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

Ivan Vulić[1] , Nikola Mrkšić[2] and Anna Korhonen[1]
[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).[1] 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,

---

[1]http://verbs.colorado.edu/∼mpalmer/projects/verbnet.html

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 lower-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
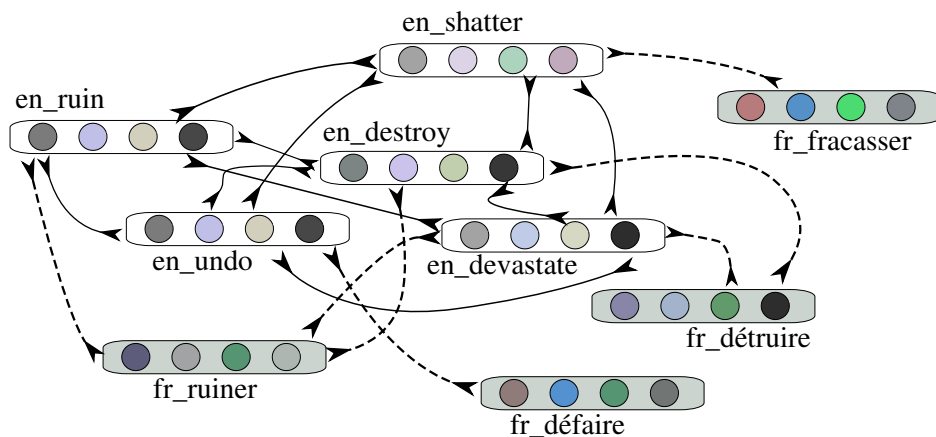
Figure 1: Transferring VerbNet information from a resource-rich to a resource-lean language through a word vector space: an *English→ French* toy example. Representations of words described by two types of ATTRACT constraints are pulled closer together in the joint vector space. (1) Monolingual pairwise constraints in English (e.g., *(en_ruin, en_shatter), (en_destroy, en_undo)*) reflect the EN VerbNet structure and are generated from the readily available verb classification in English (solid lines). They are used to specialise the distributional EN vector subspace for the VerbNet relation. (2) Cross-lingual English-French pairwise constraints (extracted from BabelNet) describe cross-lingual synonyms (i.e., translation links) such as *(en_ruin, fr_ruiner) or (en_shatter, fr_fracasser)* (dashed lines). The post-processing fine-tuning specialisation procedure based on (1) and (2) effectively transforms the initial distributional French vector subspace to also emphasise the VerbNet-style structure, facilitating the induction of verb classes in French.

fine-tuning process (thus answering Q3).[2] 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.[3] 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)[4] 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),

---

[2]On a high level, we demonstrate that a constraints-driven fine-tuning framework can specialise word embeddings to reflect VerbNet-style relations which rely not only on verb sense similarity, but also on similarity in syntax, selectional preferences, and diathesis alternations.

[3]The 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).

[4]For 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).[5] 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 $V = V_s \sqcup V_t$ 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 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:

$$O(\mathcal{B}_C) = O_C(\mathcal{B}_C) + R(\mathcal{B}_C) \qquad (1)$$

The first term of the method's cost function (i.e., $\mathcal{O}_C$) 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:

$$O_C(\mathcal{B}_C) = \sum_{(x_l, x_r) \in \mathcal{B}_C} \big[\, \tau \left(\delta_{att} + \mathbf{x}_l \mathbf{t}_l - \mathbf{x}_l \mathbf{x}_r \right)$$
$$+ \tau \left(\delta_{att} + \mathbf{x}_r \mathbf{t}_r - \mathbf{x}_l \mathbf{x}_r \right) \big] \qquad (2)$$

$\tau(x) = \max(0, x)$ is the standard rectified linear unit or the hinge loss function (Tsochantaridis et al., 2004; Nair and Hinton, 2010). δ_{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 $\mathbf{t}_i$ for each word x_i in any ATTRACT pair is always the vector closest to $\mathbf{x}_i$ taken from the pairs in the current mini-batch, distinct from the other word paired with x_i, and x_i itself.[6]

The second $R(\mathcal{B}_C)$ 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 $\mathbf{x}_i^{init}$ refer to the initial distributional vector of the word x_i and let $\mathcal{V}(\mathcal{B}_C)$ be the set of all word vectors present in the given mini-batch. If λ_{reg} denotes the L2 regularisation constant, this term can be expressed as:

$$R(\mathcal{B}_C) = \sum_{\mathbf{x}_i \in \mathcal{V}(\mathcal{B}_C)} \lambda_{reg} \left\| \mathbf{x}_i^{init} - \mathbf{x}_i \right\|_2 \qquad (3)$$

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:

LEARN-14	BREAK-45.1	ACCEPT-77
(learn, study)	(break, dissolve)	(accept, understand)
(learn, relearn)	(crash, crush)	(accept, reject)
(read, study)	(shatter, split)	(repent, rue)
(cram, relearn)	(break, rip)	(reject, discourage)
(read, assimilate)	(crack, smash)	(encourage, discourage)
(learn, assimilate)	(shred, splinter)	(reject, discourage)
(read, relearn)	(snap, tear)	(disprefer, understand)

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

a) cross-lingual (translation) links between languages, and **b)** available VerbNet annotations in a resource-rich language transformed into pairwise constraints. Cross-lingual constraints such as *(pl_wojna, it_guerra)* are extracted from BabelNet (Navigli and Ponzetto, 2012), a large-scale resource which groups words into cross-lingual BABEL synsets (and is currently available for 271 languages). The wide and steadily growing coverage of languages in BabelNet means that our proposed framework promises to support the transfer of VerbNet-style information to numerous target languages (with increasingly high accuracy).

To establish that the proposed transfer approach is in fact independent of the chosen cross-lingual information source, we also experiment with another cross-lingual dictionary: PanLex (Kamholz et al., 2014), which was used in prior work on cross-lingual word vector spaces (Duong et al., 2016; Adams et al., 2017). This dictionary currently covers around 1,300 language varieties with over 12 million expressions, thus offering support also for low-resource transfer settings.[7]

VerbNet constraints are extracted from the English VerbNet class structure in a straightforward manner. For each class VN_i from the 273 VerbNet classes, we simply take the set of all n_i verbs $CL_i = \{v_{1,i}, v_{2,i}, \ldots, v_{n_i,i}\}$ associated with that class, including its subclasses, and generate all unique pairs (v_k, v_l) so that $v_k, v_l \in CL_i$ and $v_k \neq v_l$. Example VerbNet pairwise constraints are shown in Tab. 1. Note that VerbNet classes in practice contain verb instances standing in a variety of lexical relations, including synonyms, antonyms, troponyms, hypernyms, and the class membership is determined on the basis of connections between the syntactic patterns and the underlying semantic relations (Kipper et al., 2006, 2008).

2.2 Clustering Algorithm

Given the initial distributional or specialised collection of target language vectors $\mathbf{V_t}$, 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_{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 cutoff 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>.[8] FR and IT word vectors were trained on the standard frWaC and itWaC corpora (Baroni et al., 2009), and vectors for other target languages were trained on the corpora of similar style and size: HR vectors were trained on the hrWaC corpus (Ljubešić and Klubička, 2014), PT

---

[7]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.

[8]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: .025, subsampling rate: $1e-4$. Similar trends in results persist with $d = 100, 500$.

	French:FR	Portuguese:PT	Italian:IT	Polish:PL	Croatian:HR	Finnish:FI
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 (BNet)	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 Ada-Grad (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:

$$\text{MPUR} = \frac{\sum_{C \in \textbf{Clust}, n_{prev(C)} > 1} n_{prev(C)}}{\#\text{test\_verbs}} \quad (4)$$

Here, each cluster C from the set of all K_{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 $n_{prev(C)}$. 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:
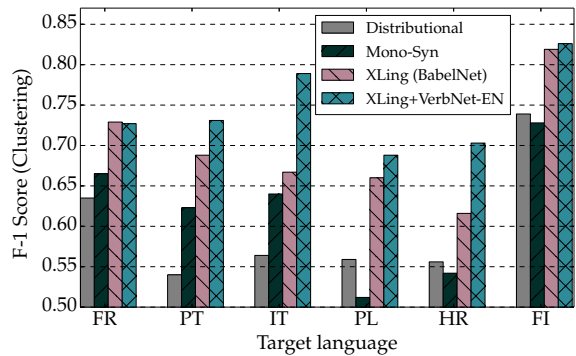
$$\text{WACC} = \frac{\sum_{C \in \textbf{Gold}} n_{dom(C)}}{\#\text{test\_verbs}} \quad (5)$$

---

[9]https://github.com/nmrksic/attract-repel
[10]Clusters with $n_{prev(C)} = 1$ are discarded from the count to avoid an undesired bias towards singleton clusters (Sun and Korhonen, 2009; Sun et al., 2010).
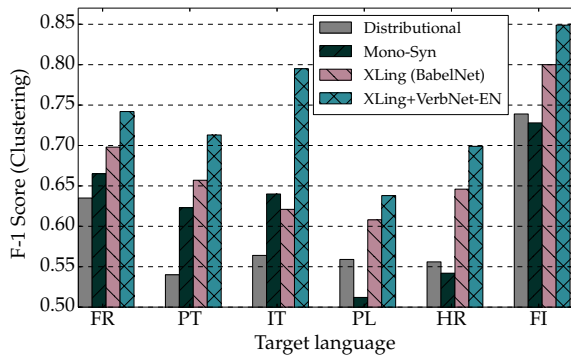
(a) EN Vectors: SGNS-BOW2

(b) EN Vectors: SGNS-DEPS

Figure 2: *F-1* scores in six target languages using the post-processing specialisation procedure from Sect. 2.1 and different sets of constraints: *Distributional* refers to the initial vector space in each target language; *Mono-Syn* is the vector space tuned using monolingual synonymy constraints from BabelNet; *XLing* uses cross-lingual EN-TARGET constraints from BabelNet (TARGET refers to any of the six target languages); *XLing+VerbNet-EN* is a fine-tuned vector space which uses both cross-lingual EN-TARGET constraints plus EN VerbNet constraints. Results are provided with (a) SGNS-BOW2 and (b) SGNS-DEPS source vector space in English for the *XLing* and *XLing+VerbNet* variants, see Sect. 3.

For each cluster C from the set of gold standard clusters **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 $n_{dom(C)}$. 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 monolin-

gual 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.

	French		Italian	
Vector space	*XLing*	*XLing+VN*	*XLing*	*XLing+VN*
EN+FR	0.698	0.742	–	–
EN+IT	–	–	0.621	**0.795**
EN+FR+IT	**0.728**	0.745	0.650	0.768
EN+FR+PL	0.697	0.717	–	–
EN+IT+PL	–	–	0.658	0.765
EN+FI+FR+IT	0.719	**0.751**	0.662	0.777
EN+FR+IT+PT	0.710	0.718	**0.688**	0.760

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*).

	VC:XLing	*Sim*:XLing	*Sim*:XLing+VN
EN-Dist	0.484	0.275	–
+FR	0.608	0.556	0.481
+PT	0.633	0.537	0.466
+IT	0.602	0.524	0.476
+PL	0.597	0.469	0.431
+HR	0.582	0.497	0.446
+FI	0.662	0.598	0.491
+FR+IT	0.633	0.571	0.526
+FI+FR+IT	0.641	0.635	0.558
+FR+IT+PT	0.674	0.596	0.515
EN-VN	**0.956**	–	0.358

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.

tion of cross-linguistic class consistency directly through a specialised vector space, and also avoids any reliance on target-language-specific NLP tools.

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.

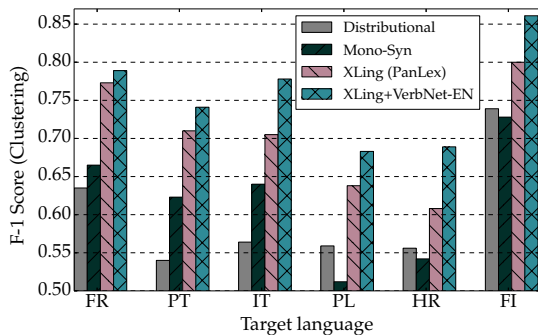**Other Cross-Lingual Sources** Replacing Babel-Net with PanLex as the alternative source of cross-

Figure 3: *F-1* scores when PanLex is used as the source of cross-lingual ATTRACT constraints (instead of BabelNet). EN Vectors: SGNS-BOW2.

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 frame-

work. 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.

References

Oliver Adams, Adam Makarucha, Graham Neubig, Steven Bird, and Trevor Cohn. 2017. Cross-lingual word embeddings for low-resource language modeling. In *Proceedings of EACL*, pages 937–947.

Rami Al-Rfou, Bryan Perozzi, and Steven Skiena. 2013. Polyglot: Distributed word representations for multilingual NLP. In *Proceedings of CoNLL*, pages 183–192.

Juan Aparicio, Mariona Taulé, and Maria Antònia Martí. 2008. AnCora-Verb: A lexical resource for the semantic annotation of corpora. In *Proceedings of LREC*, pages 797–802.

Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet project. In *Proceedings of ACL*, pages 86–90.

Simon Baker, Roi Reichart, and Anna Korhonen. 2014. An unsupervised model for instance level subcategorization acquisition. In *Proceedings of EMNLP*, pages 278–289.

Marco Baroni, Silvia Bernardini, Adriano Ferraresi, and Eros Zanchetta. 2009. The WaCky wide web: A collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation*, 43(3):209–226.

Marco Baroni, Georgiana Dinu, and Germán Kruszewski. 2014. Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors. In *Proceedings of ACL*, pages 238–247.

Vladimír Benko. 2014. Aranea: Yet another family of (comparable) Web corpora. In *Proceedings of TSD*, pages 247–254.

Chris Brew and Sabine Schulte im Walde. 2002. Spectral clustering for German verbs. In *Proceedings of EMNLP*, pages 117–124.

Susan Windisch Brown, Dmitriy Dligach, and Martha Palmer. 2011. VerbNet class assignment as a WSD task. In *Proceedings of IWCS*, pages 85–94.

Jinxiu Chen, Donghong Ji, Chew Lim Tan, and Zhengyu Niu. 2006. Relation extraction using label propagation based semi-supervised learning. In *Proceedings of ACL*, pages 129–136.

Ronan Collobert, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel P. Kuksa. 2011. Natural language processing (almost) from scratch. *Journal of Machine Learning Research*, 12:2493–2537.

John C. Duchi, Elad Hazan, and Yoram Singer. 2011. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, 12:2121–2159.

Long Duong, Hiroshi Kanayama, Tengfei Ma, Steven Bird, and Trevor Cohn. 2016. Learning crosslingual word embeddings without bilingual corpora. In *Proceedings of EMNLP*, pages 1285–1295.

Maud Ehrmann, Francesco Cecconi, Daniele Vannella, John Philip Mccrae, Philipp Cimiano, and Roberto Navigli. 2014. Representing multilingual data as linked data: The case of BabelNet 2.0. In *Proceedings of LREC*, pages 401–408.

Ingrid Falk, Claire Gardent, and Jean-Charles Lamirel. 2012. Classifying French verbs using French and English lexical resources. In *Proceedings of ACL*, pages 854–863.

Manaal Faruqui, Jesse Dodge, Sujay Kumar Jauhar, Chris Dyer, Eduard Hovy, and Noah A. Smith. 2015. Retrofitting word vectors to semantic lexicons. In *Proceedings of NAACL-HLT*, pages 1606–1615.

Christiane Fellbaum. 1998. *WordNet*.

Daniela Gerz, Ivan Vulić, Felix Hill, Roi Reichart, and Anna Korhonen. 2016. SimVerb-3500: A large-scale evaluation set of verb similarity. In *Proceedings of EMNLP*, pages 2173–2182.

Ana-Maria Giuglea and Alessandro Moschitti. 2006. Semantic role labeling via FrameNet, VerbNet and PropBank. In *Proceedings of ACL*, pages 929–936.

Jeffrey Gruber. 1976. *Lexical structure in syntax and semantics*.

Zellig S. Harris. 1954. Distributional structure. *Word*, 10(23):146–162.

Felix Hill, Roi Reichart, and Anna Korhonen. 2015. SimLex-999: Evaluating semantic models with (genuine) similarity estimation. *Computational Linguistics*, 41(4):665–695.

Ray S. Jackendoff. 1972. *Semantic interpretation in generative grammar*.

Ray S. Jackendoff. 1992. *Semantic structures*.

Eric Joanis, Suzanne Stevenson, and David James. 2008. A general feature space for automatic verb classification. *Natural Language Engineering*, 14(3):337–367.

David Kamholz, Jonathan Pool, and Susan M. Colowick. 2014. PanLex: Building a resource for panlingual lexical translation. In *Proceedings of LREC*, pages 3145–3150.

Daisuke Kawahara, Daniel W. Peterson, and Martha Palmer. 2014. A step-wise usage-based method for inducing polysemy-aware verb classes. In *Proceedings of ACL*, pages 1030–1040.

Douwe Kiela, Felix Hill, and Stephen Clark. 2015. Specializing word embeddings for similarity or relatedness. In *Proceedings of EMNLP*, pages 2044–2048.

Joo-Kyung Kim, Marie-Catherine de Marneffe, and Eric Fosler-Lussier. 2016a. Adjusting word embeddings with semantic intensity orders. In *Proceedings of the 1st Workshop on Representation Learning for NLP*, pages 62–69.

Joo-Kyung Kim, Gokhan Tur, Asli Celikyilmaz, Bin Cao, and Ye-Yi Wang. 2016b. Intent detection using semantically enriched word embeddings. In *Proceedings of IEEE SLT*.

Karin Kipper. 2005. *VerbNet: A broad-coverage, comprehensive verb lexicon*. Ph.D. thesis.

Karin Kipper, Hoa Trang Dang, and Martha Stone Palmer. 2000. Class-based construction of a verb lexicon. In *Proceedings of AAAI*, pages 691–696.

Karin Kipper, Anna Korhonen, Neville Ryant, and Martha Palmer. 2006. Extending VerbNet with novel verb classes. In *Proceedings of LREC*.

Karin Kipper, Anna Korhonen, Neville Ryant, and Martha Palmer. 2008. A large-scale classification of english verbs. *Language Resources and Evaluation*, 42(1):21–40.

Anna Korhonen. 2010. Automatic lexical classification: Bridging research and practice. *Philosophical Transactions A of the Royal Society*, 368:3621–3632.

Beth Levin. 1993. *English verb classes and alternation, A preliminary investigation*.

Omer Levy and Yoav Goldberg. 2014. Dependency-based word embeddings. In *Proceedings of ACL*, pages 302–308.

Jianguo Li and Chris Brew. 2008. Which are the best features for automatic verb classification. In *Proceedings of ACL*, pages 434–442.

Thomas Lippincott, Laura Rimell, Karin Verspoor, and Anna Korhonen. 2013. Approaches to verb subcategorization for biomedicine. *Journal of Biomedical Informatics*, 46(2):212–227.

Mei-Chun Liu and Ting-Yi Chiang. 2008. The construction of Mandarin VerbNet: A frame-based study of statement verbs. *Language and Linguistics*, 9(2):239–270.

Nikola Ljubešić and Filip Klubička. 2014. {bs,hr,sr}WaC – Web corpora of Bosnian, Croatian and Serbian. In *Proceedings of the 9th Web as Corpus Workshop*, pages 29–35, Gothenburg, Sweden. Association for Computational Linguistics.

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

Edward Loper, Szu-Ting Yi, and Martha Palmer. 2007. Combining lexical resources: Mapping between PropBank and VerbNet. In *Proceedings of IWCS*.

Ulrike von Luxburg. 2007. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416.

Olga Majewska, Ivan Vulić, Diana McCarthy, Yan Huang, Akira Murakami, Veronika Laippala, and Anna Korhonen. 2017. Investigating the cross-lingual translatability of verbnet-style classification. *Language Resources and Evaluation*.

Mausam, Michael Schmitz, Robert Bart, Stephen Soderland, and Oren Etzioni. 2012. Open language learning for information extraction. In *Proceedings of EMNLP*, pages 523–534.

Marina Meila and Jianbo Shi. 2001. A random walks view of spectral segmentation. In *Proceedings of AISTATS*.

Oren Melamud, Jacob Goldberger, and Ido Dagan. 2016. Context2vec: Learning generic context embedding with bidirectional LSTM. In *Proceedings of CoNLL*, pages 51–61.

Paola Merlo, Suzanne Stevenson, Vivian Tsang, and Gianluca Allaria. 2002. A multilingual paradigm for automatic verb classification. In *Proceedings of ACL*, pages 207–214.

Cédric Messiant. 2008. A subcategorization acquisition system for French verbs. In *Proceedings of ACL Student Research Workshop*, pages 55–60.

Tomas Mikolov, Kai Chen, Gregory S. Corrado, and Jeffrey Dean. 2013a. Efficient estimation of word representations in vector space. In *Proceedings of ICLR Workshop Papers*.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Gregory S. Corrado, and Jeffrey Dean. 2013b. Distributed representations of words and phrases and their compositionality. In *Proceedings of NIPS*, pages 3111–3119.

Nikola Mrkšić, Diarmuid Ó Séaghdha, Blaise Thomson, Milica Gašić, Lina Rojas-Barahona, Pei-Hao Su, David Vandyke, Tsung-Hsien Wen, and Steve Young. 2016. Counter-fitting word vectors to linguistic constraints. In *Proceedings of NAACL*, pages 142–148.

Nikola Mrkšić, Diarmuid Ó Séaghdha, Blaise Thomson, Tsung-Hsien Wen, and Steve Young. 2017a. Neural Belief Tracker: Data-driven dialogue state tracking. In *Proceedings of ACL*.

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*.

Vinod Nair and Geoffrey E. Hinton. 2010. Rectified linear units improve restricted Boltzmann machines. In *Proceedings of ICML*, pages 807–814.

2556

Roberto Navigli and Simone Paolo Ponzetto. 2012. BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence*, 193:217–250.

Arvind Neelakantan, Jeevan Shankar, Alexandre Passos, and Andrew McCallum. 2014. Efficient nonparametric estimation of multiple embeddings per word in vector space. In *Proceedings of EMNLP*, pages 1059–1069.

Diarmuid Ó Séaghdha and Ann Copestake. 2008. Semantic classification with distributional kernels. In *Proceedings of COLING*, pages 649–656.

Sebastian Padó and Mirella Lapata. 2007. Dependency-based construction of semantic space models. *Computational Linguistics*, 33(2):161–199.

Karel Pala and Aleš Horák. 2008. Can complex valency frames be universal? In *Proceedings of RASLAN*, pages 41–49.

Martha Palmer, Paul Kingsbury, and Daniel Gildea. 2005. The Proposition Bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31(1):71–106.

Daniel W. Peterson, Jordan Boyd-Graber, Martha Palmer, and Daisuke Kawahara. 2016. Leveraging VerbNet to build corpus-specific verb clusters. In *Proceedings of *SEM*, pages 102–107.

Mohammad Taher Pilehvar and Nigel Collier. 2016. De-conflated semantic representations. In *Proceedings of EMNLP*, pages 1680–1690.

Quentin Pradet, Laurence Danlos, and Gaël de Chalendar. 2014. Adapting VerbNet to French using existing resources. In *Proceedings of LREC*, pages 1122–1126.

Laura Rimell, Thomas Lippincott, Karin Verspoor, Helen L. Johnson, and Anna Korhonen. 2013. Acquisition and evaluation of verb subcategorization resources for biomedicine. *Journal of Biomedical Informatics*, 46(2):228–237.

Andrew Rosenberg and Julia Hirschberg. 2007. V-measure: A conditional entropy-based external cluster evaluation measure. In *Proceedings of EMNLP*, pages 410–420.

Carolina Scarton and Sandra Alusio. 2012. Towards a cross-linguistic VerbNet-style lexicon for Brazilian Portuguese. In *Proceedings of the LREC 2012 Workshop on Creating Cross-language Resources for Disconnected Languages and Styles*.

Carolina Scarton, Lin Sun, Karin Kipper Schuler, Magali Sanches Duran, Martha Palmer, and Anna Korhonen. 2014. Verb clustering for Brazilian Portuguese. In *Proceedings of CICLing*, pages 25–39.

Sabine Schulte im Walde. 2006. Experiments on the automatic induction of German semantic verb classes. *Computational Linguistics*, 32(2):159–194.

Roy Schwartz, Roi Reichart, and Ari Rappoport. 2016. Symmetric patterns and coordinations: Fast and enhanced representations of verbs and adjectives. In *Proceedings of NAACL-HLT*, pages 499–505.

Lei Shi and Rada Mihalcea. 2005. Putting pieces together: Combining FrameNet, VerbNet and WordNet for robust semantic parsing. In *Proceedings of CICLing*, pages 100–111.

Rajen Subba and Barbara Di Eugenio. 2009. An effective discourse parser that uses rich linguistic information. In *Proceedings of NAACL-HLT*, pages 566–574.

Lin Sun and Anna Korhonen. 2009. Improving verb clustering with automatically acquired selectional preferences. In *Proceedings of EMNLP*, pages 638–647.

Lin Sun, Anna Korhonen, and Yuval Krymolowski. 2008. Verb class discovery from rich syntactic data. In *Proceedings of CICLing*, pages 16–27.

Lin Sun, Diana McCarthy, and Anna Korhonen. 2013. Diathesis alternation approximation for verb clustering. In *Proceedings of ACL*, pages 736–741.

Lin Sun, Thierry Poibeau, Anna Korhonen, and Cédric Messiant. 2010. Investigating the cross-linguistic potential of VerbNet-style classification. In *Proceedings of COLING*, pages 1056–1064.

Robert S. Swier and Suzanne Stevenson. 2004. Unsupervised semantic role labelling. In *Proceedings of EMNLP*, pages 95–102.

Ioannis Tsochantaridis, Thomas Hofmann, Thorsten Joachims, and Yasemin Altun. 2004. Support vector machine learning for interdependent and structured output spaces. In *Proceedings of ICML*.

Joseph P. Turian, Lev-Arie Ratinov, and Yoshua Bengio. 2010. Word representations: A simple and general method for semi-supervised learning. In *Proceedings of ACL*, pages 384–394.

Ivan Vulić and Anna Korhonen. 2016. Is "universal syntax" universally useful for learning distributed word representations? In *Proceedings of ACL*, pages 518–524.

Jorge Alberto Wagner Filho, Rodrigo Wilkens, Leonardo Zilio, Marco Idiart, and Aline Villavicencio. 2016. Crawling by readability level. In *Proceedings of PROPOR*, pages 306–318.

Joe H. Ward. 1963. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301):236–244.

John Wieting, Mohit Bansal, Kevin Gimpel, and Karen Livescu. 2015. From paraphrase database to compositional paraphrase model and back. *Transactions of the ACL*, 3:345–358.

Derry Tanti Wijaya and Tom M. Mitchell. 2016. Mapping verbs in different languages to knowledge base relations using web text as interlingua. In *Proceedings of NAACL-HLT*, pages 818–827.

Lihi Zelnik-Manor and Pietro Perona. 2004. Self-tuning spectral clustering. In *Proceedings of NIPS*, pages 1601–1608.