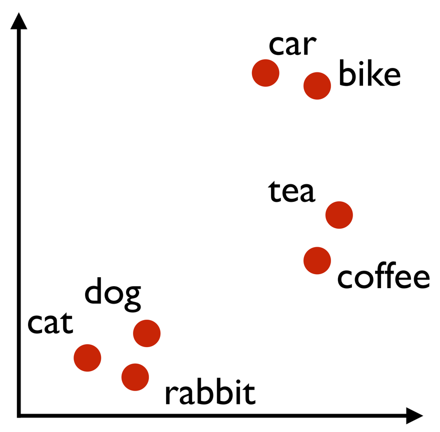
Sound Analogies with Phoneme Embeddings

Miikka Silfverberg, Lingshuang Jack Mao and Mans Hulden
miikka.silfverberg@colorado.edu

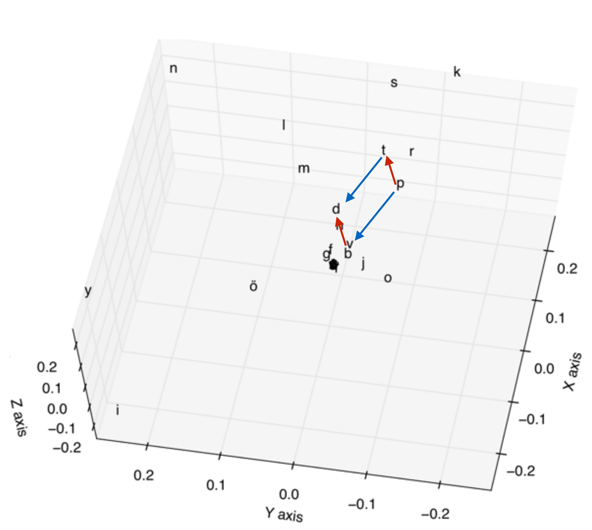
University of Colorado **Boulder**

1. Phoneme Embeddings



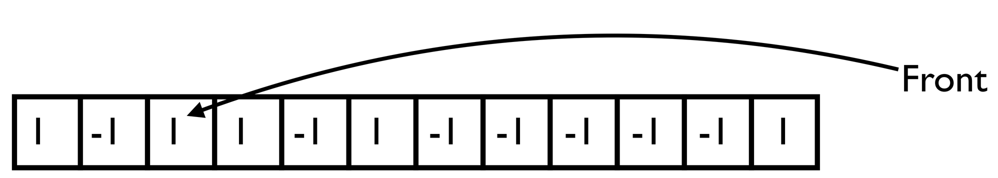
Word embeddings have attracted much attention in NLP, and their success is considered a vindication of the distributional hypothesis for lexical semantics.

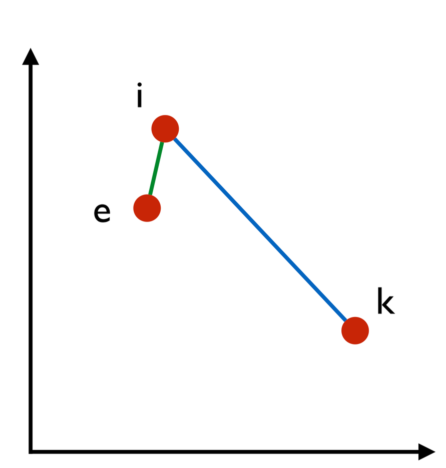
We want to investigate if distributional representations of phonemes, **phoneme embeddings**, induce a similarly coherent space as lexical items do, and if the properties of such spaces conform to linguistic expectations.

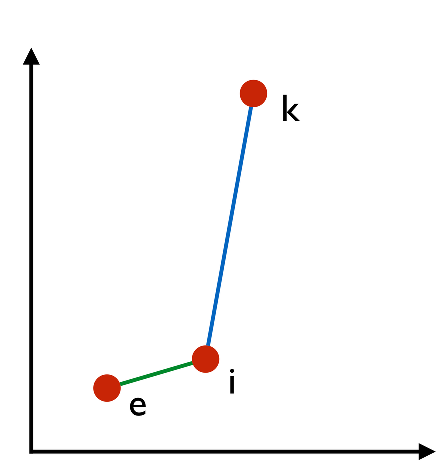


2. Similarity Correlation

Are distributional representations of phonemes congruent with commonly assumed binary phonological distinctive feature spaces?







$$\mathrm{sim}(u, v) = \frac{u^\top v}{|u| \cdot |v|}$$

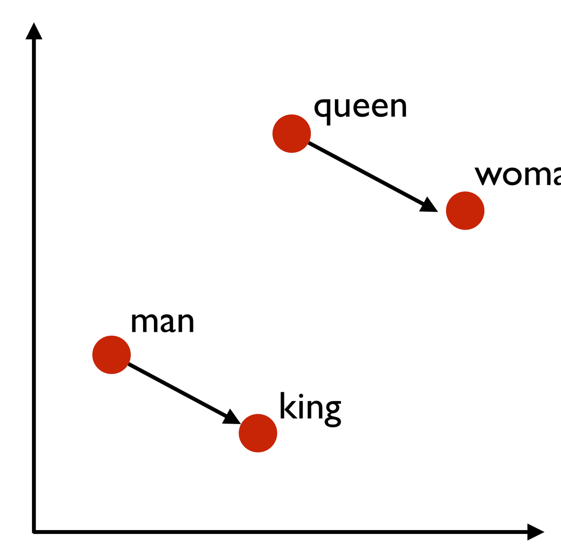
We use **Pearson's r** for measuring correlation between **cosine similarities** in feature space and embedding space. As baseline, we compute the correlation of similarities of feature representations and a random permutation of embeddings.

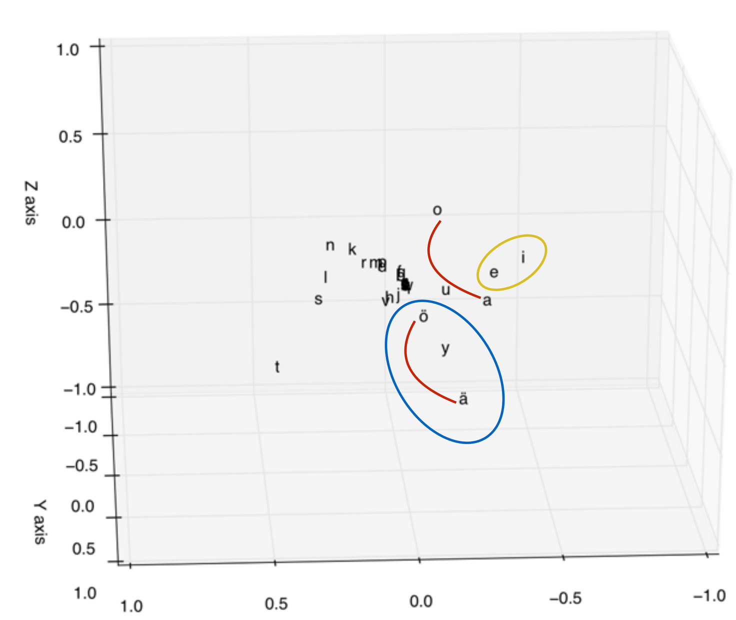
3. Analogies

Are proportional analogies of the type a:b::c:d (a is to b as c is to d) discovered in a phoneme embedding space valid analogies in a phonological distinctive feature space?

Word embeddings are known to encode semantic analogies as vector algebra. E.g.

v(woman) - v(queen) = v(man) - v(king)





We investigate whether phoneme embeddings learn corresponding analogies in phoneme space. E.g.

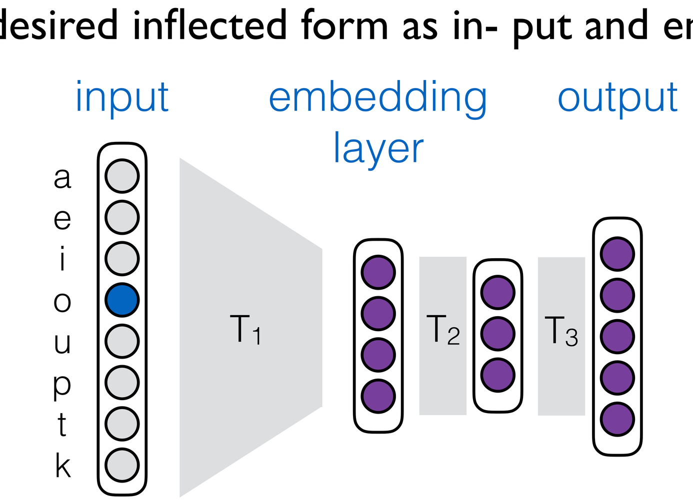
v(a) - v(o) = v(ä) - v(ö)

4. Embedding Types

PPMI+SVD These embeddings are formulated using *truncated Singular Value Decomposition* (SVD) on a matrix of *positive point-wise mutual information* (PPMI) values.

word2vec Our second model is the word2vec model introduced by Mikolov et al. (2013a) for modeling semantic relatedness of words.

RNN encoder-decoder Our final model differs from the first two in that it learns embeddings which maximize performance on a word inflection task: the system receives lemmas and the morpho- logical features of the desired inflected form as in- put and emits corresponding inflected forms.



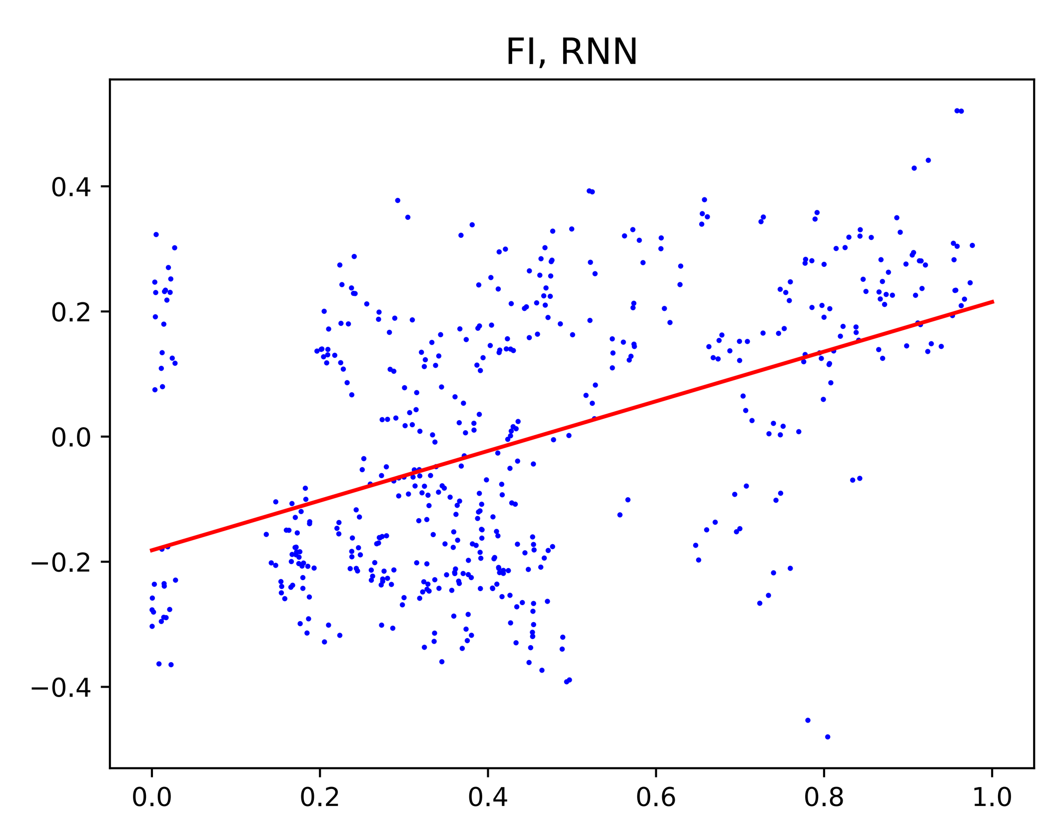
A neural network learns to map a one- hot input into an intermediate representation (the embedding layer). This transformation is tuned to perform well on an inflection task and yields a dense vector representation of segments.

5. Articulatory Space

Vowels
(Syllabic), Front, Back, High, Low, Round, Tense
Consonants
Consonantal, Sonorant, (Syllabic), Voice, Labial, Coronal, Dorsal, Pharyngeal, Lateral, Nasal, Continuant, Delayed Release, Distributed, Tap, Anterior, Strident

Table 1. Distinctive features.

6. Similarity Corr. Results



Correlation between embedding and feature similarities for the Finnish dim=30 RNN system.

PPMI+SVD			
Dim	5	15	30
Finnish	0.174	0.187	**0.204**
Turkish	0.336	0.345	**0.363**
Spanish	**0.328**	0.311	0.301

WORD2VEC			
Dim	5	15	30
Finnish	0.114	0.147	**0.157**
Turkish	**0.184**	0.178	0.177
Spanish	0.273	0.286	**0.289**

RNN ENC-DEC			
Dim	5	15	30
Finnish	0.378	0.408	**0.459**
Turkish	0.293	0.368	**0.415**
Spanish	0.279	0.318	**0.339**

7. Analogy Results

Finnish	Turkish	Spanish
ɑ is to **o** as **æ** is to **ø**	**a** is to **ɯ** as **e** is to **i**	**f** is to **θ** as **p** is to **s**
ɑ is to **æ** as **o** is to **ø**	**a** is to **e** as **ɯ** is to **i**	**k** is to **ɲ** as **t** is to **ʎ**
ɑ is to **æ** as **u** is to **y**	**a** is to **ɯ** as **e** it to **y**	**p** is to **ɾ** as **ʎ** is to **l**
a is to **y** as **o** is to **ø**	**a** is to **u** as **e** is to **i**	**l** is to **ʎ** as **ɾ** is to **p**
ɑ is to **y** as **o** is to **ø**	**b** is to **k** as **f** is to **g**	**m** is to **ʎ** as **r** is to **ɲ**

Table 2. Top analogies discovered by the system for Finnish, Turkish and Spanish.

Dim	5			15			30		
# Top Analogies	15	30	100	15	30	100	15	30	100
PPMI+SVD									
Finnish	6.40	5.83	5.50	4.07*	4.27*	4.88	4.80*	4.27*	5.26
Turkish	5.33*	4.63*	5.21*	6.87	6.43	5.97*	6.07*	6.10*	6.12*
Spanish	4.93	4.27*	4.45*	3.40*	3.53*	4.16*	**2.93***	**3.10***	**3.79***
WORD2VEC									
Finnish	4.93*	5.20	4.87	4.13*	4.07*	4.48*	3.47*	4.00*	4.47*
Turkish	4.87*	5.47*	5.74*	3.73*	4.20*	**5.11***	3.73*	4.17*	5.15*
Spanish	5.47	5.23	5.56	5.73	5.20	5.10*	5.60	5.47	5.01*
RNN ENCODER-DECODER									
Finnish	2.67*	3.70*	4.71*	**2.27***	**2.83***	**3.75***	4.00*	4.07*	4.34*
Turkish	5.00*	5.27*	5.14*	**3.00***	**4.10***	5.20*	4.60*	4.53*	5.14*
Spanish	4.47*	4.87*	4.95*	5.40	5.00*	4.83*	4.73*	4.90*	4.88*

Table 3,.The embedding space is used to generate an n-best list of a:b::c:d analogy proposals. The table shows the average number of differing distinctive features between d and X when X is calculated by the same analogy is performed in distinctive feature space, i.e. a:b::c:X, with a, b, and c given. For each language and each n, we show the best performing system in bold font. Scores which are statistically significantly better than scores for random sets of analogies are marked by an asterisk *.

8. Conclusions

Experiments on Finnish, Turkish and Spanish show that distributional properties of phonetic segments contain information about regularities in phonetic representations.

In particular, we have shown a significant correlation between embedding spaces and distinctive feature spaces.

Embeddings can be learned from plain text in an unsupervised manner but correlation is stronger when learning is directed using a word inflection task.

We also present experiments on phonological analogies. While embeddings do not perfectly capture analogies in feature space, it is still clear that phonologically significant alternations are prominent in embedding space as well.