50-best list, and semantic parses are produced for each of these. If the semantic parse for the 1-best translation does not match the source semantic parse, we take the parse from the 50-best list that most closely matches it—otherwise we discard the sentence from our evaluation, as our semantics agrees with the machine-translation.

To ensure that the evaluation focuses on the clusters, we try to exclude several other factors that might affect the results. The coverage of our CCG parsing and semantic analysis drops significantly on noisy translated sentences, and potentially acts as a language model by failing to produce any semantic parse on ungrammatical output sentences. We therefore only consider sentences where we can produce a semantic parse for the 1-best machine translation output. We also try to avoid penalizing the machine-translation system for failing to translate named entities correctly, so we do not attempt to rerank sentences where the entities from the source sentence are not present in the 1-best translation.

Human annotators were shown the source sentence, the 1-best translation, and the translation chosen by the reranker (the translations were shown in a random order). To focus the evaluation on the semantic relations we are modelling, we ask the annotators which sentence best preserves the meaning between the named entities that have different relations

|                          | Percentage of translations preferred |
|--------------------------|---------------------------------------|
| **1-best Moses translation** | 5%                                |
| **Cluster-based Reranker**   | 39%                               |
| **No preference**            | 56%                               |

Table 5: Human preference judgements for the translation reranking experiment, based on a sample of 87 sentences. Results show the percentage of sentences for which the annotators preferred the original translation, the reranked translation, or neither. As discussed in the text, results where annotators had no preference were typically due to syntactic parse errors.

in the semantic parse. This avoids our system being penalized for choosing a translation that is worse in aspects other than the relations it is modelling. An example is shown in Table 4. The data was annotated jointly by two fluent bilingual speakers, who reported high agreement on this task.

Results are shown in Table 5, and are highly encouraging, with the original Moses output being preferred to the reranked translation in only 5% of cases where our model makes a positive prediction.

Inspecting the results, we see that many of the cases where the annotators had no preference were caused by syntactic parse errors. For example, if the 1-best translation is correct, but a prepositional phrase is incorrectly attached, it will appear to have an incorrect semantics. A similar translation in the 50-best list may be correctly parsed, and consequently selected by our reranker. However, a human will have no preference between these translations. Incorporating K-Best parsing into our pipeline may help mitigate against such cases.

This preliminary experiment suggests that there is potential for future improvements in machine translation using cross-lingual distributional semantics. The system only attempts to rerank a very small proportion of sentences, but we believe the coverage could be greatly improved by including relations between common nouns (rather than just named-entities)—future work should explore this.

# 8 Related Work

Our work builds on recent progress in monolingual distributional semantics (Poon and Domingos, 2009; Yao et al., 2011; Lewis and Steedman, 2013) by

| Source | Le Princess Elizabeth arrive à Dunkerque le 3 août 1999 |
|---|---|
| **Machine translation 1-best** | Le Princess Elizabeth is to manage to Dunkirk on 3 August 1999 |
| **Reranked translation** | The Princess Elizabeth arrives at Dunkirk on 3 August 1999 |

Table 4: Example sentence that is reranked by our clusters. Human evaluators were asked which translation best preserved the meaning between *Princess Elizabeth* and *Dunkirk*.

clustering typed predicates into those which are semantically equivalent. We also show how to bootstrap semantic information about entities from the Wikipedia markup, and believe this makes it an interesting corpus for future work on monolingual distributional semantics.

Cross-language Latent Relational Analysis (Duc et al., 2011) is perhaps the most similar previous work to ours, which moves the work of Turney (2005) into a multilingual setting. Duc et al. (2011) aim to compute, for example, that the 'latent relation' between *(Obama, US)* in an English corpus is similar to that between *(Cameron, UK)* in a foreign corpus. This is solved by finding all textual patterns between the two entity-pairs, and computing their overall similarity. Like us, they compute similarity between expressions in different languages based on named-entity arguments and clustering (unlike us, they also rely on machine translation for computing similarity). A key difference is that their system aims to understand the overall relation between an entity-pair based on many observations, whereas our approach attempts to understand each sentence individually (as is required for tasks such as translation).

Various recent papers have explored the relationship between translation and monolingual paraphrases —for example Bannard and Callison-Burch (2005) create paraphrases by pivoting through a foreign translation, and Callison-Burch et al. (2006) show that including monolingual paraphrases improves the quality of translation by reducing sparsity. The success of these approaches depends on the many-to-many relationship between equivalent expressions in different languages. Our approach aims to model this relationship explicitly by clustering all equivalent paraphrases in different languages.

Current state-of-the-art machine translation systems circumvent the problem of full semantic interpretation, by using phrase-based models learnt from large parallel corpora (Brown et al., 1993). Although this approach has been very successful, it has significant limitations—for example, when translating between languages with very different word-orders (Birch et al., 2009), or with little parallel text.

Semantic machine translation aims to map the source language to an interlingual semantic representation, and then generate the target language sentence from this. Jones et al. (2012) show how this can be done on a small dataset using hyper-edge replacement grammars. A major obstacle to this is designing a suitable meaning representation, which involves choosing a set of primitive concepts which are abstract enough to be capable of expressing meaning in any language (Dorr et al., 2004). A recent proposal for this is the Abstract Meaning Representation (Banarescu et al., 2013), which uses English verbs as a set of predicates. This is a less abstract form of semantic interpretation than our proposal, as semantically equivalent paraphrases may be given a different representation. Such an approach also relies on annotating large amounts of text with the semantic representation—whereas our unsupervised approach offers a way to build such an interlingua using only a method for extracting predicates from sentences.

Whilst almost all recent work on machine-translation has relied on parallel text, there have been several interesting approaches that do not. Rapp (1999) learns to translate words based on small seed bilingual dictionary. Klementiev et al. (2012a) exploit a variety of interesting indirect sources of information to learn a lexicon—for example assuming that equivalent Wikipedia articles in different languages will use semantically similar words. The Polylingual Topic Model (Mimno et al., 2009) makes use of similar intuitions. Whilst we exploit equivalent Wikipedia articles for entity linking, we do not require aligned articles. Incorporating such techniques into our model would be a natural next

step, allowing us to learn a more complete lexicon. To our knowledge, ours is the first approach to learn to translate semantic relations, rather than words and phrases.

Several other recent papers have learnt cross-lingual word clusters, and used these to improve cross-lingual tasks such as document-classification (Klementiev et al., 2012b), parsing (Täckström et al., 2012) and semantic role labelling (Kozhevnikov and Titov, 2013) in resource-poor languages. Cross-lingual word clusters are learnt by aligning monolingual clusters on the basis of parallel text—in language-pairs where parallel text is available, this offers an interesting complement to our method of clustering based on named entities.

## 9 Conclusions and Future Work

We have demonstated that our previous work on monolingual distributional semantics can simply be extended to learn a language-independent semantics of relations from unlabelled text, and that this semantics is powerful enough to aid applications such as question answering and translation reranking.

There is much potential for future extensions to address the limitations of the process described here. As we use a flat clustering of relations, we are only able to model synonyms and not hypernyms. More sophisticated clustering techniques, such as those used by Berant et al. (2011), seem to offer a way to address this. Our system clusters relations with similar named-entity arguments, but this means it does not cluster relations whose arguments are rarely named entities. However, using cross-lingual clusters of common nouns, such as those from Täckström et al. (2012), it should be possible to cluster relations that take semantically similar common noun arguments. Embedding cluster-identifiers in a logical form allows us to also model logical operators, such as negation and quantifiers, which may help to improve the translation of these. It would also be interesting to experiment with more diverse languages types.

### Acknowledgements

## References

Kisuh Ahn, Beatrix Alex, Johan Bos, Tiphaine Dalmas, Jochen L Leidner, and Matthew B Smillie. 2004. Cross-lingual question answering with QED. In *Working Notes, CLEF Cross-Language Evaluation Forum*, pages 335–342.

Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. 2013. Abstract Meaning Representation for sembanking. In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, Sofia, Bulgaria, August. Association for Computational Linguistics.

Colin Bannard and Chris Callison-Burch. 2005. Paraphrasing with bilingual parallel corpora. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*, pages 597–604, Ann Arbor, Michigan, June. Association for Computational Linguistics.

Stephen Beale, Sergei Nirenburg, and Kavi Mahesh. 1995. Semantic analysis in the Mikrokosmos machine translation project. In *Proceedings of the 2nd Symposium on Natural Language Processing*, pages 297–307.

Jonathan Berant, Ido Dagan, and Jacob Goldberger. 2011. Global learning of typed entailment rules. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1*, HLT '11, pages 610–619. Association for Computational Linguistics.

C. Biemann. 2006. Chinese whispers: an efficient graph clustering algorithm and its application to natural language processing problems. In *Proceedings of the First Workshop on Graph Based Methods for Natural Language Processing*, pages 73–80. Association for Computational Linguistics.

Alexandra Birch, Phil Blunsom, and Miles Osborne. 2009. A Quantitative Analysis of Reordering Phenomena. In *Proceedings of the Fourth Workshop on Statistical Machine Translation*, pages 197–205, Athens, Greece, March. Association for Computational Linguistics.

D.M. Blei, A.Y. Ng, and M.I. Jordan. 2003. Latent dirichlet allocation. *the Journal of machine Learning research*, 3:993–1022.

Kurt Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. 2008. Freebase: a collaboratively created graph database for structuring hu-

man knowledge. In *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, SIGMOD '08, pages 1247–1250, New York, NY, USA. ACM.

Peter F. Brown, Vincent J. Della Pietra, Stephen A. Della Pietra, and Robert L. Mercer. 1993. The mathematics of statistical machine translation: parameter estimation. *Comput. Linguist.*, 19(2):263–311, June.

Sabine Buchholz and Erwin Marsi. 2006. Conll-x shared task on multilingual dependency parsing. In *Proceedings of the Tenth Conference on Computational Natural Language Learning*, pages 149–164. Association for Computational Linguistics.

Chris Callison-Burch, Philipp Koehn, and Miles Osborne. 2006. Improved statistical machine translation using paraphrases. In *Proceedings of the Human Language Technology Conference of the NAACL, Main Conference*, pages 17–24, New York City, USA, June. Association for Computational Linguistics.

Marie Candito, Benoît Crabbé, Pascal Denis, et al. 2010. Statistical french dependency parsing: treebank conversion and first results. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC 2010)*, pages 1840–1847.

Stephen Clark and James R. Curran. 2004. Parsing the WSJ using CCG and log-linear models. In *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, ACL '04. Association for Computational Linguistics.

S. Clark, M. Steedman, and J.R. Curran. 2004. Object-extraction and question-parsing using CCG. In *Proceedings of the EMNLP Conference*, pages 111–118.

Pascal Denis, Benoît Sagot, et al. 2009. Coupling an annotated corpus and a morphosyntactic lexicon for state-of-the-art pos tagging with less human effort. In *PACLIC*, pages 110–119.

Bonnie J Dorr, Eduard H Hovy, and Lori S Levin. 2004. Machine translation: Interlingual methods.

Nguyen Tuan Duc, Danushka Bollegala, and Mitsuru Ishizuka. 2011. Cross-language latent relational search: Mapping knowledge across languages. *Association for the Advancement of Artificial Intelligence*, pages 1237–1242.

Anthony Fader, Stephen Soderland, and Oren Etzioni. 2011. Identifying relations for open information extraction. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP '11, pages 1535–1545. Association for Computational Linguistics.

Evgeniy Gabrilovich and Shaul Markovitch. 2007. Computing semantic relatedness using wikipedia-based explicit semantic analysis. In *IJCAI*, volume 7, pages 1606–1611.

Bevan Jones, Jacob Andreas, Daniel Bauer, Karl Moritz Hermann, and Kevin Knight. 2012. Semantics-based machine translation with hyperedge replacement grammars. *Proc. COLING, 2012*.

Alexandre Klementiev, Ann Irvine, Chris Callison-Burch, and David Yarowsky. 2012a. Toward statistical machine translation without parallel corpora. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, EACL '12, pages 130–140. Association for Computational Linguistics.

Alexandre Klementiev, Ivan Titov, and Binod Bhattarai. 2012b. Inducing crosslingual distributed representations of words. In *Proceedings of the International Conference on Computational Linguistics (COLING)*, Bombay, India, December.

Philipp Koehn and Christof Monz. 2006. Manual and automatic evaluation of machine translation between European languages. In *Proceedings on the Workshop on Statistical Machine Translation*, pages 102–121, New York City, June. Association for Computational Linguistics.

Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondřej Bojar, Alexandra Constantin, and Evan Herbst. 2007. Moses: open source toolkit for statistical machine translation. In *Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions*, ACL '07, pages 177–180. Association for Computational Linguistics.

Philipp Koehn. 2005. Europarl: A parallel corpus for statistical machine translation. In *MT summit*, volume 5.

Lili Kotlerman, Ido Dagan, Idan Szpektor, and Maayan Zhitomirsky-geffet. 2010. Directional distributional similarity for lexical inference. *Nat. Lang. Eng.*, 16(4):359–389, October.

Mikhail Kozhevnikov and Ivan Titov. 2013. Crosslingual transfer of semantic role models. In *To Appear in Proceedings of the 51th Annual Meeting of the Association for Computational Linguistics*, Sofia, Bulgaria, August. Association for Computational Linguistics.

Mike Lewis and Mark Steedman. 2013. Combined Distributional and Logical Semantics. *Transactions of the Association for Computational Linguistics*, 1:179–192.

Dekang Lin and Patrick Pantel. 2001. DIRT - Discovery of Inference Rules from Text. In *In Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 323–328.

Xiao Ling and Daniel S Weld. 2012. Fine-grained entity recognition. In *Proceedings of the 26th Conference on Artificial Intelligence (AAAI)*.

691

Paul McNamee, James Mayfield, Dawn Lawrie, Douglas W Oard, and David Doermann. 2011. Cross-language entity linking. *Proc. IJCNLP2011*.

David Mimno, Hanna M Wallach, Jason Naradowsky, David A Smith, and Andrew McCallum. 2009. Polylingual topic models. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 2-Volume 2*, pages 880–889. Association for Computational Linguistics.

M. Mintz, S. Bills, R. Snow, and D. Jurafsky. 2009. Distant supervision for relation extraction without labeled data. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2*, pages 1003–1011. Association for Computational Linguistics.

Teruko Mitamura, Eric H Nyberg, and Jaime G Carbonell. 1991. An efficient interlingua translation system for multi-lingual document production.

Joakim Nivre, Johan Hall, Jens Nilsson, Atanas Chanev, Gülsen Eryigit, Sandra Kübler, Svetoslav Marinov, and Erwin Marsi. 2007. Maltparser: A language-independent system for data-driven dependency parsing. *Natural Language Engineering*, 13(2):95–135.

Hoifung Poon and Pedro Domingos. 2009. Unsupervised semantic parsing. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1 - Volume 1*, EMNLP '09, pages 1–10. Association for Computational Linguistics.

Reinhard Rapp. 1999. Automatic identification of word translations from unrelated english and german corpora. In *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*, ACL '99, pages 519–526. Association for Computational Linguistics.

Philip Resnik and Noah A Smith. 2003. The web as a parallel corpus. *Computational Linguistics*, 29(3):349–380.

Stefan Schoenmackers, Oren Etzioni, Daniel S. Weld, and Jesse Davis. 2010. Learning first-order horn clauses from web text. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, EMNLP '10, pages 1088–1098. Association for Computational Linguistics.

Mark Steedman. 2012. *Taking Scope: The Natural Semantics of Quantifiers*. MIT Press.

Oscar Täckström, Ryan McDonald, and Jakob Uszkoreit. 2012. Cross-lingual word clusters for direct transfer of linguistic structure. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, NAACL HLT '12, pages 477–487. Association for Computational Linguistics.

Peter D. Turney. 2005. Measuring semantic similarity by latent relational analysis. In *Proceedings of the 19th international joint conference on Artificial intelligence*, IJCAI'05, pages 1136–1141, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Bernard Vauquois. 1968. A survey of formal grammars and algorithms for recognition and transformation in machine translation. In *IFIP Congress*, volume 68, pages 254–260.

Hila Weisman, Jonathan Berant, Idan Szpektor, and Ido Dagan. 2012. Learning verb inference rules from linguistically-motivated evidence. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, EMNLP-CoNLL '12, pages 194–204. Association for Computational Linguistics.

Limin Yao, Aria Haghighi, Sebastian Riedel, and Andrew McCallum. 2011. Structured relation discovery using generative models. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP '11, pages 1456–1466. Association for Computational Linguistics.

Limin Yao, Sebastian Riedel, and Andrew McCallum. 2012. Unsupervised relation discovery with sense disambiguation. In *ACL (1)*, pages 712–720.