Automatic Learning of Modality Exclusivity Norms with Crosslingual Word Embeddings

Emmanuele Chersoni, Rong Xiang, Qin Lu, Chu-Ren Huang
The Hong Kong Polytechnic University, 11 Yuk Choi Road, Hong Kong (China)
{emmanuelechersoni, xiangrong0302}@gmail.com,
csluqin@comp.polyu.edu.hk, churen.huang@polyu.edu.hk

Abstract

Collecting modality exclusivity norms for lexical items has recently become a common practice in psycholinguistics and cognitive research. However, these norms are available only for a relatively small number of languages and often involve a costly and time-consuming collection of ratings.

In this work, we aim at learning a mapping between word embeddings and modality norms. Our experiments focused on *crosslingual word embeddings*, in order to predict modality association scores by training on a high-resource language and testing on a low-resource one. We ran two experiments, one in a monolingual and the other one in a crosslingual setting. Results show that modality prediction using off-the-shelf crosslingual embeddings indeed has moderate-to-high correlations with human ratings even when regression algorithms are trained on an English resource and tested on a completely unseen language.

1 Introduction

The expression *modality exclusivity norms* refers to a collection of words and their association scores with sensory modalities. The use of modality norms has been a recent trend in psycholinguistic and cognitive science, based on the observation that words with sensory meanings are associated with certain perceptual regions of the brain (Barsalou, 1999; Goldberg et al., 2006; Barros-Loscertales et al., 2012). Words are typically associated with multiple modalities: for example, the word *sweet* can be used to describe sounds, flavours, looks etc. (Connell, 2007; Lynott and Connell, 2009, 2013). Psycholinguistic studies use modality exclusivity norms to control for

the perceptual strength of lexical items used in their experiments, motivating the publication of datasets in which words are rated according to their association with each of the five senses (see the example in Table 1). Moreover, such norms have also been shown to be useful for other NLP tasks, such as metaphor detection (Wan et al., 2020a,b). Normative studies on modality for English words are relatively common (Lynott and Connell, 2009; Juhasz et al., 2011; Lynott and Connell, 2013; Lynott et al., 2019), and similar norms have also been made available for other languages such as French (Bonin et al., 2015), Serbian (Đurđević et al., 2016), Dutch (Speed and Majid, 2017), Russian (Miklashevsky, 2018), Chinese (Chen et al., 2019) and Italian (Vergallito et al., 2020). But in general, the number of languages for which they are available is still limited, and collecting modality norms is a time-consuming process, especially for low-resource languages.

word	Taste	Sight	Sound	Smell	Touch
dress	0.02	4.86	0.4	0.86	4.37

Table 1: Modality norms for the word *dress*.

Our working hypothesis for this study is that the commonalities in human perceptual cognition would lead to reliable automatic induction of modality exclusivity norms for unseen words (i.e. words without experimental data) in the same language. In addition, the five primary sensory modalities are assumed to be universal, and consequently, the prediction of norms in a new language can be carried out with the same procedure. *Crosslingual word embeddings* provide an ideal model for the prediction of modality exclusivity norms in low-resource languages, as they represent words of multiple languages in a shared feature space. It is thus

possible to train a regressor on a resource-rich language, e.g English, and predict modality ratings for words in an unseen language. We experimented with this crosslingual transfer method on Italian, Dutch and Chinese norms. Results show that a) crosslingual embeddings perform similarly to or slightly better than monolingual embeddings in a monolingual setting; b) even after training only on English data, the regressor can predict norms in a totally unseen language with moderate-to-high correlations with human judgements.

2 Related Work

Although word vectors have been a standard for word representations for almost two decades (Lenci, 2018), they became an essential ingredient for most NLP applications only after the introduction of *word embeddings* (Mikolov et al., 2013a; Pennington et al., 2014; Bojanowski et al., 2017). Differently from the first generation models using co-occurrence counting and weighting, word embeddings are estimated via neural network training with the objective of maximizing the probability of the contexts of a target word, and they gained popularity due to the availability of efficient and easy-to-use tools (Mikolov et al., 2013a). The development of research on crosslingual transfer and the availability of new benchmarks for multilingual NLP has recently led to the introduction of the so-called *crosslingual embeddings*, vector space models that represent words from multiple languages through some form of mapping from a monolingual to a multilingual space (Conneau et al., 2018; Ruder et al., 2019). A classical study by Mikolov et al. (2013b) learnt a linear projection to transform the space of a source language to the space of a target language by maximizing the similarity between the two spaces. Other approaches apply Canonical Correlation Analysis to simultaneously project words from two languages into a shared embedding space where the correlation between projected vectors are maximized (Faruqui and Dyer, 2014). Other works make use of the max-margin method such that, for embeddings projected from a source language, they maximize the margin between the correct translations and other candidates (Lazaridou et al., 2015; Joulin et al., 2018). For this study, we use the off-the-shelf crosslingual embeddings by Joulin et al. (2018) based on FastText (Bojanowski et al., 2017)

and trained on Wikipedia. [1]

Despite the success of word embeddings, a common criticism is that they are not grounded in perception, as words are only defined in relation to each other and not to entities and actions in the physical world (Glenberg and Robertson, 2000; Fagarasan et al., 2015; Li and Gauthier, 2017). To address this issue, Fagarasan et al. (2015) used a regression method to map embeddings onto the conceptual properties of the McRae norms (McRae et al., 2005). A similar approach, using feedforward neural networks for predicting properties, was recently described by Li and Summers-Stay (2019). The work by Derby et al. (2019) goes in the opposite direction: instead of predicting norms from embeddings, they combined pretrained vectors and property vectors to inject conceptual knowledge into a new type of word representations. Their *Feature2Vec* system showed a strong performance in the predicting norms of unseen words, compared to previous proposals. Finally, Utsumi (2018, 2020) proposed a similar mapping technique to exploit semantic feature norms by Binder et al. (2016) to analyze the semantic content of word embeddings in terms of neurobiologically-motivated features. Turton et al. (2020) also experimented with embeddings based on Binder features, showing that they can achieve performances comparable to Word2Vec (Mikolov et al., 2013a) and GloVe (Pennington et al., 2014) on similarity datasets.

3 Our Proposed Approach

In this work, we used regressors trained on crosslingual word embeddings to predict modality ratings in two different scenarios. In the *monolingual* scenario, we adopt a 5-fold cross-validation to predict the modality norms of an English dataset (Lynott and Connell, 2009, 2013). In the *crosslingual* scenario, a regressor is trained on a high-resource language, e.g. English, to predict the modality norms of an unseen language.

3.1 Datasets

Four modality norms datasets are used in this work: the English norms by Lynott and Connell (2009,

---

[1] We ran experiments also with the Numberbatch embeddings by Speer and Lowry-Duda (2017), which are obtained by retrofitting different types of word embeddings with a subgraph of ConceptNet (Speer et al., 2017). However, while these vectors showed a strong performance in predicting the norms in the monolingual setting, they never achieved significant correlations with human judgements in the crosslingual prediction, and thus we omitted them from the Results section.

2013) (1002 words); the Italian norms by Vergallito et al. (2020) (1121 words); the Dutch norms by Speed and Majid (2017) (485 words); the Chinese norms by Chen et al. (2019) (291 words). The latter three datasets were collected using a similar methodology, inherited by the original study on English. The questions used for data collection are typically in the following form: "To what extent you experience a (target word) by (sensory modality)". Human participants had to provide a rating from 1 to 7 for each of the five sensory modalities, *taste, sight, sound, smell* and *touch*. All the datasets contain the mean ratings (thus, five modality scores for each word), which are the target variables to be predicted by our embedding-based regressors. The proportion of the dominant modality (i.e. the modality with the highest rating score per word) for each language is depicted in Figure 1. Generally, *sight* is the most represented modality while *smell* is relatively scarce.
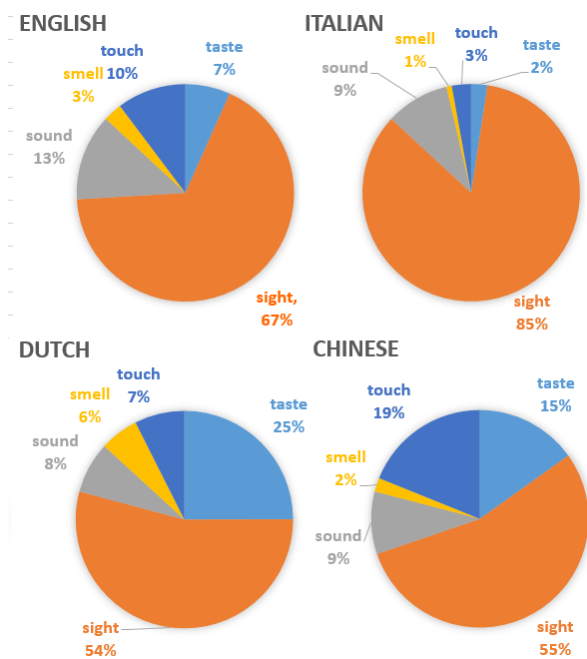

Figure 1: Percentages of dominant modality.

3.2 Models

Using the scikit-learn package (Pedregosa et al., 2011), we tested several regressors on this task. We show the results for the two top-performing ones: Multilayer Perceptron (MLP) and Ridge Regression (RR). As a result of parameter tuning, we adopt the following settings for MLP: 2 hidden layers, respectively of 50 and 10 hidden units , and the identity activation function. All the other

hyperparameters correspond to the default settings of the scikit-learn library.

In all settings, regressors are trained on the embeddings of the words to predict the modality ratings in the original norms. For the *monolingual* scenario, both monolingual and crosslingual embeddings are trained on Wikipedia with the Fast-Text library (Joulin et al., 2018). [2] As a baseline for the monolingual scenario, we also include models trained on random embeddings ($RANDOM$).

3.3 Experiment 1

In the monolingual scenario, we test whether word embeddings are capable to predict modality ratings of English. Although there were previous studies on mapping word embeddings on conceptual properties (Louwerse and Connell, 2011), the prediction on modality norms is not obvious. Firstly, most of those studies predict discrete properties (e.g. whether one concept is primarily experienced through a given sensory modality or not), and not continuous values. Secondly, modality norms represent semantic features that are learned through bodily experience, and it has yet to be tested whether they can be predicted by text-based vectors, to the best of our knowledge.

We used 5-fold cross-validation, by splitting the complete dataset into five sets. In each iteration, we leave one group out and use it as a test set, while training on the instances of all the other groups. Table 2 shows the results for Random, Monolingual and Crosslingual Embeddings, reporting the Spearman correlations, respectively, per modality and per word. [3] Concerning the differences between modalities, it can be observed that both embedding types achieve a correlation above 0.5 on all senses and are well above the random baseline, which never manages to achieve a significant correlation per modality. Smell, the least represented modality in the data, is also the least correlated while Sound and Touch are the easiest to predict. Surprisingly, we can observe that Crosslingual Embeddings perform similarly, and even slightly better than the Monolingual vectors for all modalities.

Figure 2 shows the scores for two best performing regressors, and we can observe that the Multi-

---

[2]The crosslingual embeddings can be found at https://fasttext.cc/docs/en/aligned-vectors.html and the monolingual vectors at https://fasttext.cc/docs/en/pretrained-vectors.html.

[3]Word correlations are computed between vectors of just five values, which is not ideal. Also for this reason, Random embeddings achieve relatively high values for this metric.

34

Embedding	taste	sight	sound	smell	touch	word
random	-0.084	-0.002	-0.003	-0.032	0.017	0.566
monolingual	0.502	0.543	0.652	0.464	0.641	0.781
crosslingual	**0.529**	**0.595**	**0.699**	**0.486**	**0.674**	**0.808**

Table 2: Spearman correlations per modality and per word (average across two best regressors).

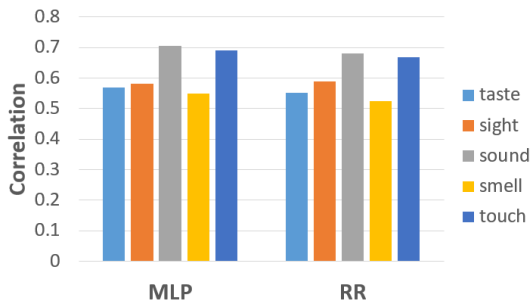layer Perceptron and the Ridge Regression perform similarly, with no significant differences. [4]

Figure 2: Monolingual Performance with Crosslingual Embeddings (Spearman correlation per modality).

Language	Avg Score
Italian	0.668
Dutch	0.663
Chinese	0.546

Table 3: Spearman correlations score per word in the crosslingual setting (scores of the two top regressors have been averaged).

3.4 Experiment 2

In the second experiment, we tested the Crosslingual Embeddings on the prediction of modality norms of Italian, Dutch and Chinese after training only on English data. The summary of the performance per sense modality is given in Figure 3 and per word in Table 3. For the error analysis, we extracted the least correlated words and reported the bottom five for each language in Table 4.

The global performance of the mappers with the Crosslingual Embeddings is not too distant from the monolingual setting. The correlations-by-modality are generally around 0.5, but there are also some notable exceptions. For example, the sound modality for Chinese seems to be particularly difficult to predict. This could be due to differences in the sensory lexicon: European languages like English and Italian have quite a lot of words where the sound is the dominant component (it is

---

the second most common dominant modality after sight), while those are rarer in Chinese (the second rarest modality after smell: see also the percentages in Figure 1). It is also noticeable that, despite being the most frequent dominant modality, sight is never the best predicted one. Actually, sight is the most internally complex modality, and recent proposals for categorizing sensory-related semantics have further divided this sense in several sub-modalities. [5] For this reason, future studies aiming at modeling this modality should probably try to adopt a more fine-grained annotation scheme.

Looking at the correlations-by-word, the values for Chinese are much lower than for the other languages. This was expected: compared to Dutch and Italian, which are both Indo-European languages, Chinese is way more distant from English.
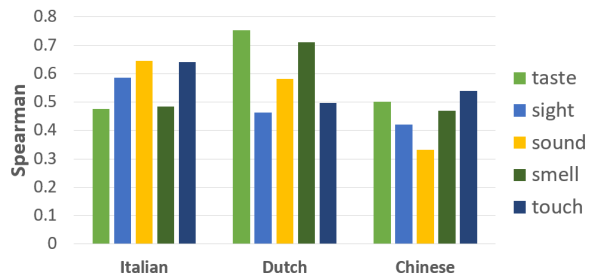

Figure 3: Spearman correlation scores per modality in the crosslingual setting (scores of the two top regressors have been averaged).

In Table 4, we can observe that many of the worst predictions are either words with a taste dominant, which are relatively rare in the English data but more common in the other languages (Dutch and Chinese), or polysemic words with strong associations with multiple senses (e.g. 'good' in Italian and 'sweet' in Chinese). Concerning this last point, we decided to test whether there is a relationship between prediction accuracy and modality exclusivity. Modality Exclusivity (ME) scores are included in the original datasets and are defined as follows:

$$ME(w) = \frac{max(w) - min(w)}{\sum (w)} \quad (1)$$

where $max(w)$ and $min(w)$ are, respectively, the mean ratings of the strongest and of the weakest sense modality for the word w. Scores close to 1 indicate that the concept described by the word is experienced only through one sensory modality,

---

5E.g. Binder et al. (2016) identifies 15 different vision-related meaning components, each one associated with a distinct neural processing system

Dataset	Error Cases
Italian	buono ('good', taste), pizza ('pizza', taste), affogare ('to drown', sight), buono ('good', smell), puzza ('stink', smell)
Dutch	parmezaan ('parmesan', taste), openhaard ('fireplace', sight), dragon ('dragon', taste), zalm ('salmon', taste), roosmarijn ('rose marin', smell)
Chinese	雜 ('mixed taste', sound), 口氣 ('tone', smell) 火鍋 ('hotpot', smell), 甘 ('sweet', taste), 苦 ('bitter', taste)

Table 4: Words with lowest correlations (English translation and worst predicted modality in brackets).

Language	Translated Words
Italian	148/1121 (13.2%)
Dutch	46/485 (9.48%)
Chinese	108/291 (37.1%)

Table 5: Number and percentage of words that are translations of English items for each dataset.

while multimodal concepts are typically associated with lower scores. Strongly multimodal concepts might be more difficult to predict, as their scores for the five modalities are generally closer than in the unimodal concepts. We tested this hypothesis by measuring the Spearman correlation between the word correlations and the modality exclusivity scores from the original dataset, but no strong evidence was found: the models showed no significant correlation for the Italian data, while finding positive weak correlations (between 0.2 and 0.3) on the Dutch and on the Chinese data.

We also needed to check whether words that are direct translations of words in the English data are predicted better than the others. The number of words translated from English and the percentages can be seen in Table 5, and in the Chinese dataset they represent more than 37% of the dataset items. A high number of these words might inflate the evaluation scores, as the models could be just memorizing the English representations, and the crosslingual transfer would be working just because the vectors are well-aligned in the target language. However, in our analysis we did not find evidence for this: we compared the word correlation scores of translations with the other items by means of a Mann-Whitney U test, without finding any significant difference for any of the datasets, with just a single exception (the Ridge Regression model on Dutch data, where translations have significantly higher scores at $p < 0.05$). In conclusion, neither modality exclusivity nor the number of translations had a big impact on our results.

4 Conclusions

In this paper, we have proposed the first study dedicated to the prediction of modality norms via word embeddings mapping. We experimented with crosslingual embeddings in order to assess the potential for crosslingual transfer.

Our results first showed that modality norms can be reliably predicted even by purely text-based vectors. This is in accordance with cognitive hypotheses claiming that various aspects of experiential information are redundantly encoded in linguistic expressions (e.g. the Symbol Interdependency Hypothesis) (Louwerse, 2008; Riordan and Jones, 2011). Moreover, crosslingual vectors turned out to be better performing than their monolingual counterparts. This result is certainly surprising, although it is unclear whether it is due to differences in the training data, or the mapping itself benefits performance by abstracting away from language-specific patterns to a more 'conceptual' space.

Even more importantly, given the availability of crosslingual embeddings for a low-resource language, it is possible to train a regressor on a high-resource language (e.g. English) and to predict the norms for the low-resource one. In our experiments, we obtained moderate-to-high correlations even in the crosslingual setting. We think this is potentially a very useful application for the research on modality norms.

In this first study we used a relatively simple methodology, but several refinements are possible for improving the prediction quality. Two possible directions would be, firstly, to exploit the presence of words that are direct translations from English to the other languages to apply retrofitting techniques (Faruqui et al., 2015; Mrkšić et al., 2016, 2017; Vulić et al., 2018) to the crosslingual space, and secondly, to tackle the task by introducing more advanced neural architectures for the representation of words in context, e.g. multilingual transformers (Devlin et al., 2019; Pires et al., 2019).

References

Alfonso Barros-Loscertales, Julio González, Friedemann Pulvermüller, Noelia Ventura-Campos, Juan Carlos Bustamante, Víctor Costumero, María Antonia Parcet, and César Ávila. 2012. Reading Salt Activates Gustatory Brain Regions: fMRI Evidence for Semantic Grounding in a Novel Sensory Modality. *Cerebral Cortex*, 22(11):2554–2563.

Lawrence W Barsalou. 1999. Perceptual Symbol Systems. *Behavioral and Brain Sciences*, 22(4):577–660.

Jeffrey R Binder, Lisa L Conant, Colin J Humphries, Leonardo Fernandino, Stephen B Simons, Mario Aguilar, and Rutvik H Desai. 2016. Toward a Brain-Based Componential Semantic Representation. *Cognitive Neuropsychology*, 33(3-4):130–174.

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.

Patrick Bonin, Alain Méot, Ludovic Ferrand, and Aurélia Bugaïska. 2015. Sensory Experience Ratings (SERs) for 1,659 French Words: Relationships with Other Psycholinguistic Variables and Visual Word Recognition. *Behavior Research Methods*, 47(3):813–825.

I-Hsuan Chen, Qingqing Zhao, Yunfei Long, Qin Lu, and Chu-Ren Huang. 2019. Mandarin Chinese Modality Exclusivity Norms. *PloS One*, 14(2).

Alexis Conneau, Guillaume Lample, Marc'Aurelio Ranzato, Ludovic Denoyer, and Hervé Jégou. 2018. Word Translation Without Parallel Data. In *Proceedings of ICLR*.

Louise Connell. 2007. Representing Object Colour in Language Comprehension. *Cognition*, 102(3):476–485.

Steven Derby, Paul Miller, and Barry Devereux. 2019. Feature2Vec: Distributional Semantic Modelling of Human Property Knowledge. In *Proceedings of EMNLP*.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of NAACL 2019*.

Luana Fagarasan, Eva Maria Vecchi, and Stephen Clark. 2015. From Distributional Semantics to Feature Norms: Grounding Semantic Models in Human Perceptual Data. In *Proceedings of IWCS*.

Manaal Faruqui, Jesse Dodge, Sujay K Jauhar, Chris Dyer, Eduard Hovy, and Noah A Smith. 2015. Retrofitting Word Vectors to Semantic Lexicons. In *Proceedings of NAACL*.

Manaal Faruqui and Chris Dyer. 2014. Improving Vector Space Word Representations Using Multilingual Correlation. In *Proceedings of EACL*.

Arthur M Glenberg and David A Robertson. 2000. Symbol Grounding and Meaning: A Comparison of High-Dimensional and Embodied Theories of Meaning. *Journal of Memory and Language*, 43(3):379–401.

Robert F Goldberg, Charles A Perfetti, and Walter Schneider. 2006. Perceptual Knowledge Retrieval Activates Sensory Brain Regions. *Journal of Neuroscience*, 26(18):4917–4921.

Armand Joulin, Piotr Bojanowski, Tomas Mikolov, Hervé Jégou, and Edouard Grave. 2018. Loss in Translation: Learning Bilingual Word Mapping with a Retrieval Criterion. In *Proceedings of EMNLP*.

Barbara J Juhasz, Melvin J Yap, Joanna Dicke, Sarah C Taylor, and Margaret M Gullick. 2011. Tangible Words Are Recognized Faster: The Grounding of Meaning in Sensory and Perceptual Systems. *Quarterly Journal of Experimental Psychology*, 64(9):1683–1691.

Angeliki Lazaridou, Georgiana Dinu, and Marco Baroni. 2015. Hubness and Pollution: Delving into Cross-Space Mapping for Zero-Shot Learning. In *Proceedings of ACL*.

Alessandro Lenci. 2018. Distributional Models of Word Meaning. *Annual Review of Linguistics*, 4:151–171.

Dandan Li and Douglas Summers-Stay. 2019. Mapping Distributional Semantics to Property Norms with Deep Neural Networks. *Big Data and Cognitive Computing*, 3(2):30.

Lucy Li and Jon Gauthier. 2017. Are Distributional Representations Ready for the Real World? Evaluating Word Vectors for Grounded Perceptual Meaning. In *Proceedings of the ACL Workshop on Language Grounding for Robotics*.

Max Louwerse. 2008. Embodied Relations Are Encoded in Language. *Psychonomic Bulletin & Review*, 15(4):838–844.

Max Louwerse and Louise Connell. 2011. A Taste of Words: Linguistic Context and Perceptual Simulation Predict the Modality of Words. *Cognitive science*, 35(2):381–398.

Dermot Lynott and Louise Connell. 2009. Modality Exclusivity Norms for 423 Object Properties. *Behavior Research Methods*, 41(2):558–564.

Dermot Lynott and Louise Connell. 2013. Modality Exclusivity Norms for 400 Nouns: The Relationship between Perceptual Experience and Surface Word Form. *Behavior Research Methods*, 45(2):516–526.

Dermot Lynott, Louise Connell, Marc Brysbaert, James Brand, and James Carney. 2019. The Lancaster Sensorimotor Norms: Multidimensional Measures of Perceptual and Action Strength for 40,000 English Words. *Behavior Research Methods*, pages 1–21.

Ken McRae, George S. Cree, Mark S. Seidenberg, and Chris McNorgan. 2005. Semantic Feature Production Norms for a Large Set of Living and Nonliving Things. *Behavior Research Methods*, 37(4):547–559.

Alex Miklashevsky. 2018. Perceptual Experience Norms for 506 Russian Nouns: Modality Rating, Spatial Localization, Manipulability, Imageability and Other Variables. *Journal of Psycholinguistic Research*, 47(3):641–661.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013a. Efficient Estimation of Word Representations in Vector Space. *arXiv preprint arXiv:1301.3781*.

Tomas Mikolov, Quoc V Le, and Ilya Sutskever. 2013b. Exploiting Similarities Among Languages for Machine Translation. *arXiv preprint arXiv:1309.4168*.

Nikola Mrkšić, Diarmuid O 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*.

Nikola Mrkšić, Ivan Vulić, Diarmuid Ó Séaghdha, Ira Leviant, Roi Reichart, Milica Gašić, Anna Korhonen, and Steve Young. 2017. Semantic Specialization of Distributional Word Vector Spaces Using Monolingual and Cross-lingual Constraints. *Transactions of the Association for Computational Linguistics*, 5:309–324.

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global Vectors for Word Representation. In *Proceedings of EMNLP*.

Telmo Pires, Eva Schlinger, and Dan Garrette. 2019. How Multilingual is Multilingual BERT? In *Proceedings of ACL*.

Brian Riordan and Michael N. Jones. 2011. Redundancy in Perceptual and Linguistic Experience: Comparing Feature-Based and Distributional Models of Semantic Representation. *Topics in Cognitive Science*, 3(2):303–345.

Sebastian Ruder, Ivan Vulić, and Anders Søgaard. 2019. A Survey of Cross-lingual Word Embedding Models. *Journal of Artificial Intelligence Research*, 65:569–631.

Laura J Speed and Asifa Majid. 2017. Dutch Modality Exclusivity Norms: Simulating perceptual Modality in Space. *Behavior Research Methods*, 49(6):2204–2218.

Robyn Speer, Joshua Chin, and Catherine Havasi. 2017. Conceptnet 5.5: An Open Multilingual Graph of General Knowledge. In *Proceedings of AAAI*.

Robyn Speer and Joanna Lowry-Duda. 2017. Conceptnet at Semeval-2017 task 2: Extending Word Embeddings with Multilingual Relational Knowledge. In *Proceedings of SemEval*.

Jacob Turton, David Vinson, and Robert Elliott Smith. 2020. Extrapolating Binder Style Word Embeddings to New Words. In *Proceedings of the LREC Workshop on Linguistic and Neurocognitive Resources*.

DF Đurđević, M Popović Stijačić, and J Karapandžić. 2016. A Quest for Sources of Perceptual Richness: Several Candidates. *Studies in Language and Mind*, pages 187–238.

Akira Utsumi. 2018. A Neurobiologically Motivated Analysis of Distributional Semantic Models. In *Proceedings of CogSci*.

Akira Utsumi. 2020. Exploring What Is Encoded in Distributional Word Vectors: A Neurobiologically Motivated Analysis. *Cognitive Science*, 44(6):e12844.

Alessandra Vergallito, Marco Alessandro Petilli, and Marco Marelli. 2020. Perceptual Modality Norms for 1,121 Italian Words: A Comparison with Concreteness and Imageability Scores and an Analysis of their Impact in Word Processing Tasks. *Behavior Research Methods*, pages 1–18.

Ivan Vulić, Goran Glavaš, Nikola Mrkšić, and Anna Korhonen. 2018. Post-specialisation: Retrofitting Vectors of Words Unseen in Lexical Resources. In *Proceedings of NAACL*.

Mingyu Wan, Kathleen Ahrens, Emmanuele Chersoni, Menghan Jiang, Qi Su, Rong Xiang, and Chu-Ren Huang. 2020a. Using Conceptual Norms for Metaphor Detection. In *Proceedings of the ACL Workshop on Figurative Language Processing*.

Mingyu Wan, Baixi Xing, Pengyuan Liu, Qi Su, and Chu-Ren Huang. 2020b. Enhancing Metaphor Detection with Sensorimotor Scores. In *Proceedings of COLING*.