Lily Kawaoto
Andrew Davis
L542 Fall 2021

Computational Phonology - Systemizing Minimal-Distance of Sounds

I. Introduction

Phonotactics is the branch of phonology that interacts with the restrictions of phonological sound patterns in a given language. The goal of our project is to create a computational model that predicts such restrictions for English, and in the process identify the core sounds compared to the peripheral sounds of English; furthermore, we want the model to predict natural classes of sounds. This paper details the steps we have taken so far, as well as our plan moving forward.

II. Methods

In this section, we first explain which sources we drew upon for the project and our reasons for choosing them. We then walk through the rationale behind certain implementations in the code.

A. Resources

We chose English as our target language due to its familiarity and the abundance of open-source resources. PanPhon (Mortensen, 2016) is one such example of a resource which conveniently maps IPA features to their respective sets of phonological features. Appendix 1 ('*ipa_all.csv*') is taken directly from PanPhon's GitHub code repository. It contains the IPA symbols along the rows and each column is filled with a 0, -, or + to represent, respectively, absence of a unary feature, the lack of binary feature, or presence of a binary feature for that row's phoneme. However, we did not agree with some of their feature choices; for example, this chart was missing the dental, Dorsal, and labiodental features, but contained columns for high and low tones (which are irrelevant in English), and did not appear to be consistent in choosing 0 or - to show the lack of a feature. Taking Professor Cavar's recommendation, we referred to the IPA chart by Hayes (2009) to edit the PanPhon table of features (see Appendix 2). There were two more modifications we made:

1. *Modification 1:* We converted 0, -, and + into numbers, since numbers are interpretable to computers while string characters like - and + are not. It follows that the most reasonable and intuitive numbers to convert - and + to are -1 and +1, respectively. 0 was kept in cases where a - feature did not make sense, such as for unary features.

2. *Modification 2:* The other modification incorporates our knowledge from feature geometry. We decided to assign different numerical weights depending on whether a + feature was unary or binary -- specifically, for features [LABIAL], [CORONAL], [DORSAL], [PHARYNGEAL], and [LARYNGEAL]. Referring to the trees in Appendix

7, we can see that binary features such as [+round] and [+labiodental] can only be possible if the unary feature is [+LABIAL].

It can thus be argued that unary features have more importance in the hierarchical structure of a feature geometry tree. To reflect this hierarchical difference, we assigned a weight of 2 to a positive feature if it was Unary; otherwise, we kept it as 1. Of course, these numbers are arbitrary, but they represent our intuition of feature geometry. The possible numerical numbers that a feature for a given IPA symbol can be associated with are therefore represented with the set {-1, 0, 1, 2}. Our final feature chart is shown in Appendix 3 ('*english_ipa.csv*').

As we will explain in the next subsection, NetworkX (Aric, 2008) was used to create visual graphs depicting the relationship between all the IPA symbols, as well as each IPA symbol and their phonological features.

In Part C we briefly talk about the transcriptions of L2 English speakers taken from the Speech Accent Archive, an online database of audio files and their corresponding narrow transcriptions from various English speakers, and how we intend to use this data for L2 applications.

B. Code

The next step bridges our abstract understanding of phonological features and the visualizations of each IPA symbol to their prominent features. We now have a way to represent features numerically; the simplest way to map a list of IPA symbols (with a known order) to their set of features is with a large list of vectors. For easier conceptualization, we can think of a list as a drawer (orange rectangle, in Appendix 4). This drawer is further divided into equal sections (which the arrows point at); each of these equal sections represents a vector, and all sections must contain the same number of items (the highlighted elements and all others within the square brackets).

Following the representation of each IPA symbol in terms of features is the graph visualization using a Python package called NetworkX. A graph, in the most rudimentary form, is simply two nodes connected by an edge. This edge shows that there exists a relationship between the nodes, and this edge can have different values depending on the strength of the relationship. In our case, each number in the feature vector from the previous step represents the weight of a feature for its associated IPA symbol. We coded this relationship as triplets in this order: [IPA symbol, feature, weight]. Our complete IPA-to-feature relationship graph therefore uses all the possible triplets for English. In other words, each IPA symbol node is connected to all the feature nodes, but the weight between that symbol and specific features vary. Having these relationship visualizations validate our qualitative data (i.e. our numerical choice of weights in the feature geometry hierarchy) while also providing a meaningful presentation to those with and without a computational or linguistic background.

The challenge in creating these graphs was to elucidate our array of data. Appendix 8 shows the progression of our arrays from three instances and illustrates the position of all the IPA and feature nodes as a result of this tug-of-war of weights. Following our in-class presentation,

we received input from fellow class phonologists. We intend to modify our code in order to provide the most validating array of our data.

(1) We are still deciding the most effective parameters to present the data in NetworkX.
(2) Note that we decided to color the edges' white in order to present the data in the most presentable manner.
(3) Essentially, we want to demonstrate the core sounds/natural classes of sounds in the data
    (a) Qualitative input welcome; see Appendix 8.

C. L2 Applications

Our goal is to create a predictive phonotactic model, and this can have various applications. Now that we have established the foundations of the project, we will focus on comparing the phonotactics of L1 and L2 English speakers (again, our target language is English). To do so, we selected phonemic transcriptions from the Speech Accent Archive, a large database containing audio files and IPA transcriptions of different English speakers. These transcriptions belong to native English speakers from the United States, native Arabic speakers, and native Japanese speakers.

First, separate transcription files were created for the L1 Arabic and L1 Japanese speakers, in which both the narrow and broad transcriptions are recorded. Then, using the broad transcriptions, we created two separate IPA phonological feature charts in which we added instances of phonemes produced by these L2 speakers but which are not found in English. See Appendix 5 and 6 for the phonological feature charts of symbols taken from transcriptions of native Arabic and Japanese speakers reading an English script. Below are examples of broad transcriptions from one native Arabic speaker, one native monolingual English speaker, and one native Japanese speaker.

Arabic - Example 1 - Female
/bəliːz kɔl stelə æsk her tu briŋ ðɪs θiŋz wɪθ her frɑm ðə stɔr sɪks spun ʌf frɛʃ sno bis faɪf θɪks əslɪp ʌf blu t͡ʃis ʌnd mebi ə snek fɔr her brʌðer bɑp wi ɔlso nid ə smɔl blæstɪk snek ɛnd ə bɪk tɔɪ frɔg fɔr ðʌ kɪts ʃi kæn skup ðɪs θɪŋs ɪntu θri rɛt bæks ɛnd wi wɪl go mit her wɛnɪzde ɛt ðə tren steʃən/

English - Example 15 - Female
/pliz kɔ stɪla ɐsk ɜ tə bɹɪŋə θiθ θɪŋz wɪð ɚ fɹəm ðə stɔɚ sɪks spɯnz əf fɹɛs sno piːz faɪf θɪk slaβz ə blu tʃiːz æ meɪbi ə snak fɔ hə bɹʌðə bɒːb wɪ also niɾ ə smɔ plasɪk sneɪk ɛn ə bɪk tɔɪ fɹɑg fɔ ðə kɪz ʃi kæn skup ðiθ θɪŋs ɪntu θʏ ɛd bægs æ wɪl go mid ɚ wɛnzde æ ðə tʃɹɛm steɪʃən/

Japanese - Example 18 - Female
/pliz kɔ stɪla ɐsk ɜ tə bɹɪŋə θiθ θɪŋz wɪð ɚ fɹəm ðə stɔɚ sɪks spɯnz əf fɹɛs sno piːz faɪf θɪk slaβz ə blu tʃiːz æ meɪbi ə snak fɔ hə bɹʌðə bɒːb wɪ also niɾ ə smɔ plasɪk sneɪk ɛn ə bɪk tɔɪ fɹɑg fɔ ðə kɪz ʃi kæn skup ðiθ θɪŋs ɪntu θʏ ɛd bægs æ wɪl go mid ɚ wɛnzde æ ðə tʃɹɛm steɪʃən/

This step is still ongoing, and we hope to continue developing more transcriptions to serve as training data for our phonotactic model.

III. Observations

Our analyses of the transcriptions were in some cases confirmed and in other cases invalidated some of our hypotheses, see below:

A. One of our predictions was accurate in terms of L2 English speakers, whose native language is Japanese, in terms of breaking-up phonemic consonant clusters:
    a. Ex. Vowel insertion: blue = /bəɾu/ (*Speech Accent Archive, Japanese 18.*)
        i. Japanese Consonant Cluster Break
    b. Ex. 2 was for the word 'three': /θʏ/ (*Speech Accent Archive, Japanese 18.*)
        i. Unexpected Interim Sound
B. Our predictions for L2 English Speakers, whose native language is Arabic, weren't as accurate:
    a. Ex. Lack of dentalization, instead voiceless plosives "... fɔr ðə kid̥s… " or "… θri ɹɛd̥ʔ" (*Speech Accent Archive, Arabic 3.*)
        i. The plosives of MSA (Modern-Standard Arabic) are dentalized, so it was interesting that this feature was not in the L2 transcription.
    b. Ex. 2 Lack of emphatic articulation carrying into L2 transcription; We don't believe this paragraph is the best elicitation, so, this would be interesting to create a separate test for emphatics in L2 English for native-Arabic speakers
C. We were unsure of how to tokenize the length marker between Narrow/Broad Transcription for a couple of reasons:
    a. Ex. in Arabic short/long vowels have considerable contrast (Mitleb, 229.) ; we want to discuss in terms of qualitative input the decision to include these in our broad transcription
    b. Some of the transcriptions used both the length marker (:) and the "doubling-up" of vowels (... *bluu*… or ... *skuup*…) within the same elicitation (*Speech Accent Archive, English 15.*)

Before moving forward, we'd like to discuss these observations further with Professor Cavar in order to standardize our qualitative input with maximum professional input.

IV. Future plans

The first steps will be to standardize a corpus of transcriptions from *The Speech Archive* for several native English speakers and several L2 English speakers whose native languages are Arabic/Japanese based on qualitative input. We'll vectorize this data in order to generate phonological relationship-graphs for the transcribed results like we did with our initial dataset in Phase I.

Then, we will introduce quantitative data via extracting phonotactic probabilities for each category of speakers using word embeddings. These vectors will be different from Phase I's vectors as they represent the words in multi-dimensional vectors that correspond with neural networks, probabilistic models, etc. like we intend to build. A productive way to conceptualize the probabilistic model is that it focuses on the frequency of phoneme *x* occurring next to

phoneme *y.* Then we will visualize these models and we predict that they will demonstrate core versus peripheral sounds in a given language, as well as natural classes, and validate this model.

Once again, we'll seek mentorship in order to compare and analyze what the probabilistic, relationship-graph shows about one's native language and how it can influence L2 acquisition/speech recognition. We will refine the model (word embeddings) as needed in order to generate a valid relationship-graph that correlates with, or improves upon, the qualitative data.

Our goal is for the model to efficiently incorporate both qualitative and quantitative data in the most accurate manner; we'll need to standardize a format that utilizes both inputs. Then, we'll use this standardized, phonotactic "score", that we've created, to analyze minimal-distances between sounds, languages, and apply it to language learning software, speech recognition software, and beyond.

We are interested in running various corpora through in order to examine the patterns that are generated. Our model will consolidate phonetic *and* phonological (qualitative & quantitative) to most successfully determine what sounds are acting as a barrier to a language-learner's success.

V. Conclusion

When learning a new language, one cannot avoid encountering a new underlying phonological system of sounds. This has been a challenge to AI, applications, language-learners, programs, professionals, and software that has the opportunity to be solved. The ability to take any dataset for an application/language-learner and provide them with the essential feedback they need to progress in their desired language based upon refined qualitative and quantitative data will, as we hypothesize, accelerate the acquisition of sounds of a speech pattern for an application or sounds of new language for L2 acquisition.

In the next course, we intend to arrive at 1) the consolidation of qualitative and quantitative phonotactic data in a productive manner and, if possible, 2) generate a preemptive minimal-distance "score" that can be portrayed on a relationship-graph for further analysis and refinement. Our program will influence the perception of sounds through these relevant, relationship-graphs (that will substantiate our methods).

Finally, through the process of generating a "score" of minimal-distance between sounds/languages, this score can be applied to a variety of NLP programs in a high-yielding manner.

References

Aric A. Hagberg, Daniel A. Schult and Pieter J. Swart, "Exploring network structure, dynamics, and function using NetworkX." *Proceedings of the 7th Python in Science Conference (SciPy2008)*, Gäel Varoquaux, Travis Vaught, and Jarrod Millman (Eds), Pasadena, CA USA, 11–15, Aug 2008

David R. Mortensen, Patrick Littell, Akash Bharadwaj, Kartik Goyal, Chris Dyer, Lori Levin (2016). "PanPhon: A Resource for Mapping IPA Segments to Articulatory Feature Vectors." *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 3475–3484, Osaka, Japan, December 11-17, 2016.

Hayes, Bruce. *Introductory Phonology*. Malden, MA: Wiley-Blackwell, 2009. Print.

Mitleb, Fares. "Vowel Length Contrast in Arabic and English: a Spectrographic Test." *Journal of Phonetics*, 12 Apr. 1984, pp. 229–233.

Speech Accent Archive, http://accent.gmu.edu/index.php.

Appendix[1]

1. *ipa_all.csv* from PanPhon

ipa	syl	son	cons	cont	delrel	lat	nas	strid	voi	sg	cg	ant	cor	distr	lab	hi	lo	back	round	velaric	tense	long	hitone	hireg
ɻ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-	-
ɻ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	+	-
ɻ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ɻ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-	+
ɻ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	+	+
ɻˠ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	+	+	0	0	0	0	-	-
ɻˠ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	+	+	0	0	0	0	+	-
ɻˠ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	+	+	0	0	0	0	0	0
ɻˤ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	+	+	0	0	0	0	-	+
ɻˤ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	+	+	0	0	0	0	+	+
ɑ	-	-	+	-	-	-	-	0	-	-	-	-	-	0	-	-	-	+	-	-	0	-	0	0
ɑ̤	-	-	+	-	-	-	-	0	-	-	-	-	-	0	-	-	-	+	-	-	0	-	0	0
ɑ̰	-	-	+	-	-	-	-	0	-	-	-	-	-	0	-	-	-	+	-	-	0	-	0	0
ɑ	-	-	+	-	-	-	-	0	+	-	-	-	-	0	-	-	-	+	-	-	0	-	0	0
ɒ̤	-	-	+	-	-	-	-	0	+	-	-	-	-	0	-	-	-	+	-	-	0	-	0	0
ɒ̰	-	-	+	-	-	-	-	0	+	-	-	-	-	0	-	-	-	+	-	-	0	-	0	0
ɑː	-	-	+	-	-	-	-	0	-	-	-	-	-	0	-	-	-	+	-	-	0	+	0	0
ɑ	-	-	+	-	-	-	-	0	+	-	-	-	-	0	-	-	-	+	-	-	0	+	0	0
ʈ	-	-	+	-	-	-	-	0	-	-	-	-	+	-	-	-	-	-	-	-	0	-	0	0
ʈ̤	-	-	+	-	-	-	-	0	-	-	-	-	+	-	-	-	-	-	-	-	0	-	0	0
ʈ̰	-	-	+	-	-	-	-	0	-	-	-	-	+	-	-	-	-	-	-	-	0	-	0	0
ʈ	-	-	+	-	-	-	-	0	-	-	-	-	+	-	-	-	-	-	-	-	0	-	0	0
ɖ	-	-	+	-	-	-	-	0	+	-	-	-	+	-	-	-	-	-	-	-	0	-	0	0
ɖ̤	-	-	+	-	-	-	-	0	+	-	-	-	+	-	-	-	-	-	-	-	0	-	0	0
ɖ̰	-	-	+	-	-	-	-	0	+	-	-	-	+	-	-	-	-	-	-	-	0	-	0	0

[1] Only the first page or so of the files and code are shown for the sake of brevity.

2. Hayes's phonological feature charts from *Introductory Phonology* (2009)

Table 4.7 Consonants I: single place of articulation

		Manner features								Laryngeal features			Place features												
		consonantal	sonorant	continuant	delayed release	approximant	tap	trill	nasal	voice	spread gl	constr gl	labial	round	labiodental	coronal	anterior	distributed	strident	lateral	dorsal	high	low	front	back	tense
bilabial	p	+	–	–	–	–	–	–	–	–	–	–	+	–	–	–	0	0	0	–	–	0	0	0	0	0
	b	+	–	–	–	–	–	–	–	+	–	–	+	–	–	–	0	0	0	–	–	0	0	0	0	0
	ɸ	+	–	+	+	–	–	–	–	–	–	–	+	–	–	–	0	0	0	–	–	0	0	0	0	0
	β	+	–	+	+	–	–	–	–	+	–	–	+	–	–	–	0	0	0	–	–	0	0	0	0	0
	m	+	+	–	0	–	–	–	+	+	–	–	+	–	–	–	0	0	0	–	–	0	0	0	0	0
	ʙ	+	+	+	0	+	–	+	–	+	–	–	+	–	–	–	0	0	0	–	–	0	0	0	0	0
labiodental	p͡f	+	–	–	+	–	–	–	–	–	–	–	+	–	+	–	0	0	0	–	–	0	0	0	0	0
	f	+	–	+	+	–	–	–	–	–	–	–	+	–	+	–	0	0	0	–	–	0	0	0	0	0
	v	+	–	+	+	–	–	–	–	+	–	–	+	–	+	–	0	0	0	–	–	0	0	0	0	0
	ɱ	+	+	–	0	–	–	–	+	+	–	–	+	–	+	–	0	0	0	–	–	0	0	0	0	0
	ʋ	–	+	+	0	+	–	–	–	+	–	–	+	–	+	–	0	0	0	–	–	0	0	0	0	0
dental	t̪	+	–	–	–	–	–	–	–	–	–	–	–	–	–	+	+	+	–	–	–	0	0	0	0	0
	d̪	+	–	–	–	–	–	–	–	+	–	–	–	–	–	+	+	+	–	–	–	0	0	0	0	0
	θ	+	–	+	+	–	–	–	–	–	–	–	–	–	–	+	+	+	–	–	–	0	0	0	0	0
	ð	+	–	+	+	–	–	–	–	+	–	–	–	–	–	+	+	+	–	–	–	0	0	0	0	0
alveolar	t	+	–	–	–	–	–	–	–	–	–	–	–	–	–	+	+	–	–	–	–	0	0	0	0	0
	d	+	–	–	–	–	–	–	–	+	–	–	–	–	–	+	+	–	–	–	–	0	0	0	0	0
	t͡s	+	–	–	+	–	–	–	–	–	–	–	–	–	–	+	+	–	+	–	–	0	0	0	0	0
	d͡z	+	–	–	+	–	–	–	–	+	–	–	–	–	–	+	+	–	+	–	–	0	0	0	0	0
	t͡ɬ	+	–	–	+	–	–	–	–	–	–	–	–	–	–	+	+	–	–	+	–	0	0	0	0	0
	s	+	–	+	+	–	–	–	–	–	–	–	–	–	–	+	+	–	+	–	–	0	0	0	0	0
	z	+	–	+	+	–	–	–	–	+	–	–	–	–	–	+	+	–	+	–	–	0	0	0	0	0
	n	+	+	–	0	–	–	–	+	+	–	–	–	–	–	+	+	–	–	–	–	0	0	0	0	0
	l	+	+	+	0	+	–	–	–	+	–	–	–	–	–	+	+	–	–	+	–	0	0	0	0	0
	ɬ	+	–	+	+	–	–	–	–	–	–	–	–	–	–	+	+	–	–	+	–	0	0	0	0	0
	ɮ	+	–	+	+	–	–	–	–	+	–	–	–	–	–	+	+	–	–	+	–	0	0	0	0	0
	ɾ	+	+	+	0	+	+	–	–	+	–	–	–	–	–	+	+	–	–	–	–	0	0	0	0	0
	ɺ	+	+	+	0	+	+	–	–	+	–	–	–	–	–	+	+	–	–	+	–	0	0	0	0	0
	r	+	+	+	0	+	–	+	–	+	–	–	–	–	–	+	+	–	–	–	–	0	0	0	0	0
palato-alveolar	t͡ʃ	+	–	–	+	–	–	–	–	–	–	–	–	–	–	+	–	+	+	–	–	0	0	0	0	0
	d͡ʒ	+	–	–	+	–	–	–	–	+	–	–	–	–	–	+	–	+	+	–	–	0	0	0	0	0
	ʃ	+	–	+	+	–	–	–	–	–	–	–	–	–	–	+	–	+	+	–	–	0	0	0	0	0
	ʒ	+	–	+	+	–	–	–	–	+	–	–	–	–	–	+	–	+	+	–	–	0	0	0	0	0
	ɹ	–	+	+	0	+	–	–	–	+	–	–	–	–	–	+	–	+	–	–	–	0	0	0	0	0

Table 4.7 (*cont'd*)

		Manner features								Laryngeal features			Place features												
		consonantal	sonorant	continuant	delayed release	approximant	tap	trill	nasal	voice	spread gl	constr gl	labial	round	labiodental	coronal	anterior	distributed	strident	lateral	dorsal	high	low	front	back	tense
retroflex	ʈ	+	–	–	–	–	–	–	–	–	–	–	–	–	–	+	–	–	–	–	–	0	0	0	0	0
	ɖ	+	–	–	–	–	–	–	–	+	–	–	–	–	–	+	–	–	–	–	–	0	0	0	0	0
	ʂ	+	–	+	+	–	–	–	–	–	–	–	–	–	–	+	–	–	+	–	–	0	0	0	0	0
	ʐ	+	–	+	+	–	–	–	–	+	–	–	–	–	–	+	–	–	+	–	–	0	0	0	0	0
	ɳ	+	+	–	0	–	–	–	+	+	–	–	–	–	–	+	–	–	–	–	–	0	0	0	0	0
	ɭ	+	+	+	0	+	–	–	–	+	–	–	–	–	–	+	–	–	–	+	–	0	0	0	0	0
	ɽ	+	+	+	0	+	+	–	–	+	–	–	–	–	–	+	–	–	–	–	–	0	0	0	0	0
	ɻ	+	+	+	0	+	–	–	–	+	–	–	–	–	–	+	–	–	–	–	–	0	0	0	0	0
fronted velar	k̟	+	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0	0	0	–	+	+	–	+	–	0
	ɡ̟	+	–	–	–	–	–	–	–	+	–	–	–	–	–	–	0	0	0	–	+	+	–	+	–	0
	x̟	+	–	+	+	–	–	–	–	–	–	–	–	–	–	–	0	0	0	–	+	+	–	+	–	0
	j	–	+	+	0	+	–	–	–	+	–	–	–	–	–	–	0	0	0	–	+	+	–	+	–	+
velar	k	+	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0	0	0	–	+	+	–	0	0	0
	ɡ	+	–	–	–	–	–	–	–	+	–	–	–	–	–	–	0	0	0	–	+	+	–	0	0	0
	ŋ	+	+	–	0	–	–	–	+	+	–	–	–	–	–	–	0	0	0	–	+	+	–	0	0	0
	k͡x	+	–	–	+	–	–	–	–	–	–	–	–	–	–	–	0	0	0	–	+	+	–	0	0	0
	x	+	–	+	+	–	–	–	–	–	–	–	–	–	–	–	0	0	0	–	+	+	–	0	0	0
	ɣ	+	–	+	+	–	–	–	–	+	–	–	–	–	–	–	0	0	0	+	+	+	–	0	0	0
	ʟ	+	+	+	0	+	–	–	–	+	–	–	–	–	–	–	0	0	0	–	+	+	–	0	0	0
back velar	k̠	+	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0	0	0	+	+	+	–	–	+	0
	ɡ̠	+	–	–	–	–	–	–	–	+	–	–	–	–	–	–	0	0	0	–	+	+	–	–	+	0
	x̠	+	–	+	+	–	–	–	–	–	–	–	–	–	–	–	0	0	0	–	+	+	–	–	+	0
	ɣ̠	+	–	+	+	–	–	–	–	+	–	–	–	–	–	–	0	0	0	–	+	+	–	–	+	0
uvular	q	+	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0	0	0	–	+	–	–	–	+	0
	ɢ	+	–	–	–	–	–	–	–	+	–	–	–	–	–	–	0	0	0	–	+	–	–	–	+	0
	χ	+	–	+	+	–	–	–	–	–	–	–	–	–	–	–	0	0	0	–	+	–	–	–	+	0
	ʁ	+	–	+	+	–	–	–	–	+	–	–	–	–	–	–	0	0	0	–	+	–	–	–	+	0
	ɴ	+	+	–	0	–	–	–	+	+	–	–	–	–	–	–	0	0	0	–	+	–	–	–	+	0
	ʀ	+	+	+	0	+	–	+	–	+	–	–	–	–	–	–	0	0	0	–	+	–	–	–	+	0
pharyn	ħ	+	–	+	+	–	–	–	–	–	–	–	–	–	–	–	0	0	0	–	+	–	+	–	+	0
	ʕ	+	–	+	–	–	–	–	–	+	–	–	–	–	–	–	0	0	0	–	+	–	+	–	+	0
glottal	ʔ	+	–	–	–	–	–	–	–	–	–	+	–	–	–	–	0	0	0	–	–	0	0	0	0	0
	h	–	–	+	+	–	–	–	–	–	+	–	–	–	–	–	0	0	0	–	–	0	0	0	0	0
	ɦ	–	–	+	+	–	–	–	–	+	+	–	–	–	–	–	0	0	0	–	–	0	0	0	0	0

Table 4.8 Consonants II: complex segments

		Manner features								Laryngeal features			Place features													
		consonantal	sonorant	continuant	delayed release	approximant	tap	trill	nasal	voice	spread gl	constr gl	labial	round	labiodental	coronal	anterior	distributed	strident	lateral	dorsal	high	low	front	back	tense
labial-back velar	w	−	+	+	0	+	−	−	−	+	−	−	+	+	−	−	0	0	0	−	+	+	−	−	+	+
	ʍ	−	−	+	+	−	−	−	−	−	+	−	+	+	−	−	0	0	0	−	+	+	−	−	+	+
labial-velar	k͡p	+	−	−	−	−	−	−	−	−	−	−	+	−	−	−	0	0	0	−	+	+	−	0	0	0
	g͡b	+	−	−	−	−	−	−	−	+	−	−	+	−	−	−	0	0	0	−	+	+	−	0	0	0
labial-front velar	ɥ	−	+	+	0	+	−	−	−	+	−	−	+	+	−	−	0	0	0	−	+	+	−	+	−	+
alveolopalatal	t͡ɕ	+	−	−	+	−	−	−	−	−	−	−	−	−	−	+	+	+	+	−	+	+	−	+	−	0
	d͡ʑ	+	−	−	+	−	−	−	−	+	−	−	−	−	−	+	+	+	+	−	+	+	−	+	−	0
	ɕ	+	−	+	+	−	−	−	−	−	−	−	−	−	−	+	+	+	+	−	+	+	−	+	−	0
	ʑ	+	−	+	+	−	−	−	−	+	−	−	−	−	−	+	+	+	+	−	+	+	−	+	−	0
palatal	c	+	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	+	−	−	+	+	−	+	−	0
	ɟ	+	−	−	−	−	−	−	−	+	−	−	−	−	−	+	−	+	−	−	+	+	−	+	−	0
	ç	+	−	+	+	−	−	−	−	−	−	−	−	−	−	+	−	+	−	−	+	+	−	+	−	0
	ʝ	+	−	+	+	−	−	−	−	+	−	−	−	−	−	+	−	+	−	−	+	+	−	+	−	0
	ɲ	+	+	−	0	−	−	−	+	+	−	−	−	−	−	+	−	+	−	−	+	+	−	+	−	0
	ʎ	+	+	+	0	+	−	−	−	+	−	−	−	−	−	+	−	+	−	+	+	+	−	+	−	0

4.10.3 Vowels

The basic features for vowels (shown in table 4.9 by the basic IPA symbols rather than the diacritics) are [round], [high], [low], [front], and [back]. [labial] is predictable, occurring only in [+round] vowels. All other features are invariant; unless overridden by a diacritic, all vowels are [+syllabic, –consonantal, +sonorant, +continuant, 0delayed release, +approximant, –tap, –trill, –nasal, +voice, –spread glottis, –constricted glottis, –labiodental, –coronal, 0anterior, 0distributed, 0strident].

Table 4.9 Vowels

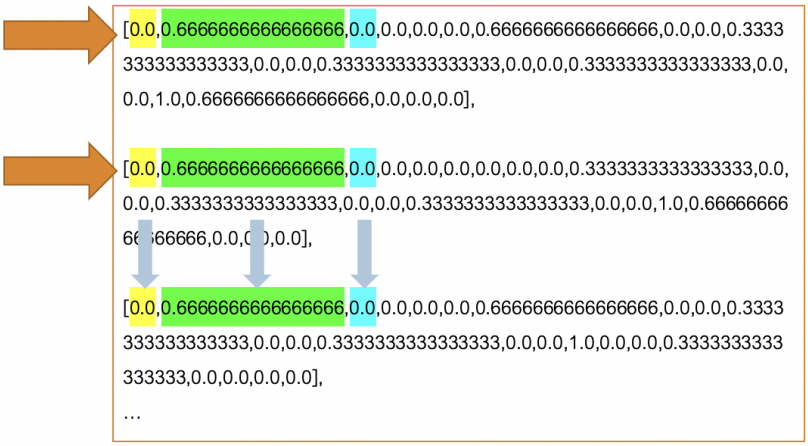
	high tense						high lax			mid tense					
	i	y	ɨ	ʉ	ɯ	u	ɪ	ʏ	ʊ	e	ø	ɘ	ɵ	ɤ	o
[high]	+	+	+	+	+	+	+	+	+	–	–	–	–	–	–
[low]	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
[tense]	+	+	+	+	+	+	–	–	–	+	+	+	+	+	+
[front]	+	+	–	–	–	–	+	+	–	+	+	–	–	–	–
[back]	–	–	–	–	+	+	–	–	+	–	–	–	–	+	+
[round]	–	+	–	+	–	+	–	+	+	–	+	–	+	–	+

	mid lax						low				
	ɛ	œ	ə	ɞ	ʌ	ɔ	æ	Œ	a	ɑ	ɒ
[high]	–	–	–	–	–	–	–	–	–	–	–
[low]	–	–	–	–	–	–	+	+	+	+	+
[tense]	–	–	–	–	–	–	0	0	0	0	0
[front]	+	+	–	–	–	–	+	+	–	–	–
[back]	–	–	–	–	+	+	–	–	–	+	+
[round]	–	+	–	+	–	+	–	+	–	–	+

3. *english_ipa.csv*, our PanPhon-and-Hayes-inspired CSV file with relevant English IPA symbols and their respective modified features

ipa	syl	cons	son	cont	delrel	nas	voi	lat	strid	LAR	sg	cg	COR	ant	distr	LAB	round	labiodental	DOR	hi	lo	back	tense
g	-	+	-	-	-	-	+	-	-	0	-	-	0	-	-	0	-	-	+	+	-	-	-
k	-	+	-	-	-	-	-	-	-	0	-	-	0	-	-	0	-	-	+	+	-	-	-
b	-	+	-	-	-	-	+	-	-	0	-	-	0	-	-	+	-	-	0	-	-	-	-
d	-	+	-	-	-	-	+	-	-	0	-	-	+	+	-	0	-	-	0	-	-	-	-
p	-	+	-	-	-	-	-	-	-	0	-	-	0	-	-	+	-	-	0	-	-	-	-
t	-	+	-	-	-	-	-	-	-	0	-	-	+	+	-	0	-	-	0	-	-	-	-
d͡ʒ	-	+	-	-	+	-	+	-	+	0	-	-	+	-	+	0	-	-	0	-	-	-	-
t͡ʃ	-	+	-	-	+	-	-	-	+	0	-	-	+	-	+	0	-	-	0	-	-	-	-
ʃ	-	+	-	+	+	-	-	-	+	0	-	-	+	-	+	0	-	-	0	-	-	-	-
ʒ	-	+	-	+	+	-	+	-	+	0	-	-	+	-	+	0	-	-	0	-	-	-	-
f	-	+	-	+	+	-	-	-	-	0	-	-	0	-	-	+	-	+	0	-	-	-	-
s	-	+	-	+	-	-	-	-	+	0	-	-	+	+	-	0	-	-	0	-	-	-	-
v	-	+	-	+	+	-	+	-	-	0	-	-	0	-	-	+	-	+	0	-	-	-	-
z	-	+	-	+	-	-	+	-	+	0	-	-	+	+	-	0	-	-	0	-	-	-	-
ð	-	+	-	+	-	-	+	-	-	0	-	-	+	+	+	0	-	-	0	-	-	-	-
θ	-	+	-	+	+	-	-	-	-	0	-	-	+	+	+	0	-	-	0	-	-	-	-
ʔ	-	+	+	-	-	-	-	-	-	+	-	+	0	-	-	0	-	-	0	-	-	-	-
ŋ	-	+	+	-	-	+	+	-	-	0	-	-	0	-	-	0	-	-	+	+	-	-	-
m	-	+	+	-	-	+	+	-	-	0	-	-	0	-	-	+	-	-	0	-	-	-	-
n	-	+	+	-	-	+	+	-	-	0	-	-	+	+	-	0	-	-	0	-	-	-	-
h	-	+	-	+	+	-	-	-	-	+	+	-	0	-	-	0	-	-	0	-	-	-	-
j	-	-	+	+	-	-	+	-	-	0	-	-	0	-	-	0	-	-	+	+	-	-	+
w	-	-	+	+	-	-	+	-	-	0	-	-	0	-	-	+	+	-	+	+	-	+	+
ɹ	-	-	+	+	-	-	+	-	-	0	-	-	+	-	+	0	-	-	0	-	-	-	-
l	-	+	+	+	-	-	+	+	-	0	-	-	+	+	-	0	-	-	+	-	-	-	-
ɑ	+	-	+	+	-	-	+	-	-	0	-	-	0	-	-	-	-	-	+	-	+	+	-
e	+	-	+	+	-	-	+	-	-	0	-	-	0	-	-	-	-	-	+	-	-	-	+
i	+	-	+	+	-	-	+	-	-	0	-	-	0	-	-	-	-	-	+	+	-	-	+
o	+	-	+	+	-	-	+	-	-	0	-	-	0	-	-	+	+	-	+	-	-	+	+
u	+	-	+	+	-	-	+	-	-	0	-	-	0	-	-	+	+	-	+	+	-	+	+
æ	+	-	+	+	-	-	+	-	-	0	-	-	0	-	-	-	-	-	+	-	+	-	-
ɔ	+	-	+	+	-	-	+	-	-	0	-	-	0	-	-	+	+	-	+	-	-	+	-
ə	+	-	+	+	-	-	+	-	-	0	-	-	0	-	-	-	-	-	+	-	-	-	-
ɛ	+	-	+	+	-	-	+	-	-	0	-	-	0	-	-	-	-	-	+	-	-	-	-
ɪ	+	-	+	+	-	-	+	-	-	0	-	-	0	-	-	-	-	-	+	+	-	-	-
ʊ	+	-	+	+	-	-	+	-	-	0	-	-	0	-	-	+	+	-	+	+	-	+	-
ʌ	+	-	+	+	-	-	+	-	-	0	-	-	0	-	-	-	-	-	+	-	-	+	-

4. *norm_vecs.json*[2] - the first three feature vectors representing the phonemes. See the explanation under Section II.A 'Methods' for color codes.



5. *L2_arb_ipa.csv*, our PanPhon-and-Hayes-inspired CSV file with relevant Modern-Standard Arabic IPA symbols and their respective modified features

L2_arb_ipa																									
ipa	syl	cons	son	cont	delrel	nas	voi	lat	strid	LAR	sg	cg	COR	ant	distr	dental	LAB	round	labiodental	DOR	hi	lo	back	tense	PHAR
b	-	+	-	-	-	-	+	-	-	0	-	-	0	-	-	-	+	-	-	0	-	-	-	-	0
f	-	+	-	+	+	-	-	-	-	0	-	-	0	-	-	-	+	-	+	0	-	-	-	-	0
m	-	+	+	-	-	+	+	-	-	0	-	-	0	-	-	-	+	-	-	0	-	-	-	-	0
w	-	-	+	+	-	-	+	-	-	0	-	-	0	-	-	-	+	+	-	+	+	-	+	+	0
θ	-	+	-	+	+	-	-	-	-	0	-	-	+	+	+	+	0	-	-	0	-	-	-	-	0
ð	-	+	-	+	-	-	+	-	-	0	-	-	+	+	+	+	0	-	-	0	-	-	-	-	0
ðˤ	-	+	-	+	-	-	+	-	-	0	-	-	+	+	+	+	0	-	-	0	-	-	-	-	+
t	-	+	-	-	-	-	-	-	-	0	-	-	+	+	-	-	0	-	-	0	-	-	-	-	0
t̪	-	+	-	-	-	-	-	-	-	0	-	-	+	+	-	+	0	-	-	0	-	-	-	-	0
tˤ	-	+	-	-	-	-	-	-	-	0	-	-	+	+	-	-	0	-	-	0	-	-	-	-	+
d	-	+	-	-	-	-	+	-	-	0	-	-	+	+	-	-	0	-	-	0	-	-	-	-	0
d̪	-	+	-	-	-	-	+	-	-	0	-	-	+	+	-	+	0	-	-	0	-	-	-	-	0
dˤ	-	+	-	-	-	-	+	-	-	0	-	-	+	+	-	-	0	-	-	0	-	-	-	-	+
s	-	+	-	+	-	-	-	-	+	0	-	-	+	+	-	-	0	-	-	0	-	-	-	-	0
z	-	+	-	+	-	-	+	-	+	0	-	-	+	+	-	-	0	-	-	0	-	-	-	-	0
sˤ	-	+	-	+	-	-	-	-	+	0	-	-	+	+	-	-	0	-	-	0	-	-	-	-	+
n	-	+	+	-	-	+	+	-	-	0	-	-	+	+	-	-	0	-	-	0	-	-	-	-	0
l	-	+	+	+	-	-	+	+	-	0	-	-	+	+	-	-	0	-	-	+	-	-	-	-	0
lˤ	-	+	+	+	-	-	+	+	-	0	-	-	+	+	-	-	0	-	-	+	-	-	-	-	+
r	-	+	+	+	-	-	+	-	-	0	-	-	+	+	-	-	0	-	-	0	-	-	-	-	0
d͡ʒ	-	+	-	-	+	-	+	-	+	0	-	-	+	-	+	-	0	-	-	0	-	-	-	-	0
ʃ	-	+	-	+	+	-	-	-	+	0	-	-	+	-	+	-	0	-	-	0	-	-	-	-	0
j	-	-	+	+	-	-	+	-	-	0	-	-	0	-	-	-	0	-	-	+	+	-	-	+	0
k	-	+	-	-	-	-	-	-	-	0	-	-	0	-	-	-	0	-	-	+	+	-	-	-	0
q	-	+	-	-	-	-	-	-	-	0	-	-	0	-	-	-	0	-	-	+	-	-	+	-	0

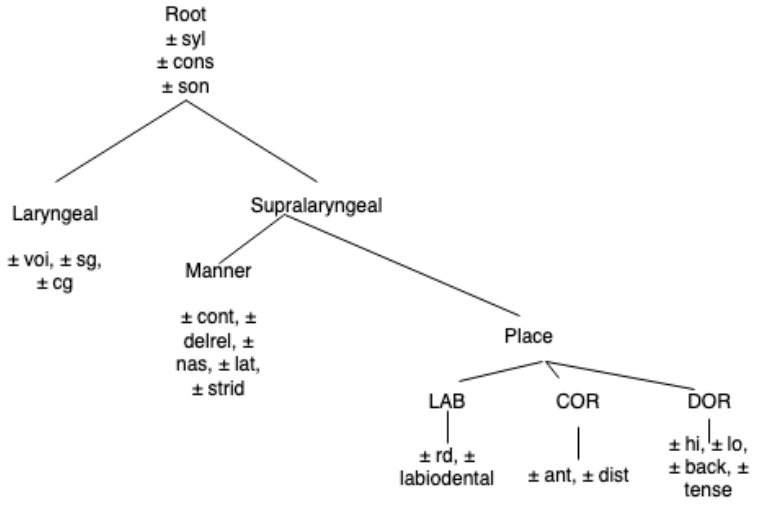
[2] Note that the vectors above have been normalized to be between 0 and 1; that means that previously, the lowest possible value of -1 is now 0, and the highest possible value of 2 is now 1. Scaling numbers to be within this range is common practice when doing Machine Learning, as it makes the results compatible both with computers and among programmers who may initially use different numbers to represent feature weights.

6. *L2_jpn_ipa.csv*, our English PanPhon-and-Hayes-inspired CSV file with relevant IPA symbols produced by native Japanese speakers and their respective modified features

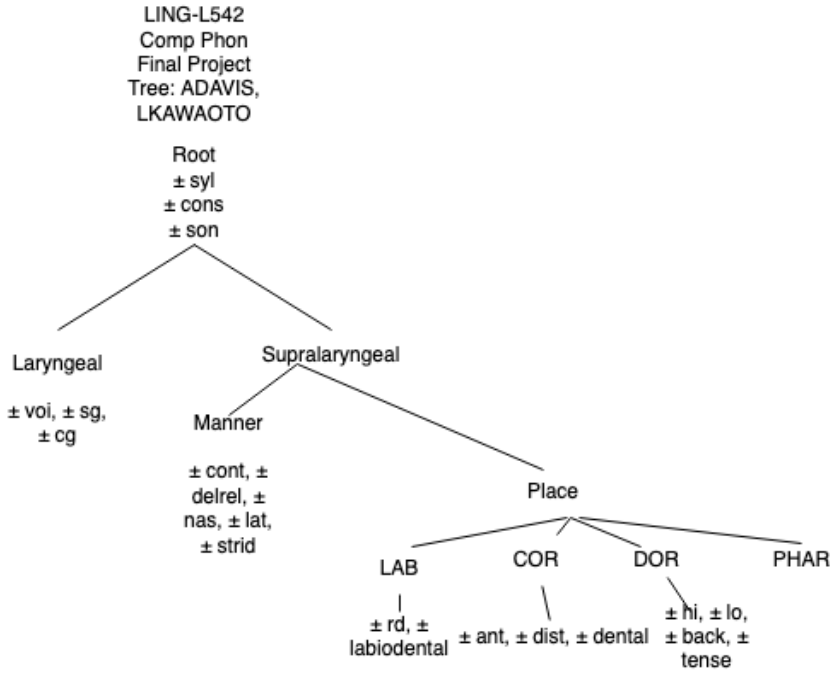
L2_jpn_ipa																							
ipa	syl	cons	son	cont	delrel	nas	voi	lat	strid	LAR	sg	cg	COR	ant	distr	LAB	round	labiodental	DOR	hi	lo	back	tense
g	-	+	-	-	-	-	+	-	-	0	-	-	0	-	-	0	-	-	+	+	-	-	-
k	-	+	-	-	-	-	-	-	-	0	-	-	0	-	-	0	-	-	+	+	-	-	-
b	-	+	-	-	-	-	+	-	-	0	-	-	0	-	-	+	-	-	0	-	-	-	-
d	-	+	-	-	-	-	+	-	-	0	-	-	+	+	-	0	-	-	0	-	-	-	-
p	-	+	-	-	-	-	-	-	-	0	-	-	0	-	-	+	-	-	0	-	-	-	-
t	-	+	-	-	-	-	-	-	-	0	-	-	+	+	-	0	-	-	0	-	-	-	-
d͡ʒ	-	+	-	-	+	-	+	-	+	0	-	-	+	-	+	0	-	-	0	-	-	-	-
t͡ʃ	-	+	-	-	+	-	-	-	+	0	-	-	+	-	+	0	-	-	0	-	-	-	-
ʃ	-	+	-	+	+	-	-	-	+	0	-	-	+	-	+	0	-	-	0	-	-	-	-
ʒ	-	+	-	+	+	-	+	-	+	0	-	-	+	-	+	0	-	-	0	-	-	-	-
f	-	+	-	+	+	-	-	-	-	0	-	-	0	-	-	+	-	+	0	-	-	-	-
s	-	+	-	+	-	-	-	-	+	0	-	-	+	+	-	0	-	-	0	-	-	-	-
v	-	+	-	+	+	-	+	-	-	0	-	-	0	-	-	+	-	+	0	-	-	-	-
z	-	+	-	+	-	-	+	-	+	0	-	-	+	+	-	0	-	-	0	-	-	-	-
ð	-	+	-	+	-	-	+	-	-	0	-	-	+	+	+	0	-	-	0	-	-	-	-
θ	-	+	-	+	+	-	-	-	-	0	-	-	+	+	+	0	-	-	0	-	-	-	-
ʔ	-	+	+	-	-	-	-	-	-	+	-	+	0	-	-	0	-	-	0	-	-	-	-
ŋ	-	+	+	-	-	+	+	-	-	0	-	-	0	-	-	0	-	-	+	+	-	-	-
m	-	+	+	-	-	+	+	-	-	0	-	-	0	-	-	+	-	-	0	-	-	-	-
n	-	+	+	-	-	+	+	-	-	0	-	-	+	+	-	0	-	-	0	-	-	-	-
h	-	+	-	+	+	-	-	-	-	+	+	-	0	-	-	0	-	-	0	-	-	-	-
j	-	-	+	+	-	-	+	-	-	0	-	-	0	-	-	0	-	-	+	+	-	-	+
w	-	-	+	+	-	-	+	-	-	0	-	-	0	-	-	+	+	-	+	+	-	+	+
ɹ	-	-	+	+	-	-	+	-	-	0	-	-	+	-	+	0	-	-	0	-	-	-	-

7. Feature Geometry Trees

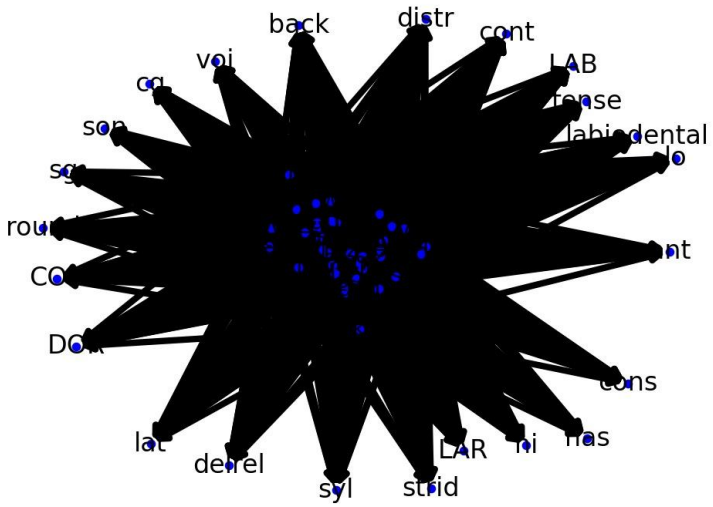
a. **Version 1:**



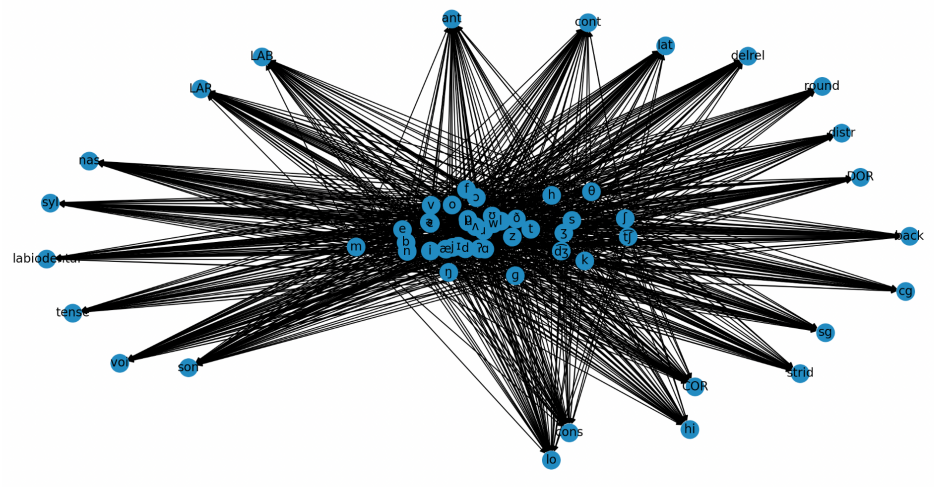
b. **Version 2:**



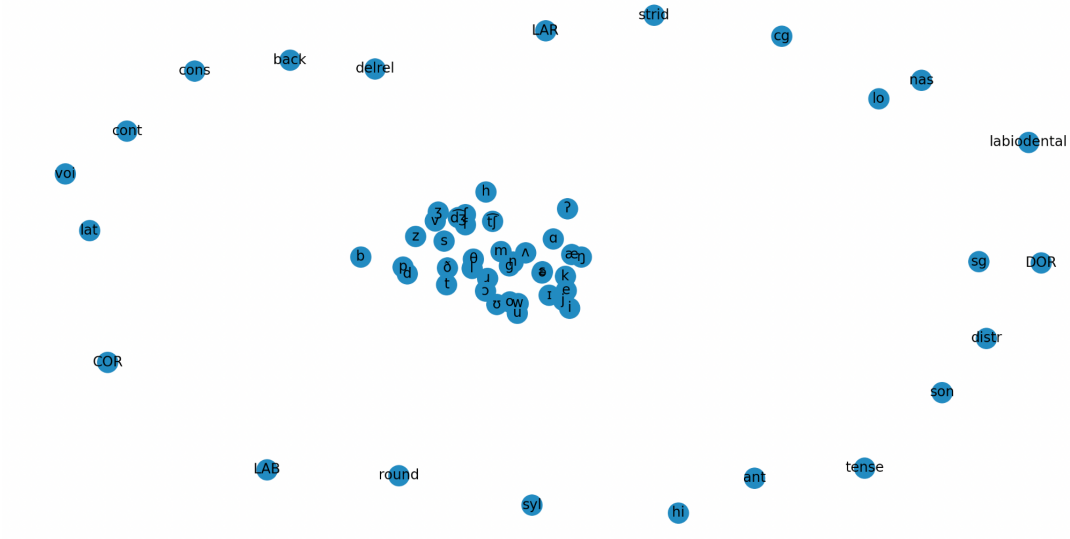
8. Progression of graphs from NetworkX



a.



b.



c.

9. NetworkX Coding Refinements

ntx.py — ~/Desktop/FinalProjPhon

ntx.py

```
import networkx as nx
import matplotlib.pyplot as plt
import json

TRIPLETS_PATH = './triplets.json'

# Create Graph
G = nx.DiGraph()
ipa_nodes = []
feature_nodes = []

# Load the triplets
with open(TRIPLETS_PATH, 'r') as trip:
    tl = json.load(trip)
#print(tl)

options = {
    'node_color': 'blue',
    'node_size': 10,
    'width': 3,
    'arrowstyle': '-|>',
    'arrowsize': 10,
    'with_labels': True,
}

G.add_weighted_edges_from(tl, color='w')
# use nested for loop like e.g. for i in range of the len
#for i in range(len(tl)):
#    if i == "0.0":
#        G.remove_edges(i)

nx.draw(G, arrows=True, with_labels=True, edge_color="w")
plt.show()
```

9.1 norm_vecs.json - intermediate step to create the graphs in NetworkX

[[0.0, 0.6666666666666666, 0.0, 0.0, 0.0, 0.0, 0.6666666666666666, 0.0, 0.0, 0.3333333333333333, 0.0, 0.0, 0.3333333333333333, 0.0, 0.0, 0.3333333333333333, 0.0, 0.0, 1.0, 0.6666666666666666, 0.0, 0.0, 0.0],
[0.0, 0.6666666666666666, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.3333333333333333, 0.0, 0.0, 0.3333333333333333, 0.0, 0.0, 0.3333333333333333, 0.0, 0.0, 1.0, 0.6666666666666666, 0.0, 0.0, 0.0], [0.0, 0.6666666666666666, 0.0, 0.0, 0.0, 0.0, 0.6666666666666666, 0.0, 0.0, 0.3333333333333333, 0.0, 0.0, 0.3333333333333333, 0.0, 0.0, 1.0, 0.0, 0.0, 0.3333333333333333, 0.0, 0.0, 0.0, 0.0],
[0.0, 0.6666666666666666, 0.0, 0.0, 0.0, 0.0, 0.6666666666666666, 0.0, 0.0, 0.3333333333333333, 0.0, 0.0, 1.0, 0.6666666666666666, 0.0, 0.3333333333333333, 0.0, 0.0, 0.3333333333333333, 0.0, 0.0, 0.0, 0.0],
[0.0, 0.6666666666666666, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.3333333333333333, 0.0, 0.0, 0.3333333333333333, 0.0, 0.0, 1.0, 0.0, 0.0, 0.3333333333333333, 0.0, 0.0, 0.0, 0.0],
[0.0, 0.6666666666666666, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.3333333333333333, 0.0, 0.0, 1.0, 0.6666666666666666, 0.0, 0.3333333333333333, 0.0, 0.0, 0.3333333333333333, 0.0, 0.0, 0.0, 0.0],
[0.0, 0.6666666666666666, 0.0, 0.0, 0.6666666666666666, 0.0, 0.6666666666666666, 0.0, 0.6666666666666666, 0.3333333333333333, 0.0, 0.0, 1.0, 0.0, 0.6666666666666666, 0.3333333333333333, 0.0, 0.0, 0.3333333333333333, 0.0, 0.0, 0.0, 0.0],
[0.0, 0.6666666666666666, 0.0, 0.0, 0.6666666666666666, 0.0, 0.0, 0.0, 0.6666666666666666, 0.3333333333333333, 0.0, 0.0, 1.0, 0.0, 0.6666666666666666, 0.3333333333333333, 0.0, 0.0, 0.3333333333333333, 0.0, 0.0, 0.0, 0.0],
[0.0, 0.6666666666666666, 0.0, 0.6666666666666666, 0.6666666666666666, 0.0, 0.0, 0.0, 0.6666666666666666, 0.3333333333333333, 0.0, 0.0, 1.0, 0.0, 0.6666666666666666, 0.3333333333333333, 0.0, 0.0, 0.3333333333333333, 0.0, 0.0, 0.0, 0.0],
[0.0, 0.6666666666666666, 0.0, 0.6666666666666666, 0.6666666666666666, 0.0, 0.6666666666666666, 0.0, 0.6666666666666666, 0.3333333333333333, 0.0, 0.0, 1.0, 0.0, 0.6666666666666666, 0.3333333333333333, 0.0, 0.0, 0.3333333333333333, 0.0, 0.0, 0.0, 0.0],
…]

9.2 triplets.json[3] - intermediate step to create the graphs in NetworkX

[["\u0261", "syl", 0.0], ["\u0261", "cons", 0.6666666666666666], ["\u0261", "son", 0.0], ["\u0261", "cont", 0.0], ["\u0261", "delrel", 0.0], ["\u0261", "nas", 0.0], ["\u0261", "voi", 0.6666666666666666], ["\u0261", "lat", 0.0], ["\u0261", "strid", 0.0], ["\u0261", "LAR", 0.3333333333333333], ["\u0261", "sg", 0.0], ["\u0261", "cg", 0.0], ["\u0261", "COR", 0.3333333333333333], ["\u0261", "ant", 0.0], ["\u0261", "distr", 0.0], ["\u0261", "LAB", 0.3333333333333333], ["\u0261", "round", 0.0], ["\u0261", "labiodental", 0.0], ["\u0261", "DOR", 1.0], ["\u0261", "hi", 0.6666666666666666], ["\u0261", "lo", 0.0], ["\u0261", "back", 0.0], ["\u0261", "tense", 0.0], ["k", "syl", 0.0], ["k", "cons", 0.6666666666666666], ["k", "son", 0.0], ["k", "cont", 0.0], ["k", "delrel", 0.0], ["k", "nas", 0.0], ["k", "voi", 0.0], ["k", "lat", 0.0],

[3] Some characters are shown in ASCII format due to the way the computer interpreted these non-standard symbols. There was no need to convert IPA symbols into X-SAMPA because JSON has an ASCII-ensured feature that matches the input and output format of the characters it accepts.

["k", "strid", 0.0], ["k", "LAR", 0.3333333333333333], ["k", "sg", 0.0], ["k", "cg", 0.0], ["k", "COR", 0.3333333333333333], ["k", "ant", 0.0], ["k", "distr", 0.0], ["k", "LAB", 0.3333333333333333], ["k", "round", 0.0], ["k", "labiodental", 0.0], ["k", "DOR", 1.0], ["k", "hi", 0.6666666666666666], ["k", "lo", 0.0], ["k", "back", 0.0], ["k", "tense", 0.0], ["b", "syl", 0.0], ["b", "cons", 0.6666666666666666], ["b", "son", 0.0], ["b", "cont", 0.0], ["b", "delrel", 0.0], ["b", "nas", 0.0], ["b", "voi", 0.6666666666666666], ["b", "lat", 0.0], ["b", "strid", 0.0], ["b", "LAR", 0.3333333333333333], ["b", "sg", 0.0], ["b", "cg", 0.0], ["b", "COR", 0.3333333333333333], ["b", "ant", 0.0], ["b", "distr", 0.0], ["b", "LAB", 1.0], …]