

Cross-lingual word embeddings and the structure of the human bilingual lexicon

Paola Merlo

University of Geneva

Paola.Merlo@unige.ch

Maria A. Rodriguez

University of Geneva

Maria.AnduezaRodriguez@unige.ch

Abstract

Research on the bilingual lexicon has uncovered fascinating interactions between the lexicons of the native language and of the second language in bilingual speakers. In particular, it has been found that the lexicon of the underlying native language affects the organisation of the second language. In the spirit of interpreting current distributed representations, this paper investigates two models of cross-lingual word embeddings, comparing them to the shared-translation effect and the cross-lingual coactivation effects of false and true friends (cognates) found in humans. We find that the similarity structure of the cross-lingual word embeddings space yields the same effects as the human bilingual lexicon.

1 Introduction

Research on the bilingual lexicon has uncovered fascinating interactions between the L1 (native language) and L2 (second language) lexicons showing that both production and comprehension coactivate lexical items in both languages, indicating that bilinguals store lexical representations from their native and their second language in the same space (Kroll and Dijkstra, 2012; Williams, 2014).¹

This paper presents the first bilingual investigation of models of cross-lingual word embeddings and asks whether the bilingual spaces they define

¹Throughout this paper, and following the current literature on the topic, we use the term ‘bilingual’ loosely, to refer to any speaker of more than one language. Although there has been much research on all aspects of bilingualism, and at all stages of proficiency, the effects we model here have been found in experiments testing speakers who began to learn their second language after their first, usually in a school context, and who are at an advanced level of proficiency. The form German-English below will indicate, for example, a native speaker of German who learnt English as a second language (Williams, 2014).

have similar properties to the human bilingual lexicon. Among the many questions and results in the vast bilingualism literature, we concentrate on coactivation effects in items with shared translations. We also study interference or facilitatory effects in form-meaning mapping, the case of *false friends*, words that share form but differ in meaning across the two languages, and *true friends*, words that share both form and meaning. We find that the similarity structure of cross-lingual word embeddings matches well with known experimental findings of the human bilingual lexicon.

2 The structure of the bilingual lexicon

The core findings about the bilingual lexicon confirm that the two languages occupy an integrated space and they interact with each other (Wolter, 2001). For example, both in the monolingual and bilingual lexicon the best predictor of the time to recognise a word is the number of similarly spelled words, within and across languages (Johnson and Pugh, 1994; van Heuven et al., 1998).² This implies that, functionally, the bilingual lexicon is an integrated system. Specifically, it has been proposed that languages do not simply share graphemes or phonemes, but that the lexicon, monolingual, bilingual or multi-lingual, is a space of distributed word representations where word forms from different languages map onto a common abstract conceptual code (Van Hell and de Groot, 1998). This general structural and functional assumption explains many findings. We concentrate here on two sets of coactivation effects.

²Monolingual work has provided a finer-grained picture of this result, modulated by number of word senses and the semantic closeness of sense extensions, but the main result remains valid (Rodd et al., 2002).

2.1 The shared translation effect

Polysemous words that have many translations, such as English *bank* translated as *banca* (financial institution) or *riva* (river bank) in Italian, coactivate the correspondences for all the word senses, with various effects. One-to-many translations have been shown to slow down acquisition and processing for Italian-English bilinguals (Degani and Tokowicz, 2010), and to slow down response times of German-English speakers in anomaly detection tasks (Elston-Güttler and Williams, 2008).³

We are mainly interested in the result of Degani et al. (2011), as it concerns similarity spaces in shared translations. Degani et al. (2011) asked Hebrew-English bilinguals to rate the semantic relatedness of English word pairs that shared a translation in Hebrew (e.g., *tool* and *dish* both translated into Hebrew *kli*). Compared to both English pairs with different Hebrew translations, and to ratings by monolingual English speakers, bilinguals judged shared-translation pairs as more related in meaning (the shared-translation effect).⁴

2.2 Form-meaning mappings in translation

Competition (and facilitation) effects have been found both in comprehension and production depending on convergence and divergence of form-meaning mappings in translation. Recall that *false friends* are cross-linguistically similar in form but not in meaning, such as the English-Italian *estate*, which in Italian means *summer*, and *true friends* are words that share both (orthographic or phonological) form and meaning, such as English-French *glucose* or *danger*, in translation.

False friends effect In a cross-modal picture decision task, Weber and Cutler (2004) find that Dutch-English speakers are slower in matching an English word (*desk*) with the corresponding picture if the target picture's word form matches the Dutch form of one of the alternative pictures (*deskel* = *lid*). It should be noted, however, that

³German-English speakers, compared to monolingual English speakers, are slower in recognising, for example, that the word *bubble* is infelicitous in contexts where the word *blister* is required, due to the fact that these two words are translated as the same word *Blase* in German.

⁴Notice that this effect is robust as it was also replicated for English-Hebrew bilinguals, who learned Hebrew as an L2. Moreover, Degani et al. (2011) used as stimuli semantically unrelated word pairs, extending previously established results for sense-related words, such as *home-house* (Jiang, 2002).

while the decision time was slower, the decision accuracy was not. Bilingual speakers do know which is the right word-picture match and perform accurately. Also, for English-Dutch false friends like *rust* (*rest* in Dutch) lexical decision times are slower than expected, if the list in which they are embedded also contains words from the other language (Dijkstra et al., 1998; Smits et al., 2006).

True friends effects In recognition, Dutch-English bilinguals performing a lexical decision task in English were found to be faster than expected for words like *type*, a near true friend with a slight difference in pronunciation (Dijkstra et al., 1999; Smits et al., 2006; Dijkstra et al., 2010). Similar cognate facilitation effects also occur in production tasks, such as picture-naming. If an advanced Catalan-Spanish bilingual is asked to name pictures in Spanish, they are faster to do so for true friends such as *gato* (*gat* in Catalan 'cat') than for non-cognates. The effect, although smaller, can also be obtained when pictures are to be named in the L1 (Costa et al., 2000). Similar effects have also been obtained for Japanese-English bilinguals, despite the difference in the scripts (Hoshino and Kroll, 2008).

3 Predictions

In this work, we ask if the structure of cross-lingual word embeddings spaces have the properties that would be expected given human bilingual behaviour. Assuming the distributed, integrated model of the lexicon proposed in the bilingualism literature, the underlying linking hypothesis is that coactivation effects (whether expressed as similarity judgments or measured as reaction times) are the expression of greater or smaller proximity in a multi-dimensional space. On this basis, several hypotheses are proposed. The first hypothesis aims to establish whether cross-lingual word embeddings are sensitive to a bilingual situation and generate an integrated cross-lingual space. Secondly, we test if cross-lingual word embeddings show the shared-translation effect. Finally we test the cross-linguistic competition/priming of lexical forms from the L1 to the L2 language, comparing cross-lingual to monolingual spaces in true friends and false friends scenarios.

We will often talk about a word and its *translation*. By this term, we mean the pair of words that a bilingual dictionary would indicate as equivalent lexical entries, a translations pair.

3.1 The integrated bilingual lexicon

We test the idea that the bilingual lexicon is an integrated system by looking at effects of such a system in one-to-one mappings and one-to-many mappings.

The simplest and most basic prediction that a model of the integrated bilingual lexicon needs to be able to confirm is that words in the bilingual lexicon are “closer” to each other than word mappings across two aligned mono-lingual lexicons.

HYPOTHESIS 1 Given an L1 word w_1 and its translation w_2 in L2, the similarity between the word embeddings pair in a cross-lingual space (w_1^{cr}, w_2^{cr}) is higher than their similarity between their aligned monolingual counterparts (w_1^{m1}, w_2^{m2}) .

$$sim(w_1^{cr1}, w_2^{cr2}) > sim(w_1^{m1}, w_2^{m2}) \quad (1)$$

The second prediction is based on the finding of shared translations, where a shared translation in L1 affects L2 similarity judgments.

HYPOTHESIS 2 Given an L1 word w_1 and its translations w_{2a} and w_{2b} in L2, the similarity between the cross-lingual embeddings of the translation pair will be greater than the similarity between their monolingual counterparts.

$$sim(w_{2a}^{cr2}, w_{2b}^{cr2}) > sim(w_{2a}^{m2}, w_{2b}^{m2}) \quad (2)$$

3.2 Form-meaning competitions in the biligual lexicon

The following experiments investigate the competition faced by words with a high level of lexical similarity. Simplifying, words across two languages can be similar in form or meaning, or both or neither. For the following predictions, then, we define five different types of word pairs. Examples are shown in Figure 3.

FALSE FRIENDS: words that share the same form, but are semantically different.

REAL TRANSLATIONS of the false friends: the real L2 translations of the L1 word that also has a false friend.

TRUE FRIENDS: words sharing form and meaning.

NORMAL TRANSLATIONS: words semantically equivalent, but with a different form.

UNCORRELATED WORDS: words lexically and semantically uncorrelated.

False and true friends coactivation In bilingual speakers, false friends show an inhibitory effect of the L1 meaning in L2 tasks, but they do not affect the final accuracy of the task completion. Consequently, in our cross-lingual vectorial space, false friends should not have higher similarity score than their real translations, but they should be included in the top translations, i.e. the difference in similarity score between the real translation and the false friends should be smaller than the difference between the real translations and other words (appropriately matched to the false friends). This in turn can be demonstrated by two expected inequalities: false friends are not closer than real translations but false friends are significantly more similar than (matching) uncorrelated words.

Precisely, given an L1 word w_1 , and its real L2 translation w_2 and the false friend w_{2ff} in cross-lingual space, we expect the similarity score between the pair (w_1, w_2) not to be lower than the similarity score between the pair (w_1, w_{2ff}) .

HYPOTHESIS 3 Real translations have a better or equal similarity score than their corresponding false friends.

$$sim(w_1, w_2) \geq sim(w_1, w_{2ff}) \quad (3)$$

Moreover, false friends (w_1, w_{2ff}) have a similarity score that is higher than uncorrelated words (w_1, w_{2nc}) . This is because the false friends pair (w_1, w_{2ff}) shares similarity of form even if it is, in fact, semantically uncorrelated.

HYPOTHESIS 4 False friends have a better similarity score than pairs with no correlation.

$$sim(w_1, w_{2ff}) > sim(w_1, w_{2nc}) \quad (4)$$

Lexical similarity can also work in the opposite direction. L1 words that are similar to the L2 word both in form and meaning, *true friends*, have been shown, in bilinguals speakers, to facilitate tasks in L2. In cross-lingual word embeddings, we expect that true friends (w_{1tf}, w_{2tf}) have a higher similarity score than normal translation pairs (w_{1n}, w_{2n}) , whose interaction is not enhanced by lexical or morphological resemblances.

HYPOTHESIS 5 True friends have a better similarity score than normal translation pairs.

$$sim(w_{1tf}, w_{2tf}) > sim(w_{1n}, w_{2n}) \quad (5)$$

HYP. 1 Cross-lingual word embeddings pairs are more similar than their aligned monolingual counterparts	$sim(w_1^{cr1}, w_2^{cr2}) > sim(w_1^{m1}, w_2^{m2})$
HYP. 2 For two L2 words sharing a translation in L1, cross-lingual word embeddings are more similar than monolingual word embeddings	$sim(w_{2a}^{cr2}, w_{2b}^{cr2}) > sim(w_{2a}^{m2}, w_{2b}^{m2})$
HYP. 3 Real translations are more similar than their corresponding false friends	$sim(w_1, w_2) \geq sim(w_1, w_{2ff})$
HYP. 4 False friends are more similar than uncorrelated pairs	$sim(w_1, w_{2ff}) > sim(w_1, w_{2nc})$
HYP. 5 True friends are more similar than normal translation pairs	$sim(w_{1tf}, w_{2tf}) > sim(w_{1n}, w_{2n})$
HYP. 6 Normal translation pairs are more similar than real translations of false friends	$sim(w_{1n}, w_{2n}) > sim(w_1, w_2)$

Figure 1: The six experimental predictions.

Another hypotheses can also be formulated that follows logically from these preceding ones. Consider the pair (w_1, w_2) where w_2 , as seen before, is the real translation of w_1 in a pair that also has a false translation. In this case, it is important to remember that w_1 has a false friend w_{2ff} , so we know that accessing w_2 is more effortful since, for a bilingual speaker, w_{2ff} will also be activated. Consequently, we can assume that a normal pair of words (w_{1n}, w_{2n}) , a pair of translated words that have no false friend, are closer in space than (w_1, w_2) precisely because (w_{1n}, w_{2n}) is not inhibited by a false translation as in the case of (w_1, w_2) .

HYPOTHESIS 6 Normal translation pairs have a higher similarity score than real translations of false friends.

$$sim(w_{1n}, w_{2n}) > sim(w_1, w_2) \quad (6)$$

The predictions are summarised in Figure 1. If confirmed, they give us a fairly detailed view of the structure of the lexicon conceived as a multi-dimensional, integrated multilingual space. In particular, they inform us on the respective importance of formal and meaning properties of words in this cross-lingual similarity space.

4 Experiments 1 and 2

We test our hypotheses using two different cross-lingual word embeddings models. (The numbering of the experiments corresponds to the numbering of the hypotheses.)

One model is VECMAP, a word-level cross-lingual word embedding method, developed by Artetxe et al. (2018), which offers different op-

translation pairs	shared translation pairs
wood-legno	legno bosco
wood-bosco	
grade-grado	grado voto
grade-voto	
block-blocco	blocco ceppo
block-ceppo	blocco bloccare
block-bloccare	blocco ostacolare
block-ostacolare	ceppo bloccare
	ceppo ostacolare

Figure 2: Sample of translation pairs and sample of shared translation pairs used in experiments 1 and 2.

tions, ranging from fully supervised to weakly supervised or unsupervised, the state-of-the-art for bilingual lexicon induction. This method, unlike other models, does not use an existing dictionary for initialization. The values of the word vectors in both the source and the target distribution are sorted, vectors that have similar permutations are identified as possible translations and are used to initialize a dictionary that is then further improved by self-iterative training.

We also test M2VEC, a weakly-supervised, concept-based adversarial model (Wang et al., 2019). This method is based on the idea that languages use similar words to express similar concepts (Søgaard et al., 2015). The adversarial training uses concepts, drawn from Wikipedia, rather than words, to learn competitive cross-lingual word embeddings. The alignments are learnt by a generative adversarial networks (GAN) adapted to the cross-lingual mapping objective.

FALSE FRIENDS		REAL TRANSLATIONS		TRUE FRIENDS		NORMAL TRANSLATIONS		UNCORRELATED PAIRS	
arrange	arrangiare	arrange	disporre	family	famiglia	jam	marmellata	arrange	tagliare
		arrange	sistemare	fantastic	fantastico	january	gennaio	attend	guardare
		arrange	organizzare	future	futuro	journey	viaggio	attic	luna
attend	attendere	attend	frequentare	general	generale	keep	tenere	attitude	canale
		attend	assistere	generation	generazione	kind	tipo	barrack	tazza
bald	baldo	bald	calvo	guide	guida	leave	partire	brave	forchetta
		bald	pelato	historial	storica	light	luce	camera	traduzione
brave	bravo	brave	coraggioso	industry	industria	mean	significare	caution	muro
		brave	valoroso	local	locale	mood	umore	code	pistola
canteen	cantina	canteen	mensa	melody	melodia	overview	panoramica	confetti	vitamine
		canteen	borraccia	minor	minore	pattern	modello	confidence	treno

Figure 3: Sample of five word pair types used in experiments 3, 4, 5, and 6.

4.1 Data

Word-embeddings data For training VECMAP, we use the English-Italian portion of the data used in Artetxe et al. (2018), which is based on the dataset described in Dinu et al. (2015). The English-Italian dataset provided by Dinu et al. (2015) contains 300-dimensional CBOW monolingual word embeddings for a total of 200K words trained on the WacKy crawling corpora.⁵ The English word embeddings use 2.8 billion tokens (ukWAC + Wikipedia + BNC) and the Italian word embeddings use 1.6 billion tokens (itWAC). An English-Italian gold standard bilingual dictionary built from Europarl⁶ word alignments is also provided with a training set of 5000 entries ordered by English frequency. For M2VEC, 300-dimensional monolingual word embeddings are trained with FastText. The training corpus is taken from a Wikipedia dump.⁷ The word embeddings are augmented to include concept-aligned articles extracted from the Linguatools Wikipedia comparable corpus.⁸ VECMAP uses word-level embeddings and M2VEC character-level embeddings.

Word lists For testing, we build lists of English words that have multiple translations in Italian. The lists of polysemous words were validated with the support of the Cambridge Bilingual Dictionary⁹ and by two bilingual speakers. Some example pairs are shown in Figure 2. Word lists are available as supplementary materials.

⁵<https://wacky.sslmit.unibo.it/>

⁶<https://www.statmt.org/europarl/>

⁷<https://dumps.wikimedia.org/>

⁸<http://linguatools.org/tools/corpora/wikipedia-comparable-corpora/>

⁹<https://dictionary.cambridge.org/dictionary/english-italian/>

		Mono lingual	Cross-lingual	
			VECMAP	M2VEC
E1	Mean	0.02	0.371	0.345
	Variance	0.03	0.044	0.044
	T(104)		-17.059	-15.937
	<i>p</i>		0.000	0.000
E2	Mean	0.153	0.163	0.184
	Variance	0.019	0.025	0.044
	T(109)		-2.050	-2.835
	<i>p</i>		0.021	0.003

Table 1: Results of experiments 1 and 2.

4.2 Results

The results are shown in Table 1. Recall the two hypotheses, hypothesis 1 and hypothesis 2 summarised in Figure 1.

The results confirm both hypotheses for both cross-lingual models and are statistically significant under a one-tailed pairwise t-test with $\alpha = 0.25$. Cross-lingual word embeddings have a higher mean similarity score than aligned monolingual word embeddings. Also, cross-lingual word embeddings, like humans, show a shared-translation effect. Both cross-lingual models show higher mean similarity scores for L2-words that share a common L1 source than the monolingual model.

5 Experiments 3, 4, 5 and 6

The next four experiments test whether the hypotheses concerning true and false friends in word embeddings are confirmed. The numbering of the experiments corresponds to the numbering of the hypotheses.

5.1 Data

Word embedding data In this set of experiments, we use the same data from the previous two experiments (the dataset described in Dinu

et al. (2015)) with the addition of the FastText pre-trained word embeddings for English and Italian (Bojanowski et al., 2017).¹⁰ These publicly available vectors are obtained by a 5-word window, for 300 resulting dimensions, on CommonCrawl and Wikipedia data using the Skip-gram model. Every word is represented as an n -grams of characters, for n training between 3 and 6. Each n -gram is represented by a vector and the sum of these vectors forms the vector representing the given word. So we have three cross-lingual models: two versions of VECMAP, one trained on CBOW (word-level) and the other on FastText (character-level), and the concept-based adversarial model M2VEC, which also uses character-based FastText representations. The inclusion of FastText embeddings is important for these experiments, as the embeddings need to be sensitive to character-level sequences to detect similarity of form in false and true friends. VECMAP was used to obtain the cross-lingual word embeddings. Having two versions of VECMAP, one that use Skip-gram and one that uses CBOW is based on their different performance on rare words. The CBOW model predicts a word from its context and is better in accuracy for frequent words, but encounters problems with rare words, while the Skip-gram model predicts the context from a target word, and so has good representation of rare words or phrases (Mikolov et al., 2013a,c). In the figures they will be indicated, respectively, as Vecmap, FastText and Concepts.

Word lists The data required for the following experiments comprise five lists of word pairs, defined in section 3.2: false friends, real translations, true friends, normal translations and uncorrelated words. Examples are shown in Figure 3 and the complete lists are available in the appendix, with their similarity scores.

These word lists have been constructed from various online resources, adding also words that were found serendipitously by the authors in different texts.¹¹ The normal translations and uncorrelated word pairs were built by one of the authors.

¹⁰<https://github.com/facebookresearch/fastText>

¹¹The false friends list was built starting from <http://www.lifemilan.it/en/false-friends-a-must-learn-list/> and <https://www.reference.tjtaylor.net/false-friends/>. The true friends list was started from <https://takelessons.com/blog/italian-grammar-cognates-z09>.

Judge 2	Judge 1					Total
	FF	TF	RT	NT	UN	
FF	93	1	0	0	0	94
TF	3	127	0	0	0	130
RT	0	2	130	8	1	141
NT	1	3	7	133	0	144
UN	0	0	0	0	97	97
Total	97	133	137	141	98	606

Table 2: Inter-judge partition of the five lists of words, rounded to closest integer. TF = true friends; FF = false friends; RT = real translations; NT = normal translations; UN = uncorrelated.

Normal translations were selected by the bilingual dictionary excluding those that had true or false friends. The uncorrelated words were selected ensuring that they were entirely uncorrelated. They were validated with the help of the Cambridge Bilingual dictionary,¹² where each translation of each pair of words was checked to ensure correctness and complete disjunction in meaning. The online English-Italian dictionary was not comprehensive: not all the meanings were reported, unlike the English-Spanish or English-French dictionaries. Therefore, sometimes a false friend was not reported by the English-Italian dictionary, but was found with the help of the other bilingual dictionaries.

Once the lists were constructed, we run an inter-judge agreement validation. The words were shuffled and the two authors, who master both languages well, classified them in the five types discussed above. Cohen’s Kappa was 94.6, showing very high agreement. Fractional numbers were distributed in the few cases of multiple classification by one or both judges. The main source of disagreement were those words that are, at the same time, false friends and true friends depending on the context. The inter-judge agreement table is shown in Table 2.

As can be seen in Figure 3, the real translations list (151 pairs of words) is larger than the false friends list (97 pairs of words) due to the different meanings that an L1 word can have in L2.

As for the uncorrelated list, the same L1 words from the false friends list have been used for a direct comparison between the false friends similarity and the uncorrelated words similarity. Since

¹²<https://dictionary.cambridge.org/dictionary/english-italian/>

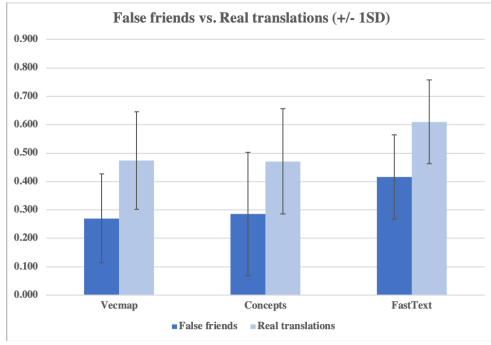


Figure 4: Experiment 3: false friends compared to real translations.

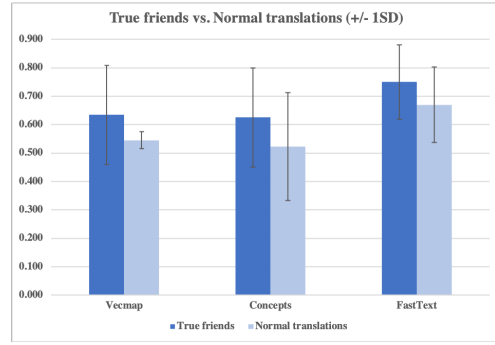


Figure 6: Experiment 5: true friends compared to normal translations.

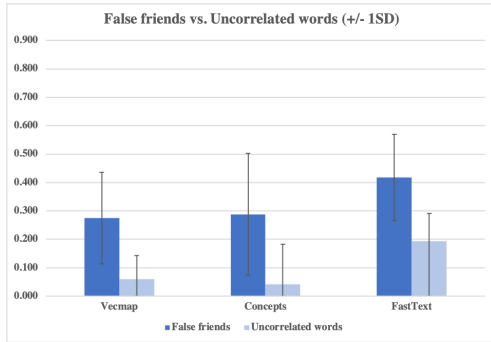


Figure 5: Experiment 4: false friends compared to pairs without correlation.

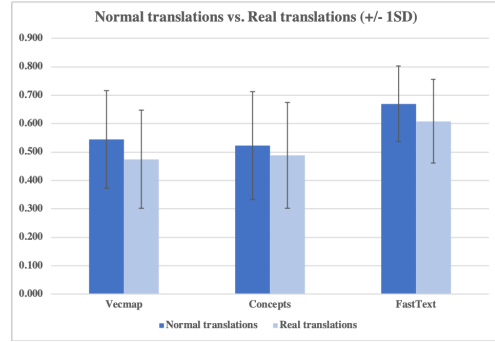


Figure 7: Experiment 6: Normal translations compared to real translations of false friends.

the real translations list contains the largest number of pairs, for the comparison between real translations and normal translations, the real translations have been sampled to have equal size.

5.2 Results

As shown in the Figures 4 to 7, overall the three methods are consistent as they show the same pattern of results. In terms of cosine similarity, the Vecmap+FastText word embeddings tend to have higher means than the other methods.

HYPOTHESIS 3 The first hypothesis concerning false friends —real translations have a better similarity score than their corresponding false friends— is confirmed. Figure 4 shows a cosine similarity that is 0.2 points higher for real translations pairs ($p < 0.0001$ in all three cases for a paired one-tailed t-test).

HYPOTHESIS 4 Based on the scores shown in Figure 5, the second hypothesis concerning false friends is also confirmed. False friends are significantly more similar in the cross-lingual space than uncorrelated words, as the similarity scores for false friends, in all three systems, are higher

by more than 0.2 points ($p < 0.0001$ in all three cases).

HYPOTHESIS 5 The hypothesis that true friends have a better similarity score than normal translation pairs is also confirmed. Figure 6 shows that the mean true friends similarity is higher than the mean similarity of the normal translations ($p < 0.001$ in all three cases).

HYPOTHESIS 6 The hypothesis that normal pairs of words have a higher similarity score than real translations of false friends is confirmed, see Figure 7. The mean similarity score for real translations is lower than the mean score for normal translations. This difference is statistically significant for Vecmap and FastText ($p < 0.01$). For the concept-based system, on the other hand, we do not have enough evidence to reject the null hypothesis as $p > 0.025$ and $T(296) < t\text{-value}$. Thus, in this case, we must conclude that the normal translations are as similar as the real translations.

BONFERRONI CORRECTION As Hypotheses 3 and 4, hypotheses 3 and 6 and hypotheses 5 and 6 use the same data (respectively false friends, real translations and normal translations), we run

the Bonferroni correction not to incur in α inflation. All hypotheses are further confirmed, except for hypothesis 6 where the concept-based system again shows a non-significant result (Bonferroni correction higher than α ($0.106 > 0.025$)).

6 Discussion

Cross-lingual word embeddings show a lexical structure matching the bilingual lexicon for the three properties that we tested.

They are an integrated multi-lingual space and exhibit the shared-translation effect. This suggests that the initial monolingual embeddings, in our case the Italian ones, are affected by the cross-lingual mapping as word vectors are changed to accommodate cross-lingual interactions. Intuitively, this is similar to the behavior of bilinguals when they gradually learn a new language and semantic links are created by associating the new words to an existing one in their native language.

Cross-lingual word embeddings did not show interference effects of false friends. They are not as affected by the lexical or morphological level when there is a semantic correlation between words. However, false friends show more similarity than uncorrelated words, showing that, like for humans, they have a different status from orthographically and semantically uncorrelated words. True friends in cross-lingual space behave like the lexicon of bilinguals, when the semantic, lexical and morphological level are aligned.

The last hypothesis shows varying results, as only two systems accept the bilingual hypothesis: Vecmap and FastText have a higher similarity score for normal translations, confirming the assumption that real translations are inhibited by false friends. Notice that this result argues again that cross-lingual word embeddings are indirectly affected by false friends, as there is no other difference between real translations and normal translations. Interestingly, on the other hand, the concept-based system is not sensitive to the indirect effect of false friends in the cross-lingual space. This is not unexpected as the concept-based model is less affected by surface form. The difference between word-based models and the concept-based model could yield an hypothesis to be tested for the human lexicon with precise underlying formal models.

Because our predictions are confirmed, they also confirm that the similarity structure defined

by current cross-lingual word embedding models is promising as a view of the structure of the lexicon conceived as a multi-dimensional, integrated multilingual space. In particular, our results inform us on the respective importance of formal and meaning properties of words in this cross-lingual similarity space. Notice that the pairwise results described so far define, by transitivity, a total order of similarity: true friends $>$ normal translations $>$ real translations $>$ false friends $>$ uncorrelated pairs. True friends match both in form and meaning, normal and real translations match only in meaning, and false friends match only in form. Thus, this order clearly indicates that while both form and meaning matter, similarity based on meaning is more important than similarity based on form.

7 Related work

The related work for the investigation reported here comprises work on the human bilingual lexicon, cross-lingual word embeddings models and computational models of the bilingual lexicon. As the relevant work on the first topic has already been discussed, we concentrate here on the latter two.

Vectors of words that are semantically or syntactically similar have been shown to be close to each other in the same space (Mikolov et al., 2013a,c; Pennington et al., 2014), making them widely useful in many natural language processing tasks such as machine translation and parsing (Bansal et al., 2014; Mi et al., 2016), both in a single language and across different languages.

Mikolov et al. (2013b) first observed that the geometric positions of similar words in different languages were related by a linear relation. Zou et al. (2013) showed that a cross-lingually shared word embedding space is more useful than a monolingual space in an end-to-end machine translation task. However, traditional methods for mapping two monolingual word embeddings require high quality aligned sentences or dictionaries (Faruqui and Dyer, 2014; Ammar et al., 2016).

Reducing the need for parallel data, then, has become the main issue for cross-lingual word embedding mapping. Methods that rely on sentence-alignments and also document-alignments have been proposed. Hermann and Blunsom (2014) present a method that, given enough data, train bilingual word embeddings from a sentence-

aligned corpus. Luong et al. (2015) propose a model, BiSkip, that takes as input a parallel corpus with both sentence and word-level alignment. Unlike other methods, BiSkip tries to learn not only target word representations from source words but also source word representations from target words. Vulic and Moens (2016) induces bilingual word embeddings from document-aligned comparable data that have been merged and shuffled producing a pseudo-bilingual document.

Some recent work aiming at reducing resources has shown competitive cross-lingual mappings across similar languages, using a pseudo-dictionary, such as identical character strings between two languages (Smith et al., 2017), or a simple list of numerals, thanks to a self-learning iterative framework (Artetxe et al., 2017). Furthermore, as indicated in section 4, Artetxe et al. (2018) extend their self-learning framework to unsupervised models, and build the state-of-the-art for bilingual lexicon induction. Another weakly-supervised model is proposed by Wang et al. (2019), a weakly-supervised concept-based adversarial method, used in our experiments, as also indicated in section 4.

Several computational models of human bilingualism exist, see Li (2013) for an overview. More relatedly to the current work that uses distributional approaches, aspects of the bilingual lexicon have been proposed for word associations (Matussevych et al., 2018). These associations are different in bilingual and monolingual speakers. For example, cognates, collocations and phonological responses are produced more frequently by non-native speakers. This work proposes a model of word association in bilinguals, implemented as a semantic network paired with a retrieval mechanism. Computational models of the influence of the native language on second language learning have also been investigated in Matussevych (2016), specifically for argument structure.

8 Conclusion

In the spirit of better understanding distributed representations and how well they match what we know about the structure and processing of language in humans (Linzen et al., 2016), this paper investigates two models of cross-lingual word embeddings comparing them to the shared translation effect and cross-lingual coactivation effects involving true and false friends found in humans.

We find that predictions about cross-lingual word embeddings are mostly confirmed, making them promising functional models of at least some aspects of the bilingual lexicon, despite their structural simplicity.

9 Acknowledgments

We thank Haozhou Wang for giving us access to his model M2VEC and to the embeddings it produces, and to Suzanne Stevenson and Yevgen Matussevych for pointing us to relevant work.

References

- Waleed Ammar, George Mulcaire, Yulia Tsvetkov, Guillaume Lample, Chris Dyer, and Noah A. Smith. 2016. *Massively Multilingual Word Embeddings*. *arXiv preprint arXiv:1309.4168*.
- Mikel Artetxe, Gorka Labaka, and Eneko Agirre. 2017. Learning bilingual word embeddings with (almost) no bilingual data. In *55th Annual Meeting of the Association for Computational Linguistics*, pages 451–462, Vancouver, Canada.
- Mikel Artetxe, Gorka Labaka, and Eneko Agirre. 2018. A robust self-learning method for fully unsupervised cross-lingual mappings of word embeddings. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 789–798, Melbourne, Australia. Association for Computational Linguistics.
- Mohit Bansal, Kevin Gimpel, and Karen Livescu. 2014. Tailoring Continuous Word Representations for Dependency Parsing. In *52nd Annual Meeting of the Association for Computational Linguistics*, pages 809–815, Baltimore, Maryland, USA.
- Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. Enriching Word Vectors with Subword Information. *Transactions of the Association for Computational Linguistics*, 5:135–146.
- Albert Costa, Alfonso Caramazza, and Nuria Sebastian-Galles. 2000. The cognate facilitation effect: Implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(5):1283–1296.
- Tamar Degani, Anat Prior, and Natasha Tokowicz. 2011. Bidirectional transfer: The effect of sharing a translation. *Journal of Cognitive Psychology*, 23(1):18–28.
- Tamar Degani and Natasha Tokowicz. 2010. *Ambiguous words are harder to learn*. *Bilingualism: Language and Cognition*, 13(3):299–314.

- Ton Dijkstra, Jonathan Grainger, and Walter J.B. van Heuven. 1999. Recognition of cognates and interlingual homographs: The neglected role of phonology. *Journal of Memory and Language*, 41(4):496 – 518.
- Ton Dijkstra, Koji Miwa, Bianca Brummelhuis, Maya Sappelli, and Harald Baayen. 2010. How cross-language similarity and task demands affect cognate recognition. *Journal of Memory and Language*, 62(3):284 – 301.
- Ton Dijkstra, Henk Van Jaarsveld, and Sjoerd Ten Brinke. 1998. Interlingual homograph recognition: Effects of task demands and language intermixing. *Bilingualism: Language and Cognition*, 1(1):5166.
- Georgiana Dinu, Angeliki Lazaridou, and Marco Baroni. 2015. Improving zero-shot learning by mitigating the hubness problem. In *3rd International Conference on Learning Representations*, pages 1–10, Toulon, France.
- K. Elston-Güttler and J. N. Williams. 2008. L1 polysemy affects L2 meaning interpretation: evidence for L1 concepts active during L2 reading. *Second Language Research*, 24:167187.
- Manaal Faruqui and Chris Dyer. 2014. Improving Vector Space Word Representations Using Multilingual Correlation. In *14th Conference of the European Chapter of the Association for Computational Linguistics*, pages 462–471, Gothenburg, Sweden.
- Karl Moritz Hermann and Phil Blunsom. 2014. Multilingual models for compositional distributed semantics. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 58–68, Baltimore, Maryland. Association for Computational Linguistics.
- Walter J.B. van Heuven, Ton Dijkstra, and Jonathan Grainger. 1998. Orthographic neighborhood effects in bilingual word recognition. *Journal of Memory and Language*, 39(3):458 – 483.
- Noriko Hoshino and Judith F. Kroll. 2008. Cognate effects in picture naming: Does cross-language activation survive a change of script? *Cognition*, 106(1):501 – 511.
- Nan Jiang. 2002. Form-meaning mapping in vocabulary acquisition in a second language. *Studies in Second Language Acquisition*, 24(4):617637.
- N.F. Johnson and K.R. Pugh. 1994. A cohort model of visual word recognition. *Cognitive Psychology*, 26(3):240 – 346.
- Judith F. Kroll and Ton Dijkstra. 2012. The bilingual lexicon. In *The Oxford Handbook of Applied Linguistics*. Oxford University Press.
- Ping Li. 2013. Computational modeling of bilingualism: How can models tell us more about the bilingual mind? *Bilingualism: Language and Cognition*, 16(2):241245.
- Tal Linzen, Emmanuel Dupoux, and Yoav Goldberg. 2016. Assessing the ability of LSTMs to learn syntax-sensitive dependencies. *Transactions of the Association for Computational Linguistics*, 4:521–535.
- Thang Luong, Hieu Pham, and Christopher D. Manning. 2015. Bilingual word representations with monolingual quality in mind. In *Proceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing*, pages 151–159, Denver, Colorado. Association for Computational Linguistics.
- Yevgen Matushevych. 2016. *Learning constructions from bilingual exposure. Computational studies of argument structure acquisition*. Ph.D. thesis, Tilburg University.
- Yevgen Matushevych, Amir Ardalan Kalantari Dehaghi, and Suzanne Stevenson. 2018. Modeling bilingual word associations as connected monolingual networks. In *Proceedings of the 8th Workshop on Cognitive Modeling and Computational Linguistics (CMCL 2018)*, pages 46–56, Salt Lake City, Utah. Association for Computational Linguistics.
- Haitao Mi, Baskaran Sankaran, Zhiguo Wang, and Abe Ittycheriah. 2016. Coverage Embedding Models for Neural Machine Translation. In *2016 Conference on Empirical Methods in Natural Language Processing*, pages 955–960, Austin, Texas, USA.
- Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013a. Efficient Estimation of Word Representations in Vector Space. *arXiv preprint arXiv:1309.4168*.
- Tomas Mikolov, Quoc V Le, and Ilya Sutskever. 2013b. Exploiting Similarities among Languages for Machine Translation. *arXiv preprint arXiv:1309.4168*, (Computation and Language):1–10.
- Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013c. Distributed Representations of Words and Phrases and their Compositionality. In *26th International Conference on Neural Information Processing Systems*, pages 3111–3119, Lake Tahoe, Nevada, USA.
- Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. Glove - Global Vectors for Word Representation. In *2014 Conference on Empirical Methods in Natural Language Processing*, pages 1532–1543, Doha, Qatar.
- Jennifer Rodd, Gareth Gaskell, and William Marslen-Wilson. 2002. Making sense of semantic ambiguity: Semantic competition in lexical access. *Journal of Memory and Language*, 46(2):245 – 266.

- Samuel L. Smith, David H. P. Turban, Steven Hamblin, and Nils Y. Hammerla. 2017. Offline bilingual word vectors, orthogonal transformations and the inverted softmax. In *15th International Conference on Learning Representations*, pages 1–10, Toulon, France.
- Erica Smits, Heike Martensen, Ton Dijkstra, and Dominiek Sandra. 2006. Naming interlingual homographs: Variable competition and the role of the decision system. *Bilingualism: Language and Cognition*, 9(3):281297.
- Anders Søgaard, Zeljko Agić, Héctor Martínez Alonso, Barbara Plank, Bernd Bohnet, and Anders Johannsen. 2015. Inverted indexing for cross-lingual NLP. In *53rd Annual Meeting of the Association for Computational Linguistics and 7th International Joint Conference on Natural Language Processing*, pages 1713–1722, Beijing, China.
- Janet G. Van Hell and Annette M. B. de Groot. 1998. Conceptual representation in bilingual memory: Effects of concreteness and cognate status in word association. *Bilingualism: Language and Cognition*, 1(3):193211.
- Ivan Vulic and Marie-Francine Moens. 2016. Bilingual distributed word representations from document-aligned comparable data. *J. Artif. Int. Res.*, 55(1):953–994.
- Haozhou Wang, James Henderson, and Paola Merlo. 2019. Weakly-supervised concept-based adversarial learning for cross-lingual word embeddings. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing*, Hong Kong.
- Andrea Weber and Anne Cutler. 2004. Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, 50(1):1 – 25.
- John N. Williams. 2014. The bilingual lexicon. In *The Oxford Handbook of the Word*. Oxford University Press.
- Brent Wolter. 2001. Comparing the 11 and 12 mental lexicon. *Studies in Second Language Acquisition*, 23(1):41–69.
- Will Y Zou, Richard Socher, Daniel Cer, and Christopher D. Manning. 2013. Bilingual Word Embeddings for Phrase-Based Machine Translation. In *Conference on Empirical Methods in Natural Language Processing*, pages 1393–1398, Seattle, Washington, USA.