Sound Analogies with Phoneme Embeddings

Miikka Silfverberg and **Lingshuang Jack Mao** and **Mans Hulden**
Department of Linguistics
University of Colorado
first.last@colorado.edu

Abstract

Vector space models of words in NLP—*word embeddings*—have been recently shown to reliably encode semantic information, offering capabilities such as solving proportional analogy tasks such as man:woman::king:queen. We study how well these distributional properties carry over to similarly learned *phoneme embeddings*, and whether phoneme vector spaces align with articulatory distinctive features, using several methods of obtaining such continuous-space representations. We demonstrate a statistically significant correlation between distinctive feature spaces and vector spaces learned with word-context PPMI+SVD and word2vec, showing that many distinctive feature contrasts are implicitly present in phoneme distributions. Furthermore, these distributed representations allow us to solve proportional analogy tasks with phonemes, such as **p** is to **b** as **t** is to **X**, where the solution is that **X = d**. This effect is even stronger when a supervision signal is added where we extract phoneme representations from the embedding layer of an recurrent neural network that is trained to solve a word inflection task, i.e. a model that is made aware of word relatedness.

1 Introduction

Distributional word representations, or word embeddings, have attracted much attention in NLP, and their success is considered a vindication of the distributional hypothesis for lexical semantics espoused much earlier by the likes of Wittgenstein,[1] Firth,[2] Harris,[3] and other contemporaries. Often overlooked is that this hypothesis among linguists has extended itself much wider to include phonology and grammar: "all elements of speech (phonological, lexical, and grammatical) are now to be defined and classified in terms of their relations to one another" (Haas, 1954, p.54).

Given the successes of distributional models not only in specifying semantic similarity, but also addressing proportional analogy tasks (Turney and Pantel, 2010; Mikolov et al., 2013a,b; Levy et al., 2014), we want to investigate if distributional representations of phonemes induce a similarly coherent space as lexical items do, and if the properties of such spaces conform to linguistic expectations, a question posed in another form as early as in Fischer-Jørgensen (1952). In particular, we address two questions: (1) whether learned vector representations of phonemes are congruent with commonly assumed binary phonological distinctive feature spaces, and (2) whether a proportional analogy of the type a:b::c:d (**a** is to **b** as **c** is to **d**) discovered in a phoneme embedding space is also a valid analogy in a phonological distinctive feature space, where phonemes are represented in a space of standard articulatory binary features (Mielke, 2008) such as ±continuant, ±voice, ±high, ±coronal, etc.

We address these questions using three different methods of obtaining vector representations of phonemes; two unsupervised models that associate phonemes and the contexts (neighboring phonemes) they occur in (word2vec, PPMI+SVD), and one model where we are given lemmas and inflected forms as supervised data for a recurrent neural network (RNN) encoder-

[1] "the meaning of a word is its use in the language" (Wittgenstein, 1953, p.43)

[2] "The complete meaning of a word is always contextual" (Firth, 1937, p.37)

[3] "distributional statements can cover all of the material of a language" (Harris, 1954, p.34)

Proceedings of the Society for Computation in Linguistics (SCiL) 2018, pages 136-144.
Salt Lake City, Utah, January 4-7, 2018

decoder model that is trained to perform such inflections from where we extract the vector representations in the embedding layer. The latter method has a weak supervision signal in that the system knows which words are truly related forms, since it is trained only on word pairs where one of the words is a result of transforming the other one according to inflectional features. Hence, this latter model can presumably better pick up on phonological alternations along distinctive feature lines that occur between phonemes; for example, knowledge of relatedness helps in determining that two forms exhibit a voicing alternation, as in the Finnish pair ma**t**o ('worm', nominative) ∼ ma**d**on ('worm', genitive), and hence, the segments **t** and **d** should align themselves in the embedding space parallel to other voiceless/voiced pairs that alternate similarly. We perform experiments in three languages: Finnish, Turkish, and Spanish.

2 Related Work

Local co-occurrence of phonemes and (in writing) phonemic graphemes has been widely explored in the literature for unsupervised discovery of phonological features. The observation of Markov (1913, 2006) that vowels and consonants tend to alternate in a statistically robust way in phonemic writing systems is an early observation that some articulatory features can be recovered in an unsupervised way. Algorithms for cryptographic decipherment often take advantage of such patterns (Guy, 1991; Sukhotin, 1962, 1973). Local co-occurrence counts have also been analyzed through spectral methods, such as singular value decomposition (Moler and Morrison, 1983; Goldsmith and Xanthos, 2009; Thaine and Penn, 2017), revealing that significant latent structure can be recovered, mainly with respect to vowels and consonants. Recent works along the same lines of inquiry include Kim and Snyder (2013) that presents a Bayesian approach that simultaneously clusters languages and reveals consonant/vowel/nasal distinctions in an unsupervised manner. Hulden (2017) shows that an algorithm based on the obligatory contour principle (Leben, 1973) and an additional assumption of phonological tiers being present (Goldsmith, 1976) robustly reveals at least consonant/vowel, coronal/non-coronal, and front/back distinctions from unlabeled phonetic data or orthographic data from phonemic writing systems.

Learning features directly from waveform representations (see e.g. Lin (2005))—while not addressed in this paper—is also highly relevant to the current study, and is indeed a question to which some lower-level, speech signal-based form of distributed representations may be adapted.

The idea explored in this paper—that phonemes (or graphemes) might exhibit linguistically apt correlations in an embedding space—has been implied by earlier research, for example Faruqui et al. (2016). In that work, a neural encoder-decoder model (Cho et al., 2014; Sutskever et al., 2014) was trained to perform a transformation of words from their citation forms to a 'target' inflected form and, after training, the vowels in the embedding layer of the long-short term memory (LSTM) neural model trained for Finnish were found to clearly group themselves according to known harmony patterns in the language. Li et al. (2016) take advantage of phoneme transcriptions in a neural speech synthesis application, showing improvements on this task and indicating that similar phonemes in a bidirectional LSTM (Bi-LSTM) embedding layer map close to each other. More closely related to the current work, Dunbar et al. (2015) investigate how well phonetic feature representations in English align with vector representations learned from local contexts of sound occurrence using both a neural language model and also a matrix factorization model. Their (surprisingly) negative results can be explained by the fact that the experiment was set up to compare two spaces with respect to all and only minimally differing pairs (such as: is the distinctive feature vector offset from **p** to **b** like the offset from **t** to **d** in the embedding space?) using a small specific fixed number of phonological features. By contrast, in our experiments, we learn a vector space model and examine if all phoneme-pair distances correlate globally between the embedding space and a known phonological distinctive feature space, and also whether analogies deemed to be 'good' in the embedding spaces are also 'good' in the distinctive feature space. This is less sensitive to an assumption that *all* distinctive features have correlates in the embedding space. For example, Figure 3 shows a space induced by one of our methods, where the offset from **ɑ** to **æ** (low vowels) is similar to that of **o** to **ø** (high vowels), both being back-to-front vowel transformations, but where the equivalent harmonic corre-

spondence **u** to **y** is not represented equally prominently.

3 Models and methods

We consider three different models for learning phoneme embeddings.

PPMI+SVD These embeddings are formulated using *truncated Singular Value Decomposition* (SVD) on a matrix of *positive point-wise mutual information* (PPMI) values (Bullinaria and Levy, 2007; Levy and Goldberg, 2014). For the definition of PPMI, see Equation 1. We first compute PPMI values for co-occurrences of center phonemes and context phonemes in a five character sliding window over the training data. We then arrange the PPMI values into an $n \times n$ matrix M and apply SVD to get the factorization $M = U\Sigma V^\top$, where U and V are orthonormal, and Σ is diagonal. Let U_d denote the $n \times d$ matrix derived from $U\Sigma$ by truncating all rows to length d. Then our d-dimensional phoneme embeddings are the rows of U_d.

PPMI is not n x n.

Isl x lcl where s is segment and c is context.

$$\mathrm{PPMI}(x, y) = \max(\log \frac{p(x, y)}{p(x)p(y)}, 0) \quad (1)$$

The probabilities $p(x, y)$, $p(x)$, and $p(y)$ are derived from simple counts by maximum likelihood estimation.

word2vec Our second model is the word2vec model introduced by Mikolov et al. (2013a) for modeling semantic relatedness of words. Like the PPMI+SVD model, the word2vec model captures distributional information about phonemes. However, it explicitly constructs a language model and trains embeddings which perform well on the language modeling task. Nevertheless, Levy et al. (2014) show that the models are in a sense the same: like PPMI+SVD, the skip-gram variant of word2vec with negative sampling is also implicitly performing a factorization of a matrix of shifted PMI values of words and their context words. Although the models are similar, we decided to include word2vec because of claims that it sometimes tends to give better results on analogy tasks (Levy et al., 2014).

In our experiments, we use standard word2vec embeddings generated with the `gensim` toolkit.[4] We use the skipgram model and negative sampling with window size 1.

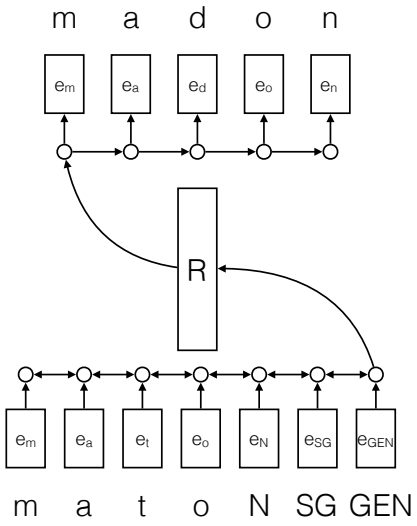

Figure 1: A word inflection system implemented as an RNN encoder-decoder. The system first encodes the input (the citation form) and the desired output morphological features "mato+N+SG+GEN" 'worm' into a single vector R using a bidirectional LSTM encoder. A decoder LSTM network then uses R to generate the output word form "madon". The system learns the appropriate phoneme embeddings, for example e_m and e_a, to aid in the inflection process. We use these as our vector representations of phonemes.

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 morphological features of the desired inflected form as input and emits corresponding inflected forms. We formulate the system as an RNN encoder-decoder (Cho et al., 2014). Our system is identical to the system presented in Kann and Schütze (2016) except that it does not incorporate attention. The encoder is realized using a bidirectional LSTM model which operates on character/phoneme embeddings (see Figure 1). The same embeddings are also used by the decoder as explained more thoroughly in Kann and Schütze (2016). We first train the system and then extract the phoneme embeddings (see Figure 2) and use them as our phoneme vector representations.

The performance of RNN encoder-decoder models is known to be sensitive to the random initialization of parameters during training. In all ex-

[4] https://radimrehurek.com/gensim/

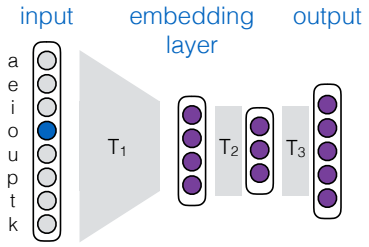

Figure 2: A neural network learns to map a one-hot input into an intermediate representation (the embedding layer). This transformation is tuned to perform well on an inflection task and yields a dense vector representation of segments.

periments, we therefore train five separate models using five different random initializations of parameters and compute similarity scores and analogy scores as averages of the scores given by individual models.

4 Data and Resources

We train all models using Finnish, Spanish and Turkish data sets from the SIGMORPHON 2016 shared task for morphological re-inflection (Cotterell et al., 2016). Each line in the data sets contains an inflected word form, its associated lemma and morphological features. For Finnish, the training data consists of 12,692 lines, for Spanish, 12,575 lines and, for Turkish, 12,336 lines. We learn embeddings for orthographic symbols occurring more than 100 times in the respective data sets. For Finnish, this set includes 25 symbols, for Spanish, 28 symbols and, for Turkish, 27 symbols.

The PPMI+SVD and word2vec models only use word forms for training. In contrast, the RNN encoder-decoder is trained on all parts of the training set: word forms, lemmas and morphological features. For all three languages, we use the training data for subtask 1 of the shared task.

There is a near one-to-one correspondence between Finnish and Turkish graphemes and phonemes. For Spanish, the correspondence between the orthographic and phonetic representation of the language is, however, less straightforward. We therefore perform a number of transformations on the training data in order to bring it closer to a phonetic representation of the language. Specifically, we transform voiced stops **b**, **d** and **g** to the voiced fricatives with the same place of articulation postvocalically (**β**, **ð**, **ɣ**). We additionally replace **ll** with **ʎ**, **r** with **ɾ**, and **c** with **θ**.[5]

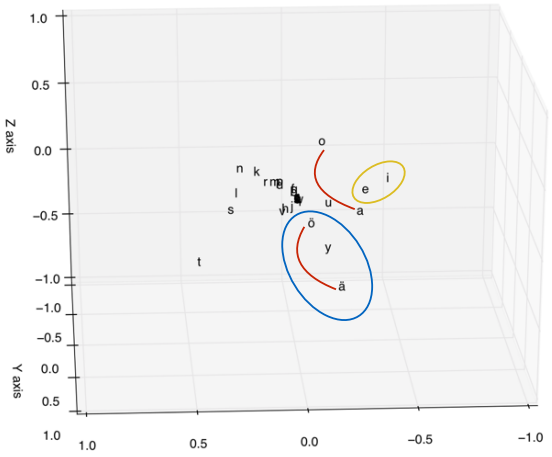

Figure 3: A vowel space for Finnish learned by collecting local phone(me)-context counts (window size 5), followed by a PPMI transform, followed by SVD, truncated to three dimensions. Neutral vowels are **e, i** (circled in yellow) and the harmony correspondences are **a** (=IPA **ɑ**) ∼ **ä** (= IPA **æ**), **o** ∼ **ö** (= IPA **ø**) and **u** ∼ **y**. The front harmonic group is circled in blue.

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: Features used in manually crafted articulatory representations.

We compare embeddings discovered by different systems to manually crafted articulatory representations of phonemes/allophones based on standard IPA descriptions in Hayes (2011). The list of the phonetic features we use is given in Table 4. We realize the representations as vectors $v \in \{0,1\}^n$ in a *distinctive feature space*, where n is the number of distinctive features in the description (22 in our model). Each dimension in feature space corresponds to a phonetic feature such as *continuant*, *syllabic* and *voice*. Entry i in a feature vector is 1 if the corresponding phoneme is positive for the given feature. Otherwise, it is 0.

[5]Our code is available at s://github.com/mpsilfvehttp/phonembedding

	PPMI+SVD				WORD2VEC				RNN ENCODER-DECODER		
Dim	5	15	30	Dim	5	15	30	Dim	5	15	30
Finnish	0.174	0.187	**0.204**	Finnish	0.114	0.147	**0.157**	Finnish	0.378	0.408	**0.459**
Turkish	0.336	0.345	**0.363**	Turkish	**0.184**	0.178	0.177	Turkish	0.293	0.368	**0.415**
Spanish	**0.328**	0.311	0.301	Spanish	0.273	0.286	**0.289**	Spanish	0.279	0.318	**0.339**

Table 2: Correlation between feature similarities $\mathrm{sim}(\mathrm{feat}(x), \mathrm{feat}(\mathrm{y}))$ and embedding similarities $\mathrm{sim}(\mathrm{emb}(x), \mathrm{emb}(y))$ for all unordered pairs of phonemes $\{x, y\}$ (where $x \neq y$). All correlations are significantly higher (with p-value < 0.01) than ones obtained using a random assignment of embedding vectors to phonemes.

Word2Vec the weakest

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*

Lower is better

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 *.

5 Experiments

Correlation Our first experiment investigates the relationship between the geometries of embedding space and the distinctive feature space.

Let the embedding for phoneme p be $\mathrm{emb}(p)$, its distinctive feature vector $\mathrm{feat}(p)$, and cosine similarity of vectors u and v be given by Equation 2.

$$\mathrm{sim}(u, v) = \frac{u^\top v}{|u| \cdot |v|} \quad (2)$$

We measure the linear correlation of $\mathrm{sim}(\mathrm{emb}(p), \mathrm{emb}(q))$ and $\mathrm{sim}(\mathrm{feat}(p), \mathrm{feat}(q))$ over all unordered pairs of phonemes $\{p, q\}$ (where $p \neq q$) using Pearson's r. As a baseline, we compute the correlation of similarities of feature representations and random embeddings $\mathrm{remb}(p)$. These are derived by randomly permuting the embeddings of phonemes. That is, $\mathrm{remb}(p) = \mathrm{emb}(q)$ for some random phoneme q.

Analogy Our second experiment investigates phoneme analogies. We first score four-tuples (a, b, c, d) of phonemes using cosine similarity in embedding space as defined by Equation 3. This corresponds to a proportional analogy a:b::c:d.

$$\mathrm{score}(a, b, c, d) = \mathrm{sim}(\mathrm{emb}(b) - \mathrm{emb}(a), \mathrm{emb}(d) - \mathrm{emb}(c)) \quad (3)$$

We then evaluate the top 15, 30 and 100 four-tuples w.r.t. phonological analogy in distinctive feature space. Our evaluation is based on applying the transformation defined by the first two phonemes a and b on the third phoneme c and measuring the Hamming distance of the result and the feature representation of d. For example, given tuple (**p**,**b**,**t**,**d**), we get Hamming distance 0. This happens because **p** is transformed to **b** by changing the value of feature *voice* from 0 to 1. When the same transformation is applied to **t**, the result is **d**, which obviously has Hamming distance 0 with

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 4: Top 5 analogies (in IPA) discovered by the best model for each languages: Finnish, Turkish and Spanish.

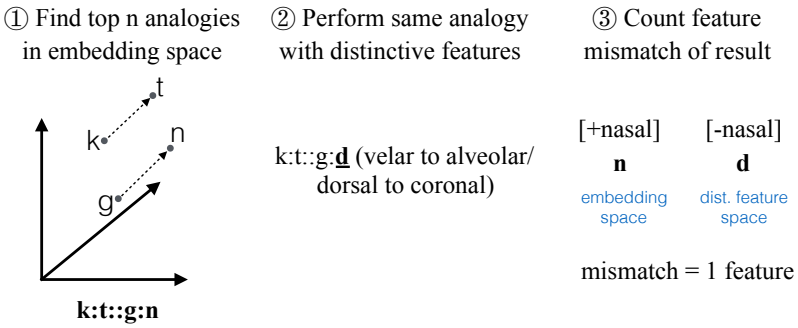

Figure 4: Illustration of the evaluation of the analogy coherence. This procedure is repeated for the top 15, 30, and 100 four-tuples in the embedding space and the average distance of the resulting analogy when performed in the distinctive feature space is reported.

Figure 5:

the fourth phoneme of the tuple in the embedding analogy.[6] We restrict tuples in two ways: (1) all phonemes in the tuple have to be distinct symbols, and (2) all phonemes in the tuple have to be consonants or all of them have to be vowels.

As baseline, we randomly select 15, 30 or 100 phoneme tuples (a, b, c, d). We then apply the transformation defined by a and b onto c and then compute the Hamming distance of the transformed image of c and the phoneme d. We restrict these random tuples as explained above.

6 Results

Table 2 shows results for linear correlation measured by Pearson's r for the similarity between phonetic representations and similarity of corresponding embedding vectors. Overall, the RNN encoder-decoder with embedding dimension 30 gives the best results. The correlation is the weakest for word2vec. However, all methods give a statistically significant positive correlation compared with random embeddings with p-value < 0.01 for appropriately chosen embedding dimension. For all three models: PPMI+SVD, word2vec and RNN encoder-decoder, there seems to be a tendency that higher dimension gives better correlation. This is not the case for PPMI+SVD for Spanish or word2vec for Turkish, However, in these cases, the results for all embedding dimensions are very similar. Figure 6 shows the correlation between cosine similarities of phoneme embeddings and the corresponding phonological feature representations.

Table 3 shows results for analogies as measured by average Hamming distance. Results are presented for the top 15, 30 and 100 analogies discovered by each of the systems. Overall, there is a strong trend that average Hamming distance increases in distinctive feature space when more (lower-ranked) analogies are considered in the embedding space. This is to be expected if the two spaces are coherent—as we include lower and lower ranked analogies and evaluate them, we expect them to be less fitting in the distinctive fea-

[6]Note that, we can only apply a transformation in co-ordinate i if the ith co-ordinates of the first and third phoneme in the tuple match. If this is not the case for some i, we do not apply any transformation for that co-ordinate. For example, if the first phoneme is [+voice], the second [-voice], and the third also [-voice], changing the third phoneme from $+$ to $-$ voice is not well defined.

Significant but not compelling.

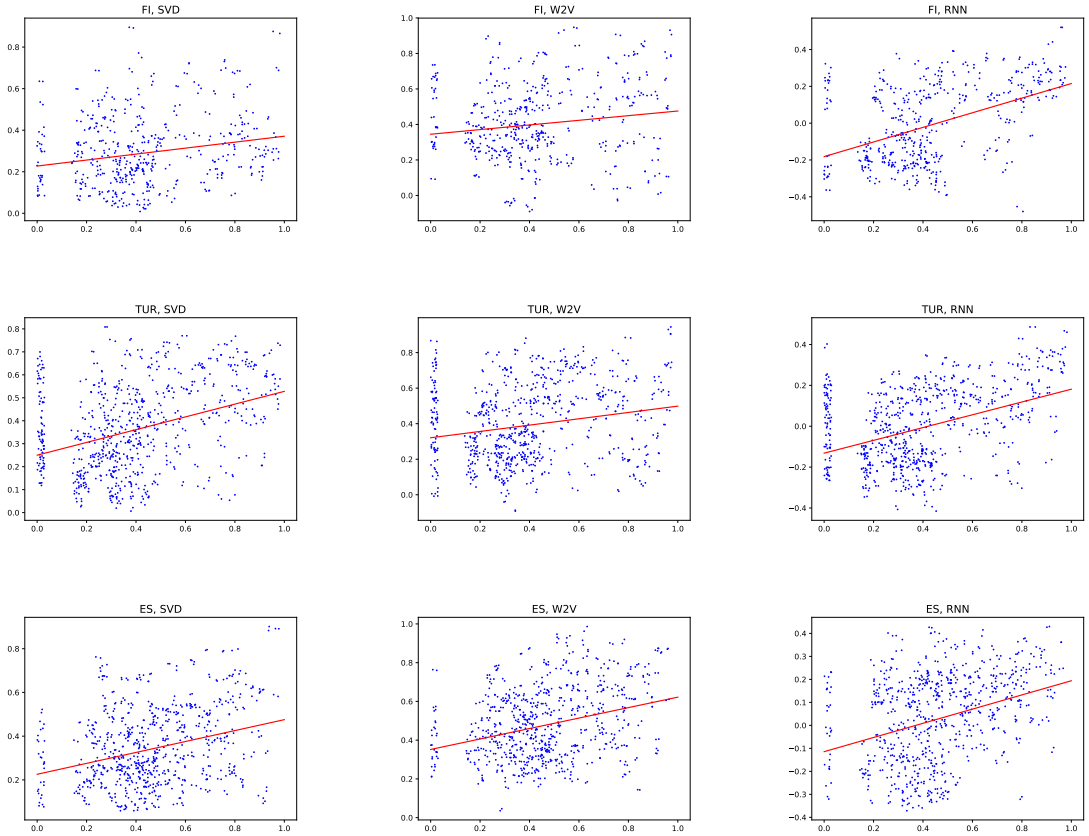

Figure 6: Scatter plots of cosine similarities of phonemes in feature space (x-axis) and embeddings space (y-axis). The figures present results for 30 dimensional PPMI+SVD, word2vec and RNN embeddings for Finnish, Turkish and Spanish, respectively. The red line represents the regression line.

ture space somewhat monotonically. The best results for Hamming distance are delivered by the 30 dimension RNN encoder-decoder for Finnish and Turkish and the 30 dimension PPMI+SVD system for Spanish. Table 4 shows a selection of top analogies for each language.

7 Discussion

The results both in comparing the geometry of the spaces learned and the alignment of analogies to distinctive features show a clear effect of distinctive features being aligned and discovered by distributional properties. The strength of the alignment appears to be somewhat language-dependent; in both Finnish and Turkish, vowel harmony effects are quite prominent and come out as many of the top-ranking analogies in an embedding space. In Spanish, by contrast, the correlation of the space is less robust, probably because there are fewer symmetrical phonological alternations witnessed in the data, although ±continuant alternation is a prominent one (**b/β**, **d/ð g/ɣ**). Likewise, non-symmetric alternations in the data may distort the vector space to not align perfectly along distinctive feature lines. For example, while Finnish exhibits a **t/d** alternation (katu/kadun; 'street' nominative/genitive) the corresponding analogical labial alternation in the embedding space is **p/v** (apu/avun; 'help' nominative/genitive), not **p/b**, as one would assume by distinctive features. This is an interesting discovery since, while the analogy in the embedding space in this case does not correlate to the analogy in the feature space, this distortion of the embedding space of phonemes is arguably more "correct" than the feature-based expected one where t:d::p:b. In fact, the /b/-phoneme is only present in loanwords in the Finnish data, and the spirantization seen in **p/v** was historically present for the alveolar stop as well (**t/ð**). This analogy it-

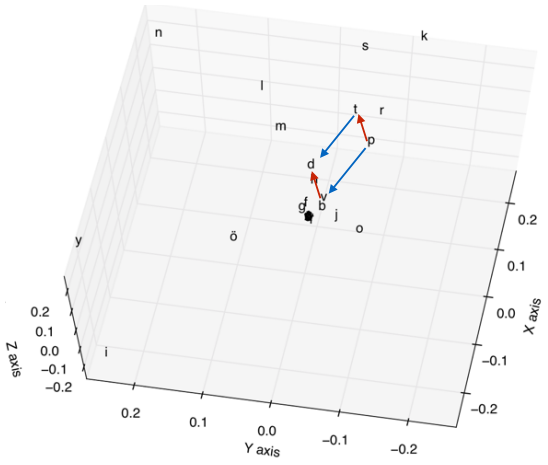

Figure 7: The phoneme space, focusing on consonants, for Finnish learned by collecting local phone(me)-context counts (window size 5), a PPMI transform, followed by SVD, truncated to three dimensions. Marked here is the (correct from a Finnish speaker point-of-view) analogy of consonant gradation **t:d::p:v**.

self is an example of Finnish consonant gradation which manifests itself through many idiosyncratic alternations (Karlsson, 2008), some of which are clearly captured in symmetries in the embedding space. Hence, although such mappings are often present and prevents many analogies from being 'perfect' along distinctive feature lines, embedding spaces where such seemingly 'incorrect' analogies are drawn are in fact good representations for learning tasks such as morphological inflection, since they yield generalization power to task learning, i.e. learning of phonological alternations. This flexibility to learn a vector space representation that does not always strictly conform to distinctive features is then an advantage of the representations and partly explains their recent success (Cotterell et al., 2016, 2017) in learning inflectional patterns from examples.

8 Conclusion

We have presented a set of experiments on three languages that examine how distributional properties of phonetic segments contain information about regularities in the distinctive feature alternations present in the language. In particular, we have shown a significant correlation between embedding spaces learned from either co-occurrence and distinctive feature spaces. While such embeddings can be learned from raw data without any supervision, this correlation is consistently stronger if embeddings are learned and extracted from a recurrent neural network in conjunction with a supervised task of learning to inflect word forms. Apart from a holistic inspection of the embedding spaces, we also developed an experiment that measures how well phonological analogies can be performed using the embeddings learned. While the analogies do not perfectly correlate with similar analogies in distinctive feature space, it is clear that those distinctive features that play a part in prominent phonological alternations are also latently present in co-occurrence generalizations and can be seen in the learned embedding space.

References

John A. Bullinaria and Joseph P. Levy. 2007. Extracting semantic representations from word co-occurrence statistics: A computational study. *Behavior research methods* 39(3):510–526.

Kyunghyun Cho, Bart van Merrienboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning phrase representations using RNN encoder–decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, Doha, Qatar, pages 1724–1734.

Ryan Cotterell, Christo Kirov, John Sylak-Glassman, Géraldine Walther, Ekaterina Vylomova, Patrick Xia, Manaal Faruqui, Sandra Kübler, David Yarowsky, Jason Eisner, and Mans Hulden. 2017. The CoNLL-SIGMORPHON 2017 shared task. In *CoNLL-SIGMORPHON 2017 Shared Task*.

Ryan Cotterell, Christo Kirov, John Sylak-Glassman, David Yarowsky, Jason Eisner, and Mans Hulden. 2016. The SIGMORPHON 2016 shared task—morphological reinflection. *Proceedings of the 14th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology* pages 10–22.

Ewan Dunbar, Gabriel Synnaeve, and Emmanuel Dupoux. 2015. Quantitative methods for comparing featural representations. In *Proceedings of the 18th International Congress of Phonetic Sciences*.

Manaal Faruqui, Yulia Tsvetkov, Graham Neubig, and Chris Dyer. 2016. Morphological inflection generation using character sequence to sequence learning. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, San Diego, California, pages 634–643.

J. R. Firth. 1937. The technique of semantics. *Transactions of the Philosophical Society* pages 36–72.

Eli Fischer-Jørgensen. 1952. On the definition of phoneme categories on a distributional basis. *Acta linguistica* 7(1-2):8–39.

John Goldsmith. 1976. An overview of autosegmental phonology. *Linguistic analysis* 2:23–68.

John Goldsmith and Aris Xanthos. 2009. Learning phonological categories. *Language* 85(1):4–38.

Jacques B. M. Guy. 1991. Vowel identification: an old (but good) algorithm. *Cryptologia* 15(3):258–262.

William Haas. 1954. On defining linguistic units. *Transactions of the Philosophical Society* pages 54–84.

Zellig S. Harris. 1954. Distributional structure. *Word* 10(2-3):146–162.

Bruce Hayes. 2011. *Introductory Phonology*. John Wiley & Sons.

Mans Hulden. 2017. A phoneme clustering algorithm based on the obligatory contour principle. In *Proceedings of The 21st SIGNLL Conference on Computational Natural Language Learning (CoNLL)*. Association for Computational Linguistics, Vancouver, Canada.

Katharina Kann and Hinrich Schütze. 2016. MED: The LMU system for the SIGMORPHON 2016 shared task on morphological reinflection. In *SIGMORPHON*.

Fred Karlsson. 2008. *Finnish: An essential grammar*. Routledge.

Young-Bum Kim and Benjamin Snyder. 2013. Unsupervised consonant-vowel prediction over hundreds of languages. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Sofia, Bulgaria, pages 1527–1536.

William Ronald Leben. 1973. *Suprasegmental Phonology*. Ph.D. thesis, Massachusetts Institute of Technology.

Omer Levy and Yoav Goldberg. 2014. Neural word embedding as implicit matrix factorization. In *Proceedings of the 27th International Conference on Neural Information Processing Systems*. MIT Press, Cambridge, MA, USA, NIPS'14, pages 2177–2185.

Omer Levy, Yoav Goldberg, and Israel Ramat-Gan. 2014. Linguistic regularities in sparse and explicit word representations. In *CoNLL*. pages 171–180.

Xu Li, Zhiyong Wu, Helen Meng, Jia Jia, Xiaoyan Lou, and Lianhong Cai. 2016. Phoneme embedding and its application to speech driven talking avatar synthesis. *Interspeech 2016* pages 1472–1476.

Ying Lin. 2005. *Learning features and segments from waveforms: A statistical model of early phonological acquisition*. Ph.D. thesis, University of California Los Angeles.

A. A. Markov. 1913. Primer statisticheskogo issledovaniya nad tekstom "Evgeniya Onegina", illyustriruyuschij svyaz ispytanij v cep. *Izvestiya Akademii Nauk* Ser. 6(3):153–162.

A. A. Markov. 2006. An example of statistical investigation of the text "Eugene Onegin" concerning the connection of samples in chains. *Science in Context* 19(4):591–600.

Jeff Mielke. 2008. *The Emergence of Distinctive Features*. Oxford University Press.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013a. Distributed representations of words and phrases and their compositionality. In *Neural and Information Processing Systems (NIPS)*.

Tomas Mikolov, Wen-tau Yih, and Geoffrey Zweig. 2013b. Linguistic regularities in continuous space word representations. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, Atlanta, Georgia, pages 746–751.

Cleve Moler and Donald Morrison. 1983. Singular value analysis of cryptograms. *American Mathematical Monthly* pages 78–87.

Boris V. Sukhotin. 1962. Eksperimental'noe vydelenie klassov bukv s pomoshch'ju EVM. *Problemy strukturnoj lingvistiki* pages 198–206.

Boris V. Sukhotin. 1973. Méthode de déchiffrage, outil de recherche en linguistique. *T. A. Informations* pages 1–43.

Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. 2014. Sequence to sequence learning with neural networks. In *Advances in Neural Information Processing Systems*. pages 3104–3112.

Patricia Thaine and Gerald Penn. 2017. Vowel and consonant classification through spectral decomposition. *Proceedings of the Workshop on Subword and Character Level Models in NLP (SCLeM)* .

Peter D. Turney and Patrick Pantel. 2010. From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research* 37:141–188.

Ludwig Wittgenstein. 1953. *Philosophical Investigations [German original: Philosophische Untersuchungen]*. Blackwell, Oxford.