Cross-Lingual Induction and Transfer of Verb Classes Based on Word Vector Space Specialisation

Ivan Vulić1, Nikola Mrkšić2 and Anna Korhonen1

1 Language Technology Lab, University of Cambridge, UK
2 Dialogue Systems Group, University of Cambridge, UK
{iv250,nm480,alk23}@cam.ac.uk

Abstract

Existing approaches to automatic VerbNet-style verb classification are heavily dependent on feature engineering and therefore limited to languages with mature NLP pipelines. In this work, we propose a novel cross-lingual transfer method for inducing VerbNets for multiple languages. To the best of our knowledge, this is the first study which demonstrates how the architectures for learning word embeddings can be applied to this challenging syntactic-semantic task. Our method uses cross-lingual translation pairs to tie each of the six target languages into a bilingual vector space with English, jointly specialising the representations to encode the relational information from English VerbNet. A standard clustering algorithm is then run on top of the VerbNet-specialised representations, using vector dimensions as features for learning verb classes. Our results show that the proposed cross-lingual transfer approach sets new state-of-the-art verb classification performance across all six target languages explored in this work.

1 Introduction

Playing a key role in conveying the meaning of a sentence, verbs are famously complex. They display a wide range of syntactic-semantic behaviour, expressing the semantics of an event as well as relational information among its participants (Jackendoff, 1972; Gruber, 1976; Levin, 1993, inter alia).

Lexical resources which capture the variability of verbs are instrumental for many Natural Language Processing (NLP) applications. One of the richest verb resources currently available for English is VerbNet (Kipper et al., 2000; Kipper, 2005). Based on the work of Levin (1993), this largely hand-crafted taxonomy organises verbs into classes on the basis of their shared syntactic-semantic behaviour. Providing a useful level of generalisation for many NLP tasks, VerbNet has been used to support semantic role labelling (Swier and Stevenson, 2004; Giuglea and Moschitti, 2006), semantic parsing (Shi and Mihalcea, 2005), word sense disambiguation (Brown et al., 2011), discourse parsing (Subba and Di Eugenio, 2009), information extraction (Mausam et al., 2012), text mining applications (Lippincott et al., 2013; Rimell et al., 2013), research into human language acquisition (Korhonen, 2010), and other tasks.

This benefit for English NLP has motivated the development of VerbNets for languages such as Spanish and Catalan (Aparicio et al., 2008), Czech (Pala and Horák, 2008), and Mandarin (Liu and Chiang, 2008). However, end-to-end manual resource development using Levin’s methodology is extremely time consuming, even when supported by translations of English VerbNet classes to other languages (Sun et al., 2010; Scarton et al., 2014). Approaches which aim to learn verb classes automatically offer an attractive alternative. However, existing methods rely on carefully engineered features that are extracted using sophisticated language-specific resources (Joanis et al., 2008; Sun et al., 2010; Falk et al., 2012, i.a.), ranging from accurate parsers to pre-compiled subcategorisation frames (Schulte im Walde, 2006; Li and Brew, 2008; Messiant, 2008). Such methods are limited to a small set of resource-rich languages.

It has been argued that VerbNet-style classification has a strong cross-lingual element (Jackendoff, 1992; Levin, 1993). In support of this argument, Majewska et al. (2017) have shown that English VerbNet has high translatability across different,
even typologically diverse languages. Based on this finding, we propose an automatic approach which exploits readily available annotations for English to facilitate efficient, large-scale development of VerbNets for a wide set of target languages.

Recently, unsupervised methods for inducing distributed word vector space representations or word embeddings (Mikolov et al., 2013a) have been successfully applied to a plethora of NLP tasks (Turian et al., 2010; Collobert et al., 2011; Baroni et al., 2014, i.a.). These methods offer an elegant way to learn directly from large corpora, bypassing the feature engineering step and the dependence on mature NLP pipelines (e.g., POS taggers, parsers, extraction of subcategorisation frames). In this work, we demonstrate how these models can be used to support automatic verb class induction. Moreover, we show that these models offer the means to exploit inherent cross-lingual links in VerbNet-style classification in order to guide the development of new classifications for resource-lean languages. To the best of our knowledge, this proposition has not been investigated in previous work.

There has been little work on assessing the suitability of embeddings for capturing rich syntactic-semantic phenomena. One challenge is their reliance on the distributional hypothesis (Harris, 1954), which coalesces fine-grained syntactic-semantic relations between words into a broad relation of semantic relatedness (e.g., coffee:cup) (Hill et al., 2015; Kiela et al., 2015). This property has an adverse effect when word embeddings are used in downstream tasks such as spoken language understanding (Kim et al., 2016a,b) or dialogue state tracking (Mrkšić et al., 2016, 2017a). It could have a similar effect on verb classification, which relies on the similarity in syntactic-semantic properties of verbs within a class. In summary, we explore three important questions in this paper:

(Q1) Given their fundamental dependence on the distributional hypothesis, to what extent can unsupervised methods for inducing vector spaces facilitate the automatic induction of VerbNet-style verb classes across different languages?

(Q2) Can one boost verb classification for low-resource languages by exploiting general-purpose cross-lingual resources such as BabelNet (Navigli and Ponzetto, 2012; Ehrmann et al., 2014) or bilingual dictionaries such as PanLex (Kamholz et al., 2014) to construct better word vector spaces for these languages?

(Q3) Based on the stipulated cross-linguistic validity of VerbNet-style classification, can one exploit rich sets of readily available annotations in one language (e.g., the full English VerbNet) to automatically bootstrap the creation of VerbNets for other languages? In other words, is it possible to exploit a cross-lingual vector space to transfer VerbNet knowledge from a resource-rich to a resource-lean language?

To investigate Q1, we induce standard distributional vector spaces (Mikolov et al., 2013b; Levy and Goldberg, 2014) from large monolingual corpora in English and six target languages. As expected, the results obtained with this straightforward approach show positive trends, but at the same time reveal its limitations for all the languages involved. Therefore, the focus of our work shifts to Q2 and Q3. The problem of inducing VerbNet-oriented embeddings is framed as vector space specialisation using the available external resources: BabelNet or PanLex, and (English) VerbNet. Formalised as an instance of post-processing semantic specialisation approaches (Faruqui et al., 2015; Mrkšić et al., 2016), our procedure is steered by two sets of linguistic constraints: 1) cross-lingual (translation) links between languages extracted from BabelNet (targeting Q2); and 2) the available VerbNet annotations for a resource-rich language. The two sets of constraints jointly target Q3.

The main goal of vector space specialisation is to pull examples standing in desirable relations, as described by the constraints, closer together in the transformed vector space. The specialisation process can capitalise on the knowledge of VerbNet relations in the source language (English) by using translation pairs to transfer that knowledge to each of the target languages. By constructing shared bilingual vector spaces, our method facilitates the transfer of semantic relations derived from VerbNet to the vector spaces of resource-lean target languages. This idea is illustrated by Fig. 1.

Our results indicate that cross-lingual connections yield improved verb classes across all six target languages (thus answering Q2). Moreover, a consistent and significant boost in verb classification performance is achieved by propagating the VerbNet-style information from the source language (English) to any other target language (e.g., Italian, Croatian, Polish, Finnish) for which no VerbNet-style information is available during the
fine-tuning process (thus answering Q3). We report state-of-the-art verb classification performance for all six languages in our experiments. For instance, we improve the state-of-the-art F-1 score from prior work from 0.55 to 0.79 for French, and from 0.43 to 0.74 for Brazilian Portuguese.

2 Methodology: Specialising for VerbNet

Motivation: Verb Classes and VerbNet VerbNet is a hierarchical, domain-independent, broad-coverage verb lexicon based on Levin’s classification and taxonomy of English (EN) verbs (Levin, 1993; Kipper, 2005). Verbs are grouped into classes (e.g. the class PUT-9.1 for verbs such as place, position, insert, and arrange) based on their shared meaning components and syntactic behaviour, defined in terms of their participation in diathesis alternations, i.e., alternating verb frames that are related with the same or similar meaning. VerbNet extends and refines Levin’s classification, providing more fine-grained syntactic and semantic information for individual classes. Each VerbNet class is characterised by its member verbs, syntactic frames, semantic predicates and typical verb arguments. The current version of VerbNet (v3.2) contains 8,537 distinct English verbs grouped into 273 VerbNet main classes.

The inter-relatedness of syntactic behaviour and meaning of verbs is not limited to English (Levin, 1993). The basic meaning components underlying verb classes are said to be cross-linguistically valid (Jackendoff, 1992; Merlo et al., 2002) and therefore the classification has a strong cross-lingual dimension. A recent investigation of Majewska et al. (2017) show that it is possible to manually translate VerbNet classes and class members to different, typologically diverse languages with high accuracy.

The practical usefulness of VerbNet style classification both within and across languages has been limited by the fact that few languages boast resources similar to the English VerbNet. Some VerbNets have been developed completely manually from scratch, aiming to capture properties specific to the language in question, e.g., the resources for Spanish and Catalan (Aparicio et al., 2008, 3). VerbNet is further accentuated by available mappings (Loper et al., 2007) to a number of other verb resources such as WordNet (Fellbaum, 1998), FrameNet (Baker et al., 1998), and PropBank (Palmer et al., 2005).

3The usefulness of VerbNet is further accentuated by available mappings (Loper et al., 2007) to a number of other verb resources such as WordNet (Fellbaum, 1998), FrameNet (Baker et al., 1998), and PropBank (Palmer et al., 2005).

4For example, Levin (1993) notes that verbs in Warlpiri manifest analogous behavior to English with respect to the conative alternation. In another example, Polish verbs have the same patterns as EN verbs in terms of the middle construction.
Czech (Pala and Horák, 2008), and Mandarin (Liu and Chiang, 2008). Other VerbNets were created semi-automatically, with the help of other lexical resources, e.g., for French (Pradet et al., 2014) and Brazilian Portuguese (Scarton and Aluísio, 2012). These approaches involved substantial amounts of specialised linguistic and translation work. Finally, automatic methods have been developed, e.g., for French (Sun et al., 2010; Falk et al., 2012) and Brazilian Portuguese (Scarton et al., 2014), with insufficient accuracy (as emphasised in Sect. 4). Until now, work in this area has been limited to a small number of languages, due to the large requirements in terms of human input and/or the availability of mature NLP pipelines which exist only for a few resource-rich languages (e.g., English, German).

In this work, we propose a novel, fully automated approach for inducing VerbNets for multiple languages - one based on cross-lingual transfer. Unlike earlier approaches, our method does not require any parsed data or manual annotations for the target language. It encodes the cross-linguistic validity of Levin-style verb classifications into the vector-space specialisation framework (Sect. 2.1) driven by linguistic constraints. A standard clustering algorithm is then run on top of the VerbNet-specialised representations using vector dimensions as features to learn verb clusters (Sect. 2.2). Our approach attains state-of-the-art verb classification performance across all six target languages.

2.1 Vector Space Specialisation

Specialisation Model Our departure point is a state-of-the-art specialisation model for fine-tuning vector spaces termed PARAGRAM (Wieting et al., 2015). The PARAGRAM procedure injects similarity constraints between word pairs in order to make their vector space representations more similar; we term these the ATTRACT constraints. Let \( V = V_s \cup V_t \) be the vocabulary consisting of the source language and target language vocabularies \( V_s \) and \( V_t \), respectively. Let \( C \) be the set of word pairs standing in desirable lexical relations; these include: 1) verb pairs from the same VerbNet class (e.g. \((en\_transport, en\_transfer)\) from verb class \textsc{send-11.1}); and 2) the cross-lingual synonymy pairs (e.g. \((en\_peace, fi\_rauha)\)). Given the initial distributional space and collections of such ATTRACT pairs \( C \), the model gradually modifies the space to bring the designated word vectors closer together, working in mini-batches of size \( k \). The method’s cost function can be expressed as:

\[
O(B_C) = O_C(B_C) + R(B_C)
\]

The first term of the method’s cost function (i.e., \( O_C \)) pulls the ATTRACT examples \((x_l, x_r) \in C\) closer together (see Fig. 1 for an illustration). \( B_C \) refers to the current mini-batch of ATTRACT constraints. This term is expressed as follows:

\[
O_C(B_C) = \sum_{(x_l, x_r) \in B_C} \left[ \tau (\delta_{att} + x_l t_i - x_r s_i) + \tau (\delta_{att} + x_r t_i - x_l s_i) \right] \tag{2}
\]

\( \tau(x) = \max(0, x) \) is the standard rectified linear unit or the hinge loss function (Tsochantaridis et al., 2004; Nair and Hinton, 2010). \( \delta_{att} \) is the “attract” margin: it determines how much vectors of words from ATTRACT constraints should be closer to each other than to their negative examples. The negative example \( t_i \) for each word \( s_i \) in any ATTRACT pair is always the vector closest to \( x_l \) taken from the pairs in the current mini-batch, distinct from the other word paired with \( x_l \), and \( x_r \) itself.\(^6\)

The second \( R(B_C) \) term is the regularisation which aims to retain the semantic information encoded in the initial distributional space as long as this information does not contradict the used ATTRACT constraints. Let \( x_i^{\text{init}} \) refer to the initial distributional vector of the word \( x_i \) and let \( V(B_C) \) be the set of all word vectors present in the given mini-batch. If \( \lambda_{reg} \) denotes the L2 regularisation constant, this term can be expressed as:

\[
R(B_C) = \sum_{x_i \in V(B_C)} \lambda_{reg} \left\| x_i^{\text{init}} - x_i \right\|_2 \tag{3}
\]

Linguistic Constraints: Transferring VerbNet-Style Knowledge The fine-tuning procedure effectively blends the knowledge from external resources (i.e., the input ATTRACT set of constraints) with distributional information extracted directly from large corpora. We show how to propagate annotations from a knowledge source such as VerbNet from source to target by combining two types of constraints within the specialisation framework:

\(^5\)The original PARAGRAM model as well as other fine-tuning models from prior work inject pairwise linguistic constraints into existing vector spaces in order to improve their ability to capture semantic similarity/paraphrasability. In this work, we demonstrate that the same generic specialisation framework can be used to transform vector spaces to capture other types of lexical relations such as VerbNet relations.

\(^6\)Effectively, this term forces word pairs from the in-batch ATTRACT constraints to be closer to one another than to any other word in the current mini-batch.
### 2.2 Clustering Algorithm

Given the initial distributional or specialised collection of target language vectors $V_t$, we apply an off-the-shelf clustering algorithm on top of these vectors in order to group verbs into classes. Following prior work (Brew and Schulte im Walde, 2002; Sun and Korhonen, 2009; Sun et al., 2010), we employ the MNCut spectral clustering algorithm (Meila and Shi, 2001), which has wide applicability in similar NLP tasks which involve high-dimensional feature spaces (Chen et al., 2006; von Luxburg, 2007; Scarton et al., 2014, i.a.). Again, following prior work (Sun et al., 2010, 2013), we estimate the number of clusters $K_{\text{Clust}}$ using the self-tuning method of Zelnik-Manor and Perona (2004). This algorithm finds the optimal number by minimising a cost function based on the eigenvector structure of the word similarity matrix. We refer the reader to the relevant literature for further details.

### 3 Experimental Setup

#### Languages

We experiment with six target languages: French (FR), Brazilian Portuguese (PT), Italian (IT), Polish (PL), Croatian (HR), and Finnish (FI). All statistics regarding the source and size of training and test data, and linguistic constraints for each target language are summarised in Tab. 2.

Automatic approaches to verb class induction have been tried out in prior work for FR and PT. To the best of our knowledge, our cross-lingual study is the first aiming to generalise an automatic induction method to more languages using an underlying methodology which is language-pair independent.

#### Initial Vector Space: Training Data and Setup

All target language vectors were trained on large monolingual running text using the same setup: 300-dimensional word vectors, the frequency cut-off set to 100, bag-of-words (BOW) contexts, and the window size of 2 (Levy and Goldberg, 2014; Schwartz et al., 2016). All tokens were lowercased, and all numbers were converted to a placeholder symbol $<$NUM$>$.

---

Table 1: Example pairwise ATTRACT constraints extracted from three VerbNet classes in English.

<table>
<thead>
<tr>
<th>LEARN-14</th>
<th>BREAK-45.1</th>
<th>ACCEPT-77</th>
</tr>
</thead>
<tbody>
<tr>
<td>(learn, study)</td>
<td>(break, dissolve)</td>
<td>(accept, understand)</td>
</tr>
<tr>
<td>(learn, relearn)</td>
<td>(crash, crush)</td>
<td>(accept, reject)</td>
</tr>
<tr>
<td>(read, study)</td>
<td>(shatter, split)</td>
<td>(reject, rue)</td>
</tr>
<tr>
<td>(read, relearn)</td>
<td>(break, rip)</td>
<td>(reject, discourage)</td>
</tr>
<tr>
<td>(cram, relearn)</td>
<td>(crack, smash)</td>
<td>(encourage, discourage)</td>
</tr>
<tr>
<td>(learn, relearn)</td>
<td>(shred, splinter)</td>
<td>(reject, discourage)</td>
</tr>
<tr>
<td>(read, relearn)</td>
<td>(snap, tear)</td>
<td>(disprefer, understand)</td>
</tr>
</tbody>
</table>

#### Similar to BabelNet, the translations in PanLex were derived from various sources such as glossaries, dictionaries, and automatic inference from other languages. This results in a high-coverage lexicon containing a certain amount of noise.

---

* Other SGNS parameters were also set to standard values (Baroni et al., 2014; Vulić and Korhonen, 2016): 15 epochs, 15 negative samples, global learning rate: $0.2$, subsampling rate: $1e^{-4}$. Similar trends in results persist with $d = 100, 500$. 

---

* Similar to BabelNet, the translations in PanLex were derived from various sources such as glossaries, dictionaries, and automatic inference from other languages. This results in a high-coverage lexicon containing a certain amount of noise.
Training Corpus frWaC brWaC itWaC Araneum hrWaC fiWaC  
Corpus size (in tokens) 1.6B 2.7B 2B 1.2B 1.4B 1.7B  
Vocabulary size 242,611 257,310 178,965 373,882 396,103 448,863  
Constraints 
# monolingual VerbNet-EN 220,052 220,052 220,052 220,052 220,052 220,052  
# mono TARGET (MONO-SYN) 428,329 292,937 362,452 423,711 209,626 377,548  
# cross-ling EN-TARGET (BNNet) 310,410 245,354 258,102 219,614 160,963 284,167  
# cross-ling EN-TARGET (PLex) 225,819 187,386 216,574 154,159 201,329 257,106  
Test 
# Verbs (# Classes) 169 (16) 660 (17) 177 (17) 258 (17) 277 (17) 201 (17)  
Coverage of test instances 94.1% 95.5% 96.6% 93.4% 98.2% 84.6%  

Table 2: Statistics of the experimental setup for each target language: training/test data and constraints. Coverage refers to the percentage of test verbs represented in the target language vocabularies.

vectors on ptWaC (Wagner Filho et al., 2016), fi vectors on fiWaC (Ljubešić et al., 2016), and pl vectors on the Araneum Polonicum Maius Web corpus (Benko, 2014). Note that we do not utilise any VerbNet-specific knowledge in the target language to induce and further specialise these word vectors.

Source EN vectors were taken directly from the work of Levy and Goldberg (2014): they are trained with SGNS on the cleaned and tokenised Polyglot Wikipedia (Al-Rfou et al., 2013) containing ~75M sentences, ~1.7B word tokens and a vocabulary of ~180k words after lowercasing and frequency cut-off. To measure the importance of the starting source language space as well as to test if syntactic knowledge on the source side may be propagated to the target space, we test two variant EN vector spaces: SGNS with (a) BOW contexts and the window size 2 (SGNS-BOW2); and (b) dependency-based contexts (SGNS-DEPS) (Padó and Lapata, 2007; Levy and Goldberg, 2014).

Linguistic Constraints We experiment with the following constraint types: (a) monolingual synonymy constraints in each target language extracted from BabelNet (Mono-Syn); (b) cross-lingual EN-TARGET constraints from BabelNet; (c) cross-lingual EN-TARGET constraints plus EN VerbNet constraints (see Sect. 2.1 and Fig. 1). Unless stated otherwise, we use BabelNet as the default source of cross-lingual constraints for (b) and (c).

Vector Space Specialisation The PARAGRAM model’s parameters are adopted directly from prior work (Wieting et al., 2015) without any additional fine-tuning: $\delta_{att} = 0.6$, $\lambda_{reg} = 10^{-9}$, $k = 50$. We train for 5 epochs without early stopping using AdaGrad (Duchi et al., 2011). PARAGRAM is in fact a special case of the more general ATTRACT-REPEL specialisation framework (Mrkšić et al., 2017b); we use this more recent and more efficient TensorFlow implementation of the model in all experiments.9

Test Data The development of an automatic verb classification approach requires an initial gold standard (Sun et al., 2010): these have been developed for FR (Sun et al., 2010), PT (Scarton et al., 2014), IT, PL, HR, and FI (Majewska et al., 2017). They were created using the methodology of Sun et al. (2010), based on the EN gold standard of Sun et al. (2008) which contains 17 fine-grained Levin classes with 12 member verbs each. For instance, the class PUT-9.1 in French contains verbs such as accrocher, déposer, mettre, répartir, réintégrer, etc.

Evaluation Measures We use standard evaluation measures from prior work on verb clustering (Ó Séaghdha and Copestake, 2008; Sun and Korhonen, 2009; Sun et al., 2010; Falk et al., 2012, i.a.). The mean precision of induced verb clusters labelled modified purity (MPUR) is computed as:

$$
MPUR = \frac{\sum_{C \in \text{Clust}} n_{\text{prev}(C)} > 1 \cdot n_{\text{prev}(C)}}{\# \text{test verbs}}
$$

Here, each cluster $C$ from the set of all $K_{\text{Clust}}$ induced clusters Clust is associated with its prevalent class/cluster from the gold standard, and the number of verbs in an induced cluster $C$ taking this prevalent class is labelled $n_{\text{prev}(C)}$. All other verbs not taking the prevalent class are considered errors.10 #test_verbs denotes the total number of test verb instances. The second measure targeting recall is weighted class accuracy (WACC), computed as:

$$
WACC = \frac{\sum_{C \in \text{Gold}} n_{\text{dom}(C)}}{\# \text{test verbs}}
$$

9https://github.com/nmrksic/attract-repel

10Clusters with $n_{\text{prev}(C)} = 1$ are discarded from the count to avoid an undesired bias towards singleton clusters (Sun and Korhonen, 2009; Sun et al., 2010).
We can draw several interesting conclusions. First, XLing uses cross-lingual constraints which balances between homogeneity (precision) and completeness (recall). We have also experimented with V-measure (Rosenberg and Hirschberg, 2007), another standard evaluation measure which balances between homogeneity (precision) and completeness (recall); we do not report these scores for brevity as similar trends in results are observed.

For each cluster \( C \) from the set of gold standard clusters \( \text{Gold} \), we have to find the dominant cluster from the set of induced clusters: this cluster has the most verbs in common with the gold cluster \( C \), and that number is \( n_{\text{dom}}(C) \). As measures of precision and recall, MPUR and wACC may be combined into an F-1 score, computed as the balanced harmonic mean, which we report in this work.\(^{11}\)

## 4 Results and Discussion

**Cross-Lingual Transfer Model** F-1 verb classification scores for the six target languages with different sets of constraints are summarised in Fig. 2. We can draw several interesting conclusions. First, the strongest results on average are obtained with the model which transfers the VerbNet knowledge from English (as a resource-rich language) to the resource-lean target language (providing an answer to question Q3, Sect. 1). These improvements are visible across all target languages, empirically demonstrating the cross-lingual nature of VerbNet-style classifications. Second, using cross-lingual constraints alone (\(XLing\)) yields strong gains over initial distributional spaces (answering Q1 and Q2). Fig. 2 also shows that cross-lingual similarity constraints are more beneficial than the monolingual ones, despite a larger total number of the monolingual constraints in each language (see Tab. 2). This suggests that such cross-lingual similarity links are strong implicit indicators of class membership. Namely, target language words which map to the same source language word are likely to be synonyms and consequently end up in the same verb class in the target language. However, the cross-lingual links are even more useful as means for transferring the VerbNet knowledge, as evidenced by additional gains with \(XLing+VerbNet-EN\).

The absolute classification scores are the lowest for the two Slavic languages: PL and HR. This may be partially explained by the lowest number of cross-lingual constraints for the two languages covering only a subset of their entire vocabularies (see Tab. 2 and compare the total number of constraints for HR and PL to the numbers for e.g. FI or FR). Another reason for weaker performance of these two languages could be their rich morphology, which induces data sparsity both in the initial vector space estimation and in the coverage of constraints.

**State-of-the-Art** A direct comparison of previous state-of-the-art classification scores available for FR (Sun et al., 2010) and PT (Scarton et al., 2014) on the same test data exemplifies the extent of improvement achieved by our transfer model. F-1 scores improve from 0.55 to 0.75 for FR and from 0.43 to 0.73 for PT. Scarton et al. (2014) explain the low performance by “the lower quality NLP tools”. This issue is largely mitigated by our VerbNet transfer model, which exploits the assump-

\(^{11}\)We have also experimented with V-measure (Rosenberg and Hirschberg, 2007), another standard evaluation measure which balances between homogeneity (precision) and completeness (recall); we do not report these scores for brevity as similar trends in results are observed.
Table 3: The effect of multilingual vector space specialisation. Results are reported for FR and IT using: a) cross-lingual constraints only (XLing); and b) the VerbNet transfer model (XLing+VN).

<table>
<thead>
<tr>
<th>Vector space</th>
<th>French</th>
<th>Italian</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>XLing</td>
<td>XLing+VN</td>
</tr>
<tr>
<td>EN+FR</td>
<td>0.698</td>
<td>0.742</td>
</tr>
<tr>
<td>EN+IT</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>EN+FR+IT</td>
<td>0.728</td>
<td>0.745</td>
</tr>
<tr>
<td>EN+FR+PL</td>
<td>0.697</td>
<td>0.717</td>
</tr>
<tr>
<td>EN+IT+FR+IT</td>
<td>0.719</td>
<td>0.751</td>
</tr>
<tr>
<td>EN+IT+FR+PT</td>
<td>0.710</td>
<td>0.718</td>
</tr>
</tbody>
</table>

Table 4: Comparison of verb classification (VC) and verb semantic similarity (Sim) for English. VC is measured on the EN test set of Sun et al. (2008). Sim is measured on SimVerb-3500 (Gerz et al., 2016). The scores are Spearman’s ρ correlation scores. EN-Dist is the initial distributional English vector space: SGNS-BOW2; EN-VN is the same space transformed using monolingual EN VerbNet constraints only, an upper bound for the specialisation-based approach in EN.

<table>
<thead>
<tr>
<th>VC:XLing</th>
<th>Sim:XLing</th>
<th>Sim:XLing+VN</th>
</tr>
</thead>
<tbody>
<tr>
<td>EN-Dist</td>
<td>0.484</td>
<td>0.275</td>
</tr>
<tr>
<td>+FR</td>
<td>0.608</td>
<td>0.556</td>
</tr>
<tr>
<td>+PT</td>
<td>0.633</td>
<td>0.537</td>
</tr>
<tr>
<td>+IT</td>
<td>0.602</td>
<td>0.524</td>
</tr>
<tr>
<td>+PL</td>
<td>0.597</td>
<td>0.469</td>
</tr>
<tr>
<td>+HR</td>
<td>0.582</td>
<td>0.497</td>
</tr>
<tr>
<td>+FI</td>
<td>0.662</td>
<td>0.598</td>
</tr>
<tr>
<td>+FR+IT</td>
<td>0.633</td>
<td>0.571</td>
</tr>
<tr>
<td>+HR+FR+IT</td>
<td>0.641</td>
<td>0.635</td>
</tr>
<tr>
<td>+HR+IT+PT</td>
<td>0.674</td>
<td>0.596</td>
</tr>
<tr>
<td>EN-VN</td>
<td><strong>0.956</strong></td>
<td>–</td>
</tr>
</tbody>
</table>

**Starting Source Vector Space** Fig. 2a and Fig. 2b enable a brief analysis of the influence of the starting EN vector space on the results for each target language. We observe small but consistent gains with SGNS-DEPS, which utilises syntactic information stemming from a dependency parser on the source side, over SGNS-BOW for the XLing variant. The improvements are +2.1 points on average, visible for 5 out of 6 target languages. We again see an increase in performance with the XLing+VerbNet model, but we do not observe any major difference between the two starting source spaces now: average slight score difference of 0.3 is in favour of SGNS-BOW2, which outperforms SGNS-DEPS for 3 out of 6 target languages. This finding indicates that VerbNet-based linguistic constraints are more important for the final classification performance, and mitigate the artefacts of the starting distributional source space.

**Bilingual vs. Multilingual** The transfer model can operate with more than two languages, effectively inducing a multilingual vector space. We analyse such multilingual training based on the results on FR and IT (Tab. 3). On average, the results with XLing improve with more languages (see also the results for EN in Tab. 4), as the model relies on more constraints for the vector space specialisation. Yet additional languages do not lead to clear improvements with XLing+VerbNet: we hypothesise that the specialisation procedure becomes dominated by cross-lingual constraints which may diminish the importance of VerbNet-based EN constraints. The language configuration in the multilingual vector space also makes a difference: e.g., combining PL with the Romance languages degrades the performance, while FI surprisingly boosts it slightly. For brevity, we only report the results for FR and IT. Similar trends are observed when making the transition from bilingual to multilingual vector spaces for other target languages.

**Clustering Algorithm** Since vector space specialisation is detached from the application of the clustering algorithm, our framework allows straightforward experimentation with other algorithms. Following prior work (Brew and Schulte im Walde, 2002; Sun et al., 2010), we also test K-means clustering. Results for the six languages using the EN SGNS-BOW2 source space and XLing+VerbNet-EN are on average 3.8 points lower than the ones reported in Fig. 2a. K-Means is outperformed for each target language, confirming the superiority of spectral clustering established in prior work, e.g., (Scarton et al., 2014). On the other hand, we find results with another clustering algorithm, hierarchical agglomerative clustering with Ward’s linkage (Ward, 1963), on par with spectral clustering (1.4 points on average in favour of spectral, which is better on 4 out of 6 languages). We believe that further gains in verb class induction could be achieved by additional fine-tuning of the clustering algorithm.

**Other Cross-Lingual Sources** Replacing BabelNet with PanLex as the alternative source of cross-
lingual information again leads to large gains with the cross-lingual transfer model, as is evident from Fig. 3. This suggests that the proposed approach does not depend on a particular source of information - it can be used with any general-purpose bilingual dictionary. We mark slight improvements for 3/6 target languages when comparing the results with the ones from Fig. 2a. The new state-of-the-art F-1 scores are 0.79 for FR and 0.74 for PT.

**Verb Classification vs. Semantic Similarity**

An interesting question originating from prior work on verb representation learning, e.g., (Baker et al., 2014) touches upon the correlation between verb classification and semantic similarity. Due to the availability of VerbNet constraints and a recent similarity evaluation set (SimVerb-3500; it contains human similarity ratings for 3,500 verb pairs) (Gerz et al., 2016), we perform the analysis on English: the results are summarised in Tab. 4. They clearly indicate that cross-lingual synonymy constraints are useful for both relationship types (compare the scores with XLing), with strong gains over the non-specialised distributional space. However, the inclusion of VerbNet information, while boosting classification scores for target languages and (trivially) for EN, deteriorates EN similarity scores across the board (compare XLing+VN against XLing in Tab. 4). This suggests that the VerbNet-style class membership is definitely not equivalent to pure semantic similarity captured by SimVerb.

## 4.1 Further Discussion and Future Work

This work has proven the potential of transferring lexical resources from resource-rich to resource-poor languages using general-purpose cross-lingual dictionaries and bilingual vector spaces as means of transfer within a semantic specialisation framework. However, we believe that the proposed basic framework may be upgraded and extended across several research paths in future work.

First, in the current work we have operated with standard single-sense/single-prototype representations, thus effectively disregarding the problem of verb polysemy. While several polysemy-aware verb classification models for English were developed recently (Kawahara et al., 2014; Peterson et al., 2016), the current lack of polysemy-aware evaluation sets in other languages impedes this line of research. Evaluation issues aside, one idea for future work is to use the ATTRACT-REPEL specialisation framework for sense-aware cross-lingual transfer relying on recently developed multi-sense/prototype word representations (Neelakantan et al., 2014; Pilehvar and Collier, 2016, inter alia). Another challenge is to apply the idea from this work to enable cross-lingual transfer of other structured lexical resources available in English such as FrameNet (Baker et al., 1998), PropBank (Palmer et al., 2005), and VerbKB (Wijaya and Mitchell, 2016). Other potential research avenues include porting the approach to other typologically diverse languages and truly low-resource settings (e.g., with only limited amounts of parallel data), as well as experiments with other distributional spaces, e.g. (Melamud et al., 2016). Further refinements of the specialisation and clustering algorithms may also result in improved verb class induction.

## 5 Conclusion

We have presented a novel cross-lingual transfer model which enables the automatic induction of VerbNet-style verb classifications across multiple languages. The transfer is based on a word vector space specialisation framework, utilised to directly model the assumption of cross-linguistic validity of VerbNet-style classifications. Our results indicate strong improvements in verb classification accuracy across all six target languages explored. All automatically induced VerbNets are available at: [github.com/cambridgeltl/verbnets](https://github.com/cambridgeltl/verbnets).

## Acknowledgments

This work is supported by the ERC Consolidator Grant LEXICAL: Lexical Acquisition Across Languages (no 648909). The authors are grateful to the entire LEXICAL team, especially to Roi Reichart, and also to the three anonymous reviewers for their helpful and constructive suggestions.
References


Nikola Ljubešić, Tommi Pirinen, and Antonio Toral. 2016. Finnish Web corpus fiWaC 1.0. Slovenian language resource repository CLARIN.SI.


Nikola Mrkšić, Ivan Vulić, Diarmuid Ó Séaghdha, Ira Leviant, Roi Reichart, Milica Gašić, Anna Korhonen, and Steve Young. 2017b. Semantic specialisation of distributional word vector spaces using monolingual and cross-lingual constraints. Transactions of the ACL.


