Mining Cultural Differences of Named Entities from Large Text

Abstract

Opinions toward famous persons, places and organizations can be very different from culture to culture. Recognizing such cultural differences is important in effective cross-lingual and cross-cultural communications. In this paper, we present a framework that automatically computes such cultural differences of a given entity between English and Chinese by mapping it into two embedding vector spaces. The two embeddings are made comparable by either learning a linear transformation model, or by merging the two spaces into a higher-dimensional space based on word-to-word translations. Our experiments showed that these methods outperform the baseline methods by substantial margins in detecting culturally different named entities.

1 Introduction

Opinions about a certain named entity, such as a famous person, a world-wide organization, or a place, may differ from culture to culture. For example, Kashmir is a large mountainous region on the China-India border. Due to decades of border disputes between China and India about that region, to the Chinese people, this region is synonymous to military conflicts and political struggles. On the contrary, that same region is considered as a picturesque travel destination by the westerners due to its perfect location in the Himalayas, since the border dispute between China and India is hardly their concern. This type of cultural differences are evident from the most popular images about Kashmir on the English and Chinese search engines (Figure 1).

Another type of cultural differences exist when an entity is interesting to one culture but largely ignored by another culture. Figure 2 shows such an example where Indian People's Party (a.k.a. Bharatiya Janata Party or BJP), currently the largest political party in India, shows up frequently on Chinese web with vivid rally images, because India is a major world power that sits next to China. On the other hand, the most popular English web images about BJP are merely party logos, suggesting very little public interest about this entity from the English world.

Our goal in this paper is to identify an entity with significantly different cultural understanding, which can contribute to applications such as instant messenger or machine translator, to avoid culturally sensitive mentions or translations. Apart from these applications, a list of such entities with cultural differences in its own right is a valuable resource for multicultural studies. However, understanding subtle cultural differences requires not only perfect understanding of the two languages, but also devouring large volumes of bilingual texts to sufficiently observe how they are mentioned in each culture and how they differ.

Figure 1: Popular Images about Kashmir on Chinese web (top) and English web (bottom)

Figure 2: Popular Images about BJP on Chinese web (top) and English web (bottom)

We propose to transform this problem into a computational task, by proposing a quantitative evaluation metric measuring the cultural similarity between two cultures of a given named entity. To calculate such scores, we proposed two algorithms to compute the cultural similarity scores based on the quantitative representation for the semantic meaning of all words in mono-lingual corpus. First solution leverages word-embedding techniques such as word2vec, to mine the notion of similarity from English and Chinese corpus respectively. We then connect the two results using linear transformation so that we can directly compute the cosine similarity between the English name and the Chinese name of a certain named entity.

Another way is to construct a higher-dimensional vector space. Every dimension of this space is representing a pair of words, which are an English word and its corresponding Chinese translation. We call this space "translation space". The values in the English name vector of a certain entity are the cosine similarities between this entity's English name and each of other English words. We similarly compute the Chinese name vector. Because an English word may has many different Chinese translations, we duplicate the cosine similarities into several dimensions in translation space. After constructing this new comprehensive vector space, we can simply calculate cosine similarity between the English and Chinese vector of a certain entity.

This paper makes the following contributions:

- to the best of our knowledge, we are the first to study the problem of mining cultural differences of named entities and to present a data-driven approach to solve it;
- we applied our approach on large news corpora in English and Chinese and computed the cultural difference scores for 4,212 named entities including persons, locations and organizations from Wikipedia;
- our evaluation shows that the computed scores correlated well with human perception of cultural differences, with an average precision of 0.617 and 0.612, outperforming a strong baseline method using the popularities of entities by significant margins.

2 Approach

Our overall approach is illustrated in Figure 3. Our starting point is a set of cross-lingual named entities harvested from English and Chinese Wikipedia/Wikidata, as well as a set of translation pairs of ordinary words from Bing translator API (Section 2.1). These two sets serve as our vocabulary. Then we conduct entity linking, of connecting named entities harvested with text mentions in the English and Chinese corpora (Section 2.2). This step enables to understand named entities in the distributional semantic space, by creating English and Chinese word vector spaces respectively using skip-gram model, for both named entities an ordinary words. Finally, in these two separate methods (Section 2.4), we compute the cultural similarity scores for each cross-lingual entity pair by either linearly transforming words from the Chinese vector space into English or by merging the two spaces into a new, higher-dimension translation space. Next, we explain each step in more details.

2.1 Vocabulary Building

Our vocabulary has two parts: i) a set of named entities of interest drawn from a standard ontology and ii) a set of ordinary English-Chinese word pairs. The obvious choice of

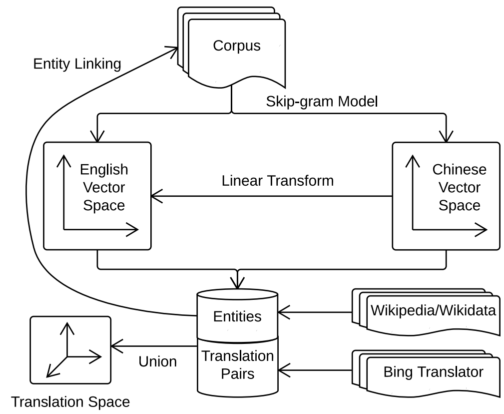


Figure 3: Overall Workflow

this ontology is Wikipedia which keeps an unique identifier for every documented named entity. Many of these entities in Wikipedia have both English and Chinese instances and thus make up the first part of the vocabulary. The ordinary words from the two languages can be connected through online dictionary or translation APIs. We discuss each step in further details below.

Named Entities We focus on three categories of named entities, namely people, locations and organizations. We ensure that an entity is a person if it belongs to the Wikipedia category "Births by year" [1]. We consider an entity to be a location, if its Wikipedia page contains longitude-latitude coordinates. An entity is considered as an organization, if it appears under the subcategories of "Organization" in Wikidata while it carries a Wikipedia page. We use the interlanguage links offered by Wikipedia to make sure all named entity exist both in English and Chinese Wikipedia.

Translation Pairs To construct the set of translation pairs of ordinary words, we first collect common English words from a large lemmatized English corpus and translate these words into Chinese translations using online dictionary and translation APIs, specifically Bing API [2] in this work. As each English word can be translated into multiple Chinese words, and a Chinese into multiple English words, this phase generates a many-to-many translation mapping.

2.2 Entity Linking

The purpose of entity linking is to link mentions of entities of our interest in a large text corpus to our vocabulary. This enables to project a named entity in the distributional semantic space, together with ordinary words.

However, this step is tricky, as the same entity is represented as many surface forms: For example, both "William Jefferson Clinton" and "Clinton" may refer to entity "Bill Clinton" while "Clinton" may also refer to "Hillary Clinton." Though entity linking has been actively studied to ad-

[1] `https://en.wikipedia.org/wiki/Category:Births_by_year`

[2] `http://www.bing.com/translator`

dress this problem, we could not adopt existing solution "as is", due to the following two reasons: First, pursuing the balance of precision and recall, they tend to generate too many links for our task requiring high precision. Second, existing approaches building on linguistic signals cannot apply equally to English and Chinese entity linking.

We devise a simple alternative achieving the comparable precision target. We obtain the set of possible surface forms for an entity by collecting all anchor texts of the entity in the Wikipedia corpus. In addition, we leverage a redirect system of Wikipedia, redirecting "UK" to "United Kingdom" in Wikipedia. We merge all entities that redirect to each other as one entity. We can also compute anchor-entity linking frequency from the Wikipedia corpus.

Our entity linking algorithm starts by looking for potential anchors in the plain text corpus. We adopt a longest match strategy here that prefers longer anchors. This is because we assume longer anchors are more reliable and bring about higher precision. In Entity linking, we aim for high precision rather than recall because even if an entity is not recognized in the text, its constituent words will still be captured later in the ordinary word vector space and contribute to the semantics of other entities or words.

For a recognized anchor a in the plain text, typically this string has been linked to multiple entities in Wikipedia pages before. All these linked entities are considered candidates C for linking a presently. We conduct the following checks to select an entity in C to link. First, if a is identical to the Wikipedia title of some entity $e \in C$, then we directly link a to e. Second, if $e \in C$ has already been linked to another anchor in the current text, then we link a to e. If multiple e in C have been linked already in the text, pick the nearest one in the text. Otherwise, we link a to an e in C if e satisfying three conditions:

$$\frac{freq(a,e)}{\sum_{c_i \in C} freq(a,c_i)} > 0.5 \tag{1}$$

$$freq(a,e) \geq 5 \tag{2}$$

$$|C| \leq 5 \tag{3}$$

where $freq(a,e)$ is the number of times a is linked to e in Wikipedia corpus. The first condition holds when e is a dominant sense (link) for a, which means there is a high prior probability that e is the right entity for a. The second condition ensures that a links to e considerable number of times in Wikipedia, which filter out mistakes by editors and rare anchors. The last condition is designed for anchors that are highly ambiguous. For example, most of the time, "Singh" refer to "Manmohan Singh" in Wikipedia, the former Prime Minister of India, we can not confidently link "Singh" to "Manmohan Singh" when we first see the anchor "Singh" in a plain text article because "Singh" is a very common Indian name that can be linked to many entities. But if there is already an anchor that links to "Manmohan Singh" in the same article, the confidence is much higher. In practice, the above heuristics have been shown to significantly enhance the precision of entity linking.

2.3 Skip-gram Model

Entity linking enables us to adopt Skip-gram model introduced by Mikolov et al. (2013) to train word embeddings to include both named entities and ordinary words, from English and Chinese corpus separately. After this step, two vector representations are generated for every word from English and Chinese vocabulary. Note that both ordinary words and entities are considered "words" here.

Skip-gram is an unsupervised learning model where the basic idea is to predict co-occurring words in a corpus.

In our experiments, we set the window size as 5 and the size of "input" and "output" vectors to 150. Further we discard any word that occurs less than 5 times in the whole corpus and use negative sampling.

After we train this model on each monolingual corpus, vector representations are generated for every word and every entity from our vocabulary. We thus have two vector spaces: one for English corpus and the other for Chinese corpus.

2.4 Cultural Similarity Computation

Next we introduce two algorithms for computing cultural similarity between the English vector and the Chinese vector of the same entity. The cultural difference can then be readily induced from the similarity.

Linear-transformation Algorithm English and Chinese vector spaces trained from the Skip-gram model are not directly comparable due to unknown meaning of each dimension. However, Mikolov et al. (2013) have shown that the relationship between these vector spaces can be captured by rotation and scaling, represented by a linear transformation matrix W. In this paper, we borrow this idea and train this matrix using a number of human annotated "seed entities" with *little* cultural difference and using the following optimization problem:

$$\underset{W}{\operatorname{argmin}} \sum_{i=1}^{n} ||Wx_i - t_i||^2, \tag{4}$$

where x_i is a word in Chinese while t_i is its corresponding translation in English and n is the size of training samples.

Thus we train a linear transformation matrix from Chinese to English spaces and map each Chinese word vector to the English space. This transform enables to visualize English and Chinese vectors in the same coordinate. Figure 4 shows an illustrative example of how we linearly transformed the embedding space of one language to match with that of another language. This example shows that, after the transformation, both Chinese and English word vectors are in the same coordinate, while the angle between Chinese and English version of "Dalai Lama" is larger than that between Chinese and English version of "Roger Federer". This suggests that Dalai Lama, a controversial political figure, has a larger cultural difference in Chinese and English, than Roger Federer, a famous tennis player.

Despite the power of linear transformation, its performance strongly depends on the quality of the seed entities. However, obtaining high quality seed entities requires time

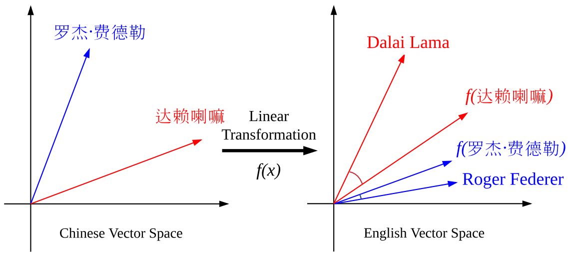


Figure 4: Linear transformation from Chinese to English

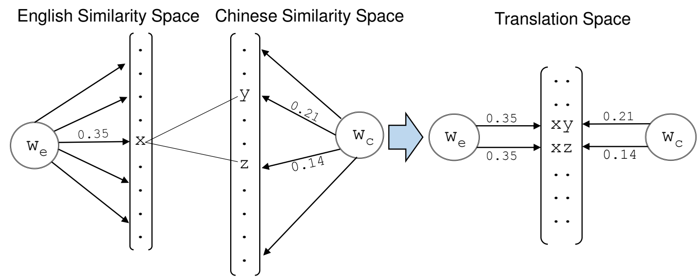


Figure 5: Example for Translation Space Algorithm

and bilingual annotators. We thus propose an alternative unsupervised approach, called *translation space algorithm*.

Translation Space Algorithm This section combines English and Chinese semantic spaces into a rich and higher dimensional space , leveraging many-to-many translation pairs created in Section 2.1.

Specifically, we first represent each entity in a word vector space by its cosine similarity with all words or entities in the same space, including itself. Suppose we want to compute the cultural similarity score for a pair of entities w_e and w_c in the English and Chinese vector spaces respectively. We first represent w_e by a *similarity vector* of size l_e where l_e is the total number of words/entities in the English space, and each dimension of this vector is the cosine similarity between w_e and all words in English. The cosine between w_e and itself is 1. We represent w_c similarly. Because the English and Chinese vocabularies are of different sizes, these two similarity vectors are of different sizes, too. Furthermore, since the translation is many to many, the two vectors are not directly comparable.

Our solution is to "expand" these two vector spaces in a higher dimensional space, where each dimension represents a translation from one English word to the corresponding Chinese word. As such, the new space, known as the "translation space", is k-dimensional, where k is equal to the total number of translation pairs or edges between the two vocabularies. In Figure 5, as an example, consider a word x in the similarity vector of w_e. If x is translated to y and z in Chinese, without prior information, we assume x is translated to y and z with equal probability. As a result, the dimension for x is then expanded into two new dimensions, namely xy and xz, where each dimension stores the same value as the value for x.

At this point, the similarity vectors of w_e and w_c are mapped to the new translation space and are now comparable. Now we can calculate the cosine similarity between w_e and w_c pairs in the translation space as the cultural similarity score between the two entities.

3 Evaluation

This section evaluates the performance of our approach to mine cultural differences of named entities from large text. First of all, we introduce the details in building the English and Chinese corpus. Then, we present how we construct the ground truth from human annotation. Finally, we report the evaluation metrics we use to evaluate our approach as well as the experiment results.

3.1 Data Preparation

To build English corpus, we crawled news articles from Daily Mail[3] and New York Times[4] published between Jan 1st, 2012 to Aug 5th, 2016, for these two sources are among the most representative news media of western cultures. Similarly, we crawled China News[5] and iFeng News[6] in the same time period to build our Chinese corpus. In total, there are 1,857,581 English news and 673,655 Chinese news. An average English news has 558.2 words while the average length of a Chinese news is 507.3 words.

Our vocabulary contains 45,740 English terms and 47,854 Chinese terms, including 4,212 terms representing the named entities common to two term sets. For the purpose of implementing the Translation Space algorithm, we build 122,284 translation pairs between the two term sets.

3.2 Ground Truth

As shown in Section 1, cultural differences of a given entity is visible from the most popular images in the image search results. It is because that people from different cultures have different views on the same entity so the kind of images that they search or create on the Internet are very different, too. With the help of the online image search engine such as Bing, we can get the most interesting images of a given named entity in western culture with Bing's global site [7] and in Chinese culture with its Chinese site [8].

Thus we obtain manual labels of 885 named entities by showing human annotators the top 20 pictures of a certain named entity from global Bing image search and the Chinese Bing image search respectively. This set of 885 entities is the intersection of the 2000 most frequent entities in the English corpus and Chinese corpus respectively.

We invited 14 annotators from different cultures (both Chinese and international students) to judge whether the two

[3]http://www.dailymail.co.uk/
[4]http://www.nytimes.com/
[5]http://www.chinanews.com/
[6]http://news.ifeng.com/
[7]http://global.bing.com
[8]http://cn.bing.com

sets of image search results of a given entity are visually different, without considering the actual meaning of the entity.

We choose 497 entities that most annotator agree on as our evaluation ground truth. The inter-annotator agreement by Cohen's kappa coefficient among these annotator is 0.6. Among these annotated pairs, we set aside 100 entities for which all annotators consider culturally similar. These are used as the training set for the linear transformation model. Consequently, our final test dataset consists of 397 entities, out of which 173 are labeled as culturally different and 224 labeled as culturally similar.

3.3 Baselines

We compare our approach with two baseline methods:

Biased Random Classifier To judge whether a named entity is culturally different or not is actually a classification problem. Thus, a biased random classifier with a prior probability computed by the ratio of the number of culturally different entities to the total number of entities in the training data can be a simple baseline.

Ranking by Popularity A stronger baseline is to rank the entities in test dataset by the sum of the relative frequency of this entity in the English corpus and Chinese corpus. The reason why it is stronger than the former one is that when a named entity is more likely to occur in different cultures, it has more chance to be viewed in different ways. If an object is not very common in different cultures, it has almost no opportunity to be exposed to multiple cultural views. Based on this assumption, we consider this baseline is a stronger competitor to our two algorithms.

3.4 Experimental Results

Entity Linking Accuracy In order to see the performance of our entity linking method, we randomly sampled 50 pieces of English news and Chinese news respectively. On the whole, there are 1,530 links in English samples, 50 among which are incorrect. Chinese samples contain 436 links and there are 23 errors. We thus achieve accuracies of **96.7%** and **94.7%**, respectively.

Word-embedding Results To evaluate the correctness of our word embedding results of entities, we show qualitative results of high cosine similarity neighbors. To illustrate, Table 1 shows the top similar entities of "Adolf Hitler" in the two cultures, including similar semantic information with Benito Mussolini, Nazi Germany and the word "dictator".

Precision, Recall and F1-score at Top k Figure 6a reports quantitative comparison of our two algorithms with the two baselines. Note the accuracy of Expected Random Classifier baseline is fixed as $173/379$ and its recall-at-k as $k/379$, shown as a dotted line. In the figure, our two algorithms consistently outperform the two baselines, until k reaches 150 where all algorithms converge. Translational space performs comparably to Linear transform requiring seed annotation, and even outperforms when $k < 20$ or $k > 100$.

Our algorithms, focusing on precision, are comparable in terms of recall with baselines as shown in Figure 6b, such that in terms of F1-measure, we outperform the baselines in Figure 6c.

Table 2 reports the average precision (Schütze 2008). Biased random as a baseline achieves 0.456, which is improved by our two proposed algorithms by 35.3% and 34.2% respectively.

Table 1: Top 7 cosine similar terms (except itself) to the named entity "Adolf Hitler". (The Chinese terms in the table have already been translated into English. The words in italic are named entities in our vocabulary.)

English Space	Cosine	Chinese Space	Cosine
Hitler	0.929	*Nazi Germany*	0.869
Benito Mussolini	0.827	Nazi	0.811
Fuhrer	0.817	*Nazi Party*	0.769
Stalin	0.798	Napoleon	0.753
Nazi Germany	0.790	Stalin	0.729
Nazi	0.774	*Benito Mussolini*	0.716
Heinrich Himmler	0.751	dictator	0.704

Table 2: Average Precision

Method	AP
Biased Random	0.456
Popularity Ranking	0.543
Linear Transform	0.612
Translation Space	**0.617**

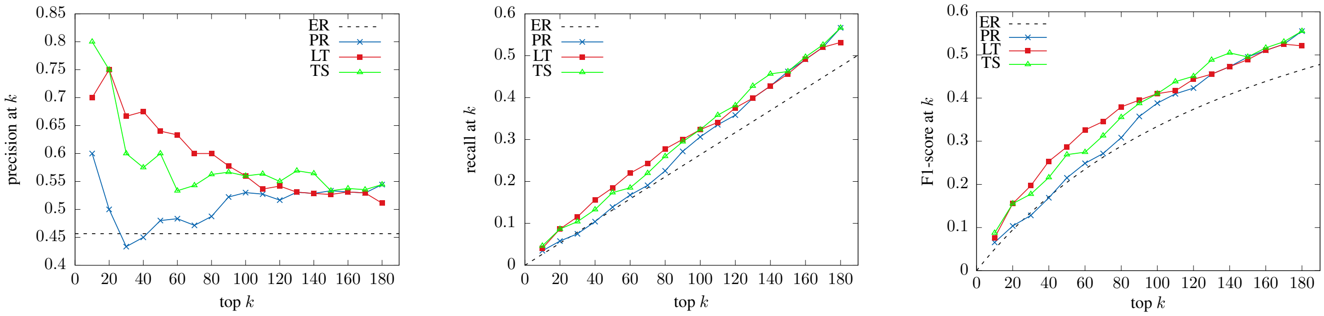
The Most Culturally Different Entities Table 3 shows the most culturally different entities we mined from our two algorithms. As discussed in Section 1, entities in the list include entities located in China or neighboring countries (e.g., Bihar, Sichuan, CCTV and Korean Central News Agency), for which the volume of interests is significantly different in the two cultures. In the case of more common entities such as Beijing, it carries more political connotations for the westerners but is instead more of a cultural and geographic landmark for the Chinese people, which shows different directions of interest.

Table 3: Most culturally different named entities.

Linear Transform	Translation Space
Bihar	Baltimore
Sichuan	Human Rights Watch
Gujarat	APEC
China Central Television	Beijing
West Bengal	Greenpeace
Madhya Pradesh	China Central Television
Korean Central News Agency	Korean Central News Agency
Bharatiya Janata Party	African Union

4 Related Work

The field of intercultural research has received significant cross-fertilization from many academic disciplines suggested by (Triandis 1994). For instance, (Weber and Hsee



(a) Precision at top k (b) Recall at top k (c) F1-score at top k

Figure 6: The precision, recall and F1-score at top k of the 4 methods. (ER=Expected Random, PR=Popularity Ranking, LT=Linear Transform, TS=Translation Space)

1998) demonstrates that apparent differences in risk preference exist in different cultures.

EL(Entity Linking) is a hot topic in these years. The improvement of EL technique enhances a big range of issues such as name transliteration (Lin et al. 2016). Solutions for EL problem usually have to trade off between precision and recall. In addition to traditional local and global models, recently (Gruetze et al. 2016) proposes an efficient linking method that uses a random walk strategy to combine a precision-oriented and a recall-oriented classifier. Probabilistic approach making use of an effective graphical model(Ganea et al. 2016) is also a good solution for EL. Comparing with these state-of-art models, our method is much simpler but achieve a reasonable performance in our task.

Cognitive linguistic studies (Kovecses 2006) have shown that equivalent terms in different languages may have very different meanings. This phenomenon proves to hold between English and Chinese (Chen 2007; Tavassoli 1999; Krifka 1995). In computational linguistics, discovery of relationships across languages is an emerging topic. There are generally two research directions: graph-based knowledge network or distribution-based vector representation.

BabelNet (Navigli and Ponzetto 2012) and Yago3 (Mahdisoltani, Biega, and Suchanek 2014) are representatives of graph-based knowledge network, with an ambition to construct a unified multilingual knowledge base. They integrate resources such as WordNet and Wikipedia to achieve this goal. The knowledge base thus built can be used to calculate relatedness across languages. However, both of them rely on existing structured resources to create the networks, which limit their scale and extendability.

For distributional models, the predominant approach to represent the semantics of words is word embedding. The embeddings are usually trained using co-occurrence matrix, matrix factorization (Lebret and Collobert 2013; Levy and Goldberg 2014; Li et al. 2015) or neural network (Mikolov et al. 2013). Traditionally, these vectors are trained on monolingual corpus and the vector spaces of different languages are not directly comparable with each other. To solve this, some researchers try to train unified representations from multilingual corpus (Klementiev, Titov, and Bhattarai 2012; Hermann and Blunsom 2014; Vulic and Moens 2015) or construct a mapping between the vector spaces of different languages (Mikolov, Le, and Sutskever 2013). These vectors are then evaluated in tasks such as bilingual lexicon induction or cross-lingual word sense disambiguation, and have shown to achieve state-of-art performance.

Our task is similar to bilingual lexicon induction, though we want to detect differences for named entities instead of finding similar words. Tomas Mikolov (Mikolov, Le, and Sutskever 2013) shows the potential to detect errors in bilingual dictionary with Word2Vec and linear transformation among different vector spaces. In this paper, we implement their idea (linear transformation) and compare with ideas of ours.

5 Conclusion

In this paper, we develop a framework to compute cultural differences between named entities from English and Chinese news articles. Our first conclusion is that the accuracy of entity linking performance is extremely crucial to our results. Our strategy of aiming for high precision rather than recall pays off. The second conclusion is that popular images on search engines for different cultures are a good source for detecting cultural difference. We use these images to create the ground truth of our evaluation. Third, the reliability of the human annotators can be improved by using more annotators and from more diverse cultures. This may be possible through crowdsourcing. Finally, the cultural differences were mined from news articles which often reflect the official opinions rather than the opinion of the masses. It would be interesting to do similar research using other text resources, such as the social media data. Obviously this poses new challenges as social media data is a lot noisier and more ambiguous.

References

Chen, J.-Y. 2007. Do chinese and english speakers think about time differently? failure of replicating boroditsky (2001). *Cognition* 104(2):427–436.

Ganea, O.-E.; Ganea, M.; Lucchi, A.; Eickhoff, C.; and Hofmann, T. 2016. Probabilistic bag-of-hyperlinks model for entity linking. In *Proceedings of the 25th International Conference on World Wide Web*, 927–938. International World Wide Web Conferences Steering Committee.

Gruetze, T.; Kasneci, G.; Zuo, Z.; and Naumann, F. 2016. Coheel: Coherent and efficient named entity linking through random walks. *Web Semantics: Science, Services and Agents on the World Wide Web* 37:75–89.

Hermann, K. M., and Blunsom, P. 2014. Multilingual models for compositional distributed semantics. *arXiv preprint arXiv:1404.4641*.

Klementiev, A.; Titov, I.; and Bhattarai, B. 2012. Inducing crosslingual distributed representations of words.

Kovecses, Z. 2006. Language, mind, and culture.

Krifka, M. 1995. 11 common nouns: A contrastive analysis of chinese and english. *The generic book* 398.

Lebret, R., and Collobert, R. 2013. Word emdeddings through hellinger pca. *arXiv preprint arXiv:1312.5542*.

Levy, O., and Goldberg, Y. 2014. Neural word embedding as implicit matrix factorization. In *Advances in Neural Information Processing Systems*, 2177–2185.

Li, Y.; Xu, L.; Tian, F.; Jiang, L.; Zhong, X.; and Chen, E. 2015. Word embedding revisited: A new representation learning and explicit matrix factorization perspective. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence*, 3650–3656.

Lin, Y.; Pan, X.; Deri, A.; Ji, H.; and Knight, K. 2016. Leveraging entity linking and related language projection to improve name transliteration. *ACL 2016* 1.

Mahdisoltani, F.; Biega, J.; and Suchanek, F. 2014. Yago3: A knowledge base from multilingual wikipedias. In *7th Biennial Conference on Innovative Data Systems Research*.

Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G. S.; and Dean, J. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, 3111–3119.

Mikolov, T.; Le, Q. V.; and Sutskever, I. 2013. Exploiting Similarities among Languages for Machine Translation. *arXiv.org*.

Navigli, R., and Ponzetto, S. P. 2012. BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artif. Intell.* 193:217–250.

Schütze, H. 2008. Introduction to information retrieval. In *Proceedings of the international communication of association for computing machinery conference*.

Tavassoli, N. T. 1999. Temporal and associative memory in chinese and english. *Journal of Consumer Research* 26(2):170–181.

Triandis, H. C. 1994. Culture and social behavior.

Vulic, I., and Moens, M.-F. 2015. Bilingual word embeddings from non-parallel document-aligned data applied to bilingual lexicon induction. 719–725.

Weber, E. U., and Hsee, C. 1998. Cross-cultural differences in risk perception, but cross-cultural similarities in attitudes towards perceived risk. *Management science* 44(9):1205–1217.