

COLLEGE

Wendy Wen she/her/hers pwen1@haverford.edu

Mentor: Benjamin Zinszer Collaborators:

Selena She and Gabrielle Ma

Choosing a Semantic Space to Invite the Other Language in

Now we have the relationships between all the words (say, chair = 0.7^* yizi + 0.2^* shafa + 0.1*dengzi).

But we'd still need to fill in the values.

So, we generated Mandarin word vectors for the names given by the participants using Word2Vec model, and then calculated the English word vectors which are now in the same space as the Mandarin words. Vice versa if we use English W2V vectors as the base.

Establishing the Equations -Using Images as Mediators

We think of all the words and images of objects as vectors-they each represents a certain meaning in the semantic space.

The objects are never perfectly equal to any of the words: we see them as linear combinations of word vectors.

The vector representations of objects can be obtained by either adding up the weighted English word vectors or the weighted Mandarin word vectors. Thus, equations are established, of which quantity is the number of images.



No 1-1 Dictionaries between Languages!

Dictionaries provide translations:

3 : a reference book listing alphabetically the words of one language and showing their meanings or translations in another language





- Speaking of untranslatability, we'd usually think of culturally dense words or phrases like hygge or wabi-sabi.
- But few of us realizes even simply day-to-day nouns like sofa and cup are hard to translate, too.
- So, does nature even have joints? After all, different languages carve them very differently.

airplane plane		
airplane		helicopter
airplane feiji		
helicopter	feiji	



Challenges

 We cannot directly ask bilingual speakers to compare meanings of words, because they may not consciously realize their subtle differences until they see actual objects; also, bilingual participants are rarer.

Current statistical word-embedding models like Word2Vec only work within one language.

Could use Affine Transformation to map one language's semantic space to another. But it still requires the numbers of words to be equal on both sides.

Thus, a more indirect path was taken: We asked participants to name objects, and used the objects as a mediator.