

Homework: Japanese Phonotactics

due in class Tues. 2/15

1. Assignment

This problem is meant to give you practice in working with corpus data, an increasingly common approach in phonological research.

Download from the course web site the file **JapaneseWords.txt**. This consists of 20,000 Japanese words from a electronic dictionary, in an ad hoc phonetic transcription. An attempt was made to remove possible compounds, but this could only be done in a very rough way. The original source of the data was the EDICT project, at http://www.csse.monash.edu.au/~jwb/edict_doc.html

Examine this corpus and write an essay describing the phonotactic regularities of Japanese. Your essay should be on the scale of the previous homeworks; if it looks like it's going to be longer or take more time, then figure out a way to limit your essay to a size that is feasible. A good strategy is to pick just a few topics and cover them in more detail.

Include both inviolable and gradient phonotactics in your description, and in the latter case give (briefly) representative statistics illustrating the rarity of the semi-banned sequences. See Hints section below for examples.

1.1. Software option

An option you can try—if you like—is to run the phonotactic learning software posted at <http://www.linguistics.ucla.edu/people/hayes/Phonotactics/index.htm><http://www.linguistics.ucla.edu/people/hayes/Phonotactics/>. If you want to try this, let me know and I will give you a feature system. Report clearly how you proceeded and what the software found.

2. Symbol chart

The phonetic symbols are meant to avoid font problems and are hopefully pretty intuitive. Their IPA equivalents (in the non-obvious cases) are as follows:

Vowels

- A [a], low central unrounded vowel
U [u] or [ɯ], depending on speaker. You can assume [u] if you like.
AA, EE, II, OO, UU [a:, e:, i:, o:, u:]

Consonants

BB, DD, etc.	geminate (long consonants). You can treat them as units if you like.
CH	[tʃ] (affricate)
CCH	[tʃʃ] (geminate affricate; as is normal, there's just one affricated release.)
J	[dʒ] (affricate)
JJH	[ddʒ] (geminate affricate)
TS	[ts] (affricate)
TTS	[tts] (geminate affricate)
F	[ɸ], i.e. bilabial not labiodental as in English.
NG	[ŋ]
R	[ɾ]. Some speakers have various other liquids, often including lateral flaps.
SH	[ʃ]
SSH	[ʃʃ]
Y	[j] in IPA; the symbol matches English spelling.

3. Hints

- I recommend that for solving this problem you also download the following Excel spreadsheet: **JapaneseNGrams.xls**. This was created by chopping the words into “bigrams” (sequences of two sounds) and “trigrams” (sequences of three sounds) and counting up how many of each type there are. Navigate the spreadsheet using the tabs at the bottom, and sort on the various columns to find what you need.
 - Note: in the spreadsheet, “O/E” means “observed divided by expected”. This is a casual statistic sometimes used to give perspective to claims that a sequence is rare. Example: suppose 10% of all sounds are /t/ and 10% are /a/. Suppose 2% of all bigrams are /ta/. Then O/E for /ta/ is 2, namely $.02 / (.1 * .1)$. In other words, /ta/ is twice as common as you'd expect it to be given the frequencies of /t/ and /a/.
- Here are some sample observations you might make:
 - “I propose that the syllable template is [(C)(j)V(:)C], where C can only be the first half of a geminate, a word final [n], or a nasal homorganic to a following consonant.”
 - “Geminates are always medial, usually intervocalic, but occasionally they occur / V __ j V.
 - “With a single exception, /w/ may only precede the non-low back vowels /a, a:, o, o:/. The exception is a single instance of /we/.”
 - “Palato-alveolars tend to avoid a following /e/. Here are the only instances:

CH	E	25
CH	EE	4
SH	EE	4
SH	E	3
J	EE	1
SSH	E	1

- You don't have to read anything for this homework, but if you really want to I can recommend two fine reference sources that cover Japanese phonotactics.
 - Junko Itô and Armin Mester (1994) Japanese phonology: constraint domains and structure preservation. In Goldsmith, J. (ed.), *The Handbook of Phonological Theory*, Blackwell, 817-838. On line at http://people.ucsc.edu/~mester/papers/1995_ito_mester_japanese_phonology.pdf
 - Vance, Timothy J. (1987) *An introduction to Japanese phonology*. Albany: State University of New York Press.