INTRODUCTION TO MORPHOLOGY

Rutgers and McGill

Draft – for class use

CHAPTER E

"EXOTIC" WORD FORMATION

7.1 Is Morphology Just About Stems and Affixes?

So far, we have presented a theory of morphology based on the idea that complex words are always made up of discrete morphemes. These morphemes can be of two types: roots and affixes, the difference being whether the morpheme can form the basis of a word (a root) or whether it is required to be in a particular position with respect to other material (an affix). In essence, this is saying that there are only two kinds of word formation: affixation, which adds an affix to a stem, and compounding, which combines two stems. Some prototypical examples are repeated here:

(1)			
	Prefix + Stem:	joy – **en**joy	Suffix + Stem: joy – joy**ful**
		write – **re**write	write – writ-**er**
	German**:**	Enkel – **Ur**enkel	Enkel – Enkel**in**
	'grandson'-'greatgrandson'		'grandson'-'granddaughter'
	Stem + Stem:	grandson	
		typewriter	

The technical term for these methods of creating new words by simple processes of adding identifiable morphemes is *concatenation.* And indeed the large majority of the word formation in languages of the world is clearly of this type.

Nevertheless, there are types of word formation that do not fit so nicely into this picture. There seem to be other ways in which words can be related to each other, for which it is harder to clearly distinguish a discrete affix. These are sometimes called *nonconcatenative morphology* or morphological *processes*. Several kinds have been identified; they constitute the exotica that

are often included in introductory textbooks to impress beginning students with the strange things that can happen in languages of the world.

For example, in addition to prefixes and suffixes, there exist some infixes in some languages. These are sequences of sounds that appear inside the root, rather than before it or after it. Tagalog (a language of the Philippines) is a famous example of a language with infixes:

(2)

present	past passive	Tagalog
lakad	linakad	'walk'
pili?	pinipili?	'choose'
sulat	sinulat	'write'
hanap	hinanap	'seek'
basa	binasa	'xxx'

Another, more radical case of nonconcatenative morphology is what is called reduplication. In reduplication, a new form of a word is created by doubling some part of the original stem. Tagalog has instances of reduplication as well, as in the following future tense forms of the verb:

(3)

present	future	TAGALOG
takbuh	tatakbuh	'run'
lakad	lalakad	'walk'
pili?	pipili?	'choose'
sulat	susulat	'write'
hanap	hahanap	'seek'

There is a clear regularity in this data: the future tense involves repeating the first consonant and vowel of the stem/root. On the basis of this regularity, you could reliably create the future form of new Tagalog root that you had never seen before. (What, for example, is the future form of *basa*? It would be *babasa*.) However, it is difficult to isolate a well-defined morpheme that expresses the future meaning. In some cases, what is added is *ta-*, in others *la-*, in others *pi-*; it all depends on the root that you start with.

In still other cases, one seems to make new words not by adding a morpheme, but rather by transforming the sounds of the root in a systematic way. Here is a case in point from Nivkh:

(4) Nivkh verbs (spoken on Sakhalin Island, Russian Far East) ɖ = palatalized /d/

Intransitive Verb	Transitive Verb	
tʌŋzʌlʌɖ	rʌŋzʌlʌɖ	'weigh (s.th.)'
q^havuɖ	χavuɖ	'warm (s.th.) up'
kesqoɖ	ɣesqoɖ	'burn something/oneself'
pʌkzɖ	vʌkzɖ	'lose something / get lost'

Again, there is a regularity: the intransitive verbs all begin with a stop consonant, whereas the transitive verbs begin with a continuant produced at the same place of articulation. However, it is hard to see the transitive verbs as containing an affix that is not present in the intransitive verb, even though we clearly want to say that the roots are the same in both forms.

As a final example, consider the following verb forms from Classical Arabic (the language of the Koran).

(5) Classical Arabic (all forms 3sg masc subject)

	"write"		"earn"	
	ACTIVE	PASSIVE	ACTIVE	PASSIVE
present (X-s)	katab	kutib	daraj	durij
cause to X	kattab	kuttib	darraj	durrij
X for ea. other	kaatab	kuutib	daaraj	duurij
make X	?aktab	?uktib	?adraj	?udrij

A careful look at this table of forms reveals that all the verbs that contain the meaning of 'write' contain the consonants *k-t-b*, always in that order. In contrast, verbs that contain the meaning 'earn' do not have these consonants; rather they contain *d-r-j* in the same positions. Meanwhile, all of the active forms have (only) the vowel "a", whereas the passive forms have the vowels "u" and "i", in that order. There are no special vowels or consonants associated with the meaning 'to cause to X'. Nevertheless, the notion is expressed in a systematic way: the middle consonant is doubled if and only if this meaning is present. Now clearly we cannot say that the passive morpheme *u-i* is a prefix or suffix to the verb root *k-t-b* 'write'. Rather, it seems like the two morphemes are fused, or interleaved in a discontinuous way. More generally, the lexical content of the word is given by the consonants, the voice of the verb (whether it is active or passive) is given by the vowels, and notions like causative and reciprocal are expressed by exactly how these consonants are arranged. For this reason, Arabic is said to have a "root and pattern" system. It seems quite remote from the ordinary world of affixation and compounding.

For all of these cases that seem to go beyond the bounds of affixation, the most obvious thing to do is to formulate transformational rules. These rules would take the string of phonemes that constitutes the root and transform it in some systematic way into a new string of phonemes. For example, the morphological processes we have seen so far could be expressed as follows:

(6) a. To make the passive form of the verb, add *in* after the first consonant. (Tagalog)

b. To make the future form of the verb, repeat the first consonant and vowel of the root. (Tagalog).

c. To make the transitive form of the verb, change the first consonant of the root into a +continuant sound. (Nivkh)

d. To make a causative verb, double the middle consonant of the root. (Arabic)

e. To make a passive verb, replace the vowels *a* and *a* in the root with *u* and *i*. (Arabic)

Morphological transformations of this kind have been known as *Word Formation Rules*.

If we simply added the notion of word formation rules to our theory as it stands, the result would be a somewhat eclectic, hybrid theory. We would be saying that there are really two fundamentally different kinds of morphology: concatenation and transformation. The two would have no obvious principles in common. This would not be a very attractive view. Rather, it is appropriate to ask whether one type of morphology can be explained as a special case of the other type.

In fact, it is easy to recast standard cases of affixation as word formation rules. We could state the following, for example:

(7) a. To form the plural of a noun, add /s/ to the end. (English)

b. To form the past tense of a verb, add /d/ to the end. (English)

c. To form a transitive verb from a noun, add /en/ to the beginning. (English)

These rules are quite simple, and will work perfectly smoothly. From this perspective, ordinary affixation is no more than a special case of word formation rules: they are word formation rules that add fixed sounds to the beginning or end of a word. This could give a unified approach to morphology—but one that is very different from the one we have developed so far. In this approach, the idea of a morpheme would lose most or all of its significance, being replaced with the idea of a rule.

At first glance, it seems harder to accomplish the opposite reduction: it is hard to imagine how the "exotic" word formations could be seen as special kinds of affixation. However, there is a sense in which this difficulty is a good thing. The reason why it was so easy to recast affixations as word formation rules is because word formation rules can do anything. Any regular pattern whatsoever can in principle be stated as a word formation rule. Therefore, it is hardly surprising that affixation can be stated in this way. However, it is not the case that any regular pattern can be the basis of productive morphology in natural human languages. For example, the following are perfectly reasonable word formation rules:

(8) a. To make the future form of a verb, repeat the first consonant and the last vowel.

b. To make the causative form of a verb, reverse the order of the sounds.

c. To make the transitive form of a verb, cut off the first syllable of the word and place it at the end.

If Tagalog, for example, had morphological processes like this, they would result in data like the following:

(9)	present	future	causative	transitive
	lakad	lalakad	dalak	kadla
	pili?	pipili?	?ilip	li?pi
	sulat	sasulat	talus	latsu
	takbuh	tutakbuh	hubkat	buhtak

Consciously constructed language games sometimes behave a bit like this: for example, Pig Latin is a little bit like the third word formation rule. But no actual spoken language has morphological processes like these. This is true even though the word formation rules are not intrinsically more complex or difficult to understand than the attested ones in (6). Thus, the unconstrained use of word formation rules is not an appropriate theory of what is and is not possible in the morphological systems of natural languages.

Overall, then, our simple affixation theory is a bit too restrictive; it cannot account for all the kinds of word formation found in languages of the world as stated. On the other hand, the word formation rule theory is much too permissive; it cannot explain why some systematic changes are never used by natural languages. In principle, we could proceed by trying to restrict the kinds of word formation rules that are possible in some substantive ways. However, in practice it has been more productive to go the other way—to loosen up the affixation theory slightly, so that it can account for things like reduplication and infixation as well. If we can do this, then we can preserve the inherent restrictiveness of the affix-based theory in an attractive way. This then will be the fundamental goal of this chapter:

GOAL:

To show that *non-concatenative morphology* really is *concatenative* (just with more going on in the phonology).

This goal will become achievable once we integrate fully some independently motivated ideas from phonology. We will see that "exotic" morphemes are just like ordinary prefixes or suffixes, but they have slightly different phonological representations in their lexical entries.

E.2 Infixation

Let us begin by looking more carefully at the case of infixation. Our first example of this was from Tagalog, the forms of which are repeated here.

(10)	present	past passive	Tagalog
	lakad	linakad	'walk'
	pili?	pinipili?	'choose'
	sulat	sinulat	'write'
	hanap	hinanap	'seek'

basa binasa ‘xxx’

Similar infixes are found in many other Austronesian languages. These infixes provide us with a good starting point for two reasons. On the one hand, this is perhaps the easiest case to fit into our affixation framework. On the other hand, it already introduces the main theme of this chapter: that these processes are not so exotic at all morphologically, but the complications come from the phonology.

Our morpheme-based theory is built around three basic assumptions, which we state explicitly in (11).

(11) Aspects of a morpheme-based analysis:

a. Complex words can be divided up into discrete morphemes.

b. Each morpheme can be given a consistent lexical entry, including (at least) a phonological representation and a meaning (or set of syntactic features).

c. The morphemes are combined in the form of a word structure tree, in which the attachment properties of the morphemes are satisfied, and which determines the features of the whole word by way of the Node Labelling Conventions.

Now infixation already has most of these properties without doing anything too fancy. It is easy to see that a form like *sinulat* ‘was written’ comes in two distinguishable parts: *sulat* and *in*. Moreover *sulat* and *–in-* both have a consistent phonological shape and a consistent meaning. Therefore, it is not difficult to give them ordinary-looking lexical entries:

(12)

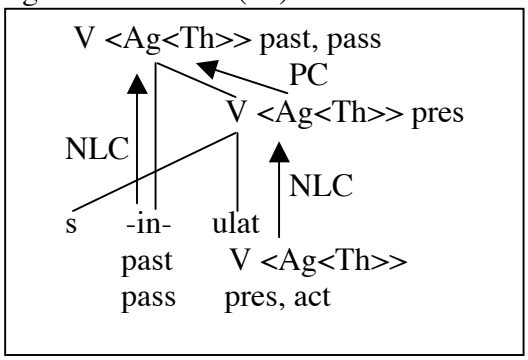
Label/phonology	sulat	in
Meaning	write	was Xed
Attachment properties	root	???
Features	V, <Agent <Theme>>, present, active	Ø, Ø, past, passive

The only cell that is difficult to fill in is the attachment properties of *–in-*. And we do already know something about this. Clearly we want to treat *–in-* as an affix, rather than as a root. There are several reasons for this. First, *-in-* clearly cannot serve as the basis for a well-formed word on its own; rather it must attach to some other morpheme. Indeed, it must attach to a verb root; this is the kind of category-specific selection that is typical of affixes cross-linguistically.

Second, if we treat *–in-* as an affix, then we will expect its features to take priority over the features of the root in determining the features of the derived word, in accordance with the Node Labelling Convention. And that would be a useful thing to say, since *–in-* changes the tense and voice features of the root to past and passive. The only real difficulty, then, is to state exactly where *–in-* attaches in a tractable way, and to draw a well-formed word structure tree correspondingly.

Maybe the most intuitive way to draw the word structure tree is with *–in-* inside of *sulat*, capturing directly the fact that *–in-* is an infix. Then we could draw a word structure tree something like the one in (13):

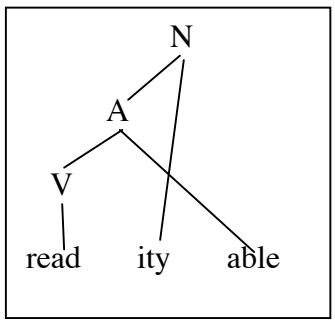
(13)



Corresponding to this tree, we might try stating the attachment properties of *–in-* by putting some notation like "appears with verb on both sides of it" in the relevant cell of the lexical entry. The innovation here is that we are allowing the branches of a word structure tree to cross, a possibility that we have so far made no use of. This might seem like a small price to pay to get our theory to work for infixation.

However, the smallest initial price does not always lead to the smallest total cost in the end. Allowing representations with crossing branches will cause trouble for our theory in other cases. For example, back in chapter d we gave an explanation for why the English word *readability* is well-formed, but the word **readityable*, containing the same morphemes, is not. The explanation was that since *–able* is a suffix that comes after *–ity* in the second example, it must be the last affix to attach to the word. *–Ity* creates nouns, and *–able* does not attach to nouns, so the example is ill-formed. But this mode of explanation becomes unavailable if we allow the branches of a word structure tree to cross. Then the following representation would be legitimate, in which all of the attachment properties of the affixes are respected:

(14)



Therefore, allowing word structure tree branches to cross undermines our explanation of this and many similar morphological facts. We would no longer be able to infer anything from the visible position of morphemes about the order in which they are attached—and we do not want this. This cure seems worse than the original disease: we would rather not have an account of infixes (which are fairly rare) than to have an account that undermines our theory of usual cases.

There are other reasons to be suspicious of an account in terms of crossing word structure trees. First, such an account may handle infixation, but it will make no contribution toward understanding the other types of exotic word formation. If there is a general theme to be discerned in the various exotic cases outlined in section one, this theory does not reveal it. Rather, we would have to make different adjustments for each type of exotic word formation if we start down this path.

Finally, note that the attachment statement "goes inside the verb root" is not something we should be satisfied with even if it was theoretically desirable. It is much too crude. The infix *-in-* cannot be placed just anywhere in the verb root: one cannot say *sulinat*, for example, nor *suinlat*, nor *sulaint*. Rather, the infix *–in-* must come immediately after the first consonant (or consonant cluster) of the root. But the morphology proper does not really know about consonants and consonant clusters: these are phonological notions, not morphological ones. This then is a hint that what is really going on with infixation might have more to do with phonology than with morphology.

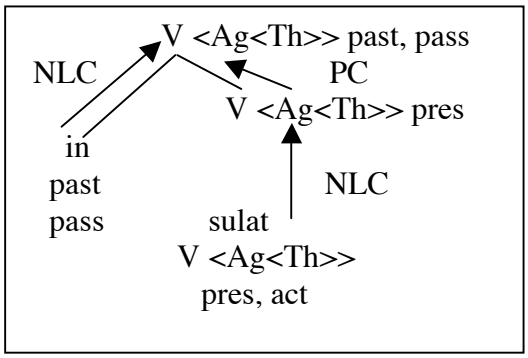
Suppose then we bite the bullet and say that there are no true infixes, only prefixes and suffixes. Which then is *–in-*? Clearly if one is forced to choose, we want to say that it is a prefix, because it is attached closer to the beginning of the word than to the end. More precisely, its attachment site is defined with respect to the beginning edge of the word, just as prefixes are. So then we complete the lexical entry of *–in-* as follows:

(15)

Label/phonology	sulat	in
meaning	write	was Xed
Attachment properties	root	*Prefix to verbs*
Features	V, <Agent <Theme>>, present, active	Ø, Ø, past, passive

On this assumption, the word structure tree of examples like *sinulat* will be a perfectly normal one, as follows:

(16)



This is all perfectly straightforward and unproblematic—except for the little detail that Tagalog speakers say *sinulat*, not *insulat*.

However, this may not be such a big problem after all. Our study of allomorphy has made us familiar with the idea that what people say is not identical to the output of the word structure tree; rather phonological rules can come in and adjust the representation in various ways. Thus, English speakers do not really say "walked" either; they actually say "walkt", with the past tense morpheme /d/ changing to a voiceless consonant [t] under the influence of the adjoining voiceless consonant /k/. Suppose then that it is a phonological rule that takes the underlying form *insulat* and produces from it *sinulat*. The phonological rule that we need to do this can be stated as follows:

(17) V1 – C1 – C2 – V2 → C2 – V1 – C1 – V2 / # ____

(morphologically sensitive)

Rules that change the order of phonemes are known in the phonological literature as *metathesis rules*. They are not particularly common, compared to rules of assimilation—but then infixation is not very common compared to suffixation either.

At first, it might seem that this proposal simply passes the buck: it takes a morphological oddity, and repackages it as a phonological oddity. One might ask what is ultimately gained by this. The glib answer is that now it is not our job to worry so much about infixation, given that this is a book about morphology, not phonology. But there is a more serious and responsible version of this answer as well. There are several signs that *–in-* infixation in Tagalog is actually a phonological phenomenon. We have mentioned one of these signs already: the fact that the exact place where *–in-* ends up is a phonologically defined location, not a morphologically defined on. *–In-* tucks in after the first *consonant* (cluster), not after the first *morpheme*. Another signs that phonology is at work comes from comparing *–in-* to a second infix that is found in Tagalog. *–in* is the past passive affix; there is also a second infix *–um-* that forms the simple past of the verb. It is illustrated in (18):

(18)	present	past (active)	Tagalog [Re Check: for declension class effects]
	lakad	lumakad	'walk'
	pili?	pumipili?	'choose'
	sulat	sumulat	'write'
	hanap	humanap	'seek'
	basa	bumasa	'xxx'

There are two interesting things to note here. First, the infix *–um-* appears in exactly the same place inside the verb root as *–in-* does, immediately after the first consonant (cluster). Second, the infix *–um-* itself has a phonological representation that is very similar to that of *–in-*: both consist of a single (high) vowel, followed by a single (nasal) consonant. Clearly, we want to treat *–um-* as a prefix too, for all the same reasons that we want to say –in- is a prefix. Because of these similarities, the same phonological rule that restructures *insulat* to *sinulat* 'was written' can be used to restructure *umsulat* into *sumulat* 'wrote'. The metathesis rule applies not just to one morpheme, but to a class of morphemes that match a certain phonological profile. This is one of our best tests for the presence of phonological allomorphy. It is precisely the same criterion as that says *walkd* becomes *walkt* by a phonological rule because the same alternation is seen with a different morpheme when *catz* becomes *cats*.

Most phonological rules are responses to some kind of phonological pressure, to repair some kind of phonological defect in the underlying form. The voicing assimilation rule in English fits this pattern: *cats* is a better phonological representation than *catz*, because it is bad to have consonant clusters that differ in their voicing (especially if the voiceless one is sandwiched between the voiced consonant and the voiced vowel). Tagalog's metathesis rule also fits this pattern. In this case, *sinulat* is a better phonological representation than *insulat*, because its syllable structure is better. Universally languages like to have syllables that begin with a consonant and end in a vowel. *Insulat* deviates from this ideal in two respects: the first syllable begins with a vowel and ends with a consonant. The metathesis rule repairs both defects, creating *sinulat*, a form with perfect syllables. This too is consistent with the claim that infixation is really a result of phonology.

The clincher comes from looking at forms in which this phonological pressure does not exist. Consider what happens with vowel initial roots like *aral* 'teach' and *ibig* xxx appear in the past tense or the past passive. In this case, the context for the phonological rule is not met: there is no sequence of VCCV, but only an alternating (easily syllabified) series of vowels and consonants. Thus our theory predicts that there should be no infixation in this case: *-in-* and *–um-* should show up as ordinary prefixes. And this is exactly what happens:

(19)	root	past	past passive	
	aral	umaral	inaral	'teach'
	ibig	umibig	inibig	

These forms provide direct justification for our claim that *um* and *in* are really prefixes inherently, and that the factors that cause it to often show up as an infix are phonological ones. [Also add that the Chamorro cognates of these affixes show up as infixes EXCEPT with sonorant-initial roots, in which case they show up as prefixes?]

Other instances of infixation can be handled in the same way. [Discuss punctual aspect in Tuscarora here: it is a glottal stop. It shows up as an infix with consonant final stems, but as a suffix with verb final stems. It is a suffix, plus there is a metathesis rule: C? → ?C (C? being an impossible syllable coda.)

E.2 Reduplication

The previous section can be looked at not only as a solution for the problem of infixation, but as a pattern for dealing with other kinds of exotic word formation as well. The strategy is to treat the exotic processes as ordinary morphemes, prefixes or suffixes, which enter into perfectly normal word structure trees. However, there is something about the phonological representation of the morpheme that triggers some kind of phonological rule, creating the exotic effect. In this section, we will consider how cases of reduplication can be handled in a similar way.

At first glance, reduplication seems harder to fit into the morpheme-based mold than infixation. For infixation, it was relatively easy to isolate two separate morphemes and to assign consistent properties to each; problems only arose when it came time to draw a word structure tree. For reduplication, problems arise even at the first step of isolating a discrete morpheme. The basic data for a simple reduplication process in Tagalog is repeated here:

(xx)	root	future tense	
	takbuh	tatakbuh	'run'
	lakad	lalakad	'walk'
	pili?	pipili?	'choose'
	sulat	susulat	'write'
	hanap	hahanap	'seek'

It is clear where the root is in these examples, and it is clear that new material is being added to the beginning of the root. In this respect, this reduplication seems to be an instance of prefixation. The problem is that what is added seems to be inconsistent: sometimes it is *ta-*, sometimes *pi-*, sometimes *su-*, sometimes *ha-*, depending on the root. Put another way, it seems like Tagalog does not have just one future prefix, but many, one for each root in the language.

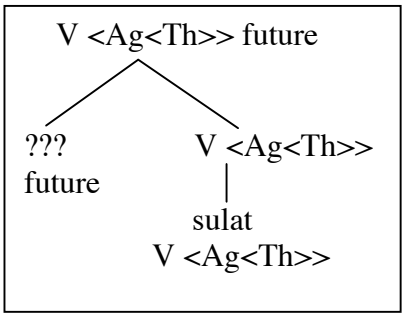
We can pinpoint exactly where the trouble lies by trying to write a lexical entry for the future morpheme and seeing where we get stuck. Most of the entry is rather easy to fill in, as follows:

(xx)

Label/phonology	sulat	??
meaning	write	will X
Attachment properties	root	Prefix to verbs
Features	V, <Agent <Theme>>, present, active	Ø, Ø, future

The meaning of the affix is clear: it is the same as the English future marker "will". Its position is also clear: it attaches to the beginning of a verb stem. Nor are its features problematic: like other inflectional morphemes, it has no category or argument structure. However, it is a realization of the tense feature "future". Given all this, we can draw an ordinary word structure tree for a word like *susulat* 'will write'.

(xx)



The troubles with reduplication arise at one point, and one point only: what is the phonological representation of the morpheme? There seems to be nothing consistent to put in the first cell of the lexical entry.

Although we have localized the problem in one particular part of the lexical entry, it is by no means a small problem. A morpheme is an element that has both a consistent sound and a consistent meaning, by definition. If we cannot isolate a consistent phonological representation that goes along with the meaning of future in Tagalog, our very notion of morpheme is undermined. And there is no consistent phonological representation to be found in this case.

Or is there? While it is true that there is much that changes in the future prefix, there is one point of commonality: the future forms always start with a single consonant and a single vowel. How the first consonant and the first vowel are pronounced varies from form to form, but there is invariably one consonant followed by one vowel. This is true regardless of the syllable structure of the root. When people first see the Tagalog example, they often describe what is happening by saying "you repeat the first syllable of the root" (which would be a kind of word formation rule). However, this is not exactly correct. For a root like *sulat*, the intuitive rule works fine: the first syllable of *sulat* is *su*, and doubling it gives the correct form *susulat*. But the intuitive rule does not work for examples like *takbuh* 'run' that have two consonants in the

middle. This time the first syllable is *tak*, and repeating it would give **taktakbuh*. But the correct form is *tatakbuh*. Apparently, the prefix must consist of no more than one consonant and one vowel, regardless of the syllable structure of the root. This consistency could be the basis for completing our lexical entry for the future morpheme. The phonological representation could be CV—a sequence of a consonant and a vowel—where the exact features of pronunciation of both the consonant and the vowel are left open.

This proposal looks better grounded when we are aware of some aspects of modern phonological representation, which are independently motivated by the attempt to state phonological rules in the simplest and most general way. The basic insight is that instead of stating phonological rules over unstructured bundles of features (phonemes) that are put into a linear sequence, phonological representations need to be factored into (at least) two distinct dimensions. One dimension involves the basic structure of the string in terms of its pattern of consonants and vowels. The second dimension consists of the particular features of pronunciation that realize this pattern. Many phonological rules refer directly to these features, such as [±voice], [dorsal], [±nasal] and [±back]. There is, however, a whole class of rules that crucially refer only to the pattern of consonants (C) or vowels (V), regardless of the features involved.

For example, in one of the assignments we saw a syncope rule in Icelandic that worked this way:

(2) Syncope: $V \rightarrow \emptyset \; / \; \_\_\_ \; C \; V$

[-stress]

This interacted with other cyclic stress rules, and was responsible for alternations such as:

akur ‘acre’

akri ‘acre [DATIVE]’ from: / a k u r / + AFFIX

fifil ‘dandelion’

fifli ‘dandelion [DATIVE]’ from: / f i f i l / + AFFIX

In order for these rules to operate correctly, it is useful to have a representation which is split into two parts: one with all of the features or SEGMENTS, the other with just the SKELETON, i.e., the Cs and Vs:

(3)

‘dandelion’	‘dandelion [DATIVE]’	
f í f i l	f í f i l + i	< Segmental Level (or Tier)
| | | | |	| | | | | |	
C V C V C	C V C V C V	< Skeletal Level (or Tier) < Syncope looks here

The partial independence of these two levels can also be seen in the fact that sometimes the elements of each are not in perfect one-to-one correspondence with each other. For example, a single element of pronunciation can be associated with two different C positions or V positions. The result is a geminate consonant or a long vowel in languages that (unlike English) distinguish these from single consonants and short vowels. Finnish is one such language, as shown by the following minimal pairs:

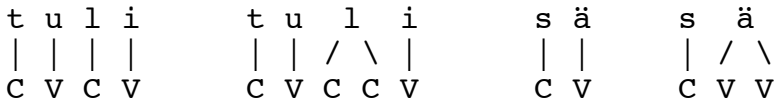
(4) tuli 'fire' tulli 'customs'

kuka 'who' kukka 'flower'

meri 'sea' Meeri 'Mary'

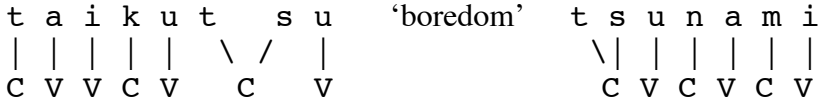
sä 'you' sää 'weather'

These examples correspond to representations like the following:



Essentially, this says that in the Finnish word for 'customs', the +lateral (and other elements of the l pronunciation) lasts for two beats, whereas in the word for 'fire' they only last for one beat. Based on this musical analogy, the CV skeleton is also sometimes known as the timing units, and the segments as the "melody" of the word. Phonological rules stated in terms of Cs and Vs often take long vowels or geminate consonants to behave as sequences of two vowels or two consonants, even though they have only one set of features.

There are also cases in which two distinct aspects of pronunciation can be associated with a single position in the CV skeleton. A single C with two elements of pronunciation is what is traditionally known as an affricate. Similarly, a single V with two elements of pronunciation is (one kind of) diphthong. Affricates like ch and j count for some phonological purposes as if they are a single unit, even though they clearly have two pieces. For example, Japanese does not allow consonant clusters in onset position as a very general feature of the phonology of the language. As a result, English words like "grizzly" are borrowed into Japanese as *gurizuri* since the *gr* and *zl* sequences are impossible in that language. But the ts sequence does seem possible, as in words like taikutsu 'boredom' and tsunami. The reason is because these are affricates: the –cont feature of t and the +con feature of s are both linked to a single C position:



Thus, the Japanese rule that one can never have more than one (nonnasal) consonant in a row is satisfied by these representations, because that rule looks at the CV skeleton tier only. Similarly, Mohawk has only voiceless consonants underlyingly, but it has a phonological rule that voices a

single consonant when it come between two vowels. Thus --- is pronounced -- , whereas *akta* 'near' remains [akta] since the k and the t protect each other from becoming voiced. However, the word for fish is pronounced [kvdzu], not [kvtsu]. Why don't the t and the s protect each other from voicing? Because again they are an affricate, and hence count as only a single consonant, even though they have two elements of pronunciation.

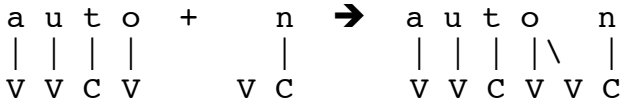
Knowing that phonology is organized in this manner pays dividends in morphology even before we get back to reduplication. Consider the following examples, which illustrate the so-called illative case in Finnish (which means "to X").

(5)

auto 'car'	autoon 'to the car'
koulu 'school'	kouluun 'to school'
päivä 'day'	päivään 'to the day'
kuva 'picture'	kuvaan 'to the picture'
bussi 'bus'	bussiin 'to the bus'

The question is, what is the phonological representation of the illative case suffix? It is easy to see the pattern: illative case is indicated by the last vowel of the stem, followed by the consonant *n*. By distinguishing the segments from the CV skeleton and allowing multiple-linking, we can easily diagram this as follows:

(6)



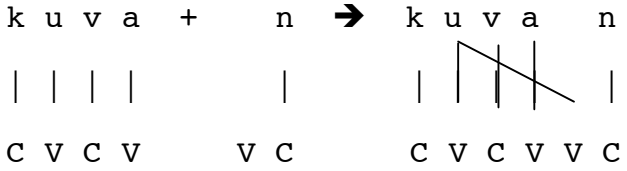
More generally, the phonological representation of the illative suffix will be Vn, where the empty vowel slot gets filled in by borrowing the pronunciation of the last vowel of the stem. The fact that is must be the last vowel that fills in is crucial; otherwise we could get incorrect forms like *kuvaun 'to the picture' or *bussiun 'to the bus'. In general, we want only the closest vowel pronunciation to fill the empty V slot. This can be enforced by adding the following convention, which restricts the kinds of relationships that one can have between the segments and the CV skeleton:

Important: The NO CROSSING LINES principle.

While many-to-one and one-to-many mappings are possible, the lines which link the segmental tier to the skeletal tier MAY NEVER CROSS.

All of the desirable representations that we have seen so far obey this principle, but the illegitimate forms like *kuvaun do not.

(7)



The no crossing lines principle is thus an essential feature of these types of representations. We shall see that this principle, more than anything single aspect, is what makes the morpheme-based account more restrictive than a theory that uses word formation rules. It will be the cornerstone of what enables us to explain why not every regularly describable pattern can be a legitimate morphological process.

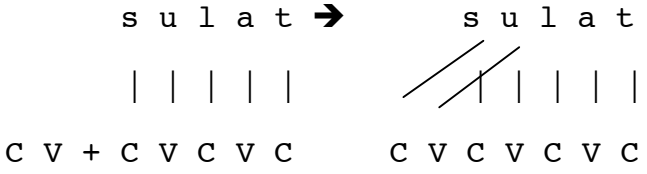
Now we are ready to come back to reduplication in Tagalog, examples of which are repeated here as (9):

(8)	takbuh	tatakbuh	'run'
	lakad	lalakad	'walk'
	pili?	pipili?	'choose'
	sulat	susulat	'write'
	hanap	hahanap	'seek'

Recall that what we wanted to say was that the future morpheme in this language is a prefix with a phonological representation something like CV, where the specific pronunciations of both the C and the V are somehow borrowed from the root. Given what we have seen in Finnish, this does not look so strange any more. First, we know that sequences of Cs and Vs do exist as legitimate phonological representations, quasi-independently of any segmental tier. The Tagalog prefix is precisely this: a piece of a CV skeleton. Furthermore, the Finnish illative affix shows that it is possible for a morpheme to have positions in its CV tier that are not inherently specified for any pronunciation, but that get filled in with the pronunciations from the root. This is just what we want to have happen in Tagalog too, except that Tagalog is a more extreme case: in Finnish only one element in the affix gets filled in with pronunciations, whereas in Tagalog every element in the affix needs to be filled in in this way. There is a kind of continuum here, from English prefixes like *re-* where the pronunciations of the affix are fully specified, through Finnish –Vn where they are partially specified, to Tagalog CV- where they are not specified at all.

Whereas this solves the fundamental morphological problem of reduplication in Tagalog, there is more to say about the phonological process by which future tense verbs get pronounced. We cannot simply do the linking we want directly, the way we did in Finnish, because this would violate the No-Crossing Lines Constraint:

(9)



Thus, we need to state a phonological rule that will help to interpret this type of representation. We can state the rule in three steps as follows:

1. Copy the segments of the entire root over the affix.
2. Link the segments to the skeleton one to one, left to right.
3. Delete all unlinked / leftover elements.

Consider how this applies to an example like *susulat* 'will write'. The first step of copying segments gives the following:

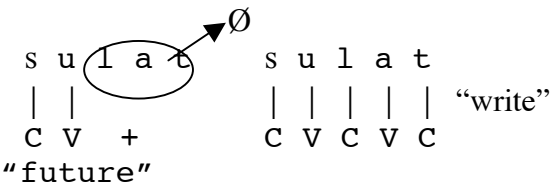
```
 s u l a t     s u l a t
               | | | | | "write"
 C V  +        C V C V C
"future"
```

Next we can link the copied segments to the empty spots in the skeleton one to one, left to right, as follows:

```
 s u l a t     s u l a t
  | |          | | | | | "write"
 C V  +        C V C V C
"future"
```

Notice that crossing lines is no longer a problem, because of the copying. However, the simple copying rule did not know how much material would be needed to fill in the affix, and as a result it copied more segments than we need. Therefore, the last step is to apply the convention that leftover segments are deleted (or at least not pronounced):

(10)



This correctly gives us *susulat*, as the future tense for *sulat*. Exactly the same sequence of steps applied to *takbuh* 'run' will give us *tatakbuh* 'will run', as the reader can easily check.

Whereas the copying and deleting parts of the reduplication rule are quite straightforward, the linking step contains some "fine print" that is necessary to ensure that we get the correct forms in all cases. First, it is stipulated that linking goes from left to right, not from right to left. The need for this is obvious. If we did not add this requirement, then the future tense of basa could be *sabasa, rather than the correct form babasa.

(11)

```
 b a s a       b a s a
  | |          | | | |
```

```
 C V  +       C V C V
"future"
```

The other stipulation is that the linking must start with the segments and match them to the CV-skeleton, rather than the other way around. Which way you do it does not make any crucial difference in the examples that we have seen so far. However, the need becomes evident in cases where the CV structure of the root and the CV structure of the affix are significantly different. To see this, consider the future tense form of a vowel-initial root like *aral* 'to teach.' The representation immediately prior to linking looks like this:

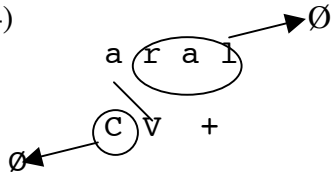
(12)
```
      a r a l   a r a l
               | | | |   "teach"
 C V  +        V C V C
"future"
```

Here it is not so obvious how to begin, because the first segment is a vowel-type pronunciation which does not match the first element of the CV skeleton. Therefore, one of these elements must take priority over the other. Suppose that we insisted that the first consonant position must be filled first. The first available pronunciation is the r. Now the initial a cannot link to the V position, because that would create a crossing line. Therefore the second a must link here, giving the following representation:

(13)
```
 a r a l      a r a l
  \ /         | | | |   "teach"
 C V  +       V C V C
"future"
```

After the leftovers are deleted, this would give us raaral as the future tense of 'teach'—which is not correct. Suppose then that we make the opposite assumption, saying that the linking must start with the first element of pronunciation a. This can only associate with the V position of the affix, as follows:

(14)




```
              a r a l
              | | | |   "teach"
              V C V C
```

Now, no other linking is possible. In particular, the r pronunciation cannot link to the first C of the future affix, because this would violate the no crossing lines principle. (Otherwise we would get *raaral again.) Neither, obviously, can the l. Thus, the initial C remains unlinked, and falls prey to the deletion convention, as do the last three elements in the pronunciation. This way of doing things, then, predicts that the future tense form of 'teach' will be *aaral*. This is not quite

right either: the correct form is *a?aral*. However, this time we were almost right. The fault is easily fixed with a straightfoward rule of consonant epenthesis, which can be stated as follows:

Extra Phonology: Ø → ʔ / V_V

This, then, illustrates the importance of beginning the linking with the first segment, rather than with the first member of the CV skeleton. With this fine print in place, we now correctly account for all of the Tagalog forms with a single phonological representation for the future morpheme together with a kind of phonological rule.

E. Varieties of Reduplication

Tagalog is certainly not the only language in the world that has reduplication; indeed, this is a relatively common process, as exotic word formations go. In this section, we will see how the concepts developed for Tagalog in the previous section can account for this variety. No major changes in the approach will be needed; all that is necessary is to discover the correct underlying form of the affix in terms of Cs and Vs. By the end of the section, we will be in a position to see another, deeper pattern here as well that underlies which kinds of CV templates are used for reduplication in natural languages and which kinds are not.

As a first example, consider the following morphological paradigm for perfective verbs in Ancient Greek.

(15) Greek:

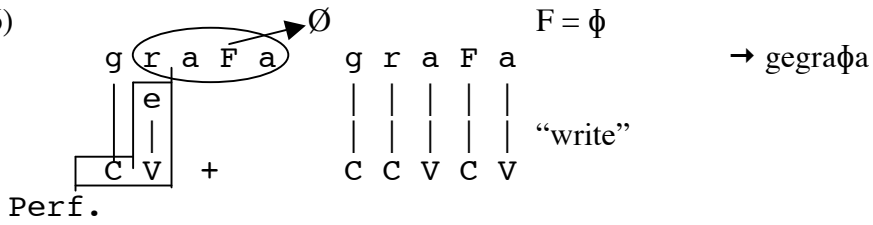
Verb Root	Perfective	
thɛɛka	tethɛɛka	'place'
klɛɛka	keklɛɛka	'call'
luka	leluka	'lose'
bouleka	bebouleka	'consider'
grapha	gegrapha	'write'

The pattern should be quite clear. The examples are like Tagalog in as much as the verb root is the same as the perfective form minus the initial consonant and vowel. Thus, the perfective morpheme in Greek must be CV prefix that subcategorizes for a verb root and bears the aspect feature "perfective" (whatever that is). As in Tagalog, the initial consonant of the added syllable is a copy of the initial consonant of the original root. The difference is that the vowel is not copied from the root. Rather, the vowel of the prefix is always "e", regardless of whether the verb root begins with E, u, o, or a. The way to express this is to say that the Greek perfective prefix is a CV skeleton with the C left open (as in Tagalog), but with the V filled by the segment e, as shown in xx.

Perfective: e-
|
C V

The copying, linking, and deleting aspects of the reduplication rule then apply in a straightforward way. The first segment of the copied melody links to the C position at the beginning of the affix. There is no other open position in the affix, so linking stops there. The rest of the segments copied from the root are not pronounced, giving us gegraFa, as desired.

(16)



This example is similar in many ways to the example of –Vn suffixation in Finnish. In both cases the affix has a partial specification, but is also partly left open, to be filled in by elements borrowed from the root. It reconfirms that there is something of a continuum between conventional affixation and reduplication. The difference is that in Finnish the consonant of the affix was specified and the vowel was left open, whereas it is the other way around in Greek.

Consider next the following pairs of nouns in Agta, a language of the Philippines, related to Tagalog.

'body'	bari	barbari	'whole body'
'leak'	saddu	sadsaddu	'leak in many places'
'lost'	wakay	wakwakay	'many things lost'
'leg'	takki	taktakki	'legs'
'head'	ulu	ululu	'heads'

Once again, there is a new syllable at the beginning of the pluralized word, constructed from the phonological materials present in the root. But a bit more is copied in Agta than in Tagalog. Tagalog copies only the first CV of the word, but in Agta the second consonant is copied as well. Thus the plural prefix in Agta must have the shape CVC—a bigger affix than the Tagalog one. (xx) sketches out how the reduplication rule applies to a typical example.

Diagram

The last example, ululu 'heads' seems like a counterexample to the general pattern, because only a VC is copied, not a CVC. Thus, one might be tempted to revise the phonological representation of the Agta plural prefix to say that the first consonant is optional, notated perhaps by putting the first C in parentheses. Then the phonological representation in the lexical entry of the plural prefix would be (C)VC. But in fact there is no need for this revision. The correct pronunciation for ululu already follows from assuming that the prefix is CVC together with the reduplication rule as given. It is another illustration of the importance of the linking starting with the segments and matching them to the CV skeleton, rather than the other way around. In this

case, the first segment copied from the root happens to be u, a vowel-type pronunciation. This cannot link to the initial C of the prefix, so it links to the V instead. Next we link the "l" segment. This is qualified to link to the first C of the prefix, but doing so would violate the no-crossing lines rule. Hence this "l" links to the second C of the prefix instead. That is all the linking that can take place, and the leftovers are not pronounced—including the first C of the root.

We don't have to say explicitly that this C is optional; rather it follows from the theory that there is no way to pronounce it in this particular example. Indeed, it would be wrong to say that the first C position is optional, as a property of the affix. It is a fact about the root ulu that it does not begin with a consonant pronunciation that drives this example, not a fact about the plural morpheme. If we said that the C of the plural prefix were optional, that would imply that the plural of 'legs' could be aktakki as well as taktakki (or maybe ttakki, depending on how one defined the rules). And that is false. In general one wants to keep the basic representation of the reduplicative morpheme simple and straightforward and leave it to the phonological rules to handle just how it is realized in specific cases.

The Agta example also shows once again the advantages of the morpheme-based approach as opposed to a word formation rule like "repeat the first syllable of the root". The syllable based rule works fine for saying that the plural of takki is taktakki. But it fails when applied to bari 'body'. The first syllable of bari is ba, so the syllable-copy rule would expect the plural to be babari. In fact, it is barbari. So in some cases in Agta, one copies more than the first syllable of the root, just as in Tagalog one sometimes copies less than the first syllable of the root. It is not the syllable patterns of the root that guide the extent of the copying in both languages, but rather the needs of the affix. One copies enough to fill in just a CV in Tagalog, even if that is less than the first syllable of the root; one copies enough to fill in a CVC in Agta, even if that is more than the first syllable of the word. The contrast between these two cases shows clearly that the affix has a phonological consistency in both languages that is independent of the properties of the root. And that is exactly what we said must be the case given the definition of morpheme as something with a substantive—although not fully specified—phonological representation.

Consider next the reduplication found in Chukchee, a language spoken in Siberia.

PLAIN	ABSOLUTIVE
jil?e 'gopher'	jil?ejil
nute 'ground'	nutenut
inu 'reindeer leg'	inuin

(add two or three more)

This clearly differs from the other examples that we have seen so far in that the action here is at the end of the word, rather than at the beginning. This is not so obvious in examples like (xx), in which the entire root is reduplicated; in this case it is hard to tell which of the two is the original and which is the copy. (xxc) is unambiguous, however. There is more phonological material associated with the first xxx than with the second one; therefore the second one must be the copy. This then is our first case of a reduplicative suffix. Apparent from this minor innovation, reduplication in Chukchee is identical to reduplication in Agta. In particular, the most that gets copied is a consonant-vowel-consonant series, implying that the lexical entry is –CVC. Here is the complete lexical entry, and the derivation of a typical example:

Practice exercise. Note that the absolutive singular form of inu ‘land’ is inuin. Explain why, and compare this to the examples seen above in Tagalog and Agta.

The last illustrative example that we will give here is from the Australian language Rembarrnga. Consider the following examples:

	MEANING	BASIC FORM	"be X-ing" FORM
(1)	go	riya	riyariya
(2)	forget	winan	winawinan
(3)	hear	Nawaniñ	NawaNawaniñ
(4)	leave	papiñ	papipapiñ
(5)	descend	toto?miñ	totototo?miñ
(6)	lack	tumpal?miñ	tumpatumpal?miñ
(7)	rub	palarmiñ	palapalarmiñ
(8)	smell	popnan	popnapopnan
(9)	finish	tey?puwa	tey?putey?puwa
(10)	follow	wawiñ	wawiwawiñ
(11)	catch	palkmiya	palkmipalkmiya

Here we are back into the domain of prefixation, because when one of the two copies is longer than the other it is always the second copy that is longer—like Agta but not like Chukchee. But this time the prefix is bigger even then the Agta prefix. In Rembarrnga one copies roughly two syllables, rather than just one, as shown in xxx.

But one does not copy exactly two syllables. The first two syllables of xxx are xx and xx, but the plural of xxx is not xxxx, but rather xxx. One copies slightly less than two syllables here. The case is very much like CV reduplication in Tagalog, except that the cut off point is partway through the second syllable in Rembarrnga, whereas it is in the first syllable in Tagalog. We get the correct patterns if we say that the plural prefix in Rembarrnga has the phonological representation CVCCCV-. Here is its complete lexical entry, and a representative derivation.

Practice exercise: (1) Why must the affix be CVCCCV-? Why couldn't it be simply CVCCV-? Or even CVC(C)(C)V?

(2) What about the following data? How can they be accomodated?

(12)	get	ma	mamama
(13)	see	na	nanana

This completes our survey of reduplication processes in languages of the world. We have certainly not exhausted the kinds of reduplication. But these are some of the most common types. Most other reduplications in the world are a minor variation on one of these.

And that is somewhat remarkable, if you think about it. One of our goals in this chapter—and indeed in the whole book—is to seek a theory of morphology that captures the restrictions on what kinds of word formation one finds in natural languages as well as the richness and diversity. That is why we opt for underspecified morphemes, copying rules, and the no-crossing lines principle rather than word formation rules. Such a theory can explain why there are reduplicative formations in languages of the world, but not reversive ones that give sulattalus from sulat. But here we have a hint that reduplication is even more restrictive than the CV skeleton theory would expect. Here is a summary of the kinds of reduplicative morphemes that we have seen so far:

CV-	Tagalog, Greek
CVC-	Agta, Chukchee
CVCCCV-	Rembarrnga

Several of these we have already seen more than once, even in our small sample. But of course there are many other types of morphemes that could in principle be defined in terms of sequences of Cs and Vs. Here are some generalizations:

(i) Reduplication is never less than a single syllable.

There are no morphemes like C or CC. (compare –s in English)

(ii) Reduplication is never more than two syllables (unless it is total word reduplication)

There are no morphemes like CVCVCV-

(iii) Reduplications are easily syllabified.

There are no morphemes like CVCCC- (compare Rembarrnga)

(iv) Reduplications don't start with a vowel.

There are no morphemes like VC

The first three statements should be self-explanatory, but the fourth is worth lingering on a bit. Obviously in terms of CV sequences, a VC morpheme is no more complex or mysterious than a CV morpheme is. And it could give well-formed results. When attached to a vowel initial root, it will give the same results as the attested reduplication in Agta.

U l ulu

Things wouldn't go as smoothly if such a morpheme were attached to a consonant initial root like bari. Here the initial segment of the root could not attach to the V at the beginning of the affix. It would have to attach to the C slot instead, and no other linking would be possible. (This is the exact reverse of the derivation of a'aral in Tagalog, xx, with Cs in the place of Vs and vice versa). So we would get.something like bbari.

Now this does not make well-formed syllables. But that should not necessarily be a fatal objection, meaning that such a morpheme could not exist. One might expect that a phonological repair would happen, perhaps epenthesizing a neutral vowel (a schwa perhaps, or I) between the two consonants to give bEbari. Again, this is just like what happens in a'aral in Tagalog. Or the worst case scenario, if there is no repair, is that such a morpheme would be limited to attaching to vowel initial roots, and some other allomorph would be selected for consonant initial roots.

We know that this kind of allomorphy is abundant in languages of the world; see xxx for an example in Russian. So the absence of VC morphemes counts as mysterious within the terms that we have seen so far.

Another fact relevant to (iv) comes from loan words in Tagalog and Agta. These languages did not originally have words that began with consonant clusters; words could begin with zero consonants or one consonant, but not with more than one consonant. But in modern times words that begin with more than one consonant were borrowed into Tagalog from Spanish and English; an example is trabajar 'work'. Now suppose that a Tagalog speaker wanted to use this brand new word in the future. Taking our theory so far very literally, they would have a problem. The first t would link to the C position, as usual, but the second segment r, would have no place to link. Linking should then stop, leaving the other segments unpronounced.

The result should be ttrabajar. Perhaps the unwieldy beginning would then be repaired by epenthesis to something like titrabajar, or perhaps Tagalog speakers would decide that this verb cannot be used in the future. But neither is what happens. Rather, Tagalog speakers unhesitatingly produce the future form tratrabajar. This is exactly the result one would expect if the prefix were CCV-, not CV- as we assumed before. Now trabajar works fine:

Examples with one consonant—or none—also work fine; the extra consonant(s) simply go unlinked, and so are automatically not pronounced by general principles.

But this gives a paradoxical result. It suggests that the Tagalog prefix was CCV- all along … even though no Tagalog speaker ever had experience with a CC cluster before contact with the Spanish. And that is absurd. Rather, what seems to be going on is that this observation should be combined with the one in the previous paragraph into the following generalization, a revision of (iv).

One syllable reduplication morphemes do not care how many consonants come

before the vowel.

They do care how many consonants come after the vowel.

So a-, ba-, and tra- are all possible realizations of the future in Tagalog. And if it borrowed splater, then we would expect splasplater. But ar- is impossible (*araral from 'aral' teach) and so is xx. In contrast, ul-, bar-, and trab- are possible reduplications in Agta, but ba- and wakw- are not. Something similar can be seen in the Rembarrnga reduplication morpheme that we gave above as CVCCCV-. This does not contrast with reduplication in other Australian languages which is only CVCCV- or even CVCV- and hence has problems with examples like xxx. If an Australian language allows xxx at all, then its reduplication will be something like xxx (not xxx,

or xxx). The number of consonants between the two vowels is not crucial; the number of consonants at the end is.

Why should the possible reduplication morphemes be constrained in this way? The answer (due to McCarthy and Prince) seems to be that reduplication morphemes make use of exactly the same phonological resources as stress systems do. Recall the English stress rule, which we discussed in xxx. We stated it as follows:

(Not counting the last syllable in nouns and derived adjectives…)

Stress the last syllable if it is heavy;

Stress the second to last syllable if the last syllable is light.

Light syllables=syllables that end with a short vowel.

Heavy syllables=syllables that end with a long vowel or a consonant.

(See also Palestinian Arabic, if we use that one.) Notice that the concepts used in this rule are strikingly like those used in reduplication. The difference between heavy and light syllables is important in stress systems, and this is defined by whether the syllable ends in a consonant or not. How many consonants come at the beginning of a syllable is irrelevant to the placement of stress. This parallels exactly the fact that reduplication in Philippine languages cares about consonants after the vowel—whether a heavy or light syllable is created—but not about consonants before the vowel. Second, the English stress rule pays attention to two-syllable sequences, particularly when the last one is light, but never to three-syllable sequences. (Stress is never on the third to last syllable if the last two syllables are light, for example.) Similarly, reduplication in Rembarrnga is a two syllable sequence—where the last one needs to be light, in fact—but never a three syllable sequence. And trivially the smallest unit that can bear stress in English is a syllable, which must contain some kind of vowel. In the same way, the smallest unit that can be a reduplication is a syllable, which minimally contains a vowel. The parallels between stress systems and reduplication are thus detailed and exact.

We can capture this parallelism by revising our theory as follows. Reduplication morphemes are not arbitrary sequences of Cs and Vs; rather, they are syllable templates. The vocabulary of these syllable templates is independently motivated in the theory of stress assignment. We then revise our list of reduplicating morphemes as follows:

Single light syllable: Tagalog, Greek

Single heavy syllable: Agta, Chukchee

Syllable-light syllable unit: Rembarrnga [note: i.e. a trochaic foot]

These are approximately the only kinds of reduplication possible, because they already exploit the full vocabulary of stress systems.

This revision actually leads to an interesting new prediction. The expression "light syllable" is almost identical to our original characterization of CV. But the expression "heavy syllable" is not identical to CVC. Having a consonant after the vowel is one way to make a syllable heavy, but having a long vowel is a second way, treated as equivalent by many stress systems (see the English rule). For Agta and Chukchee, the point is moot, because these languages do not have a long vowel-short vowel contrast. But suppose that there was another language that had long vowels, closed syllables, and heavy syllable reduplication. Then we would expect that the reduplicated syllable could contain either a long vowel or a final consonant, but never a short vowel. This is exactly what we see in Mokilese:

Wadek	wadwadek	read
PilOd	pilpilOd	pick breadfruit
Poki	pokpoki	beat
DOp^{w}O	dOp^{w}dOp^{w}O	pull
KookO	kookookO	grind coconut
SOOrOk	sOOsOOrOk	tear
Caak	CaaCaak	bend
Pa	paapa	weave

Notice that there is no good way to state what all the reduplicating forms have in common in terms of a CV template; in some cases, the prefix is CVC- and in others it is CVV-, two quite distinct skeletal representations. But there is a very easy way to state what all the reduplicating forms have in common in terms of stress-inspired syllable talk: they are all heavy syllables. This then is a spectacular success story for the revised theory of reduplication morphemes.

We will not discuss how exactly to fill in all the details of this syllable-talk, or how to revise the linking rules to make use of them. Doing this properly requires more phonological sophistication than we assume all of our readers will have. But two morphological points should be clear. First and foremost, there is still some consistent piece of phonological representation that can be associated with each meaning or each new syntactic feature, in support of the definition of the morpheme. Second, it is right to pursue a restrictive theory, in which not every regular transformation of a root is a possible morphological operation. Apart from Mokilese, the real weakness of our initial CV theory wasn't that it was too restrictive, but that it was not restrictive enough.

E. Mutation and Ablaut

We turn now to another kind of exotic word formation that seems at first glance quite different from reduplication. Consider the following data from Nivkh, a language spoken on Sakhalin Island, in the Russian Far East ɖʲ = palatalized /d/

Intransitive	Transitive	
tʌŋzʌlʌɖ	rʌŋzʌlʌɖ	'weigh (s.th.)'
q^havuɖ	χavuɖ	'warm (s.th.) up'
kesqoɖ	ɣesqoɖ	'burn something/oneself'
pʌkzɖ	vʌkzɖ	'lose something / get lost'

There is a systematic difference between the words in the first column and the related words in the second column. Their meanings are, of course, very similar; the difference is simply one of argument structure. The words in the first column are intransitives, having an argument structure something like <Th>, whereas the words in the second column are transitives, with the argument structure <Ag <Th>>. This is the kind of difference that is often produced by affixation; see the analysis of causative affixes described in chapter Arg. The sounds are also very similar; the only difference in pronunciation is in the first segment of the word. More specifically, the words in the first column all begin with a stop consonant, whereas the words in the second column begin with a continuant consonant made at the same point of articulation. Thus, to burn oneself starts with k, a velar stop, whereas 'to burn something' begins with a velar fricative x. To get lost begins with the labial stop p; to lose something begins with the labial fricative v. The morphological problem here is that it seems like nothing is added to the roots in the first column to form the words in the second column. Put another way, there is no consistent piece of phonological representation with which to associate the argument structure <Ag …>. Nothing is added to the second column, even though something is definitely changed. As usual this can easily be stated as a word formation rule:

> To make the transitive form of the verb, change the first consonant of the root into a +continuant sound. (Nivkh)

But that is not surprising, since any regularity can be stated as a word formation rule. The question is whether there is also a way to capture this kind of pattern using a more restrictive theory of affixation.

Again, it pays to imagine what a lexical entry for the morpheme would be like if there were one, to pinpoint exactly the problem. The meaning and syntactic features of the morpheme are clear; it would mean roughly 'cause to' with the syntactic features of <Ag <X>>. The attachment properties are also clear: the morpheme would be a prefix that attaches to verbs. We say it is a prefix, because the only other choices are root or suffix. It is not a root, because it cannot form the basis of a word on its own, and we would say it is a prefix rather than a suffix because clearly the action is taking place at the beginning of the word. So we imagine an ordinary-looking word structure tree like the one in (xx).

The problem, once again, is in the phonological representation. It is not clear what to put in the phonological representation box of the lexical entry.

Here again the idea that phonological representations are segmented into two parts or "tiers" proves helpful. The CV pattern of the word and its series of substantive phonological features are two distinct aspects of a phonological representation, and certain mismatches between the two are tolerated. Reduplication is the result of one such mismatch: reduplicative morphemes are CV skeletons with no pronunciations of their own (or equivalent syllable templates). So they have to "borrow" pronunciations from the stem. The Nivkh example can be seen as the exact opposite. It is true that no CV skeleton material added; the words in the second column have exactly the same sequence of consonants and vowels, and exactly the same syllable structure as the words in the first column. But it is possible to have a phonological representation without a CV skeleton. If a CV skeleton without phonological features is legitimate, then so should phonological features without a CV skeleton. And there is a phonological feature that is present in the second column of (xx): the feature [+continuant]. Suppose then that we complete our lexical entry of the transitivizing morpheme as follows:

(17)
[+continuant]
"transitive"
prefix

Now we need a phonological rule that can interpret a representation like the one in (xx).

(18)

	"k"	"e"	…
	-continuant -aspirated velar	+vocalic -back -high	
[+continuant]	C	V	

In fact, approximately the same linking rule that we used in reduplication will work here too. [+continuant] is a consonantal element of pronunciation, so it links to the first available C position.

	"k"	"e"	…
	-continuant -aspirated velar	+vocalic -back -high	
[+continuant]	C	V	

The "no crossing lines" principle that was essential to reduplication also plays an important role here. It ensures that it is always the first consonant of the stem that becomes a fricative; it one tried to make it the second consonant instead, a line would be crossed.

pʌkzɖ → vʌkzɖ , not *pʊxzɖ 'lose something / get lost'

representation

The one difference is that in these cases the floating features associated with the affix can link to a C position in spite of the fact that it already has a full set of features associated with it—something that is avoided in reduplication. (See the analysis of Greek perfective in (xx). The vowel pronunciation copied from the stem does not "barge in" on the features of [e], already associated with the V of the affix.) This can give rise to contradictory feature specifications. For example, the first C in (x) has both [+continuant] and [-continuant] associated with it, a contradiction. The convention that applies in this case is as follows:

(a) Features of the stem that directly contradict features of the affix are deleted.

(b) Other features of the stem are readjusted minimally so as to create a segment that is well-formed within the phonology of the language in question.

For example, in the case of pʌkzd̥ → vʌkzd̥, -continuant feature of the initial [p] is deleted because it is incompatible with the feature of the affix. The result is a segment that is a continuant, unaspirated, and bilabial. Technically, this should be a [□]. [□] is not a possible consonant in Nivkh, but [v] is a minimally different sound that is (it is a continuant, unaspirated, and labiodental). So the exact place of articulation is adjusted slightly in accordance with (b) to give the final result. As you can imagine, this last step requires knowing the details of phonological feature theory and how it is realized in particular languages to do right. As usual, we gloss over these phonological details to focus on the morphological aspects of the issue.

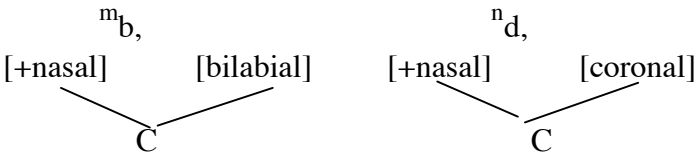
[Note: Another such detail concerns voicing in Nivkh. An aspirated stop becomes a voiceless continuant, whereas an unaspirated voiceless stop becomes a voiced continuant. This makes more sense when one knows more about laryngeal features (which includes both aspiration and voicing). One should think of the voicing alternations in terms of a feature like "aspiration" (really, Voice Onset Time). For stops, +aspirated and –aspirated have their respective values, while for fricatives, +aspirated means voiceless and –aspirated means voiced.]

A second example of this kind comes from the Bantu language Kihehe. The following data compares verb roots with verbs that are inflected for 1st person singular (subject or object).

ku-vee^{m}bela	'to mourn'	kuu-^{m}bee^{m}bela	'to mourn me'
ku-tesa	'to hurt'	kuu-nesa	'to hurt me'
ku-lima	'to cultivate'	^{n}dima	'I'll cultivate'
ku-teleka	'to cook'	neleke	'May I cook?'

Notice that this language contains "prenasalized" stops, written as ^{m}b and ^{n}d. These are single consonants that are in a sense blends of m and b or n and d; they start out nasalized and turn

unnasalized prior to being released. They can be analyzed the same way we analyzed affricates in section xxx: they consist of two sets of features that share a single "C" on the skeletal tier.



Add +voice, -nasal

Now with this in mind, what is the difference between the two columns in (xx)? They have exactly the same number of C and V positions, but the examples in the second column all have a +nasal feature that is not present in the first column. So this is like Nivkh, except that a different feature is added. The lexical entry and a simple word structure tree would be:

The nasal feature then links to the first C of the stem by (a), producing roughly the correct forms.

Then comes the fine-tuning (cf. (b)), which we only sketch enough of here to convince you that the details can be filled in. This fine-tuning needs to account for the fact that the nasalized version of a voiced fricative is a voiced prenasalized stop, and the nasalized version of a voiceless stop is a simple nasal. In order to understand these complexities, one needs to know the following general facts:

There are no simple voiced stops (b, d, g) in Kihehe, only continuants (v, l, g).

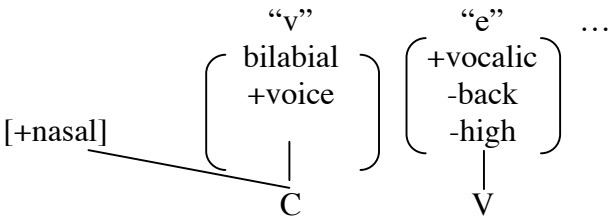
There is no such thing as a prenasalized fricative.

There is no such thing as a voiceless prenasalized stop.

The first of these statements is specific to Kihehe; you can see that it is true in the tiny fragment of data in (xx). The second two are more generally true in languages that allow prenasalization at all. It makes sense that a single consonant cannot tolerate too many contradictory feature specifications. Affricates in English, for example, can start out –cont and end +cont, but both parts must have the same voicing specification and place of articulation, so [ts] and [dz] are common sounds in languages of the world, but not [tz] or [ds]. It is the same with prenasalized stops: they can start out +nasal and end up –nasal, but all of the other features must be the same. Now nasals are (almost) always voiced stops in languages of the world. It then follows that the second half of the consonant must be voiced and a stop too. Thus, one can have mb and nd, but not mv or nl, and not mp or nt. More specifically, in a form like kuu-^{m}bee^{m}bela, the underlying form is:

Kuu- [+nasal] beembela

If the [+nasal] 1st person morpheme weren't there, then [b] would turn to [v] by a general rule. But when [+nasal] is present, it links to the [b], blocking the fricative-creating rule for the reasons discussed above. The specifications of [+nasal] and [-nasal] in a single C are not considered a contradiction in the phonological system of Kihehe, so the result is a prenasalized stop.



The first steps are the same in a form like kuu-nesa, but [+nasal] and [-nasal] is considered a contradiction in the context of a segment that is underlyingly [-voiced]. So a prenasalized stop cannot surface in this case, and the representation is reinterpreted as [n]—the simple sound in Kihehe that most resembles the original [t], but that incorporates the new [+nasal] feature.

Note that we would not be able to expect students to work through all the details of such an account for themselves without knowing more about phonology in general and Kihehe in particular. But we could expect them to recognize from the data that the 1st person singular morpheme is some kind of +nasal prefix that has no CV skeleton of its own.

In Nivkh and Kihehe, a floating consonantal feature is added to the representation. There is no general, widely-agreed on name for this kind of morphology, on a par with the term "reduplication". Rather, these processes have been given particular names in the traditions of particular languages. We will refer to this as "mutation", taking the name from the morphology of Celtic languages, which have morphological phenomenon that are quite similar to those we saw in Nivkh and Kihehe (although not, perhaps, quite as regular).

In principle, one could also imagine morphemes that consist of a single vowel feature that is not linked to any CV skeleton. Something like this seems to be at work in a corner of English, as shown in the following data:

Singular		Plural	
tuw	tooth	tiy	teeth
fUt	foot	fiyt	feet
guws	goose	giys	geese
maws	mouse	mays	mice
laws	louse	lays	lice

These are, of course, irregular plurals in English, and there are only a small number of them. House, for example, rhymes with mouse and louse, but its plural is houses, not *hice. But there is a regularity in these irregulars. The singular forms all contain some kind of back vowel, either a high back vowel (u, U), or a diphthong with a low vowel and a back glide (aw). The plural forms have the same CV-structure, but they contain the corresponding front vowel, either the

high front vowel (i), or a diphthong with a low vowel and a front glide (ay). This could be thought of in the following way: English has an allomorph of the plural affix which consists of a single vowel feature [-back]. This forms an ordinary word structure tree, and then links to the closest available V slot so that it can be pronounced, creating a front vowel. (We cannot really tell if the affix is a prefix or a suffix in this case, because all of the words concerned are only one syllable, so the vowel is equally close to the front of the word and to the back. Other allomorphs of the plural are suffixes in English, however, and it does no harm to take this one to be a suffix too.) The traditional name for vowel changes of this kind is "ablaut". Because the pattern in English is so unproductive, this might not count as a real case of exotic morphology.[1] But at least it gives us an idea of what this kind of morphology, potentially allowed by our theoretical assumptions, would look like.

E. HAVING IT ALL: ROOT AND PATTERN MORPHOLOGY

So far, we have seen how several kinds of apparently exotic word formation can be reduced to cases of ordinary affixation once we adopt a somewhat more sophisticated view of phonological representation (which is needed anyway). Reduplication then falls into place as the result of morphemes that have more CV structure than phonological features; mutation and affixation of various kinds fall into place as being morphemes with more phonological features than CV structure. Nevertheless, all of the languages that we have seen so far are fairly normal, in that the majority of their morphemes are still fully specified, consisting of a complete array of CV skeleton/syllable structure and features/pronunciations. It is natural to wonder, however, whether this is always the case. Could there be a more radical language, which took this opportunity to the limit, a language in which many, even most morphemes were pure CV skeleton or pure features of pronunciation? If so, what would such a language look like?

The answer is that such a language would look a lot like Arabic. Here is a set of verb forms from Classical Arabic (the language of the Koran); all of them are forms with a third person singular masculine subject.

---

[1] The alternative would be to treat these cases as a type of stem allomorphy. The plural affix takes the form of Ø after /MOUSE/ and the stem takes the allomorph of /mays/ before the plural. We know that there is a Ø form of the plural affix in English anyway, apart from these forms, because we see it in sheep-sheep, fish-fish, deer-deer, moose-moose. We also know that other cases of ablaut in the verb system must be handled by stem allomorphy, both because the vowel change is not systematic and because it happens in the presence of a regular tense ending: freeze-froze+n, tell-tol+d, etc. [Jonathan: whether this kind of footnote is enough or whether we want a fuller text discussion—maybe its own section—will depend on what we say elsewhere, eg. In the allomorphy chapter.

(19) Classical Arabic

	"write"		"earn"	
	ACTIVE	PASSIVE	ACTIVE	PASSIVE
present (X-s)	katab	kutib	daraj	durij
cause to X	kattab	kuttib	darraj	durrij
X for ea. other	kaatab	kuutib	daaraj	duurij
make X	?aktab	?uktib	?adraj	?udrij

Now it is very difficult to segment these words into morphemes in the traditional way—especially if we are stuck on the idea that morphemes are *contiguous* sets of sounds (sounds that are right next to each other). But our definition of morpheme does not require this; it only requires that each morpheme have a consistent phonological representation of some kind. And by now we are used to the idea that these representations can be fairly abstract. So let us try to segment the morphemes in the Arabic data without our contiguity blinders on.

What do all the forms that have the meaning 'write' share in terms of phonology, that distinguishes write-type words from (say) 'earn' type words? Just this: all of the words that containing the meaning 'write' contain that sounds k, t, and b, *in that order*. Wherever you have k-t-b, you have a meaning with something to do with writing; wherever you don't have those sounds, you don't have that meaning. Crucially in order to see this one has to ignore the vowels. Those are not consistent in the "write" forms; sometimes they are /a/, sometimes they are /u/ and /i/. These are the only three vowels in Arabic, and all of them are found with 'write' sometimes and not others. Nor do the vowels help to distinguish 'write' from 'earn', because the full range of vowels is found with 'earn' as well. The consonants are also arranged differently in the different forms: 'cause to write' has a doubled consonant that the other versions don't have for example. And one needs to overlook the initial glottal stop in the 'make X' forms. So there is a lot to ignore in isolating the 'write' morpheme. But you don't want to ignore the sequence of consonants k-t-b. By the same token, the morpheme for 'earn' has something to do with the consonants d-r-j in that order.

Consider now how the difference between active and passive is expressed in Arabic. The passive is a kind of verb form with a changed argument structure. In English it is expressed by an –en suffix, among other realizations. But the argument structure change and the affixation are illustrated in xxb.

(a) I give candy to children at Halloween.

(b) Candy is give-n to children at Halloween.

Now what about Arabic? What do all the passive forms have in common that distinguishes them from the active forms? This is where the vowels come into their own. All of the passive forms in the data share the vowels u-i, in that order. In contrast, all of the active forms have two a

vowels. The u and the I are never adjacent to each other; they are always separated by at least one consonant. Nor is their ordering with respect to the consonants completely consistent. In kutib 'he writes', for example, the u comes after the k, whereas in ?uktib 'he makes write' the u comes before the k. But there is no doubt that the vowels are the key to communicating the active-passive distinction.

Finally, what is left to express the derivational differences between the simple verbs 'write' and 'earn' and the fancier variants 'cause to write', 'write to each other', and 'make write'. This is the hardest to see right off. Since the consonants are busy expressing the root meaning and the vowels are taken up with active versus passive, it seems that there is nothing left to express these semantic differences. But there is something that all the simple verbs have in common, which distinguishes them from the 'cause to' verbs. This has to do with how the consonants and vowels are arranged. All of the simple verbs have a CVCVC pattern, with no long vowels or geminate consonants. In contrast, the 'cause to' verbs always have a long first vowel. The reciprocal verbs have two short vowels, but the middle consonant is a geminate. Finally, the 'make X' forms always have a CVCCVC pattern—identical to the 'X for each other' verbs except that the 'make X' forms always start with a glottal stop and the other consonants are shifted to the right to accommodate this. Having already seen reduplication, we can recognize what this is. These morphemes are pure CV skeletons. They have no distinctive pronunciations (except for the glottal stop in 'make X'), but they do have consistent and coherent phonological representations, namely the ones given in xx:

- All of the present tense verbs have the shape CVCVC.
- All of the cause to X forms have the shape CVCCVC
- All of the X for each other forms have the shape CVVCVC
- All of the make X forms have the shape ʔ̤ç̤VCCVC

In this respect, they are identical to the reduplication affixes that we saw in Agta, Rembarrnga, and other languages.

But if Arabic has derivational morphemes with the same kind of phonological structure as reduplication morphemes in other languages, why doesn't it have reduplication? The answer is because the verb roots are represented differently in Arabic from the way they are in Rembarrnga. Roots in Rembarrnga have their own ideas about how their consonants and vowels are arranged. These ideas conflict with those of the affix, which creates a copying of the pronunciations so there is enough for both. That copying is the essence of reduplication. But verb roots in Arabic do not have their own ideas about how they should be arranged. They are perfectly happy to follow the lead of the derivational affixes in this respect. As a result, there is no conflicting vision, and no copying is triggered. Basically what this means is that roots in Arabic are consonantal pronunciations in a certain characteristic sequence, but without any distinctive arrangement of those pronunciations into well-formed syllables. Arabic roots are consonant pronunciations with no CV-skeleton associated. In this respect they are like the mutation/ablaut morphemes seen in the previous section, only bigger. (A single consonant feature is not enough to build an interesting variety of words out of; rather one needs a few full

sets of consonant features to do this.) They get their arrangement properties entirely from the derivational morphemes listed in (xx).

Once one sees this clearly, it is not hard to fit the active/passive morphemes into the picture. They two consist of ordered phonological features—the ones that define the various vowels in Arabic—with no particular preferences about where they are pronounced. They too are features with no CV skeleton that get linked into the CV skeleton provided by the derivational affix.

We can now analyze particular examples as follows. The lexical entries for Arabic are as follows:

A word structure tree would look like this:

Then in the phonology the pieces get put together by a process of linking that is pretty much the same as what we have seen in other types of exotic word formation, especially reduplication. First the consonant pronunciations of the root are linked to the available C slots in the affix in a sequential fashion, with no crossing lines:

Then the vowel pronunciations of the active or passive are linked to the available V slots in the affix sequentially, again with no crossing lines. This gives the following phonological representations, which are approximately correct.

```
k a t a b  "write, active"        k a  t  a b  "write, active"
| | | | |                        | | / \ | |
C V C V C  [present]              C V C C V C  "cause to X"

k  a  t a b  "write, active"
| / \ | | |
C V V C V C  "for ea. oth."
```

(20)

```
k   t   b  "write"            k    t    b  "write"
|   |   |                     |   / \   |
C V C V C  [present]           C V C C V C  "cause to X"
   \ /                           |     |
    a      "active"               u     i   "passive"
```

The one significant difference between the kind of linking seen in Arabic and the kind seen in reduplication is that there are exceptions to the principle that linking should be one-to-

one. In a form like kuutib, the u links to two V slots before the I links at all. Similarly, in kattab, the t links twice before the b links. We did not see double linking of this kind in reduplication. There is an intuitively clear reason. In reduplication, there is almost always enough featural material in the root to fill up the affix, given that roots are relatively long and reduplicating affixes are fairly short. (This is especially true if one looks at what is crucial in reduplication as being the number of syllables and whether they are heavy or light. There might not always be enough material to fill up the onsets (beginnings) of the reduplicated syllables, but that is not the important part of the reduplication in our revised theory.) The situation is systematically different in Arabic. Here the CV skeletons of the derivational morphemes are relatively long, and the featural material provided by the other morphemes is limited; there will often not be enough features to spell out each C and V. But if one does not fill up the CV skeleton, then the distinctions among the different forms would be lost. The best solution, then, is one which uses multiple linking. The principle that overrides the principle of 1-1 association can be stated as follows:

(a) CV skeletons must be fully realized.

(b) When multiple linking is used, it is best to link a feature to two adjacent positions.

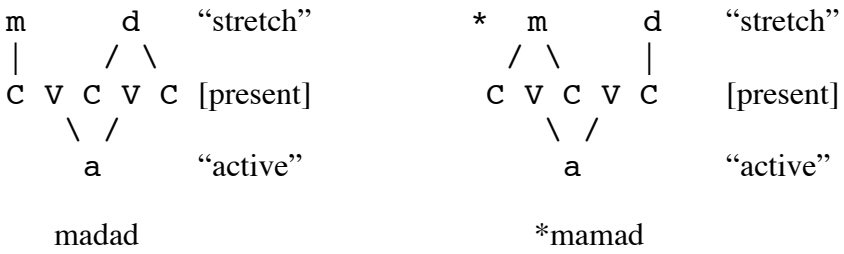
(xxa) correctly predicts that you never use [u] twice if there are only two V slots. If there are two V slots, you use u once and I once. You use [u] twice if and only if there are three V slots open, more than can be filled by two vowel pronunciations in strict 1-to-1 fashion. So we see forms like kutib and kuutib, but never *kutub. (xxb) explains why in these emergency situations one gets kuutib rather than *kuitib.

The same factors can be seen at work with consonants. An interesting comparison is kattab versus ?atkab. The CV-skeleton has four Cs in both cases; it is CVCCVC. But the ‘make X’ form provides a pronunciation with the fourth C; as a result, the consonantal pronunciations of the root can map onto the skeleton in the preferred 1-to-1 fashion:

In contrast, the ‘X for each other’ form has no pronunciations. It can only be realized fully (and so distinguished from the simple verb) if one consonantal position is used twice. Since the middle two Cs are adjacent, the best choice is to fill these with both with t, giving kattab, rather than *kaktab or *katbab.

The importance of the principle that every slot in an Arabic CV template be filled, and that this drives multiple linking, can be seen in another way. Some Arabic roots have only two distinct consonants, rather than three, like the ones we have seen so far. An example is the word for stretch, which has an m and a d in that order. Such words always show up in a certain pattern. One does not find *mada ‘he stretched’, with the third consonant position unfilled and

hence unpronounced, as one might expect. Rather, the third person simple verb from is madad, with the d double linked into the last two C positions.



This shows that principle (xxb) is a preference, but not a requirement. It is not necessary that a single feature bundle be linked to two positions in the CV skeleton that are adjacent; this is a preference only. In forms like kattab, it can be obeyed and is. But in the form above there are not enough segments for each C position, and no two C positions are adjacent. Thus, the imperative that the CV skeleton be fully realized takes precedent, and double-linking to nonadjacent C positions takes place. Note that one always gets madad, with the last segment double-linked, never *mamad, with the first of the two consonants used twice. This shows that Arabic does have a preference for linking that is left-to-right and one-to-one, just as reduplicating languages do. At the beginning of the derivation, at the start of the word, Arabic speakers are never desperate enough to resort to double-linking, although by the end of the word they might be. This explains why forms like (xxa) are found in Arabic, but forms like (xxb) are not.

The other point that needs to be discussed in Arabic is the status of the no-crossing lines principle. This principle does crucial work in the language, in that it keeps the elements of the roots and affixes in the proper order. It is not just the consonants k, t, and b that symbolize 'write' in Arabic, it is having the consonants in that order. The three consonants can come in various places with respect to other material (kuttib vs. ?uktib), but they always come in the same order with respect to each other. Scrambling the consonants to give something like buttik or tubbik results in either gibberish or an unrelated word. This result is guaranteed by the no-crossing lines principle, which rules out a representation like (xx).

The same rule can be seen with the vowels. The u and I of the passive always come in a particular order; one can have kutib, but never kitub. Again, this is a result of the no crossing lines constraint, as shown in (xx). Yet another effect comes with two consonant roots. When simple left-to-right, one-to-one linking has come to an end and there is still the final C slot to be filled, this guarantees that it is the last consonant that gets recycled, not the first, yielding madad rather than madam. The form madam can only be generated with a crossing line.

Arabic is complex, but not chaotic, and the no-crossing lines principle is a large pare of why.

However, a qualification has to be made concerning the no-crossing lines constraint that is probably obvious to you by now. It must be that the consonants and the vowels of Arabic are

on different "planes" or "dimensions" of the representation; they must link directly the CV skeleton, without having any ordering of their own. This is because consonants and vowels can be reordered with respect to each other. The u of the passive comes before the k of 'write' in ?uktib, but not in kuttib, for example. If k-t-b were part of the same dimension as u-I, one of these would have to have a crossing line:

But they are not on the same dimension, so that consonants never interfere with the linking of vowels in Arabic and vowels never interfere with the linking of consonants. These two dimensions are essentially independent of each other. This is different from other reduplicating languages, in which we have written both the vowel features and the consonant features in a single row, and have been careful to keep them from crossing. There is of course a good reason why the consonants and the vowels form two independent dimensions in Arabic: it is because they constitute different morphemes. Each morpheme has a life of its own; the order of its segments is fixed, even though the order of those segments with respect to the segments of some other morpheme may not be. In our first example, this followed rather nicely from the word structure tree. The lowest thing in the word structure tree is the root, say k-t-b, then the derivational morpheme (say CVVCVC) is added, and lastly the passive affix u-I is added. Then it is natural to say that the consonants link to the CV template from below, and the vowels link into it from above. Then there is no danger of lines crossing between these two morphemes, although there is danger of the consonants crossing the consonants and the vowels crossing the vowels.

It is important to see that, thanks to the no-crossing lines principle and the others that we have stated, the theory is still restrictive. In other words, it explains why languages like Arabic can exist (although they are rare), but other kinds do not exist. For example, a language that doesn't exist is Cibara, described as follows:

(21) For example: Cibara, a non-existent language:

*k-t-b 'write'	→	katab [present]	batak [past]
*k-t-b 'write'	→	uktib [passive]	utbik [active]

It is easy enough describe what's going on in Cibara in terms of word formation rules. The tense alternation involves reversing the order of the consonants in the root and the passive active alternation involves shifting the first consonant to the end. In some sense these rules are not much (if any) more complex than the word formation rules one would need for Arabic. But even with the rich power of a CV skeleton and the rules of linking, we could not describe such a morphological system without making use of crossing lines. We thus predict that languages such as Cibara, despite their similarity to Arabic, cannot exist. And the predication is correct: language games (like Pig Latin) sometimes permute sounds in this way, but full-fledged natural languages never do.

In summary, Arabic is the fullest expression of an exotic language from the point of view of morphology. It takes full advantage of the distinction between the CV skeleton and the pronunciation features that exists in the phonology of all languages. This distinction is exploited by a few morphemes in other languages, which stand out as special cases: the reduplication morphemes and/or the mutation morphemes. But in Arabic the bulk of the morphemes are "exotic", consisting of pronunciations without a syllable structure or of syllable structures without pronunciations. The result is what is known as a "root and pattern" system of morphology. But exotic as this is, it is still much less than what one could imagine morphology being if it were the result of unconstrained word formation rules instead. Even the morphology of Arabic is about roots and affixes, about composing words out of discrete parts that have their own well-defined semantics and phonology. It is simply that the phonological representations of these morphemes are a bit unusual, triggering some general rules of phonology. There is no need to change significantly the basic definitions of morphemes that we have had since the beginning, nor to throw away the theory of morphology that we have worked so hard to develop on this foundation here at the very last phase of the discussion. And that is comforting, we hope.