

Mystery Language Project Guidelines (Graduate)

Final Project Guidelines (Grad) For CLMS students For CLMS students, there are two options for the final project. You are encouraged to choose the ASR sub-option. If you have substantial experience or interest in language acquisition, you can choose the human language learning sub-option. In addition to the requirements for the Mystery Language, you will also be investigating a second language and comparing the sound systems of the two languages. You will choose this language yourself from the other languages you know.

Automatic speech recognition (ASR) sub-option (preferred) Predict the errors made by an automatic speech recognizer set to the wrong language based on the differences between the sound systems of the two languages. Pick one direction of recognition (e.g. predict errors in the output obtained from a German ASR engine asked to transcribe Korean audio) - you will not have the space to discuss both directions. You can test your results using an existing ASR model available online, but the test is optional. For example, you can pass your Korean audio to an existing German ASR engine and see what errors were made in the output.

Human language learning sub-option Predict the pattern of errors that result in a foreign accent through transfer of first language phonetics and phonology to a second language. Pick one direction of learning (e.g. a German learner of Korean) - you will not have the space to discuss both directions. Your comparison should center around the mistakes in pronunciation that you predict will occur including predicting substitution of one phone by another. You might (optionally) test your results if you have access to an L2 speaker of interest. ‘How to compare two sound systems’ (ch. 2 of Lado 1957) is available to you on Canvas. While the contrastive analysis hypothesis has fallen out of favor with respect to morphology and syntax, contrastive analysis of sound systems does seem to correctly predict certain aspects of a foreign language accent. (There are limits; Lado’s explanation for why English German speakers transfer /s/ as opposed to /t/ to English // does not seem convincing.) Note, you can consider yourself as the L2 speaker of the mystery language to test your results.

Part A: Transcription and Description for the second language Due 2/19/2024

A) Mystery Language:

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

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

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.) 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) 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 to your prose. Add footnotes to distinguish pairs, e.g. voiced/voiceless and rounded/unrounded. 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. 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. 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. Diphthongs (and triphthongs, if applicable). If your language has a small vowel inventory, you may put diphthongs on the same chart as monophthongs. You can put the diphthong at the location of any vowel involved, or you can draw an arrow from the first vowel to the second vowel (and to the third vowel, for triphthongs), and write the diphthong symbol at either vowel location. 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. Tone, pitch accent, etc. If your language has lexical tone, or if you have evidence of grammatical tone or pitch accent, include a chart with example words to illustrate the distinction. (Ask me (your instructor) for examples if necessary.) Below each set of charts (consonants, vowels, tones, etc.), write brief prose descriptions of all the sounds identified (1-2 paragraphs per chart). 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, ʔ]”; or “The language has bilabial, alveolar, and velar stops, both voiceless [p, t, k] and voiced [b, d, ʔ].” While finishing your first attempt at your transcriptions, charts, and descriptions, you may want to meet with the other student(s) working on languages of the same language family (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. However, you do not need to agree on particular sounds or symbol choices, and your descriptions should not be worded identically.

Turn in your transcriptions, charts, and descriptions (steps 3-5 above) in one PDF. Note that your writing style should be appropriate for an academic paper:

Not conversational It need not be extremely formal (e.g., you may use first person, active voice, contractions, etc.) You should also use complete sentences, correct spelling, clear organization, varied sentence structure, and avoid slang).

B) Second Language

- Submit a phonetic description for your second language of your choice. You need to refer to at least two sources for the phonetic inventory of second language. You cannot copy and paste the table or graph from the sources, but you can recreate them on your own. If there are any differences between the two (or more) sources, you need to describe the differences between them in prose.

Vowel Description Vowel chart(s). This should be a chart of phonemes. Preferably place symbols in a trapezoid. Prepare separate charts for monophthongs and diphthongs (these

can be placed side by side to save space). A word list illustrating each vowel phoneme in the chart or charts. This is preferably a minimal set. A long list should have some structure. Divide words into groups and label the groups as to what they show. For example, there can be separate lists for monophthongs and diphthongs. Consonant Description consonant chart (phonemes). Your chart should not include columns for places or rows for manners of articulation that are not contrastive in your languages. word list illustrating each consonant. Suprasegments (Tone, pitch accent, etc.) If the second language has lexical tone, contrastive phonation, or if you have evidence of grammatical tone or pitch accent, include a chart with example words to illustrate the distinction. (Ask me (your instructor) for examples if necessary.). References at least two references for your vowel and consonant descriptions. The references need to be cited in APA format.

Part B: Mystery Language Reveal; Submission Due 3/14/2024

- 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).

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 Type and turn in your final report. The length limit is 10 pages, double- or 1.5-spaced, including all charts, and figures (but references should be single-spaced, do not count against your page count, and the references section does not need to start on a new page). All text and data set off from the text should be in 11-12-point font (depending on default font size), although footnotes can be in a smaller font. Margins should not be smaller than 1". Should you choose to test your predictions, you can include this as an appendix that does not count towards the page limit – maximum 3 pages, including charts or figures.

Throughout the paper:

For the error predictions, please including which sounds or suprasegments may be substituted for other and why. Where applicable, group sounds together to make broader generalizations Please define any transcriptional symbols which depart from IPA usage, but you should use IPA symbols even if your sources do not Published papers could be "wrong"! If you disagree with the illustration paper or the sources for the second language in their transcription or analysis, please describe your disagreement and explain why your transcription and/or analysis would be more reasonable. You can use additional sources to support your analyses.

The paper needs to include the following:

Brief overview of the two languages:

ISO codes Genetic affiliation and related languages Geography demographics: where spoken; speakers/degree of endangerment A broad summary of the differences between the two

languages Language learning situation: who is studying the learned language as an L2? Description of ASR or learning situation EITHER: ASR situation: does an ASR engine exist on the language of interest? How accurate and widely used is it? OR: Are there many speakers of the L1 language that speak the L2 language? Consonants consonant charts (corrected for mystery language based on feedback and article) consonant descriptions for both languages (revised for mystery language) error predictions, including which sounds may be substituted for which others and why. Where applicable group sounds together to make broader generalizations Vowels vowel charts (corrected for mystery language based on feedback and article) vowel descriptions for both languages (revised for mystery language) error predictions, including which sounds may be substituted for other and why. Where applicable, group sounds together to make broader generalizations

Suprasegments (tone, phonation, pitch accent, stress, length, etc. If you believe these will be significant for ASR performance or L2 accent, include this section). Possible suprasegments that you might find in your languages (only include the sections that are relevant to your languages):

Contrastive Stress or Pitch Accent Short word list illustrating that stress/pitch accent has a contrastive function Description of acoustic or articulatory correlates of stress/pitch accent Error analysis between two languages Predictable or No Stress If the language has no stress, tone, or pitch accent whatsoever, one sentence stating so will do. For a language with predictable stress, provide the rules or a short summary thereof that predict the stress along with words that illustrate them. Extensive discussion of predictable stress is not necessary. Error analysis between two languages Contrastive Lexical Tone Tone chart (phonemic tones) Word list illustrating each tone Description of tone (e.g. whether the tones are level tone or contour tone; if there are tonal sequences, provide phonological justification for any contour tones which could be interpreted as sequences or unit phonemes) define any transcriptional symbols which depart from IPA usage, but you should use IPA symbols even if your sources do not Error analysis between two languages Contrastive Phonation Phonation chart (contrastive phonations) Word list illustrating each phonations Description of the contrastive phonations (e.g. the acoustic and articulatory properties of the phonations; whether the phonation contrasts only occur in certain phonological environment but neutralized in other environments;) Error analysis between two languages Test of Predictions (optional) You may test your predictions using an automated speech engine of your choosing, or by transcribing the speech of an L2 speaker or analyzing your errors in perception of the mystery language. This should be done after making your predictions; this project is not intended to be done in reverse where you look at learner or ASR errors due to other factors that may affect their performance. References APA author-year format (can use JIPA format if you already know it) Other formats (like MLA, Turabian, Vancouver, Chicago, etc.) will lose points Do not use numbered citations Minimum of 5 references Lado (1957) Ethnologue or similar for mystery language demographics Ethnologue or similar for chosen language demographics Article given to you for mystery language Phonetic description article for chosen language (JIPA articles are preferable if available)