

How to Make a Word List for a Linguistics 103 Term Project

1. Length

The **length limit** for your word is as follows: number of segmental phonemes, plus number of stress and tone contrasts, plus 25%. Examples: American English (in the IPA Handbook analysis) has 39 segmental phonemes and phonemic stress; so $(39 + 1) \times 1.25 = 50$ words. Cantonese (in the IPA Handbook analysis) has 42 segmental phonemes and 9 tones, so $(42 + 9) \times 1.25 = 64$ words.

2. Finding the words

You need to comb through your reference source and find words that fit the following pattern:

2.1 Minimal or near-minimal sets

Make a phonetic chart of all the consonants in the usual layout, something like (for an imaginary language):

		<i>Labial</i>	<i>Alveolar</i>	<i>Palato-Alveolar</i>	<i>Palatal</i>	<i>Velar</i>
<i>Stops</i>	<i>voiceless</i>	p	t			k
	<i>voiced</i>	b	d			g
<i>Affricates</i>	<i>voiceless</i>			$\overline{tʃ}$		
	<i>voiced</i>			$\overline{dʒ}$		
<i>Fricatives</i>	<i>voiceless</i>	f	s			
<i>Tap</i>			r			
<i>Approximants</i>	<i>lateral</i>		l			
	<i>central</i>	w			j	

Then find a minimal or near minimal set to illustrate all the consonants. Since the vowel /a/ (or something like it) is often common in a language, and tends not to produce special allophones, a good arrangement for your minimal set is consonants occurring initially before /a/. Here is an example (my pretend language has tones, marked in IPA with acute accents for high, grave accents for low).

1. [pámà] 'goat'
2. [báfà] 'sincerity'
3. [tánà] 'tiptoe'
4. [dámà] 'tooth'
5. [$\overline{tʃ}$ árà] 'moon'

6. [dʒámà] ‘dwelling’
7. [kámé] ‘sing in falsetto’
8. [gámà] ‘friend (slang word)’
9. [fámà] ‘transmigration’
10. [sámà] ‘mother’s younger brother’
11. [márà] ‘color name (blue/green region of spectrum)’
12. [námà] ‘nine’
13. [rámà] ‘urine’
14. [jánà] ‘call (as to an animal)’
15. [wámà] ‘hope’

If a consonant doesn’t occur initially, then make a set that compares with other consonants finally, like English *sim* [ˈsɪm], *sin* [ˈsɪn], *sing* [ˈsɪŋ].

Then make a chart for vowels, in the usual layout.

	<i>Front</i>	<i>Central</i>	<i>Back</i>
	<i>unrounded</i>	<i>unrounded</i>	<i>rounded</i>
<i>High</i>	i		u
<i>Mid</i>	e		o
<i>Low</i>		a	

Try to find a minimal or near-minimal set for vowels. Vowels are probably a bit harder to hear before nasals, so that might be worth avoiding. The following set uses [l] as the following sound.

16. [pìlì] ‘crust of rice left in bottom of pot after cooking’
17. [pèlì] ‘perch’
18. [pálì] ‘opium pipe’
19. [bólì] ‘friend’
20. [púlì] ‘bestow (used only in poetry)’

Then, if your language is a tone language, find a minimal set for tones. Since my hypothetical language has just high tone and low tone, this is pretty easy:

21. [má] ‘mother’
22. [mà] ‘horse’

Note that although this minimal pair is the main example for tone, I have transcribed tone on all the other examples. This is crucial, since tone is phonemic in this language. Please don’t forget to do this. (You can use the keyword method and your consultant to help get the tones right.)

If your language is a stress language, with phonemic stress, find minimal pairs or triplets. For English, a minimal pair is (more or less) *defer* [dɪˈfɜː] vs. *differ* [ˈdɪfəː].

2.2 Selected allophones

Use your reference source, or failing that, look at <http://www.linguistics.ucla.edu/people/hayes/103/FindingAllophones.pdf> and find them yourself. If your source has many, many allophones, then pick the ones you find most interesting.

To illustrate an allophone, include a word that has the phoneme with the environment for the allophone, and a very similar word that doesn't have the environment.

Example: nasalized [ɛ̃] is an allophone of /ɛ/ in English, so:

	<i>Phonemic transcription</i>	<i>Phonetic transcription</i>
<i>ten</i>	/tɛn/	[tɛ̃n]
<i>Ted</i>	/tɛd/	[tɛd]

2.3 Backups

It's completely common for a linguist to come to the speaker with a set of possible words and find that they're not all useable. Responses:

- "I don't know that word."
- "They use that word in the eastern province of my country, but I wouldn't ever say it myself."
- "That word appears in literature but I don't know how to pronounce it."
- "That word is embarrassing."

You can either have backup material on hand (like a dictionary) to fill the gaps that result, or you can come with backup words (for example, three different words that start with [pá...] for /p/. This depends on how much time you have, and how much time your consultant has to share with you.

Don't forget to jolly your consultant along. You want them to show up at recording time...

2.4 Sentence

The last item on your script will be a sentence. Get your consultant to make up a sentence that includes two or three of the words in the script. The purpose of this is that when you do your transcriptions, you can compare isolation pronunciations with in-context pronunciations. Note that often in order to get a real sentence your speaker will have to include grammatical suffixes or prefixes, as shown below.

23. pá-mà nì má tú pì-lì-sé tán-à-sì
 goat of Mother through crust of rice-locative tiptoe-past
 'Mother's goat tiptoed through the crust of rice.'