Web Tool for Phonemes Week-5

University at Buffalo The State University of New York

Agenda

- Word Data Script
- Phoneme Data Script - G2P
- Category Data Script
- Semantic Data Script
- Articulation Data
- Integration of all data

Data Format


```
{
    "word": "abandon",
    "source": { "word": "websterdictionary", "category": "conceptnet", "POS": "wordnet" },
    "category": [ "unrestraint" ],
    "POS": [ "VERB", "NOUN" ],
    "phoneme": [ "AH", "B", "AE", "N", "D", "AH", "N" ],
    "POA": [ "", "bilabial", "", "dental;alveolar", "dental;alveolar", "", "dental;alveolar" ],
    "MOA": [ "", "plosive", "", "nasal", "plosive", "", "nasal" ],
    "VOA": [ "", "voiced", "", "voiced", "voiced", "", "voiced" ]
}
```



```
{
  "word": "abandon",
  "source": {
    "word": "websterdictionary",
    "category": "conceptnet",
    "POS": "wordnet"
  },
  "category": [
    "unrestraint"
  ],
  "POS": [
    "VERB",
    "NOUN"
  ],
  "phoneme": [
    "AH",
    "B",
    "AE",
    "N",
    "D",
    "AH",
    "N"
  ],
  "POA": [
    "",
    "bilabial",
    "",
    "dental;alveolar",
    "dental;alveolar",
    "",
    "dental;alveolar"
  ],
  "MOA": [
    "",
    "plosive",
    "",
    "nasal",
    "plosive",
    "",
    "nasal"
  ],
  "VOA": [
    "",
    "voiced",
    "",
    "voiced",
    "voiced",
    "",
    "voiced"
  ]
},
```

Word Data Scraping

- Scraped around 225k+ words from the following sources :

1. Cmudict
2. Wordnet
3. ConceptNet
4. Webster English Dictionary

https://raw.githubusercontent.com/matthewreagan/WebstersEnglishDictionary/master/WebstersEnglishDictionary.txt

Phoneme Data

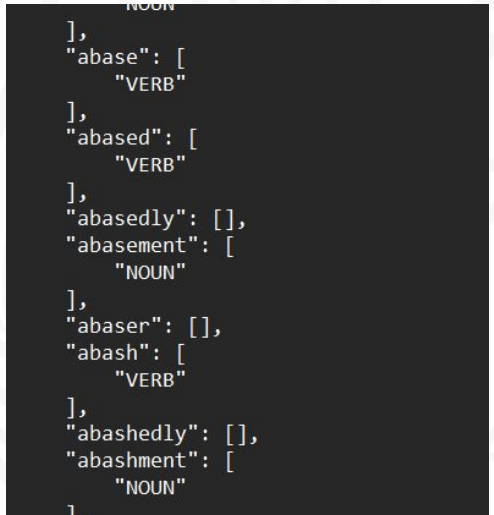
- G2P model to generate the Phonemes for the words

Category Data

- Wordnet
- ConceptNet

Semantics Data Script

- Spacy : Only context based
- wordnet : can list all possible parts of speech
- First, Looking up Synsets for a word in WordNet
- If POS not found
- Generate from Spacy, has support vectors



Articulation Data - Meeting with Alison

IPA Chart for American English Consonants

Manner of Articulation	Place of Articulation																	
	Bilabial		Labiodental		Interdental		Alveolar		Postalveolar /Palatal		Velar		Glottal	
Stops	p	b					t	d			k	g	ʔ	
Tap								ɾ						
Fricative			f	v	θ	ð	s	z	ʃ	ʒ			h	
Affricate									tʃ	dʒ				
Nasal		m						n				ŋ		
Glide		w								j				
Lateral Approximate								l						
Retroflex Approximate								ɹ						

NOTE: Shaded sounds are voiced, unshaded sounds are unvoiced.

Articulation Data

Vowel Chart

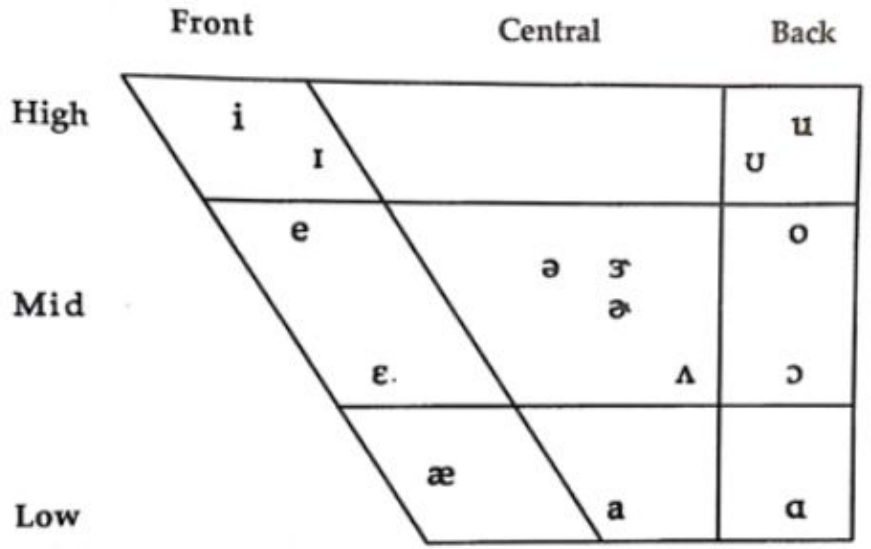


FIGURE 4.1 The vowel quadrilateral for American English vowels.

Minimal And Maximal Pairs

Phonetic Length Should be same.

Phonetic Transcription

Place of Articulation **Manner of Articulation** **Voicing**

For the phonemes differing, If 1/3 are different then they are Minimal Pairs.

For the phonemes differing, If 3/3 are different then they are Maximal Pairs.

Minimal Pairs and Maximal Pairs Script

Integration of all data

Problems Faced

- Search an Alternative for VM for computing
- Still require VM for Minimal and Maximal Pairs
-

Questions

1. How do you want Minimal And Maximal pairs to be used? Queries or anything else?
2. Categories -