

Constructing a language

LING 20001: Group project



Calvin and Hobbes by Bill Watterson for January 09, 1995

As we have seen, the primary task of a linguist is to identify and analyze patterns in data from natural languages. However, the help of linguists is often enlisted when it comes to creating invented languages, which are sometimes referred to as *conlangs* (*con(structed) lang(uage)s*). Nowadays, the producers of many big-budget TV shows and movies look to linguists to create realistic languages for the fictional characters in their stories. Some more recent examples of this include Dothraki and (High) Valyrian in *Game of Thrones*, Klingon in *Star Trek*, Na'vi in *Avatar*. J. R. R. Tolkien was a pioneer in this regard creating detailed conlangs such as the dialects of Elvish for his *The Lord of the Rings* novels. In order to have an invented language be realistic and practical, knowledge of linguistics is invaluable when it comes to designing the grammar and vocabulary of a language based on what we know about real languages.

This is your task in this group project, albeit on a much smaller scale. The job of each group is to construct and analyze a fragment of the grammar of the fictional language *Tokfewa*. We don't know much about Tokfewa other than its name and that it is spoken by humans or humanoid creatures and should therefore resemble a realistic human language in terms of its sounds and grammar. You could imagine that it is only mentioned in passing in some novel which is now being adapted for TV or film and the showrunners want to have actual lines of dialogue in this language. Given what we have learned so far about linguistic analysis, each group will write a short report detailing the core aspects of the phonology, morphology and syntax of the language. Of course, this language doesn't really exist, so it is up to you to figure out what it could look like.

This document contains detailed instructions for the project. Read these instructions **carefully**. They outline the structure and form that the final report should take.

Instructions

The report should be divided into the sections and subsections given below. My instructions for the contents of each part also follow this structure. Your report should also have this exact structure.

- Section 1** Phonology
 - Section 1.1** Sound inventory
 - Section 1.2** Phonemes and allophones
 - Section 1.3** A phonological derivation
- Section 2** Morphology
 - Section 2.1** Morphemes
 - Section 2.2** Person and number inflection
- Section 3** Syntax
- Section 4** Vocabulary list

The report should include both examples from your language and illustrations of the analysis for each relevant section. In addition, you should describe what you have presented in each part in prose, as if you were explaining it to the reader. I will provide examples of what this could look like in gray **example boxes** such as the one below. Your report should contain sentences in full prose that describe what you are showing, as in the examples I have given. The purpose of these examples is to give you an idea of the form/style and level of detail that I am expecting the report to have. You should not just minimally adapt this template for your own language with different sounds. Try to be creative and come up with your own ideas!

This is an example box.

You should only submit **one** report per group. It should be uploaded to Canvas as a single PDF and presented in a word-processed format (**not** handwritten!).

Note that you are **not** expected to invent a writing system for the language. All the examples you provide from the language should be given using IPA transcription. If you want to provide an English transliteration using Latin alphabet, you can. I have done that in my examples to show you what that would look like.

The group project constitutes **25% of your final grade**. The final grade for the project will be based on how well you follow the instructions given below and whether you demonstrate a good understanding and application of the analytical concepts we have addressed in class and the homework assignments. In addition to your final project, I will ask you to submit a **brief** explanation of how the work was divided up between the group members.

If you are unsure about anything when developing your language, you can get in touch with me, Kat or Naomi. **The deadline for the project is Friday June 3rd at midnight. I am happy to give comments on any drafts of your project that I receive by the end of Week 9 (by Sunday May 29th).**

1 Phonology

1.1 Sound inventory

The first step is to think about what sounds your language has. No language uses all the possible sounds in the IPA, so neither should yours. As we have seen, languages often have multiple sounds with the same place or manner of articulation. Also, sometimes there is just a distinction in voicing.

You should first present the basic sound inventory of the language (i.e. the sounds that its words can contain). The language should have vowels and consonants. It should have **at least 10 consonants** and **at least 5 vowels**. This includes both affricates and diphthongs, if you want to include these in your language. Furthermore, **at least two of the consonants** and **one of the vowels** should be **sounds that English does not have**. Also, since you already know that the name of the language in English orthography is *Tokfewa*, this should give you a hint of some of the sounds it could have, though the correspondence between sounds and letters could be unlike English.

The consonant inventory of the language should be presented in the form of a table similar to the IPA format with a column for the place of articulation and rows for the manner of articulation. Rows or columns that do not contain consonants in your language do not need to be included. In addition to arranging the sounds of your language in this way, you should describe what the table shows (as in the example below). If your language has affricates, they should be mentioned separately below the table. This description should make reference to **every sound in the language**.

Consonants: Tokfewa [tɔkɛβa] has a total of 12 consonants. These include four bilabial sounds: a voiceless bilabial plosive [p], a voiced bilabial plosive [b] and a voiced bilabial fricative [β]. There is also a bilabial nasal [m]. It only has one dental sound, a voiceless dental fricative [θ]. [*and so on...*]

	bilabial	dental	alveolar	postalveolar	palatal	...
Nasal	m					
Plosive	p b					
Fricative	β	θ				
...						

(Note: This table is incomplete!)

Furthermore, the language has two affricates: [kf] and [gv].

As for the vowels, these should be arranged in a similar kind of table based on the relevant articulatory features of the vowel. Since there are fewer combinations, the format of the table should look exactly as it does below (do not omit any columns or rows for the vowels). Bear in mind that the pairs of sounds differ in lip rounding (right member of the pair is rounded).

Vowels: Tokfewa has six vowels. These include three front vowels: a close unrounded high vowel [i] and its rounded counterpart [y], as well as an open-mid front unrounded vowel [ɛ]. [*and so on...*]

	Front	Central	Back
Close	i y		
Close-mid			
Open-mid			
Open			

(Note: This table is incomplete!)

Tokfewa also has two diphthongs, [aɪ] and [ɔɪ].

1.2 Phonemes and allophones

The next step is to decide which sounds are going to be phonemes and which will be allophones. Obviously, most of them will be phonemes, since we want to actually be able to express a wide range of meanings with them. You should illustrate **at least five** phonemic contrasts in your language with minimal pairs.

In Tokfewa, there are the following phonemic pairs: [a]:[ɔ], [ʃ]:[ʒ], [*and so on...*]. This is shown by the minimal pairs below.

<i>kfa</i>	[kfa]	‘this’	<i>kfo</i>	[kɸ]	‘sky’
<i>payzach</i>	[paɪʒaɰ]	‘feeling’	<i>payshach</i>	[paɪʃaɰ]	‘to see’
[<i>and so on...</i>]					

In addition to this, your language should have **at least two** cases in which a sound which is an allophone of another sound. You should think about the **conditioning environment** for each of the allophones, i.e. the context in which one shows up instead of the other. To show this, you should present a set of **at least six** examples for each allophone pair that shows the distribution of the sounds clearly. You don’t have to, but it might be helpful to put all sounds with one allophone in one column, and the others in a different column, as I have done in the example below.

In Tokfewa, [b] and [β] are allophones of /b/, as the following examples show:

<i>bunda</i>	[bʊnda]	‘something’	<i>awakfe</i>	[aβakfɛ]	‘house’
<i>gazob</i>	[gaʒɔb]	‘to listen’	<i>gazowo</i>	[gaʒɔβɔ]	‘he/she/it listens’
<i>ubmasa</i>	[ʊbmaza]	‘father’	<i>ewa</i>	[ɛβa]	‘happy’

[β] is found between vowels, while [b] surfaces in all other contexts (including word-intially, before and after consonants and word-finally). This can be captured by the following phonological rule:

$$/b/ \rightarrow [\beta] / V_V$$

Also, the following examples illustrate that [u] and [ʊ] are allophones of /ʊ/.

<i>owoosa</i>	[ʊβuza]	‘child’	<i>ushafi</i>	[ʊʃafi]	‘we two know’
<i>oozaza</i>	[uʒaʒa]	‘mountain’	<i>gvuma</i>	[gvʊma]	‘I sing’
<i>koowo</i>	[kuβɔ]	‘dog’	<i>bashub</i>	[baʃʊb]	‘to run’

[u] is only found before voiced fricatives such as [z], [ʒ], [β]. [ʊ] is found in other contexts (including before voiced stops [b], voiceless fricatives [f] and nasals [m]). The following rule accounts for this:

$$/ʊ/ \rightarrow [u] / __ [\text{voiced fricative}]$$

A word of caution: Assuming a rule like $/ʊ/ \rightarrow [u] / __ [\text{voiced fricative}]$ will have implications for every word you create in your language. It means that there should be no word in the language that has a voiced fricative following [ʊ].¹ Given what I have written here, a word like *owusa* [ʊβʊza] should not be possible in Tokfewa.

1.3 A phonological derivation

The description of your language should also include a case where **two** of the phonological rules you proposed **interact** in a derivation. Two rules ‘interact’ if the order of the rules matters, and you should show this in your writeup. Think about the possible ways in which two rules can interact: for example, one rule can create the context for the other rule, or it can remove the context for the other rule. There are other options too. In this part, you should repeat and number the two relevant rules, give a small set of relevant examples for your derivation and then provide a table showing the derivation of **at least four** output forms with these two rules. This should include one example in which **only rule 1 applies**, one example in which **only rule 2 applies**, one example in which

¹The only exception here would be if there was a rule that creates voiced fricatives and that applies after the vowel raising rule. I have not assumed any such rule here.

neither rule applies and one example where the result is **different depending on the order of the rules**. You should also show that the reverse order of rules leads to the wrong result in this last case.

To give you an example of how to go about this, we just saw that my version of the language has a rule that turns an underlying /b/ into a [β] between vowels (/b/ → [β] / V__V). It also has a rule that turns an /ʊ/ into [u] before a voiced fricative. We can just pick one of the forms we used in each data set for the allophones to include in our sample derivation. I will use [εβa] (‘happy’) for the first rule and [uʒaʒa] (‘mountain’) for the second. You should still repeat these examples in this section (see example box below). Now, we need an example with no rule applying and another where the order matters. There are cases in which neither rule will apply since its context is not met. We already saw the word [baʃʊb] (‘to run’) for the distribution of the allophones of /ʊ/. There is a /b/, but it is not before a vowel so the first rule won’t apply here. Also, it contains an /ʊ/, but this is not before a voiced fricative such as [β] here, so the second rule won’t turn it into [u]. This can be used as the third example (where no rule applies). Now, we need to find an example where the order of rules matters. This is the case when one rule creates or removes the context for the other. So, if a rule changes /ʊ/ to [u] before voiced fricatives, then another rule that creates a voiced fricative like [β] before a vowel can create a new context for that first rule to apply. Similarly, a different kind of rule might remove the context the /ʊ/ → [u] rule by changing the vowel or the consonant in some way. In the section on morphology, you will be asked to provide examples of some affixes in the language, e.g. for plural, derivation and inflection. If a vowel-initial suffix were added to [baʃʊb], the context for the first rule would be met (/b/ between two vowels). Since [baʃʊb] is a verb, I will use the suffix -ɔ for 3rd person singular to give us the underlying representation /baʃʊb-ɔ/. This now gives us a case where the order of the rules matters (as we will see in the example below).

In Tokfewa, the following two rules interact:

1. /b/ → [β] / V__V
2. /ʊ/ → [u] / __ [voiced fricative]

We will show this based on these words:

<i>ewa</i>	[εβa]	‘happy’	<i>oozaza</i>	[uʒaʒa]	‘mountain’
<i>bashub</i>	[baʃʊb]	‘to run’	<i>bashoowo</i>	[baʃuβɔ]	‘he/she/it runs’

The following derivation shows that Rule 1 must apply before Rule 2. Rule 1 turns /b/ into [β] between vowels, as in [εβa]. The underlying /ʊ/ in /ʊʒaʒa/ becomes [u] due to Rule 2. In the word [baʃʊb], neither of the rules have applied. However, when the 3rd singular suffix -ɔ is added, Rule 1 applies to give us baʃʊβɔ. The output

of Rule 1 now meets the context for Rule 2 (β before a υ) and it applies to give the surface form $[ba\upharpoonright u\beta\upsilon]$.

UR	/ $\epsilon\beta a$ /	/ $\upsilon\zeta a\zeta a$ /	/ $ba\upharpoonright\upsilon b$ /	/ $ba\upharpoonright\upsilon b-\upsilon$ /
Rule 1	$\epsilon\beta a$	—	—	$ba\upharpoonright\upsilon\beta\upsilon$
Rule 2	—	$u\zeta a\zeta a$	—	$ba\upharpoonright u\beta\upsilon$
SR	$[\epsilon\beta a]$	$[u\zeta a\zeta a]$	$[ba\upharpoonright\upsilon b]$	$[ba\upharpoonright u\beta\upsilon]$

If the rules applied in the reverse order (Rule 2 before Rule 1), we would get the wrong result for the underlying form $/ba\upharpoonright\upsilon b\upsilon$, namely $*[ba\upharpoonright\upsilon\beta\upsilon]$, as shown below.

UR	/ $\epsilon\beta a$ /	/ $\upsilon\zeta a\zeta a$ /	/ $ba\upharpoonright\upsilon b$ /	/ $ba\upharpoonright\upsilon b-\upsilon$ /
Rule 2	—	$u\zeta a\zeta a$	—	—
Rule 1	$\epsilon\beta a$	—	—	$ba\upharpoonright\upsilon\beta\upsilon$
SR	$[\epsilon\beta a]$	$[u\zeta a\zeta a]$	$[ba\upharpoonright\upsilon b]$	$*[ba\upharpoonright\upsilon\beta\upsilon]$

Bear in mind, that coming up with rules like this means making a general statement about the shape of words in your language. All else being equal, no word should have a $[\beta]$ before a $[\upsilon]$ because the rule would change it. This doesn't have to be the case though. It is equally possible that the second order (Rule 2 before Rule 1) was what we found in the language with $[ba\upharpoonright\upsilon\beta\upsilon]$ as the grammatical form. As you can see, this gives rise to what looks like an exception to Rule 1, where we do find $[\upsilon]$ before $[\beta]$ in a word. But these examples can only come from cases where we have an underlying $/\upsilon b/$. An underlying $/\upsilon\beta/$ sequence would still be changed to $[u\beta]$. Feel free to explore all of these options when creating your language, including cases where one rule removes the context for the other to apply!

2 Morphology

2.1 Morphemes

Next, we turn the morphology of the language. Given the sounds you said that your language has, you should think about its morphemes. You will need some free morphemes (roots) and bound morphemes (affixes). The vocabulary list at the end will include more examples, just select a few here for the purposes of illustration.

You should illustrate with **four pairs** of examples of how plural and singular nouns are formed. You could have a separate affix for singular and plural, or you could just have one affix for either. In addition, provide another set of four pairs illustrating a derivational affix that **changes the category** of one morpheme into another.

At least one of these affixes should be something other than a prefix or a suffix (e.g. a circumfix, infix, reduplicating affix, template). For infixes and reduplication, you should state which part of

the word is relevant (e.g. ‘the infix is placed after the first consonant in the word’, ‘the final CV sequence is reduplicated’).

Plural nouns in Tokfewa are marked by a circumfix *gv(a)-ta* (the vowel is absent before vowel-initial roots), as shown below.

<i>koowo</i>	[kuβɔ]	‘dog’	<i>gvakoowota</i>	[gvakuβɔta]	‘dogs’
<i>oozaza</i>	[uʒaʒa]	‘mountain’	<i>gvoozazata</i>	[gvuʒaʒata]	‘mountains’

[and so on...]

Tokfewa also has the suffix *-(a)χi* that turns adjectives into nouns:

<i>ewa</i>	[εβα]	‘happy’	<i>ewachee</i>	[εβαχι]	‘happiness’
<i>zooz</i>	[ʒuʒ]	‘angry’	<i>zoozachee</i>	[ʒuʒaχι]	‘angriness’

[and so on...]

2.2 Person and number inflection

Next, you should present the way person and number inflection is marked on pronouns and verbs. First, give the paradigm of pronouns for your language. You do not have to analyze this with realization rules (but you can if you want to). You might want to think about how pronouns could be constructed – is there one form for each person distinction and then a suffix for dual and plural? Maybe the singular is marked with a suffix and the plural is not. Some or all of the forms could also be irregular (perhaps not involving any affixes).

After this, present a paradigm corresponding to the inflectional affixes that appear on a Tokfewa verb depending on the kind of subject that verb has. You should include the features singular, dual and plural for number, and first, second and third for person in the paradigm. Bear in mind that this affix does not have to be a prefix or suffix (and different cells might vary in this regard). You should also provide a set of realization rules for these affixes. The paradigm should contain **at least three** distinct affixes and should contain **at least one** form that requires reference to feature decomposition (e.g. [+singular]). For the purposes of this assignment, you can treat dual as either [+singular, +plural] or [-singular, -plural] to give you more options. There should also be an affix that is **an elsewhere form**. After the realization rules, provide an example of a verb combined with each of the affixes (and provide its translation(s)).

The pronouns in Tokfewa have the following forms:

	singular	dual	plural
1st person	<i>ba</i> [ba]	<i>bazu</i> [baʒu]	<i>bacho</i> [baχɔ]
2nd person	<i>wu</i> [βu]	<i>woozu</i> [βuʒu]	<i>wucho</i> [βuχɔ]
3rd person	<i>shu</i> [ʃu]	<i>shoozu</i> [ʃuʒu]	<i>shucho</i> [ʃuχɔ]

The person and number inflection is given below. The suffix *-a* marks 1st and 2nd person singular, while the suffix *-i* just marks 1st person dual. There is also syncretism with the affixes *-ap-* and *-ɔ*. The plural is marked with the infix *-ap-*, which is placed before the final vowel in the word. 3rd person singular and dual as well as 2nd person dual is marked by *-ɔ*.

	singular [+singular, -plural]	dual [-singular, -plural]	plural [-singular, +plural]
1st person [+speaker, -addressee]	[-a]	[-i]	[-ap-]
2nd person [-speaker, +addressee]	[-ɔ]	[-ɔ]	[-ap-]
3rd person [-speaker, -addressee]	[-a]	[-ɔ]	[-ap-]

This can be analyzed by the following realization rules:

[+speaker, -singular, -plural]	→	-i
[-addressee, +singular]	→	-a
[+plural]	→	-ap-
elsewhere	→	-ɔ

These affixes are illustrated below for the verb [gvum] ('to sing'):

<i>gvum</i>	[gvum]	'to sing'
<i>gvumi</i>	[gvumi]	'we two sing'
<i>gvuma</i>	[gvuma]	'I sing or he/she/it sings'
<i>gvumo</i>	[gvumo]	'you/you two/those two sing'
<i>gvapum</i>	[gvapum]	'we/you (pl.)/They sing'

3 Syntax

Now, you can put all of this together to construct some sentences in your language. You should provide **at least five** example sentences in your language. These should be presented in the format of a **glossed three-line example**, as we have seen in class. The first line contains the words from your language, aligned with the beginning of each word is the literal translation of that word (or 'gloss'). Underneath, you should write an English translation of the sentence in 'single quotes'.

word word word
gloss gloss gloss
'translation'

The sentences you construct must contain the following things:

- A pronoun as a subject
- A plural noun as a subject
- A noun phrase containing an adjective
- A noun phrase containing a determiner
- The category changing affix you presented in section 2.1
- An intransitive, transitive and ditransitive verb

In addition you should **draw the tree** for each example you provide and **list the full set** of syntactic rules that you need for your examples. Also state what the basic word order of your language is (subject, verb and object).

If you want to include more details about your language, you are more than welcome to. Be sure to explain anything else you assume and provide an example that illustrates it. For example, your language could have different forms for nouns when used as subjects or objects, it might have a special way of forming a question. These things are optional, but include them if you wish!

Tokfewa has the following syntactic rules:

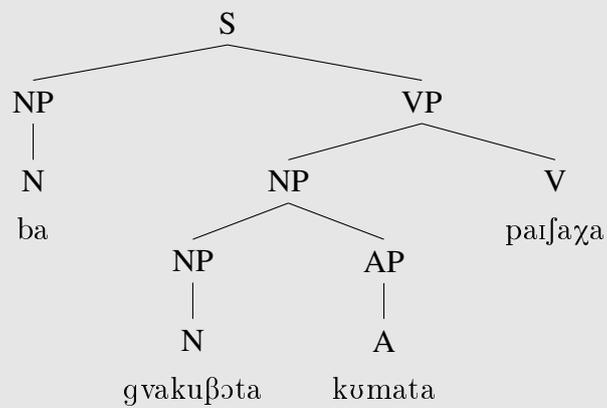
$$\begin{aligned}
 S &\rightarrow NP VP & VP &\rightarrow NP V \\
 NP &\rightarrow NP AP & NP &\rightarrow N \\
 &&& [and\ so\ on...]
 \end{aligned}$$

The basic word in the language is SOV.

Here are some examples of sentences in Tokfewa:

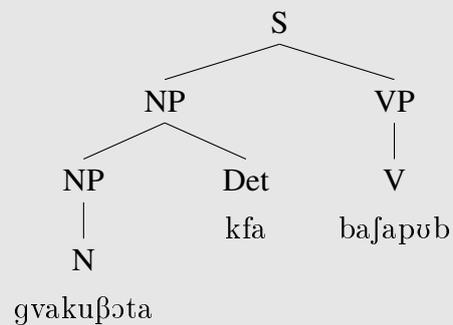
ba gvakoowota kumata payshacha.

ba gvakuβota kumata paifaxa
 I dogs black see
 'I see some black dogs'



gvakoowota kfa bashapub.

gvakuβota kfa bashapub
 dogs this run
 'these dogs run'



[and so on...]

4 Vocabulary list

The final thing to provide is a list of words in your language. Many of these will have already been mentioned so far, so you can just repeat them here again and add any others you need. The words should be grouped into categories and contain **at least ten** nouns and **at least ten** verbs, **at least five** adjectives, **at least two** determiners. You might need more categories depending on the kind of sentences you chose to create (e.g. complementizers, auxiliaries) Note that the words listed here should be root morphemes, i.e. they should **not** any contain affixes (such as plural or 3rd person singular).

<u>Nouns</u>	<u>Verbs</u>	<u>Adjectives</u>	<u>Determiners</u>				
[kuβɔ]	‘dog’	[gvɔm]	‘to sing’	[kɔmata]	‘black’	[kfa]	‘this’
[aβakɛ]	‘house’	[baʃɔb]	‘to run’	[ɛβa]	‘happy’		
[uʒaʒa]	‘mountain’	[ɔʃaf]	‘to know’	[ʒuʒ]	‘angry’		
[ɔbmaza]	‘father’						
[ɔβuza]	‘child’						

(Note: This table is incomplete!)