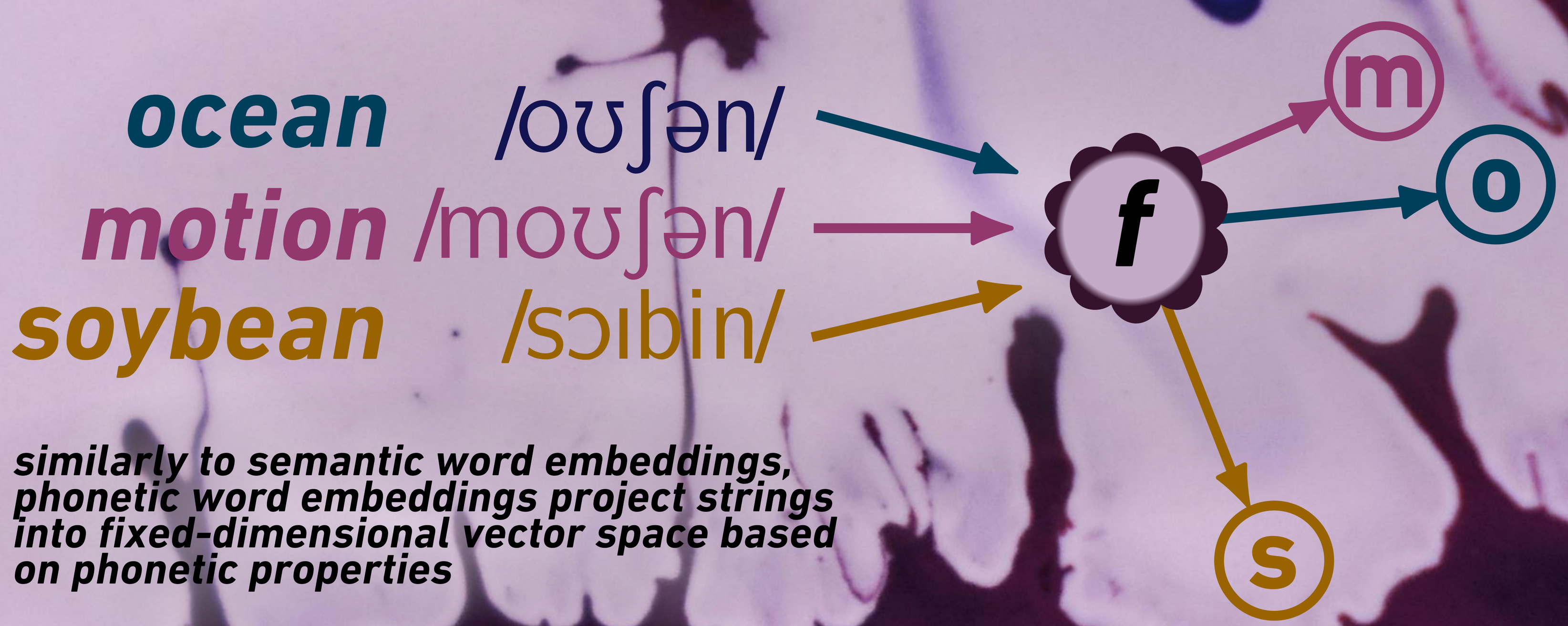
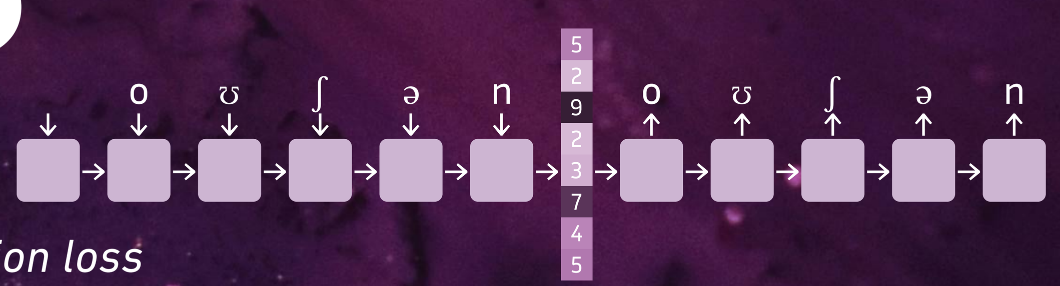
Phonetic Word Embeddings & Tasks They Facilitate



similarly to semantic word embeddings, phonetic word embeddings project strings into fixed-dimensional vector space based on phonetic properties

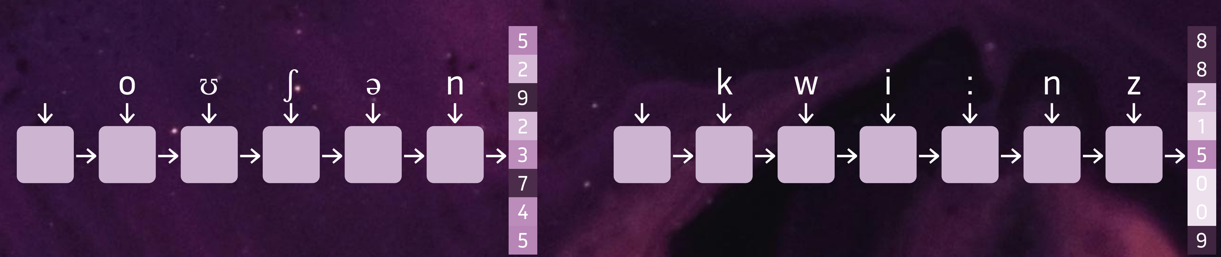
Models

Autoencoder



LSTM encoder & decoder; minimize reconstruction loss

Metric Learner



LSTM encoder; minimize distance error

we know S_A(ocean, queens) and want S_V with f such that S_V(f(ocean), f(queens)) = S_A(ocean, queens))

minimize loss $\big\| \|f(\text{ocean}) - f(\text{queens})\|_2 - S_A(\text{ocean, queens}) \big\|_2$

Results

	Model	INTRINSIC Human Sim. (Pearson)	Art. Dist. (Pearson)	Retrieval (rank perc.)	EXTRINSIC Analogies (Acc@1)	Rhyme (accuracy)	Cognate (accuracy)	OVERALL
Ours	Metric Learner	0.46	0.94	0.98	84%	83%	64%	0.78
	Triplet Margin	0.65	0.96	1.00	100%	77%	66%	0.84 ⋆
	Count-based	0.82	0.10	0.84	13%	79%	68%	0.56
	Autoencoder	0.49	0.16	0.73	50%	61%	50%	0.50
Others'	Poetic Sound Sim.	0.74	0.12	0.78	35%	60%	57%	0.53
	phoneme2vec	0.77	0.09	0.80	17%	88%	64%	0.56
	Phon. Sim. Embd.	0.16	0.05	0.50	0%	51%	52%	0.29
Semantic	BPEmb	0.23	0.08	0.60	5%	54%	66%	0.36
	fastText	0.25	0.12	0.64	2%	58%	68%	0.38
	BERT	0.10	0.34	0.69	4%	58%	63%	0.40
	INSTRUCTOR	0.60	0.12	0.73	7%	54%	66%	0.45

Table 1: Embedding method performance in our evaluation suite. Higher number is always better.

Tasks

Articulatory distance

Correlation S_A(x, x') and S_V(f(x), f(x'))

articulatory distance between strings — *S_A(sick, sit) = 0.8*
vector distance between embeddings

Human judgement

Correlation S_H(x, x') and S_V(f(x), f(x'))

human perception of sound similarity — *S_H(ocean, motion) = 0.7*
vector distance between embeddings

Rhyme detection

Accuracy of MLP classifier on ⟨f(x), f(x')⟩

each pair ⟨x, x'⟩ either rhymes or not
⟨ocean, motion⟩ → rhyme ⟨ocean, soybean⟩ → no-rhyme

Cognate detection

Accuracy of MLP classifier on ⟨f(x), f(x')⟩

each pair ⟨x, x'⟩ is either a cognate or not
⟨$plant_{EN}$, $plante_{FR}$⟩ → cognate ⟨$plane_{EN}$, $plante_{FR}$⟩ → no-cognate

Sound analogies

$w_1 : m(w_1) \leftrightarrow w_2 : m(w_2)$

for semantic embeddings man : royal(man) ↔ woman : royal(woman)
man : king ↔ woman : queen

for phonetic embeddings /sin/ : +voice(/sin/) ↔ /tin/ : +voice(/tin/)
/sin/ : /zin/ ↔ /tin/ : /din/

retrieve $f(w_1) - f(m(w_1)) + f(w_2)$ & hope for $f(m(w_2))$

Applications

Linguistic analysis • Cognate/loanword detection • Multilingual named entity recognition • Keyphrase extraction • Spelling correction • Phonotactic learning • Multimodal word embeddings • Spoken language understanding • Language identification • Poetry generation

English Hello. How are you? • **Amharic** ሀሎ. ሰላም? • **Bengali** হ্যালো, আপনি কেমন আছেন? • **French** Bonjour. Comment vas-tu? • **German** Hallo. Wie geht's? • **Polish** Cześć. Jak się masz? • **Spanish** Hola. ¿Cómo estás? • **Swahili** Habari. Habari yako? • **Uzbek** Salom. Qalaysiz?

Vilém Zouhar, Kalvin Chang, Chenxuan Cui, Nathaniel Carlson, Nathaniel R. Robinson, Mrinmaya Sachan, David Mortensen

ETH zürich • LRE Language Reasoning Education • Language Technologies Institute • Carnegie Mellon University