Zero-shot Learning for Grapheme to Phoneme Conversion with Language Ensemble

Introduction

Graphemes: a letter (or a group of letters) that symbolize a single phoneme.

Formally , grapheme is the smallest functional unit of a writing system.

Ex: CHEAP → **CH** IY1 P → **CH** is the grapheme

Phonemes : A phoneme is the smallest unit of sound in speech.

EX: CAR →**k ɑ r**

Proposed Approach

Apply **zero-shot learning** to approximate G2P models for all **low-resource** and endangered languages in GlottoLog (about 8k languages).

Approximate the G2P model of an **unseen language using those of related languages because languages related to the target language should have similar orthographic rules**

Exploiting language similarities

English speakers speaking

grapheme ‘h’ of hello will most likely

pronounce h correctly.

But ‘h’ grapheme in hola is pronounced

differently.

Language	Grapheme	Phoneme
English	hello	/hələʊ/
Mandarin	你好	/nixɑʊ/
French	bonjour	/bɔʒuʁ/
German	hallo	/halo/
Japanese	こんにちは	/konnichiwa/
Spanish	hola	/ola/

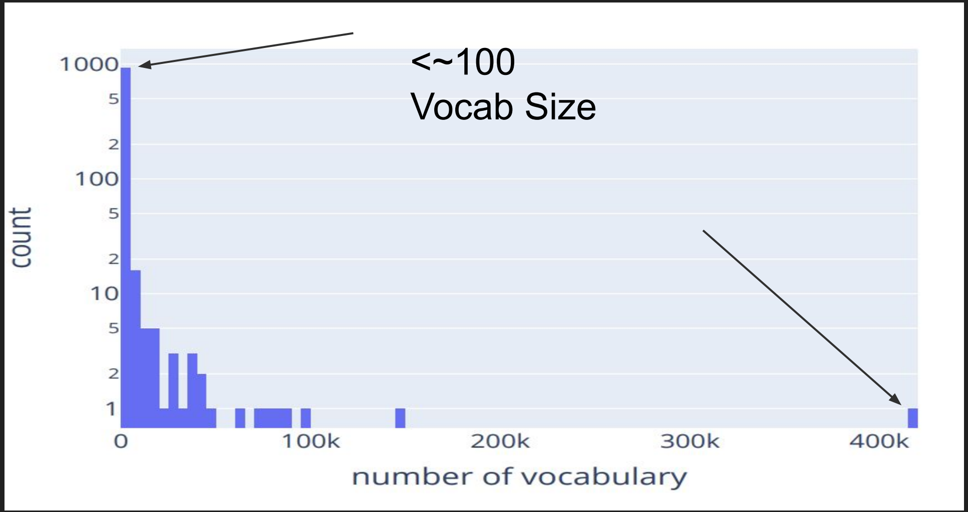
Dataset

Constructed **Wikitionary.**

Languages in Testing set are not present in

Training set (not trained upon).

Dataset	# Languages	# Vocabulary
Training set	269	1,672,444
Testing set	605	4,796
All	874	1,677,240



Model Training Architectures & Paradigms

3 models architectures were experimented with these include:

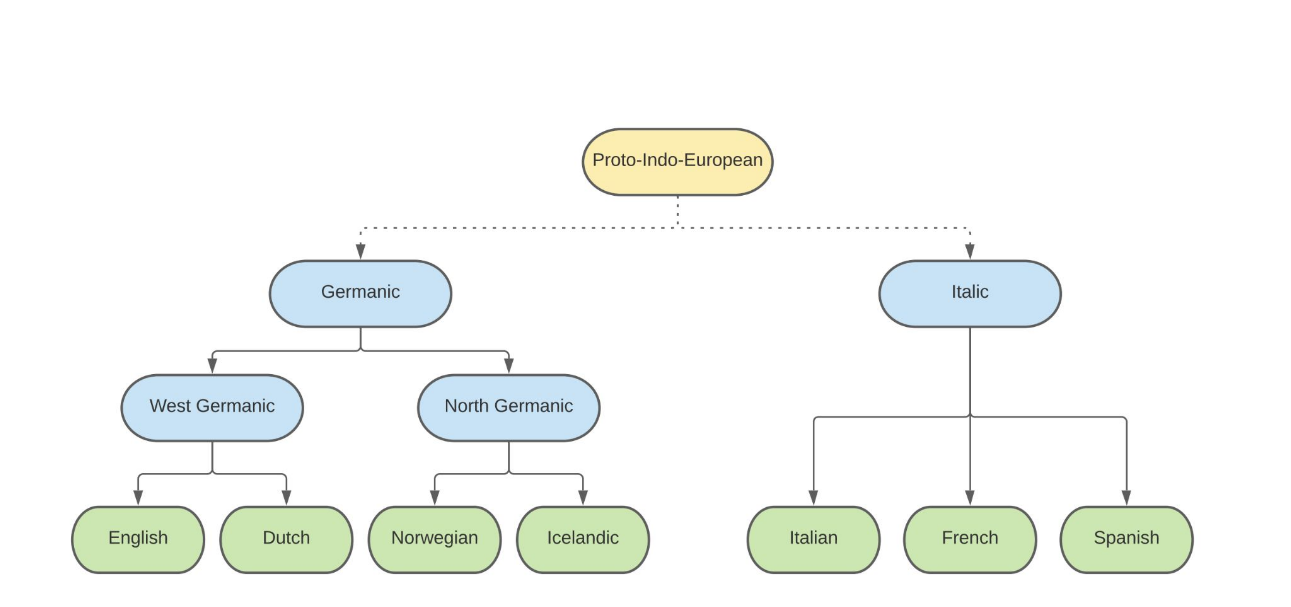
- N-gram model
- Seq2Seq Bi-Directional LSTM
- Transformer Encoder

3 training paradigms are used:

- **Fixed model :** Trained on English *Monolingual*
- **Global model :** Trained on mixture of set of training langs $T \subset L$ *Multilingual*
- **Nearest model :** Using the nearest language's model for inference. *Monolingual*
- **Ensemble Model : Ensembling of nearest models ← Focus** *Multilingual*

Phylogenetic Tree of Languages

These trees are constructed by using languages in **Glottolog** database by a Root node, and joining other languages which share similar influences.



Selecting nearest languages

Compute **nearest languages/highly related** by using **d(Lang1,Lang2)** for all training langs T.

$$d(l_1, l_2) = H(l_1) + H(l_2) - H(LCA(l_1, l_2))$$

Discussions

Nearest models have a flaw that the nearest lang d(l1,l2) could be low resource, but this is better than just an English model which might not share any linguistic similarities with the language in Test set.

The global language model suffers from the inconsistency of the training set: the same grapheme might map to different phonemes in different languages, therefore it cannot learn consistent rules across all languages.

Ensemble model relies on more than 1 language when predicting for the target language: even 1 language is a low-resource language, other languages might be able to compensate for that low-resource language. Additionally, introducing more language also reduces the variance

Ensembling

Global models are trained on many languages, so they fail to map some concrete rules and make mistakes. This introduces variance in output.

Ensembling reduces variances caused in nearest model (neighbouring language could contain low-resource language in the k nearest langs).

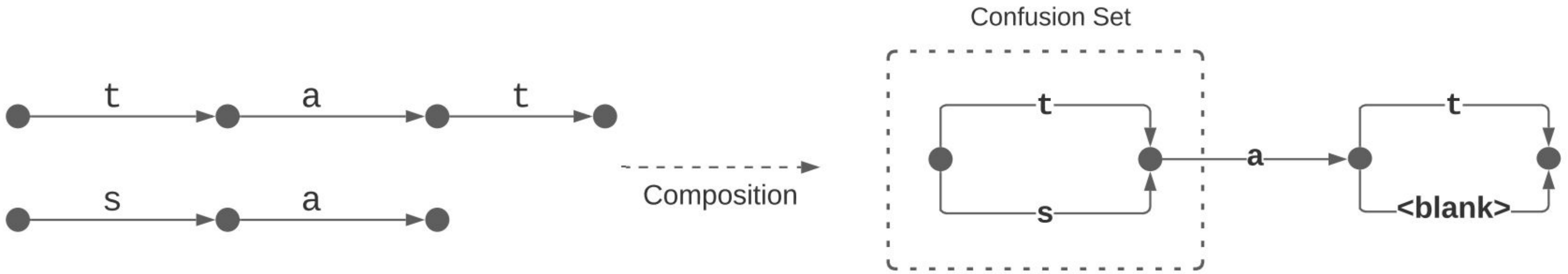
Ensembling algorithm

1. Train monolingual G2P supervised models on high resource languages.
2. Create Phylogenetic tree from **GlottoLog**
3. Given a **unseen Lang (test)** find **k nearest similar training langs** $T \subset L$ (db)
4. Use output phonemes of k nearest similar Neural Model(hypothesized phonemes given graphemes),

   and convert them to graphs.

6. Align the graphs.

7. Select the path based on voting (/t/,/s/ , /t/→/t/) or nearest lang hypo during a tie.



Confusion Network Lattice

Iteration 1 : 1-nearest & 1+1 nearest lattice aligned & composed

Iteration 2 : (1 & 2) & 3 nearest aligned and composed.

Results

	N-gram Model				LSTM Model				Transformer Model			
	PER	Add	Del	Sub	PER	Add	Del	Sub	PER	Add	Del	Sub
Fixed Model	76.0	4.52	9.39	62.1	78.1	4.53	20.4	53.2	78.5	3.2	19.0	56.2
Global Model	70.4	6.89	9.86	53.6	72.8	3.4	29.0	43.4	74.2	2.9	20.6	50.8
Nearest Model	68.4	4.51	12.4	51.5	43.8	12.1	4.0	27.6	45.4	15.8	3.6	26.1
Ensemble Model	**55.0**	0.56	23.6	30.9	**35.7**	10.0	3.4	22.2	**39.8**	13.9	3.1	22.8

DONE

:)

Top-level families not joined by a Root node

Families

Showing 1 to 100 of 427 entries (filtered from 4,815 total entries)

← Previous 1 2 3 4 5 Next →

Name	Level	Macro-area	Sub-families	Child languages	Top-level family
Search	Top-leve	--any--	Search	Search	Search
Atlantic-Congo	Top-level family	Africa, North America	899	1406	
Austronesian	Top-level family	Africa, Eurasia, Papunesia, South America	740	1271	
Indo-European	Top-level family	Africa, Australia, Eurasia, North America, Papunesia, South America	337	583	
Sino-Tibetan	Top-level family	Eurasia	303	501	
Bookkeeping	Top-level family	Africa, Australia, Eurasia, North America, Papunesia, South America	1	392	
Afro-Asiatic	Top-level family	Africa, Eurasia	229	379	
Nuclear Trans New Guinea	Top-level family	Papunesia	178	317	
Pama-Nyungan	Top-level family	Australia	152	250	
Sign Language	Top-level family	Africa, Australia, Eurasia, North America, Papunesia, South America	48	215	
Otomanguean	Top-level family	North America	92	181	
Austroasiatic	Top-level family	Eurasia	94	158	
Unclassifiable	Top-level family	Africa, Australia, Eurasia, North America, Papunesia, South America	0	121	
Tai-Kadai	Top-level family	Eurasia	55	95	
Pidgin	Top-level family	Africa, Australia, Eurasia, North America, Papunesia, South America	48	84	
Dravidian	Top-level family	Eurasia	47	82	
Arawakan	Top-level family	North America, South America	52	77	
Mande	Top-level family	Africa	56	75	
Tupian	Top-level family	South America	43	71	
Uto-Aztecan	Top-level family	North America	36	69	

Performance of ensemble

In general as N increases Add. increases,

del decreases.

More hypothesis phonemes, means more lattices, and more lattice paths.