Mystery Language Project Guidelines Ling 450/550, Wright/Freeman

You will complete an individual project on a language that you do not know. The process includes several steps that will introduce you to resources used by phoneticians and help you hone your transcription skills. See your course calendar for dates corresponding to the steps outlined below.

450 students: Follow all steps except those marked as "550."

550 students: For the graduate-level credit course, you will complete additional steps involving the comparison of your mystery language to another language you know well. There are three versions of the project – if you are in CLMS or MATESOL, be sure to follow the steps marked for your program; all other 550 students should follow the 550 LING steps.

Preparation (Catalyst WebQ)

1) Turn in the following, via the Catalyst WebQ on the course webpage:
   a) A list of languages that you know, along with short descriptions of how well you know them (your proficiency, how you learned and use them, etc.).
   b) Describe any preferences for what languages, language families, geographic areas, or linguistic features you would like to work with in your project. You will not be assigned a language that you know well. There are no guarantees that your preferences can be accommodated, so it's best to note more than one to increase your chances of getting something similar.

Part A: Transcriptions, Charts, and Descriptions (Turn in to Drop Box)

2) **Sound files.** You will be given sound files for an unlabeled language that you do not know.
   - Each sound file is one word said by a native speaker, and the file name is the English gloss (translation) for the meaning of that word.
   - Most collections are organized into folders labeled *consonants, vowels,* and *tone* when applicable. Some may have separate folders for diphthongs or other relevant distinctions, but the absence of a grouping does not necessarily imply the absence of a distinction (e.g., you may find diphthongs in your vowels folder). In your consonants folder, you will have enough evidence to identify all the consonant distinctions in your language (all phonemes and possibly some allophones, although you may not have enough evidence to determine whether a sound is an allophone). In your vowels folder, you will have at least one instance of each vowel phoneme, etc.
   - The number of files does not necessarily correspond directly to the number of phonemes in your language; i.e., not every sound file has a unique sound, but it is possible that you will only have one example of a given sound.
3) **Transcriptions.** Transcribe your sound files in a list/table format. You may want to begin doing this by hand, but the draft you turn in must be typed using correct IPA symbols.
   - It is highly recommended that you use high-quality headphones, preferably circum-aural (enclosing your ears), to listen to your files in a quiet environment while you transcribe. Computer speakers often distort sounds, and earbuds have low quality output. The Language Learning Center in the basement of Denny Hall has good headphones, and other computer labs on campus may also.

a) Your transcriptions should appear in neat tables with column headings and use Unicode fonts (e.g., Charis SIL or Doulos SIL) with standard IPA symbols and conventions. (If you use a word processor, make a table or use uniform tab stops throughout to create the appearance of columns).

b) Column headings should include, in order: word number, word gloss, your transcription. (You may add others, as appropriate; e.g., if you'd like to add second options or notes to unconfident transcriptions.)

c) You should have sections with headings for each folder (consonants, vowels, tones, etc.).

d) You have two options for how to order your sound files within each table section/folder:

   i) Present your transcriptions in the order the words appear in their folders – if the files are numbered, order your transcriptions numerically; if they are not, order them alphabetically.

   ii) Or, you may rearrange the words in each section to more clearly show phonetic distinctions. For example, you may want to put minimal pairs together or put words that illustrate each consonant sound in the order they appear in your chart. If your files are numbered, include the number in the first column of your table, even though the numbers will be out of order. (This makes it easier to find each sound file later.)

4) **Phonetic inventory charts.** From your transcriptions, build phonetic inventory charts for your language. Every symbol used in your transcriptions should appear in your charts and vice versa. Your final drafts should be clear (not blurry or cramped) and typed using Unicode IPA fonts and conventions, arranged in neat tables with row and column headings for articulatory descriptions. You should not have any empty rows/columns, and the font style and size should be uniform and of a similar size as your prose. Add footnotes to indicate how you distinguish pairs, e.g. voiced/voiceless and rounded/unrounded.

   a) **Consonants.** Follow the main IPA consonant chart, placing in the appropriate locations only those sounds you observe in your sound files. You may add rows or columns as appropriate to fully describe your inventory, or you may add more than two symbols to a cell. Examples: Add rows for affricates, implosives, palatalized consonants; add columns for labialized consonants after corresponding place columns; or just add the appropriate symbols to existing cells.

   b) **Vowels.** Follow the IPA vowels chart, placing the vowels you transcribe in locations appropriate to their articulatory descriptions. (Tip: Your chart doesn't have to be slanted for the front vowels – it can be a simple rectangular table.) Note: You only need to include tense/lax in descriptions if there are pairs that are distinguished only by this, or if you can see a pattern where tense vowels appear in different environments than lax.

      i) **Non-quality distinctions.** If your language has phonemic distinctions other than quality (length, nasalization, creakiness, breathiness, laryngealization, etc.), decide whether to create separate charts for each set. If all vowel qualities participate in the distinction, you may simply state this in words below the chart. Otherwise, it often works to add appropriate symbols next to the plain vowels in the same chart. For example, inventories with length distinctions often place long vowels next to their short counterparts. If only a subset of vowels is affected, it may be clearer to show them in a separate chart.

ii) **Diphthongs** (and triphthongs, if applicable). If your language has a small vowel inventory, you may put diphthongs on the same chart as monophthongs. If this would be crowded, create a separate chart. If you do not have a separate diphthongs folder, attempt to identify any diphthongs on your own.

c) **Tone, pitch accent, phonation, etc.** If your language has lexical tone or phonemic phonation distinctions (creaky, breathy, etc.), or if you have evidence of grammatical tone or pitch accent, include a chart with example words to illustrate the distinction. (Ask your instructor for examples if necessary.) Note that we may not cover phonation and/or tone in class before your first drafts are due. If you think you have one of these distinctions, you may want to look ahead at the corresponding readings and lecture slides (Phonation: ch. 6 beginning around p. 148 and around slide 161, Tone: ch. 10 beginning around p. 254 and around slide 200).

5) Below each set of charts (consonants, vowels, tones, etc.), write prose **descriptions** of all the sounds identified (1-2 paragraphs per chart, or more as necessary). You must use complete articulatory descriptions and list all symbols used in your charts, but sounds should be described in groups to reduce repetition. Examples: "The language has voiceless and voiced stops at three places of articulation: bilabial [p, b], alveolar [t, d], and velar [k, g]"; or "The language has bilabial, alveolar, and velar stops, both voiceless [p, t, k] and voiced [b, d, g]."

- While finishing your first attempt at your transcriptions, charts, and descriptions, you may want to meet with other students (in person or virtually through chat, email, or discussion board). Ask each other for help identifying sounds and transcribing difficult words, and look over each other's charts and descriptions for internal consistency.

6) **Turn in** your transcriptions, charts, and descriptions (steps 3-5) in one .doc or .pdf to the drop box. The instructor *may* comment (time permitting) on the completeness, consistency, style, formatting, etc. of your work. Take note of any suggestions, so you can address the feedback in your final report. Note that your writing style should be appropriate for an academic paper – not conversational – but it need not be extremely formal (e.g., you may use first person, active voice, contractions, etc., but you should also use complete sentences, correct spelling, clear organization, varied sentence structure, and avoid slang).

Part B: Comparisons and Research

7) You will receive an **article** with the phonetic description of your language that corresponds to your sound files. Read the article carefully and take notes or highlight information you'll use in your report (see below).

a) 550 CLMS and MATESOL options: You will also receive one or more articles on English. You will use this for comparison to your mystery language's phonemic inventory and phonological descriptions in order to make predictions about non-native English speakers' pronunciation (MATESOL) or errors made by recognizers trained on English (CLMS) (see below).

b) 550 LING option: You will compare your mystery language to another language that you have some experience with (other than English). To do so, you need to find linguistically sophisticated phonetic/phonological descriptions of the language on your own – see the class webpage for instructions and tips. You will choose the language, but it needs to be approved by the instructor/TA. A sign-up system will be created via Catalyst (details TBD), and you may be asked to reference or post information you've found for your top choice(s), so you should begin searching before making your request.

8) **Compare** your charts, transcriptions, and descriptions to those in the article. Where there are discrepancies, try to determine how significant they are. Did you miss a distinction entirely, or were your descriptions close? Listen to the words again to see if you can hear the distinctions in the article's transcriptions. (This will be written up in 11d.)

   a) 550 CLMS option: You will also compare your mystery language article to the article(s) on English. Take notes on differences in the phoneme inventories and any other phonological aspects. Use these to make predictions about mistakes that a speech recognizer trained on English will make when analyzing your mystery language.

      i) Consider mismatches in phonemes. Predict which English sounds the recognizer will choose for each sound in your mystery language. (You can briefly mention which sounds are the same in both languages and therefore predictably similar for the recognizer.) If a sound from the mystery language doesn't exist in English, what sound(s) might the recognizer choose instead? What if an English sound doesn't exist in the mystery language? Will the recognizer try to apply it to other sounds?

      ii) Consider mismatches in allophones. Are some sounds allophones of different phonemes between the two languages? Predict how the recognizer will assign these sounds to English phonemes.

      iii) Consider other phonetic aspects, such as VOT, nasality, airstream mechanisms, voice quality, laryngeal setting, tone, pitch accent, etc. How will differences in the contrastiveness of these aspects affect the recognizer output?

      iv) If possible, consider other phonological aspects, such as phonotactics, syllable structure, stress placement, intonation, etc.

   b) 550 MATESOL option: You will also compare your mystery language article to the article(s) on English. Take notes on differences in the phoneme inventories and any other phonological aspects. Use these to make predictions about pronunciation errors that a native speaker of your mystery language might make when speaking English.

      i) Consider mismatches in phonemes. If an English sound doesn't exist in the mystery language, what sound(s) might they use instead? Will this cause confusion for native English listeners? What if a sound from the mystery language doesn't exist in English?

      ii) Consider mismatches in allophones. Are some sounds allophones of different phonemes between the two languages? (Example: the alveolar tap is its own phoneme and sometimes an allophone of the trill in Spanish, but it is an allophone of [t] or [d] in English. What misuses of tap might a Spanish speaker use, and how will it confuse English listeners?)

      iii) Consider other phonetic aspects, such as VOT, nasality, airstream mechanisms, voice quality, laryngeal setting, tone, pitch accent, etc. How will differences in the contrastiveness of these aspects affect non-native pronunciation? What confusions might arise for listeners?

      iv) Consider other phonological aspects, such as phonotactics, syllable structure, stress placement, intonation, etc.

c) 550 LING option: You will also compare your mystery language description to the information you find on your chosen language (from 7b). Take notes on differences in the phoneme inventories and any other phonological aspects.

   i) Consider differences in phoneme inventories. First consider categorical differences. (Example: Language X has both voiced and voiceless stops, while Language Y has only voiceless.) If the inventories are similar, look at the phonemes on a more individual level.

   ii) Consider mismatches in allophones. Are some sounds allophones of different phonemes between the two languages? (Example: the alveolar tap is its own phoneme and sometimes an allophone of the trill in Spanish, but it is an allophone of [t] or [d] in English.)

   iii) Consider other phonetic aspects, such as VOT, nasality, airstream mechanisms, voice quality, laryngeal setting, tone, pitch accent, etc. Describe differences in the contrastiveness of these aspects.

   iv) Consider other phonological aspects, such as phonotactics, syllable structure, stress placement, intonation, etc.

9) Do additional **research** on your language. Some articles include some of this information, but most only give an overview. Use Ethnologue.com and other sources recommended by your instructor to find the following:

   a) Your language's three-letter ISO 639-3 code.

   b) Its genetic classification (language family, genus, and a step or two below or a lower classification that is well known).

   c) Genetic cousins (languages most closely related and/or well-known), plus comments about dialects. If your language is an isolate, try to find comments about any hypotheses about its origins, relatives, or previous classifications.

   d) The geographic area(s) where it is spoken, plus social details about its speakers, if applicable (e.g., a particular ethnicity or social class).

   e) The populations of native speakers and second-language users, plus the spheres in which it is used, if applicable/possible (e.g., used in trade, government, at home only, etc.). Note if it is endangered and any comments about efforts to revive it.

   f) 550 CLMS and MATESOL options: You may wish to do similar research for English, but it is not required.

   g) 550 LING option: Do the same research for your chosen language. If your mystery and chosen languages have any similarities, describe their relationship.

10) Keep track of your sources and compile a **bibliography in JIPA or APA format** (see the class webpage for examples and links, and run yours by the instructor if you're unsure).

   - Important: Wikipedia is not an academic source. It is not original research and should not be cited as such – it is more like a group book report. However, it can be a useful tool for finding original sources. If you find information on Wikipedia, you must verify it through the original source linked in the entry's bibliography – and then cite that source, not Wikipedia.

Part C: Final Report (Turn in to Drop Box)

11) Type and turn in your **final report** (with subject headings, 11-12 point font, 1-inch margins, double- or 1.5-spaced, as one PDF or DOC turned in to the Drop Box). Include the following, roughly in this order (or with slight modifications, as appropriate or as approved ahead of time):

    a) An **introduction** to your mystery language including the information you found in your outside research. This must be in prose (1-2 paragraphs, not a list or chart) and use in-line citations in JIPA or APA format.

        i) 550 CLMS and MATESOL options: Also include a brief introduction to your comparison of your mystery language to English and the predictions you will make.

        ii) 550 LING option: Also include your research on your chosen language and an introduction on your comparison of the two.

    b) Revised work from Part A: Your phonetic inventory **charts** (revised to address feedback from Part A), each followed by revised **descriptions**. These should be revised for completeness and accuracy in describing *only your charts*, and to address any feedback from your draft.

    c) Follow each chart + description with the **corresponding charts** from your mystery language article. You may recreate the article's charts using the same formatting as your own (but using the article's conventions); or you may copy and paste the images, but only if the image quality is high, and you must credit the source in a caption under the image (e.g., "Chart reprinted from Smith (2002)"). If your charts are very similar to those in the article, you may highlight these differences on your charts or describe them without including the article charts. For example, if you missed a symbol, you can add it to your chart using a different color or by circling it, and then explain the difference in prose below the charts. If you chose a different label for a description (e.g., place of articulation column heading, laryngealization vs. pharyngealization), you may indicate the heading on the chart (with a circle, arrow, different font color, etc.) and then explain the difference in prose.

        - 550 options: You do not need to include the charts on English or your chosen language.

    d) Summary and **discussion of differences** between your transcriptions and charts and those of the mystery language article. Some bulleted lists or charts may be appropriate to illustrate differences, but you must also discuss them in prose (1-2 pages). You should cover all large divisions (consonant place/manner distinctions, vowel features, etc.), even if your transcriptions were fairly accurate.

        i) Note categorical differences and similarities as well as smaller details (e.g., "I was able to distinguish the front round vowels, but I did not identify both central vowels, and I only noticed nasality in a few words… I described the set of consonants [t, d, l] as alveolar, but the authors describe them as dental... I missed the voiceless glottal fricative [h] at the ends of words, so I thought that [pah] and [pa] must have had a difference in vowel quality."). Also describe differences in conventions or notational systems (e.g., if you used tone numbers and they used lines, provide a table or description of how these systems correspond).

        ii) Especially discuss how the article treated the sounds that were difficult for you. How close were you? Does it use one of the options you considered?

- Note that you may choose to organize sections (b-c-d) by the folders that divided your sound files; i.e., you may write a section (b, c, d) for consonants, then a (b, c, d) for vowels, then tones, etc.

  iii) 550 CLMS and MATESOL options: **Discussion of differences** between the phonemic inventories and phonological descriptions of your mystery language and English, and **predictions** of how they might affect recognizers trained on English or non-native English pronunciation. Respond to the items in (8a) or (8b) in about 1-3 pages.

  iv) 550 LING option: **Discussion of differences** between the phonemic inventories and phonological descriptions of your mystery and chosen languages. Respond to the items in (8c) in about 1-2 pages.

e) **Conclusion**. Briefly (about a paragraph) summarize your paper. You may also add a general comment evaluating the process (what you liked/didn't, what was hardest, most enjoyable, etc.). If you worked with a classmate at all, name him/her and briefly describe how you worked together and whether it was helpful. (Did you help each other identify difficult sounds? Did you check each other's writing? If you disagreed on a transcription or symbol, did you resolve this or 'agree to disagree'? Was one of you closer to correct? etc.) You may also add constructive feedback that could be used to maintain or improve the project process for future classes (including things you think should not be changed).

f) Your original **transcriptions** with an additional column for corrections. If your transcription of a word is identical to that in the article, do not enter a correction for that word. If all of your transcriptions agree but with a consistent difference of a convention (e.g., you always used schwa whenever the article used barred-i), just describe the difference in prose below your transcriptions. Note that you can use the article's conventions for your corrections (e.g., if you used tone numbers but they used tone lines, you do not need to attempt to translate their line system to your numbers), but you should describe how the systems correspond in your discussion of differences.

g) **References** listed in JIPA or APA format. See the links and examples on the class webpage. All sources listed must be referenced in an in-line citation in your prose, and all referenced in your prose must be in your reference list.

Tips and things to look out for

Transcriptions

- Many students find it helpful to handwrite their first transcription drafts. It is faster than typing using correct IPA symbols and easier to keep track of changes and possibilities. When you go to type it up, you may find errors, patterns, or inconsistencies that you were unable to see before.
- Keep in mind that many tonal distinctions also co-occur with changes in voice quality or laryngeal setting, e.g., low dipping and super-low tones often also involve creaky voicing.
- Breathiness and nasalization can often sound similar. To distinguish the two, open your sounds in Praat to look for signs of nasalization (e.g. dampened formants, thick voice bar, etc.).

Charts and prose

- Make sure to use the IPA [g] symbol (not [g]), small-cap [ɪ] (not capital [I]), and that no symbols in your charts are automatically capitalized.

Descriptions

- Descriptors should appear in the same order as corresponding symbols, e.g.: "Language X has voiceless and voiced stops at three places of articulation: bilabial [p, b], alveolar [t, d], and velar [k, g]." (In each pair, the voiceless symbol appears first, so "voiceless" should appear before "voiced.")

Comparisons

- Note that [a] is more often considered central rather than front. So, if your article shows [a] in a central location on the vowel chart, but you used a different central symbol, you need only note the difference in symbol choices. If you used [a] but described it as front, discuss how you thought that sound was front, while they characterize it as central.

General/Report

- In your intro, refer to your language code like this: ISO 639-3 code: xxx. Replace "xxx" with your language's three-letter code, in all lowercase or all caps.
- You can type your whole document in a Unicode font (Charis SIL or Doulos SIL preferred). Using one of these ensures your teacher will be able to read your symbols in a Word doc. You can also convert your work to a PDF. In Word 2007 and beyond, you can "Save as" a PDF. If you use something else, you might try a free online converter such as http://en.pdf24.org/doc2pdf.
- Spelling: liquids and glides are types of "approximant" not "approximate"
- Use uniform line spacing throughout your document (double or 1.5).
- Don't forget in-line citations! See the tips doc on the class webpage for formatting and some ready-made citations for commonly-used sources.