

Some resources on determining the distinctive features of a language

In preparation for the FLEx Parser workshop

Short Summary:

In preparation for the FLEx Parser workshop, each participant is requested to create at least a first draft of a set of distinctive features for the language that s/he wants to parse. Instructions for doing this are provided below.

Those who want to use the HermitCrab parser are required to make a complete list of all the phonemes of the language, and to unambiguously distinguish all phonemes by their distinctive features. This may be a daunting task, as it requires a good understanding of the phonological system of the language. But even those who plan on using the XAmple parser can benefit from determining the distinctive features, as it allows a better use of natural classes, which even this parser makes use of for environments of allomorphs. Therefore we ask all FLEx-Parser workshop participants to come with a prepared set of distinctive features – at least in a first draft. This document intends to give you some instructions on how to go about it. We also hope to create a video with similar content.

What are distinctive features?

Distinctive features are widely used in phonology since the early days of Chomsky's and Halle's Generative Phonology.¹ But the idea was already developed by Prague-School forerunners Trubetzkoy and Jakobson in the 1930s. They no longer assumed that the phoneme is the smallest phonological unit, but that the phonemes of a language can be described in terms of features based on phonetic realities. These features constitute a system which not only helps to unambiguously describe each phoneme, but which also places each sound inside a network of relationships that is used by the language in its various phonological rules.

When working with phonologically determined allomorphs and/or with phonological rules in grammar, such features are very useful, and both parsers of FLEx work best when the feature system of a language is well understood.

As said above, phonological features are based on phonetic realities, and so there are widely accepted lists of features that may serve to describe all sounds of all languages in the same way. Such lists can be found in most phonology textbooks and in various places on the Internet. The FLEx-Software also comes with a pre-arranged list of features to choose from – it can also be tweaked to serve your purposes. This is likely to be necessary, as the feature system of a language is expected to deviate from the phonetically based feature inventories of such lists, as each language leaves phonetics somewhat behind in its phonological system.

How to proceed

In preparation for the workshop it is enough to create the feature set on paper or in a document, so that a consultant can have a look at it before it is implemented into the database. Those who feel confident about their analysis are of course welcome to enter everything into FLEx right away. What should be done in any case is to clean up and adapt the list of phonemes in the grammar section of fieldworks, so that only those characters are shown in their that are actually used in the data, and, just as importantly, all of them are shown.

¹ Chomsky, Noam and Morris Halle (1968). [The Sound Pattern of English](#). New York: Harper & Row.

The basis of a feature set is a complete chart of phonemes of the language, arranged by their phonetic characteristics, as in the following example from the Majang language:

	Labial	Alveolar	Postalveolar	Palatal	Velar
Voiceless stops	p	t			k
Voiced stops	b	d			g
Voiceless affricates			c		
Voiced affricates			ɟ		
Implosives	ɓ	ɗ			
Nasals	m	n		ɲ	ŋ
Trill		r			
Lateral		l			
Glides	w			j	

Such a chart based on the phonetic characteristics of the sounds is not the best possible way to display the relationships between them, as it contains a lot of phonetic detail that is not relevant to the language's phonological system. For example, in Majang there is no opposition between postalveolar and palatal sounds, and not even between stops and affricates. It is therefore more insightful to remove all distinctions that clutter up the system – a process that Pike called *matrix permutation*. This step is crucially important to find out the features behind the system, and you may want to discuss this with a phonological consultant. For Majang, it results in the following phoneme matrix:

	Labial	Alveolar	Palatal	Velar
Voiceless Obstruents	p	t	c	k
Voiced Obstruents	b	d	ɟ	g
Implosives	ɓ	ɗ		
Nasals	m	n	ɲ	ŋ
Oral Sonorants	w	r	j	
Lateral Sonorant		l		

This chart can now be carved up into the features that are needed to unambiguously distinguish all phonemes from each other. Here it is best to begin by applying the main class features consonantal, obstruent, sonorant, continuant first, followed by as many place-of-articulation features as needed. Only then you add lower-level features such as nasal, voiced, delayed release, etc, so that you don't end up with more features as needed. For Majang, this results in a surprisingly low number of features – consonantal or continuant are not even needed in this language. This is shown in the following table. It helps to identify the consonant features, which are all needed in addition to the feature *[-SYLLABIC]*.

		[+LABIAL]		[-LABIAL]	
				[-HIGH]	[+HIGH]
[+OBSTRUENT]		p	t	c	k
		b	d	ɟ	g
[-OBSTRUENT] [-SONORANT]		ɸ	ɬ		
	[+SONORANT]	w	r	j	
	[-NASAL]		l		
	[+NASAL]	m	n	ɲ	ŋ
				[-BACK]	[+BACK]

The feature syllabic is needed to distinguish all consonants from the vowels, which are classified in a different feature matrix:

		[-LABIAL]		[+LABIAL]	
[+HIGH]		i		u	
		e		o	
[-HIGH]	[+ATR]	ɛ		ɔ	
	[-ATR]		a		
		[-BACK]		[+BACK]	

[ATR] is the only vowel feature which is not also used for consonants. Majang then also needs a few prosodic features for tone and vowel length.

Based on the two feature matrices, it is now possible to define for each phoneme the set of distinctive features necessary for its unambiguous identification:

phoneme	syllabic	obstruent	sonorant	nasal	lateral	labial	back	high	voiced	ATR
p	-	+				+			-	
b	-	+				+			+	
t	-	+				-	-		-	
d	-	+				-	-		+	
c	-	+				-	+	-	-	
ɟ	-	+				-	-	+	+	
k	-	+				-	+	+	-	
g	-	+				-	+	+	+	
m	-	-	+	+		+				

phoneme	syllabic	obstruent	sonorant	nasal	lateral	labial	back	high	voiced	ATR
n	-	-	+	+		-		-		
ɲ	-	-	+	+		-	-	+		
ŋ	-	-	+	+		-	+	+		
ɸ	-	-	-			+				
ɸ'	-	-	-			-				
l	-	-	+	-	+	-		-		
r	-	-	+	-	-	-		-		
w	-	-	+	-		+				
j	-	-	+	-		-		+		
a	+					-	+	-		
ɛ	+					-	-	-		-
e	+					-	-	-		+
ɔ	+					+	+	-		-
o	+					+	+	-		+
u	+					+	+	+		
i	+					-	-	+		

Many of the cells could be left empty in this matrix, as there is a lot of redundancy with all these features. But for FLEEx it is best to specify all features fully.

Some advice

- Feel free to redefine features from what you find in the official lists if the situation in your language makes this necessary. Distinctive features are always a language-specific set which is only normally based on what you find in the literature, but not always so.
- If you know what kind of phonological rules are active in a language, this should help you with identifying the needed phonological features. All input or environment variables should be able to be expressed in simple and intuitive feature sets.
- If you can, get some help from a phonology consultant for this task.
- Insofar as they are reflected in the orthography, all prosodic variations of vowels and consonants, such as length, gemination, or tone, require separate phonemes in FLEEx, because their features differ.