Chapter 5
Morphological idiosyncrasies
/ˌmɔɹfəˈlɑdʒɨkəl/
/ˌɪdijowˈsɪŋkɹəsijz/

In the previous chapter, we saw that affixes could be picky about what kind of stem they attach to. In this chapter, we'll look at ways that stems can be picky about their affixes—*allomorphy*. We'll look at phonological kinds of pickiness, as well as more idiosyncratic, morpheme-based pickiness. Inflectional allomorphy is sometimes also called *irregularity*, and sometimes involves totally arbitrary connections between morphemes—*suppletion*. We'll see how these phenomena can tell us a lot about the history of English. We'll also see that they can tell us a lot about the way we store and produce words as we speak, and learn how *blocking* works.

In the last chapter, we saw that phonological words often can be broken down into bits—morphemes. Morphemes often have their own meanings—they're usually also listemes— and usually the meaning of the whole phonological word is composed out of the meanings of individual morphemes.

We also saw that affixes usually have particular requirements about who their stem can be—*syntactic* requirements. These requirements can be phonological (e.g. having a particular stress pattern), or categorical (e.g. being a noun or a verb).

In this chapter we'll examine some more complicated interactions between stems and their affixes. Not only can affixes *choose* a particular kind of stem to attach to—certain affixes can actually *modify* the phonology of the stems they attach to. In addition, we'll find that affixes can be even more picky about their stems than we've seen up to this point.

We've seen that affixes can systematically demand stems of a certain category, or a certain phonological shape — but they can also demand arbitrary kinds of stems, not identifiable by any defining category or phonology. This kind of arbitrary selection is called *irregularity,* and it can reveal interesting facts about the history of English

5.1 Different listemes, same meaning! Irregular suffixes

So far all the affixes we've considered have had distinct meanings, so they've been distinct listemes. We have the *-ed* affix for the past tense, the *-s* affix for the plural, etc. We've seen a couple of cases where we have homophonous affixes: the *-s* affix for forming the plural of nouns is homophonous with the *-s* affix for forming the third person singular of verbs, for example. They have the same phonology, but different semantics. These don't pose any special problems for us: it seems clear that they're distinct listemes that happen to have the same phonological representation.

There are other cases, however, where it seems like we have two clearly distinct affixes—they're phonologically completely dissimilar—but the different pronunciations don't correlate with different meanings! That is, they have different phonology but the same semantics. Consider, for instance, the following sets of singular/plural pairs:

(63)

	Singular	*Plural*
a.	dog	dogs
	cat	cats
	witch	witches
b.	alumnus	alumni
	focus	foci
	cactus	cacti
	radius	radii
c.	sheep	sheep
	fish	fish
	quail	quail
	shrimp	shrimp

	bison	bison
d.	addendum	addenda
	curriculum	curricula
	bacterium	bacteria
	millenium	millenia
	ovum	ova
	symposium	symposia
e.	analysis	analyses
	thesis	theses
	axis	axes
	diagnosis	diagnoses
	ellipsis	ellipses
f.	child	children
	ox	oxen

All of the above nouns have singular and plural forms, but the plural suffix is different from group to group. The first group is the regular English plural in *-s*. The second group is the Latin plural *-i*, which applies to certain forms ending in *-us*. The third group doesn't seem to make a distinction between singular and plural—they have the same form. The fourth group is another Latin plural, *-a*, which applies to some words ending in *-um* (though not all such words—the plural of *album* is not *alba*, for example). The fifth group is a Greek plural, *-es*, which applies to some forms ending in *-is* (but not all, e.g, not to *trellis*). Last, I've given an Old English plural, *-en*, which applies to only three roots in modern English. (Can you think of the third?)

These suffixes are all synonymous—they all mean *plural*—but they are pronounced significantly differently. They certainly meet our criteria for a defined 'listeme'—each one is an arbitrary sound-meaning pair—so we can definitely say that they are listemes. It's only when taken as a group that they seem somewhat odd. If listemes with the same pronunciation but different meanings are *homo-phones* ('same-sound'), we

could call these listemes *homo-semes* ('same-meaning')—listemes with different pronunciations but the same meaning.[55]

The choice of affix for indicating the notion *plural* in these cases cannot be ascribed to the phonology of the stems to which they attach—there's nothing phonotactically wrong with putting the regular plural suffix *-s* on these words. Rather, these idiosyncratic plurals depend on the identity of the stem to which they attach. If a noun stem belongs to a particular class, it takes a different plural suffix than normal English nouns.

When you learned the root *bison*, you memorized the fact that for this particular root, the plural is not *bisons* but simply *bison*. One thing that makes this slightly easier to remember is that this particular group of null plurals all have a semantic feature in common: they're all words referring to social animals that people raise or hunt. One might be tempted to say that the null plural on these words is conditioned by the semantics of the word, rather than by the phonology—but the semantic feature that conditions it doesn't have the same effect everywhere. The plural of *cow* is *cows*, not *cow*, even though cows are herd animals; similarly for *pigeon-pigeons, horse-horses*, etc. So the correlation must be memorized one root at a time, although the tendency for this type of plural to apply to nouns of a certain semantic class probably helps as a mnemonic.

Similarly, the Latin and Greek plurals must be learned one at a time. There's no perfectly general rule that produces plurals of stems ending in *-us* by deleting *-us* and adding *-i*, otherwise *walrus* and *circus* and *bus* would have plurals *walri, circi* and *bi*. With these, there's no

---

[55] Why don't we use the more usual term *synonym* here? The everyday use of *synonym* is not precise enough, here. Synonyms are generally very similar in meaning, but are not completely interchangeable. For instance, Merriam-Webster's online thesaurus gives *inscribe* as a synonym for *write*, but it's clear that the two have very different ranges of use and connotations—e.g. it'd be very odd to talk about *inscribing a novel*, though it's perfectly natural to *write a novel*. These plural markers, however, which are function listemes, are crucially *not* interchangeable (it's 'incorrect' to say *thesises* rather than *theses*), and *do* mean exactly the same thing, namely simply [+plural]. This difference between homosemy and synonymy reflects a deep distinction between the meanings of content listemes and those of function listemes, which we discuss more in Chapter XX, on acquisition.

semantic mnemonic to help you remember which ones it applies to. The fact that the singulars are all *-us* forms can help, but, since it's not perfectly general, relying too heavily on the *-us* clue can lead you into error. *Octopus* sounds like it ought perhaps to have a plural *octopi*, but classical scholars among you will know it's not so. *Octopus* is based on Greek roots, not Latin, *octo-* 'eight' and *'pod-'* 'foot' (as in *podiatrist*). Greek didn't use *-i* to mark plurals: the historically 'proper' plural is *octopodes*, or the regularized English plural *octopuses*.[56] It's interesting to note that enough people have made this 'mistake' that it has made it into the Merriam-Webster dictionary as a legitimate plural of *octopus*, though the more historically-oriented OED omits mention of *octopi* altogether, insisting on *octopodes* or *octopuses*. (Mistakes of this type are called *back-formations*; we'll learn more about them in a later chapter.)

In any case, it's clear that the choice between *-s* on the one hand and *-i*, *-es*, *-a* or nothing at all on the other, is determined by the particular stem one is trying to pluralize. These suffixes are different listemes, clearly, but they have the same semantics. All these suffixes— *-s, -i, -es,* etc.— have the same meaning: [+Plural]. The appearance of each one is determined by the identity of the stem the speaker is applying the [+Plural] meaning to.

Our lexical entries for these suffixes will mention each of the stems that they can attach to. The crucial difference between the *-s* plural suffix and all the other plural suffixes is that there *are* no particular stems in the listeme for *-s* — *-s* is completely free in its application; all it cares about is that it attaches to nouns. That's what makes *-s* the *regular* plural marker. All the other plural suffixes are irregular. The set of listemes with the meaning [+Plural] will look something like this. (We'll introduce a new kind of brackets to represent the idea 'one of the following': curly

---

[56] These Greek and Latin plurals illustrate a phenomenon we touched on briefly, earlier—the existence of *bound stems*. In the word *symposium*, or the word *cactus*, the plural is not formed by simply adding *–a* or *–i* to the singular form. If we did that, we'd get *symposiuma* and *cactusi*. Rather, the plural is formed by subtracting *–um* from *symposium*, or *–us* from *cactus*, and substituting *–a* or *–i*. That is, the stem of *symposium* is *symposi-*, and the stem of *cactus* is *cact-*, neither of which can occur on their own as words.

brackets, like this: {}. A list of information inside curly brackets in the syntactic section of a lexical entry indicates that the affix can apply to any one of the stems in the curly brackets I've also included ellipses, ... , to show that there are other roots in the list of items that are not included for space reasons).

(64)	*Phonology*	*Syntax*	*Semantics*
a.	/aj/	[{[alumn], [radi], ...} -i]$_{+Pl}$	+Plural
b.	/ijz/	[{[thes], [analys], [ax], ...} -es]$_{+Pl}$	+Plural
c.	/Ø/[57]	[{[sheep]$_N$, [deer]$_N$, [bison]$_N$, ...} -Ø]$_{+Pl}$	+Plural
d.	/ə/	[{[addend], [symposi], [bacteri]} -a]$_{+Pl}$	+Plural ...
e.	/ən/	[{[child]$_N$[58], [ox]$_N$, [brother]$_N$[59]} -en]$_{+Pl}$	+Plural

[57] The symbol 'Ø', the mathematical symbol for a null set, is used by linguists to indicate the pronunciation of a morpheme that has a meaning but no phonological form—a null morpheme, like the plural of *sheep* or the past tense of *hit*.

[58] Actually, *child* will have to have a stem homoseme, *childr-*, to make this rule produce the form *children*, rather than *childen.* See discussion of stem homosemes in the next section.

f. /z/[60] [[_____]$_N$ -s]$_{+Pl}$ +Plural

Of course, the particular stems in an individual English speaker's lists can vary, depending on which ones the speaker is familiar with. If it so happens that you've never heard or read the word *ox* in the plural, then *ox* won't be in your list of *-en* plurals. If you need to talk about more than one *ox*, you'll make it's plural with the unspecified plural affix *-s*: *oxes*. Similarly, if you think *octopus* has the *-us* singular suffix on it, like *alumnus* or *cactus*, you might have put a root *octop-* in your list for the *-i* plural, and produce the plural form *octopi*.

Knowledge of these kinds of irregular plurals are still considered marks of education, since they're part of the learned lexicon borrowed from Latin and Greek, and only educated people are likely to run into them often enough to learn them. It used to be that any educated person would have some grounding in classical languages, so any educated person could be expected to know, for instance, the difference in plural form between a Latin second-declension stem and a Latin fourth-declension stem. That is no longer so, and back-formed plurals like *octopi* no longer mark their user as inexperienced in academe.[61]

There are also sets of derivational-morpheme listemes that are homosemes in a similar way. One example is below:

---

[59] The form 'brother' here, and 'ax' in the *-es* plurals above, should make you notice that the identificational criteria for the relevant stems are not just phonological. Rather, it's the stem's whole listeme, complete with semantics, that determines the application of these particular plural affixes. The word 'brother' that gets the *-en* suffix isn't the word that means 'male sibling', but rather the (historically related) word that means 'monk' or 'a male member of a religious or ritualistic organization'. Similarly, the 'ax' that takes the /ijz/ plural is a bound Latin root that means something like 'line of reference', as in *axis*, not the homophonous free root *ax* that means 'instrument for chopping.'

[60] If you did Exercise 3 in the previous chapter, you should have arrived at the conclusion that the listeme for the default English plural suffix is /z/, by the same reasoning we applied to the past tense suffix /d/.

[61] Mastery of apostrophe use in *it's/its,* and of the spelling of homophones like *they're*, *their* and *there*, or *reign* and *rein*, are some of the current badges of membership among the liberal-arts-educated.

(65) *Derivational homosemes: Irregular nominalizers*

	Verb	*Noun*
a.	correspond	correspondence
	appear	appearance
	repent	repentance
	accept	acceptance
b.	reply	reply
	run	run
	cough	cough
	hit	hit
c.	condemn	condemnation
	realize	realization
	converse	conversation
	determine	determination
d.	qualify	qualification
	beautify	beautification
	apply	application
	publish	publication
e.	propel	propelling
	eat	eating
	write	writing
	mix	mixing

These suffixes— *-ance, -Ø, -ation, -cation, -ing* —are all listemes sharing a meaning like [+NounOfAction], just as *-s*, *-i,* etc. are listemes that share the meaning [+Plural]. Their lexical entries would look something like this:

(66)

	Phonology	*Syntax*	*Semantics*
a.	/əns/	$[\left\{\begin{matrix}[\text{govern}]_V \\ [\text{appear}]_V\end{matrix}\right\}$ -ance $]_N$	+NounOfAction

...

b. /Ø/ $[\{[\text{reply}]_V, [\text{run}]_V, \ldots\}\ \text{-Ø}\ ]_N$ +NounOfAction

c. /ˈejʃən/ $[\{[\text{condemn}]_V, [\text{converse}]_V, [[\_\_\_]\text{-ize}]_V, \ldots\}\ \text{-ation}]_N$ +NounOfAction

d. /ˈkejʃən/ $[\{[\text{apply}]_V, [[\_\_\_]\text{-ify}]_V, \ldots\}\ \text{-cation}]_N$ +NounOfAction

e. /ɪŋ/ $[\ [\_\_\_\__\omega]_V\ \text{-ing}]_N$ [+NounOfAction]

There are many more examples of homosemy in English, and indeed in any langugage. This is the essence of irregularity: one functional meaning is realized by several different suffixes, depending on the stem in question. In English, nouns meaning *'one who Xes/the agent of Xing'* can be formed with *-ant (assistant, contestant*) as well as the default *-er* (*writer, producer*). Adjectives meaning *'full of X/characterized by X'* are formed from *-ous* (*venomous, envious*), *-ful* (*hopeful, fearful*) and the default *-y* (*dusty, hairy*). Verbal participles are formed in *-en* (*driven, written*) and *-Ø* (*put, hit*) as well as the default *-ed*. Examples can be multiplied *ad nauseum*. Clearly, our mental lexicon is full of sets of listemes of this kind: one meaning, but multiple, arbitrarily varying pronunciations. All this irregularity is not tremendously efficient at first glance—when one invents a computer language, for instance, one usually designs it so that a single meaning is invariably represented by a single form. We'll examine the source of all this variation in English, and why it doesn't just all go away, as we continue.

It is important to realize the deep difference between phonologically conditioned allomorphy, which messes with the final pronunciation of a particular suffix, like the participial listeme /d/, and

these stem-conditioned listemes with identical meanings. For the [+Participle] meaning, for example we've got the different listemes *-Ø, –en*, and *-ed* (/d/). The *-ed* one undergoes phonologically conditioned allomorphy, and ends up pronounced as /t/ (as in *walked*), /d/ (as in *calmed*) or /ɨd/ (as in *shouted*), according to the phonology of its stem. Phonologically conditioned allomorphy is quite general, applies indiscriminately to every phonological word produced by affixation of a particular listeme, and is motivated by the phonotactic rules of the language. No one has to memorize the particular stems that each phonological allomorph of the participle suffix *-ed* attaches to, because which allomorph you get is entirely determined by the pronunciation of the final sound of the stem. On the other hand, *everyone* has to memorize which particular stems the listeme *-en* with the meaning [+Participle] goes with.

One way to think about it is that when you want to say something, you pick out particular listemes on the way from an abstract meaning to the base phonological form. Then, phonologically conditioned allomorphy happens on the way from the base phonological form to the actual pronounceable form which emerges from your lips. You could think of the whole process of producing a word like this:

(67) i. Arrive at a meaning you wish to convey. E.g. in answer to the question, "Who arranged for Bush's visit to Tucson?", you might want to convey a meaning like the following:
[[ASSIST]+ AgentOfAction]+Plural]

ii. Go to your lexicon and look up the listemes for each of these meanings in turn:
a) Look up ASSIST. You get this listeme:
[əˈsɪst]$_V$
b) With [əˈsɪst]$_V$ in mind, look up NounOfAction. You will get the *-ant* listeme, because it has ASSIST on its list, rather than the default *-er* listeme:
[[əˈsɪst]$_V$ ənt]$_N$

c) With $[[\text{ə'sɪst}]_V \text{ənt}]_N$ in mind, look up Plural. This will give you the default Plural affix /z/, since no homoseme of Plural specifies $[[\text{ə'sɪst}]_V \text{ənt}]_N$ in its list:
$[[[\text{ə'sɪst}]_V \text{ənt}]_N]\text{z}]_{Pl}$

iii. Send this off for preliminary pronunciation arrangements. Here, you'll detect that the phonological word /ə'sɪstəntz/ ends in a voiceless stop followed by a voiced fricative, violating the phonotactic rules of English. Consequently, phonologically conditioned allomorphy, is triggered, applying to the final /z/ to produce
[ə'sɪstənts]

iv. Do final fine-tuning of the pronunciation—find the right allophone for all the phonemes in the form, give it the correct intonation for the meaning you desire (assertive or questioning), and send the instructions off to your articulatory system.

We'll see some psycholinguistic evidence for this general picture of word production in chapter XX.

5.2 Root irregulars

We see a phenomenon that looks like homosemy in root morphemes, too. Consider the following lists:

(68) *Root homosemy?*

	Present Tense	*Past Tense*
a.	sink	sank
	eat	ate
	feel	felt
	sleep	slept

	make	made
	keep	kept
	write	wrote
	Verb	*Noun*
b.	induce	induction
	produce	production
	reduce	reduction
	deduce	deduction
	produce	production
	seduce	seduction
	(*conduce)	conduction
	(*subduce)	subduction
	(*abduce)	abduction

In (68), the morpheme which is pronounced differently depending on context is not the suffix, but rather the root itself. In (68)a, we see that some verb roots in English have different forms in the past tense—instead of just adding a past tense suffix, they use a form of the root with a different vowel or consonant in it. (Sometimes they also seem to add a suffix, as with *sleep-slept*; other times there is no suffix). Similarly, a number of bound Latinate root morphemes, including *-duce*, the root of all the words in (68)b, have distinct allomorphs for use when they're nominalized; in this case, /duws/ becomes /dʌkt/. There are several other Latinate root morphemes of this type; *-ceive/-cept-* and *-volve/-volu-* are two of them. The lexical entries for *eat/ate* and *-duce/-duct* will look like this:

(69)	*Phonology*	*Syntax*	*Semantics*
a.	/ejt/	$[ate]_V$	[[EAT]Past]
	/ijt/	$[eat]_V$	EAT

b.	/dʌkt/	[[__$_{Prefix}$] duct]$_V$	[[DUCE]NounOfAction]
	/duws/	[[__$_{Prefix}$] duce]$_V$	DUCE[62]

5.3 Linguistic paleontology: fossils of older forms

Homosemes, then, come in families, grouped according to the meaning that they share. Most of them are restricted to appear only in certain circumstances, but in each family, there's one listeme that can apply pretty much everywhere—one that has no restrictions at all. Among the [+pl] homosemes, *-s* is the everywhere form. Among the adjective-forming homosemes, *-y* is the everywhere form.

Why doesn't the regular listeme take over and displace its irregular sibling affixes? Wouldn't it be much more economical to just have one form for each meaning? Everyone would know what I meant if I said I was feeling very *hope-y*, rather than *hopeful*. It's clear what a child means when they say they liked what they *eated* yesterday, rather than what they *ate*. Everyone knew what Bush meant when he was talking about an *analyzation*, rather than an *analysis*. Where did the irregular listemes come from? And why do they persist?

---

[62] When you get to DUCE in your semantic analysis, you'll find a series like "DUCE='make smaller' when it occurs with RE"; "DUCE='figure out' when it occurs with DE"; etc. Recall that cran-morphs have meanings that depend on what other items are nearby. What makes them cran-morphs, then, is that they have no meaning that doesn't depend on something else.

In general, the unpredictable forms of English have four sources: incomplete application of some historical change in a English morpheme or sound pattern, or borrowing of a set of morphemes or sound patterns from another language. In this section, we'll look at examples of all four kinds. Irregulars are kind of like linguistic fossils, the last remnants of formerly productive structure in some earlier stage of development, or productive structure borrowed from some other language entirely.

5.3.1 Fossils of older forms I—Incomplete change in morpheme: a three-*legged* race.

Our first case is one particular homoseme of the adjective-forming suffix *-ed*, which appeared when the regular, default suffix underwent a sound change A few forms were left out of the change, and now the pre-change pronunciation is an irregular morpheme that shows up with only a few roots, each of which has to be listed individually in the lexical entry for that morpheme.

In Middle English, the regular form of the adjective-forming suffix *-ed* was always pronounced with a vowel, as /ɨd/, no matter what the phonology of the stem it was attached to was like. (This is why we currently spell it *-ed*.[63]) Since the suffix had its own vowel, adding the

---

[63] At least, we came to consistently spell it *-ed* with some help from Samuel Johnson's dictionary-making in the 1700s. In the 1500s, the reduction of the suffix from /ɨd/ to /d/ was often reflected in spelling. Shakespeare had a tendency to write *-'d* to indicate the reduced pronunciation, e.g. in

> I am constant as the northern star,
> Of whose true-fix'd and resting quality
> There is no fellow in the firmament,

from *Julius Caesar*, or

> ...Herein will I imitate the sun,
> Who doth permit the base contagious clouds
> To smother up his beauty from the world,
> That, when he please again to be himself,
> Being wanted, he may be more wonder'd at,

from *Henry IV, Part 1*. After Johnson published his dictionary in 1755, no one used the apostrophe'd form any longer, except occasionally for poetic effect. Johnson's standardization of the spelling to match the older pronunciation reflects an almost

suffix to a word always meant adding a new syllable to the word. (This was true not only for the derivational adjective-forming suffix *-ed*, but also for the homophonous participle-forming suffix *-ed*.) There was never any need for phonologically conditioned allomorphy of this suffix, since the phonotactic rules of English were perfectly happy with the shape of this additional syllable.

Eventually, however, the suffix was shortened, from the syllabic /ɨd/ to the simple consonant /d/. After that shortening, adding this suffix to a stem involved forcing an additional consonant into the coda of the stem's final syllable. At that point, the phonotactic rules of English jumped in to create the phonologically conditioned allomorphs of /d/ we know and love, in order to make the new more complex codas pronouncable.

This contraction in pronunciation happened to the suffix itself, so it should have happened everywhere the suffix was used. But in just a few cases, the older version of the affix has been preserved.

The /ɨd/ pronunciation was often preserved in words that were common in idioms, poems, or ritual speech, where language learners were more likely to repeat the string exactly as they heard their elders say it. In poetry, the extra syllable was often important to the meter of the poem, so reducing the suffix would hurt the poem, as in these lines of Lewis Carroll's (written long after the change occurred, but still employing the syllabic pronunciation for metric purposes):[64]

(70) I'll tell thee everything I can; there's little to relate.
I saw an agèd, agèd man, a-sitting on a gate.

This poem is arranged in iambic feet: I'll TELL thee EVeryTHING I CAN; there's LITle TO reLATE. If the adjective *aged* was pronounced

---

inescapable tendency of people to feel that older forms are 'correct' and that innovations are 'corruptions'. See the discussion of spelling in chapter XX.

[64] Syllabic pronunciation of the *-ed* suffix is sometimes indicated with a grave accent; I'll follow this convention here.

with the reduced suffix, as /ejʤd/, rather than as /ejʤɨd/, the meter of the second line would be completely off.

Other examples of adjectives that retained the old pronunciation of the suffix are *learnèd* (as in *a learnèd man*), *belovèd, accursèd* and *blessèd*. The last three of these are common in ritual speech—in the liturgy of the church—and part of the reason they were preserved is the importance people attach to the exact replication of ritual. In ritual speech, it's important to get the words exactly 'right'; this usually means pronouncing them exactly the way they were learned—even if that results in archaic-sounding speech.

The independent stems of some of these adjectives in *–ed* have since dropped out of the language: *naked* is like that—there's no independent listeme *nake* anymore. If it still exists as a morpheme at all, it's a cran-morph, like *shevel* in *disheveled.*

The adjective-forming suffix *-ed* can also apply directly to nouns, to make an adjective meaning 'having X, characterized by X' as in *toothed, moneyed, cultured, diseased, jaundiced, brown-eyed,* etc. Most of them take regular *-ed*, i.e. /d/, but there are a few adjectives formed in this way that have retained the syllabic form of the suffix: *wickèd* (related to *witch*), *raggèd, crookèd, jaggèd*. In some of these cases, like *raggèd*, or *leggèd,* (as in *a three-leggèd race* or *a one-leggèd pirate*), the root noun is still common.[65] In other cases, as with *naked* above, the original noun from which the adjective was formed has been lost, or has become uncommon: *wick* (a variant of 'witch'), *crook* and *jag* are examples of this.

The infrequency of the root in words like *naked, jagged* and *wicked* means that it might not have occurred to you before that the *-ed* in these words is a suffix at all. The combination of a cran-morph like *nake-* with the unusual homoseme of the adjectivizing suffix *-ed* may have caused your grammar to reanalyze this form as a single morpheme. If you don't run into alternative forms of the root very often, and if the suffix

---

[65] Note that the form spelled *legged* is another example that shows the adjective-forming suffix *-ed* and the past-tense suffix *-ed* are distinct. In British English, there's an idiom, 'to leg it', meaning 'to run away'. The past tense of this idiom is formed by putting the past tense suffix *-ed* on the verb *leg*, as in *I legged it*—but the resulting form is pronounced /lɛgd/, not /lɛgɨd/, like the adjective formed with *-ed* is.

itself is also rare and irregular, and if the whole word is a plausible English phonological word, your word-analysis program may decide there's just a single morpheme involved. In a moment, we'll see that this kind of frequency-related reanalysis might have played a role in causing Bush to produce the neologism *analyzation.*

In any case, the syllabic /ɨd/ pronunciation of the adjective-forming suffix has come down in the world from being the default, everywhere form, to being a very restricted allomorph. Here's how its lexical entry has changed:

(71) a. *Before contraction*:

Phonology	*Syntax*	*Semantics*
/ɨd/	$[[\_\_]_{\{N, V\}}$ -ed$]_A$	'having undergone Xing/having X'

b. *After contraction:*

/ɨd/	$[\{[age]_V, [bless]_V, [rag]_N, [leg]_N, \ldots\}$ -ed$]_A$	'having undergone Xing/having X'
/d/	$[[\_\_]_{\{N, V\}}$ -ed$]_A$	'having undergone Xing/having X'

5.3.2 Fossils of older forms II—Incomplete application of a phonotactic change: *Calves*, *wolves* and how they're *spelt*.

Above we saw a fossil of a former regular morpheme, preserved as an irregular morpheme in a tiny corner of the English vocabulary. Now we turn to a case where an irregular morpheme preserves a fossil of a former regular phonotactic rule of English, now defunct, but which has left its traces on a few common forms.

In modern English, several nouns that end in voiceless fricatives in the singular have a stem homoseme that ends in a voiced fricative in the plural, like *calf-calves*, *house-houses, mouth-mouths,* and *wife-wives*. Why

bother? Why do we say /haus/-/hauzɨz/ rather than /haus/-/hausɨz/ (as we do in the regular pair *louse-louses*)?

Old English, like other Germanic languages, used to have a quite general phonotactic rule which required fricatives to be voiced when they occurred between two voiced sounds. For example, *father* /fɑðəɹ/, cognate of Latin *pater* , has a voiced interdental fricative in the middle, rather than a voiceless one, because the fricative appears between two voiced vowel sounds. Grimm's law (see chapter 2) caused the Proto-Indo-European /t/ to become the voiceless fricative /θ/ in Germanic (as we see in the *tri-/three* cognate pair), and the P-I-E /p/ to become the voiceless fricative /f/. That rule by itself would have meant that the word for father should have been pronounced /fɑθəɹ/, not /fɑðəɹ/. It was the phonotactics of Germanic that caused the /θ/ to become its voiced counterpart /ð/ in this word, because it occurred between two vowels.

Now, it so happens that the Old English ancestor of our plural suffix, like the Old English version of our past tense suffix, used to have its own vowel: it was pronounced sort of like /ɨz/, not /z/. That meant that when the plural suffix, complete with vowel, was added to a word like *wife, knife, house,* or *wolf*, which ended in a voiceless fricative, the intervocalic voicing rule would kick in. When the suffix was added, all of a sudden the voiceless fricative /f/ (in the case of *wife*) or /s/ (in the case of *house*) was in between two vowels. Consequently, the phonotactic rule, 'intervocalic fricatives are voiced' applied, to give something like /wivɨz/ 'wives', /huzɨz/ 'houses', and /wʌlvɨz/ 'wolves'.

The voicing rule no longer applies in modern English, in which intervocalic voiceless fricatives are perfectly fine. For instance, *blessing, facile, laugher, lifer,* and *prefer* are all English words with intervocalic voiceless fricatives. Further, the plural suffix is now just /z/, so making a plural for most words doesn't involve adding an extra syllable: in /wajvz/, /wʌlvz/, /kævz/, the fricative is not even intervocalic anymore. But the intervocalic voicing rule is still with us as a fossil—in the irregular plural stem homosemes of the roots *wife, knife, wolf, life,* and *house*, which now

have to be memorized one at a time.[66] The lexical entry of *wolf*, for instance, has changed as follows:

(72) a. *Before loss of the intervocalic voicing rule:*

Pronunciation	*Syntax*	*Semantics*
/wʌlf/	$[wolf]_N$	WOLF

b. *After loss of the intervocalic voicing rule:*

/wʌlv/	$[wolv]_N$	[[WOLF]Plural]
/wʌlf/	$[wolf]_N$	WOLF

Another example of a fossilized phonological rule is the class of past-tense irregulars like *feel-felt, dream-dreamt*, *mean-meant, burn-burnt*. Some more examples of this rule, which are losing ground in written English but are still in widespread spoken use in some dialects of American English, are *spill-spilt, learn-learnt, spoil-spoilt, smell-smelt* and *spell-spelt*.

It used to be the case that the phonotactics of English required the devoiced allomorph /t/ of the past tense morpheme to appear not only after voiceless consonants, as in modern English, but also after nasals and liquids.[67] That is, the phonotactic rules for English codas didn't allow /d/ to occur after nasals and liquids, and as a result, all verbs ending with a nasal or liquid got the /t/ pronunciation of the past-tense suffix, by regular phonologically conditioned allomorphy. When the phonotactics of English changed, to allow /d/ after nasals and liquids generally, we retained a few of the more frequent forms with /t/ as memorized, irregular, homosemes. (Many of these verbs have stem allomorphs as well, their short vowels are a remnant of another former phonotactic rule.)[68] Before the change in

[66] There are plenty of words that end in /f/ or /s/ that don't have stem allomorphs in /v/ or /z/ for the plural: *belief-beliefs, grief-griefs, laugh-laughs, mess-messes* etc.
[67] This rule is also the source of the adjectival participles *gilt, pent* and *girt*.
[68] For some discussion of this rule applying in forms like *wise/wisdom*, see chapter XX.

coda phonotactics, the regular past-tense morpheme didn't have a *-t* homoseme. After codas like /nd/ and /ld/ became legal, though, the *-t* form of the suffix still appeared on a few diehard stems. The lexical entry for the past tense suffix changed as follows:

(73) a. *Before loss of the coda restriction:*

Phonology	*Syntax*	*Semantics*
/d/	$[[\_\_]_{V}\text{-ed}]_{Pst}$	+Past Tense

b. *After the loss of the coda restriction:*

/t/	$[\left\{\begin{matrix}[\text{learn}]_{V} \\ [\text{burn}]_{V} \\ [\text{mean}]_{V}\end{matrix}\right\}\text{-t}]_{Pst}$	+Past Tense
	...	
/d/	$[[\_\_\_\__{\omega}\,]\text{-ed}]_{Pst}$	+Past Tense

As time goes by, many of these are becoming more and more regularized, so alongside *dreamt* and *burnt* we now often see *dreamed* and *burned*. *Spill, learn, spoil* and *smell* occur mostly with the past tense forms *spilled, learned, spoiled* and *smelled*. However, the verbs *feel* and *mean* are still robustly hanging onto their irregular past tenses: no one talks about what they *feeled*, or what they *meaned* to say. It's possible that in the future even these fossils will erode away, and there will no longer be any record in modern English of the old phonologically conditioned allomorphy of *-ed* before nasals and liquids. In section 5.6, we'll consider what factors are involved in the retention and loss of these irregular forms.

5.3.3 Fossils of older forms III—Borrowed suffixes from another language: *-i* and *-s*, *-ity* and *-ness*

A third source of modern English homosemes was the large influx of Latinate-origin stems and affixes that were borrowed into English between 1200 and 1700 (see the next section, and chapter XX). Before the arrival of these elements, the only listemes there were were provided by

the native Anglo-Saxon stock. If you wanted to pluralize a noun, your primary choice was the regular plural suffix *-s*. If you wanted to turn an adjective into a noun, your basic option was the regular nominal-forming suffix *-ness*.

After the influx of borrowings, though, several new homosemes had entered the language, along with the stems they applied to. We've already discussed some of the homosemes for the plural (*-i, -a*, etc.). There were also new homosemes for forming nouns of action, some of which that we've looked at: *-ation*, etc., on top of the Anglo-Saxon *-ing*. There were new homosemes for forming adjectives: *-ous* on top of the Anglo-Saxon *-ful*—and new homosemes for forming nouns of quality: *-ity* on top of Anglo-Saxon *-ness*.

Many of these were restricted to occur only with certain stems, or certain other affixes. Others were more productive. They are nearly all, however, homosemes, realizing meanings that the regular Anglo-Saxon vocabulary already had forms to express.

5.3.4 Borrowing trouble: *MaLIcious magicians* with *MAlice* and *MAgic*

Finally, in some cases the borrowed Latinate suffixes have brought along special phonological rules with them. In English, these rules became memorized morphophonological rules—rules triggered by particular morphemes—rather than general phonological rules. The clearest example of this kind of affixal selection is particularly interesting because it's a complex result of a historical accident that changed English forever eight hundred years ago. Although most speakers of English are at best only vaguely aware of the historical events that led to the restructuring of the English vocabulary, we all have perfect subconscious command of the rules governing the word-building tools that English acquired as a result!

5.3.4.1 Stress-shifting and non-stress-shifting suffixes

Different derivational suffixes have different effects on stress placement in English. To see this, pronounce the following word pairs aloud to yourself, and then transcribe two of the pairs from each group, indicating the placement of main stress:

(74)

a.	regurgitate	regurgitation
	credible	credibility
	artist	artistic
	janitor	janitorial
	compliment	complimentary
	Canada	Canadian
b.	guardian	guardianship
	yellow	yellowish
	colonial	colonialism
	violin	violinist
	neighbor	neighborhood
	perish	perishable
c.	employ	employee
	kitchen	kitchenette
	Reuben	Reubenesque
	official	officialese
	tonsil	tonsilitis

Exercise 9: Transcribe three pairs of words from (74) above, indicating the main stress in each.

Recall that all English content words receive a main stress. For most multi-syllabic words in English, the stress falls on either the third-to-last ('antepenultimate') syllable, as in *PAradise, collATeral, CHICkadee, REprimand* or *MinneAPolis*, or else the second-to-last ('penultimate') syllable, as in *CANdy, TolEDo, umBRELla, piANo* or *banANa*. In one-syllable content words, of course, there's nowhere for stress to go except on the one syllable, which is both first and last.

You probably noticed that the syllable that gets main stress in the pairs in (74)a and (74)c is different for the suffixed and non-suffixed words, while the same syllable gets main stress in both kinds of words in (74)b. While the affixes in (74)a and (74)c change the placement of stress in the words they attach to, the affixes in (74)b are *stress-neutral*—they

just tack onto the end of whatever word they're in, sometimes getting their own secondary stress, but not changing the placement of main stress. These affixes are very simple to treat in our mental lexicon—we don't need to indicate any special effect for them.

In the (74)c cases, the affix itself carries its own main stress, which it brings to the word it's attaching to. The stress pattern of the root is generally maintained, albeit at a reduced level. The syllable that got main stress without the suffix now gets *secondary* stress—less stress than main stress, but more than none—and the relationship between the stem word and the suffixed word seems pretty straightforward. In the IPA, just as primary stress is represented with a high-up tick before the stressed syllable, secondary stress is represented with a low-down tick before the stressed syllable. In English orthography, I'll show secondary stress with small caps. With stress and syllabification indicated, the transcriptions for the last two examples in (74)c look like this:

(75)	a.	of.FI.cial /ə.ˈfɪ.ʃəl/	of.FI.cia.LESE /ə.ˌfɪ.ʃə.ˈlijz/
	b.	TON.sil /ˈtɑn.sɨl/	TON.si.LI.tis /ˌtɑn.sɨ.ˈlaj.təs/

These suffixes seem to be saying 'Give me main stress, and reduce the stress pattern on my stem to secondary status'.

In (74)a things are a little different. The main stress in the suffixed word doesn't fall on the suffix. Rather, it falls in the syllable *before* the suffix. This means that no matter where the main stress fell in the non-suffixed word, stress in the suffixed word must be on the syllable right before the suffix. With stress and syllabification indicated, the transcriptions for the last two examples in (74)a look like this:

(76)	a.	COM.pli.ment /ˈkɑm.plɨ.mɨnt/	COM.pli.MEN.ta.ry /ˌkɑm.plɨ.ˈmɛn.tə.rij/
	b.	CA.na.da /ˈkæ.nə.də/	Ca.NA.di.an /kə.ˈnej.dij.jən /

These suffixes seem to be saying, 'Bring the main stress to sit over here beside me, no matter what it does to the rest of the word.' Since de-stressing a syllable often involves reducing the vowel in that syllable, these stress-shifting affixes can significantly affect the pronunciation of the root. This in turn can obscure the connection between the root and the suffixed word. The pronunciation of the root in *palace* /ˈpælɪs/ and *palatial* /pəˈlejʃəl/ is distinct enough that it takes a moment of thought to recognize that the latter is derived from the former.

We can clearly see the difference between the stress-shifters and the neutral suffixes when we look at a stem that can occur with both kinds. In the word *párent*, stress falls on the first syllable. *Parent* can combine with both the adjective-forming suffix *-al* and the noun-forming suffix *-hood*. Each of these suffixes is itself just one syllable, but they result in very different stress placements: *-al* shifts the stress, so that *paRENtal* has stress on the second syllable of the stem, while *-hood* leaves the stress of the stem where it found it: *PArenthood* has stress on the first syllable.

5.3.4.2 Why some but not others?

Why does English have these distinct kinds of suffixes? Why don't all suffixes affect stress placement, or none? If we look up our stress-shifting suffixes from (74)a, *-ion, -ary, -ial, -ic,* and *-ian* in a dictionary with etymological information, like the OED, we immediately notice that they are all borrowed. They all entered English via French after 1100 A.D. The stress-bearing suffixes from (74)c are also all borrowed, mostly quite recently, after 1800 AD: *-ee* and *-ette* are from French, *-esque* and *-ese* from Italian, and *-itis* was borrowed directly from Latin.

The non-stress-shifting suffixes from (74)b, on the other hand, are a mix: *-hood, -ship,* and *-ish*[69] have been part of English from prehistoric times, while *-ism, -ist,* and *-able* are early borrowings from French. None

[69] Actually, *-ish* and *–esque* are cognate, according to the OED—Latin originally borrowed *–esque* from Old High German *–isc*, which is the source of Modern English *–ish*.

of them shift stress—but there is still a difference between the borrowed ones and the Germanic ones! The suffixes of French origin in (74)b mate happily with other suffixes from the stress-shifting list in (74)a. Nouns ending in *-ist* can usually form an adjective with *-ic*, one of the stress-shifting suffixes: *Communist-Communistic, imperialist-imperialistic*, and so on. Adjectives ending in *-able* can form a noun ending in the stress-shifter *-ity*: *perishable-perishability, readable-readability,* etc.

In contrast, our suffixes of Germanic origin, *-hood, -ish,* and *-ship*, can't be followed by suffixes of Latinate origin. We can't make a noun-form of *yellowish* by adding *-ity—*yellowishity* is definitely not a word. If we want to make an adjective out of *childhood*, we can't add the Latinate *-ial* suffix—English speakers would be very unhappy with **childhoodial*.

To make this point extra-clear, consider the difference between the suffixes *-ness* and *-ity*, both of which apply to adjectives to form nouns. The Germanic one, *-ness*, can apply to adjectives formed with the Latinate adjectival suffix *-ic*, in words like *chaoticness, rusticness,* and *causticness*. The Latinate suffix *-ity*, however, which does the same job of turning an adjective into a noun, cannot apply to adjectives formed with the Germanic adjectival suffix *-ish*: there is no *yellowishity, purplishity, freakishity*; rather, we have *yellowishness, purplishness* and *freakishness*.

We can schematize this generalization about the ordering of these classes of suffixes as in (77) below.

(77) [[[Stem]-(-LatinateAffix(es)](-GermanicAffix(es)]

The round brackets indicate optionality, as usual: most stems can occur without any derivational affixes at all, of course; they can also have Latinate affixes without any Germanic ones, and vice versa. What (77) says is that *if* a word has both Latinate and Germanic derivational affixes, the Latinate ones will occur inside the Germanic ones.

In fact, English speakers are remarkably sensitive to the fact that some of our productive suffixes 'belong' with originally borrowed vocabulary. Gene Buckley, at the University of Pennsylvania, had his introductory linguistics class collect a list of words ending in the borrowed suffix *-ize* that were created in English after 1300. Of the approximately 150 words that they found, only three are formed from stems that were

originally English—*winterize, womanize* and *weatherize*. All the others—*brutalize, compartmentalize, realize*, etc.—are formed from stems of Latinate origin. Some (like *realize*) may have been borrowed whole from the source language (in this case, French), but most of the others were formed in English by English speakers, who combined the suffix *-ize* with an independent stem. Although about 50 per cent of everyday English words are Germanic in origin, only 2 per cent of the new words formed with *-ize* from Buckley's list were formed using Germanic stems,

Many of the borrowed derivational suffixes of English are very productive (as we can see from Bush's ability to make up *securitize* and *analyzation* on the spot). Nonetheless, the Latinate suffixes, like *-ity*, can sometimes fail to attach even in places where we'd expect them to be fine—they're 'gappy', as we saw for **deservation* at the beginning of this chapter. When this happens, English plugs in an 'everywhere' Germanic suffix to fill in the gap. So while some adjectives ending in the Latinate suffix *-ous* have nouns made from *-ity* (*curious-curiosity, pompous-pomposity, viscous-viscosity*), other *-ous* adjectives reject *-ity* and prefer the more general Germanic suffix *-ness*: from *rebellious* we can't make *rebelliosity; rather, we must use *rebelliousness*; similarly for *vicious-*viciosity-viciousness* and *querulous-*querulosity-querulousness*.[70] The Germanic suffix, in this case, is the catch-all which applies when the Latinate one can't.

5.3.4.3 How do kids figure it out?

Now, in fact, no one learning English as a first language knows that some derivational suffixes were originally borrowed, and some were originally native to the English spoken 800 years ago. If you know facts like that at all, it's because you learned them in school, long after you became a competent English speaker. Yet, your knowledge of English reveals that you're aware of the existence of these two very different

[70] At the Hacker's Dictionary, http://www.mcs.kent.edu/docs/general/hackersdict/, we find that the degree to which something is *bogus* is its *bogosity*, not its *bogusness*. Using the Latinate affix here increases the humorous effect, since *bogus* isn't really suffixed in *–ous*, though it sounds like it might be.

classes of suffixes. How could you have figured this out when you were learning English?

We saw in the last chapter that infants are hyper-alert to the statistical probabilities of phoneme sequences and stress patterns. A child paying attention to the statistics will certainly notice that some lexemes never show up as phonological words by themselves. Some roots *always* need a suffix of some kind on them—they're bound, not free. Others can show up with or without suffixes. What a child learning English is sure to notice is that only certain suffixes show up next to bound roots—the ones we've been calling 'Latinate'.

So, for instance, the adjective-forming suffix /ɨbəl/—one of the originally Latinate ones, spelled sometimes as *-able* and sometimes as *-ible*—appears in words like *cap-**able**, prob-**able**, dur-**able,** incred-**ible**,* and *vis-**ible***, none of whose stems ever occur as words on their own. (There's no word *prob* in English!) Of course, *-able* is a very productive affix in English, and does occur on plenty of stems that are phonological words on their own (*washable, viewable, breakable*...)—but the crucial thing is that it occurs on some stems which aren't. The same goes for *-ity* (*author**ity**, dign**ity*** and *ent**ity*** are all formed on bound roots), *ous* (*anxi**ous**, feroci**ous*** and *frivol**ous***), and all the rest of the Latinate suffixes.

In contrast, Anglo-Saxon suffixes like *-ness* and *-ship* only go on stems that are actual phonological words on their own. Try and think of a word with one of these suffixes in it that isn't! *Childhood, friendship, happiness, ownership, callousness*...there are tremendously many, and they all have roots that are independent phonological words.[71] This distinction is a very strong clue to the existence of two classes of suffixes.

Other clues to the differences between the two kinds of suffixes are the phonological changes that some of the Latinate suffixes force on their stem, including the stress shift that we've already seen, requiring stem-forming morphemes like *-it-* in *competitive* and *competitor*. None of the Germanic suffixes alter the phonological shape of their stems like that.

---

[71] There are a few exceptions. One is *worship*, which was originally formed from *worth+ship*, but for obvious pronunciation reasons the /θ/ was lost. Others are *gormless, business, wistful, grateful, reckless* and *gruesome*.

Of course, this difference has its source in one of the major differences between languages of Latinate origin like French and languages of Germanic origin, like Dutch, or the English of 800 years ago. The former have mostly bound roots, and the latter have mostly free roots—so when Latinate vocabulary was borrowed wholesale into English, its distinct morphological properties were borrowed too. But a child learning modern English doesn't need to know that. All he needs to notice, and remember, is that some suffixes go with bound roots, and trigger phonological changes, while others never do. Then he just has to mark each kind of suffix as needing certain properties in its stem, and everything else follows.

5.3.4.4 Representing complex suffixal restricitons

Let's say the mental lexicon entries for the Germanic suffixes need to include the information that their stem has to be an independent phonological word. This kind of restriction is rather like the unstressed, open-syllable restriction that we saw for the comparative *-er*. The usual notation for 'phonological word' is the lower case Greek letter omega: ω, which we'll subscript to the blank space that stands for the stem in the syntactic part of the entry. The final entry for *-ness*, then, will look like this:

(78)	*Phonology*	*Syntax*	*Semantics*
	/nɛs/	$[\,[\,\underline{\quad}_{\omega}\,]_{\{N,\,A\}}$ -ness $]_N$	"the quality of being X"

The entries for the Latinate suffixes won't have the phonological-word restriction on them, of course, but they will place other restrictions. In particular, to explain why Latinate affixes don't attach to words derived with Germanic affixes, there must be something that the Latinate suffixes look for in their stems that the Germanic suffixes don't have.

One such something is that many Latinate suffixes attach only to stems which are also roots. The suffix *-ify* is like that. It attaches to adjectival and sometimes nominal roots to make verbs, as in the examples below:

(79)

Root	-ify *verb*	Meaning
clear	clar-ify	to make clear
yuppy	yupp-ify	to make yuppy
magn-	magn-ify	to make big
simple	simpl-ify	to make simple

The meaning of *-ify* is clear enough from the above examples; it creates verbs that mean something like 'to make X'. It is also clear that *-ify* is a Latinate suffix, since it attaches to a number of bound roots. But even when the part of speech is appropriate, and the meaning is clear, *-ify* can't attach to a stem that contains another suffix—not even another Latinate suffix:

(80)

A*djective with suffix:*	*-ify *verb:*	*What it would mean:*
act-ive	*activify	'make active'
accur-ate	*accuratify	'make accurate'
electr-ic	*electricify	'make electric'

This last example is particularly revealing, since *-ify* can attach to the root *electr-*, giving *electrify*, with exactly the meaning expected for the non-word *electricify*, above (as well as a couple of idiomatized meanings). The affix *-ify* is still a productive, independent part of the language, too—it is still used to form causative verbs in modern senses, like *yuppify* and *webify*, so these words can't sound bad because *-ify* isn't used to make new words anymore. We've just got to include the information about requiring a root in the lexical entry for *-ify*.

Note that *-ify* doesn't care if the root is bound or free (both *qualify*, formed on a bound root, and *personify*, formed on a free root, are perfectly good), as long as it doesn't have any affixes on it. The property of 'being a root', then, isn't a phonological property, like 'being a phonological word' or 'ending in an unstressed open syllable'.

We'll assume that being a 'root' is a kind of category information, like being a noun or a verb, and include it as a label on the brackets around the stem. We'll use the mathematical symbol for 'root', √, to indicate the category Root:

(81)	*Phonology*	*Syntax*	*Semantics*
	/ɪfaj/	[[____]$_{√}$ -ify]$_{√}$	"to make X"

We're not quite done with *-ify* yet, though. It's a stress-shifting suffix—the main stress of a word formed with *-ify* falls on the syllable before *-ify*, no matter where it would fall if *-ify* wasn't there. (Consider pairs like *solid* /ˈsɑlɨd/ ~ *solidify* /səˈlɪdɨfaj/.) We need to include this information in our lexical entry too.

Again, we'll represent syllables in the stem with the Greek letter sigma, σ. Round brackets indicate optionality, and the superscript *n* on the syllable symbol inside the brackets indicates that stems containing any number of syllables are possible (as long as they're roots). As usual, a high-up tick indicates the placement of main stress. We'll call the instruction to shift the stress a *readjustment rule*. The final entry for *-ify* will look like this:[72]

(82)	*Phonology*	*Syntax*	*R.Rules*	*Semantics*
	/ɪfaj/	[[____]$_{√}$ -ify]	[(σ^{n}) ˈσ] -ify	"to make X"

Other suffixes have similar restrictions. Some attach only to stems which contain a particular suffix, or one of a few particular suffixes. For example, the adjective-forming suffix *-ic* attaches to roots (as in *electric*), to verbs formed with the suffix *-ify* (as in *terrific, specific* or *horrific*), and to nouns formed with the suffix *-ist* (as in *artistic, pessimistic* or *holistic*). It doesn't attach to nouns formed with *-er*, though they're similar in meaning to nouns formed with *-ist* (**painteric, *writeric*[73]), and it doesn't

[72] Since there's a whole group of suffixes that have this stress-shifting effect, a more economical way to represent this rule would be to write it separately, as "[(σ^{n}) ˈσ] -Affix", give it a number (e.g. 'Readjustment Rule 1'), and just include a note with *-ify* and the other affixes to the effect that they are subject to Rule #1 (and any other rules of this type—there are several).

[73] These can be adjectivised with the suffix *-ly*: *painterly, writerly*.

attach to verbs formed with *-ize*, though they're similar in meaning to verbs formed with *-ify* (**colorizic, *deodorizic*[74]).

The lexical entry for *-ic*, of course, will have to encode this information:

(83)	*Phonology*	*Syntax*	*R.Rules*	*Semantics*
	/ɪk/	{[__]$_{√}$ / [__ ify]$_{V}$ / [__ ist]$_{N}$}-ic	[(σ^{n}) ˈσ] -ic	"in the manner of X"

Since *-ic* is also a stress-shifting suffix, the stress-shifting readjustment rule is also indicated in its lexical entry.

5.3.4.5 Latin phonology in modern English

As we have seen, the irregular sound changes triggered by these suffixes need to be specified suffix by suffix. All of that information is included in your mental lexical entries for those listemes.

Another of these rules was a 'softening' rule. Latinate suffixes whose spelling contains an 'i' as the first letter trigger a rule that changes non-palatal voiceless alveolar and velar obstruents, alone in the coda of the last syllable of the stem, into a palatal fricative. Consider the following pairs:

(84)	magic	/ˈmædʒɪk̲/	magician	/mæˈdʒɪʃ̲ən/
	expedite	/ˈɛkspədajt̲/	expeditious	/ˌɛkspeˈdɪʃ̲əs/
	rate	/rejt̲/	ration	/ræʃ̲ən/
	artifice	/ˈaɹtɨfɪs̲/	artificial	/ˌaɹtɨˈfɪʃ̲əl/
	malice	/ˈmælɨs̲/	malicious	/məˈlɪʃ̲əs/

In Latin and the related Romance languages, there was a regular phonotactic rule which changed /t/, /k/ and /s/ into /ʃ/ when they occurred

[74] These can be adjectivized with the (very versatile!) suffix *-ing*: *colorizing, deodorizing*

before the high front vowel /i/.[75] This rule makes sense if you consider that /i/ is the vowel-equivalent of the palatal consonant /j/—all the affected sounds are being turned into a palatal fricative. The /i/ that originally triggered the rule in Latin is missing from the borrowed English version of these suffixes, which all just contain the reduced vowel /ə/, and English is happy to have /k/, /s/ and /t/ before /i/, anyway (*keen, seen* and *teen* are all fine English words). Nonetheless, when English borrowed these suffixes it also borrowed the softening rule, treating it as a morphologically conditioned readjustment rule mentioned in the lexical entries for the relevant suffixes. The rule applies to every stem that these suffixes attach to, providing it ends in one of the 'hard' consonants covered by the rule. This allomorphic readjustment rule is a fossil of a phonologically regular requirement *from an entirely different language*.

It's worth noting, perhaps, that there's a one-way implication among these readjustment rules. If an affix triggers the softening rule, it also triggers the stress-shift rule, though not vice versa. So, for instance, *-ious,* as in *infect/infectious*, *malice/malicious* triggers both softening and stress shift, but *-ic*, which does trigger stress shift, doesn't trigger softening: *athletic,* formed from *athlete*, is pronounced /əθˈlɛtik/, not /əθˈlɛʃik/. This is the kind of thing that students of English as a Second Language have to spend time practicing and memorizing, but which just 'come naturally' to those of us who learned English as a first language.

5.4 Keeping Irregulars: Semantic clues to morphological classes

With irregulars which are fossils of old phonological rules, there are phonological clues which might help a speaker remember that a particular form is irregular—clues like an *-f* at the end of the stem, in the case of the *calf-calves* type. These clues, while they're not 100 per cent reliable, seem to be used as mnemonics by the linguistic system, to help it remember the

---

[75] A related phonological rule turned /k/ into /s/ before /i/ and /e/, as in *electric/electricity*: /əˈlɛktrɪk/~/ˌɪjləkˈtrɪsiti/. See Chapter XX on spelling for discussion of how this affected the English spelling system.

irregular pattern. For other kinds of irregular forms, there seem to be *semantic* clues that could help with the memorization of the irregular. For instance, the null plural we see in pairs like *sheep-sheep* most often applies to domestic or game animals which travel in groups. We'll look at two other examples of semantic associations among irregulars next.

5.4.1 Pluralia Tantum

The first is the 'forced plural' we see in words like *pants, scissors,* and *binoculars*. These words are made up of a root morpheme, like *pant* or *scissor*, plus the plural morpheme *-s*. They're weird because they never occur in the singular.

It's not that the singular/plural difference is simply marked with a null morpheme, as with *sheep*. Rather, there simply is no singular form of words like *pants*. We can see this when we compare examples (85)a and b with (85)c and d:

(85) a. That sheep is going *baaa*.
b. Those sheep are going *baaa*.
c. *That pant is lying on the floor.
d. Those pants are lying on the floor

In (85)a, we can see that there is one sheep, both because of the singular determiner *that* and the singular agreement on the verb *to be*, which occurs in its 3rd singular present tense form *is*. And although there is no number marking on the noun in (85)b, the 3rd plural present tense form the verb (*are*) and the plural determiner *those* gives it away: we're talking about plural sheep. In (85)c, on the other hand we can't use *pant* with a singular determiner and verb; in (85)d, we see the correct form, where determiner and verb trigger plural agreement.

What is particularly interesting about (85)d is that its *meaning* doesn't have to be plural. That is, even if there's only one pair of pants on the floor, you have to say (85)d in order to get someone to pass it to you. With *pants* there's just no way to distinguish a plural from a singular meaning, as you can in (85)a and b. Similarly, if you say *Hand me those scissors*, you might be asking for one pair of scissors, or many pairs of

scissors—the person you're addressing has to figure it out by context. If you say *Bring me that sheep*, on the other hand, the singular nature of the noun is clear from the determiner, even though there's no marking on the noun itself.

Of course, there is something that seems sort of inherently plural about the kinds of objects that occur with these mandatory plurals: *pants*, *scissors, binoculars*, *tongs* etc. They are all made of two almost-but-not-quite-separable identical parts. This is obviously not an accident, although it's a hard criterion to define precisely. The words *panties* or *briefs*, for underwear, are inherently plural, although they don't have the Siamese-twin structure of *pants*, *glasses* or *tongs*. There are a few game-names that are examples as well, where the notion of almost-separable part is really irrelevant: *billiards, skittles* and *cards*. Although these latter two have singular forms—*a card, a skittle*—the singular refers to one playing piece, not one game. The game-name *cards* can be used to refer to one game or more than one. Two other non-twin, inherently plural words of interest are *thanks* and *kudos*: while you can *give thanks*, you can't *give a thank. Thanks* and *kudos* also require plural agreement: *Thanks are in order*, not **Thanks is in order*, and *Kudos go to the director*, not *Kudos goes to the director*.[76] The twinned-item—inherent plural rule fails in the other direction, as well: there are plenty of twinned-parts items which are *not* inherent plurals. Consider (a) *bicycle, teeter-totter, yo-yo, compass* (for drawing circles), or *barbell*. The word *overall* is a singular for some English speakers but a plural (*overalls*) for others. While we have a semantic clue to the irregular items, it is only a clue, not a hard-and-fast rule. We still need to memorize the inherent plural marking for each of these roots individually.

Note that the roots, *pant-*, *scissor-* or *tong-*, can occur without the plural suffix when part of a compound: *pantleg, scissor factory, tong holder*. This shows that the *-s* suffix on these words really is the regular

[76] In fact, *kudos* is a borrowed Greek word that originally ended in /s/, and was pronounced /kuwdows/, not /kuwdowz/. People unfamiliar with its pronuncation read the *–s* as a plural suffix and pronounced it as /z/ accordingly; this pronunciation and analysis is now the standard one.

plural marker. Within compounds, singular or plural is simply not relevant. We say *lawn-mower*, not **lawns-mower*, even though any given lawn-mower could easily be intended to mow multiple lawns. So the existence of *pantleg* shows that the root *pant-* does exist independently of the suffix *-s*. The only strange thing in these cases is that the plural-marking is required even when the meaning is singular.

5.4.2 Mass nouns

To specify a singular number of any of the inherently-plural nouns that we just discussed, we have to use a 'packaging' noun that has a proper singular, like *pair*. So we talk about *a pair of pants, a pair of scissors, a game of billiards*, etc. In this respect, these nouns have a lot in common with another class of exceptional nouns that do not make a singular/plural distinction: *mass* nouns.

This group includes nouns like *flour, wheat, rice, sand, water, money, furniture, weather* and *cola*, as well as many abstract nouns like *advice, fun, information, knowledge*, and *peace*. To talk about particular quantities of any of these things, you also need a packaging noun: *two cups of flour*, *several pieces of advice, three years of peace*. The difference between these nouns and the *scissors, pants, billiards* examples on the one hand, and the *sheep, bison, fish* examples on the other, is that when we test them with subject-verb agreement, we find that they are inherently *singular*:

(86) a. That flour is infested with moths.
b. *Those flour are infested with moths.
c. That information is reliable.
d. *Those information are reliable.
e. That furniture is color-coordinated.
f. *Those funiture are color-coordinated.

We also have a semantic clue to membership in this class of morphemes: the kind of amorphous, unbounded, 'stuff' quality that the referents of many of these nouns (like *sand, water*, etc.) have. While this semantic property can act as a clue to mass-noun status, it again doesn't work as a definite rule. Consider the physical qualities of the referents of

the words *wheat* and *oats*. Despite being almost indistinguishable to the eye, *wheat* happens to be a mass noun, while *oats* is a count noun. So you can have *one oat, many oats*, but you can't have **much oats*. In contrast, you can have *much wheat* but not **one wheat* or **many wheats*. With *wheat* you need a packaging noun again: *one grain of wheat, many grains of wheat*. Again, the morphological fact of being a mass noun must be individually learned for every root, though the amorphous quality of their referents might help as a reminder.

5.5 Irregulars III: Suppletion

In all of the previous kinds of root homosemy we've seen, there was at least some phonological reason to think that the two forms of the root were related. Usually most of the consonants remained the same, even if some vowels changed: in *dream-dreamt*, for example, the consonant sequence of the root, /dr-m/, remains the same, even if the particular vowel in the middle changes.

For certain kinds of irregular roots, however, there is not even a hint that the two forms of the word are phonologically related. The primary examples of this in English are given in (87):

(87)	a.	Today, you *go*		Yesterday, you *went*
	b.	Today, you *are*		Yesterday, you *were*
		Today, he *is*		Tomorrow you will *be*.
	c.	good	better/best	(well)
	d.	bad	worse/worst	

There's not even a single phonological feature in common between the sequences /gow/ and /wɛnt/, yet they are present and past tense forms of the same verb. Similarly for /aɹ/, /wʌɹ/, /ɪz/ and /bij/, and for /gʊd/-/bɛtr̩/-/wɛl/ and /bæd/-/wʌɹs/. How could such differentiation have come about? And how could a child learning English guess that these very different-sounding sequences are different forms of the same meaning?

The easier question to answer is the one about the source of the different forms. *Went* was originally a past tense form of a verb with a meaning very similar to that of *go*: the verb *to wend* (as in the expression

to wend one's way). In the Middle English period, the past tense form *went* gradually came to displace a different past tense for *go*, *ēode*, and before long, *wend* had become quite infrequent and *went* was never used in any other context.

Similarly, the different forms of *be* in English are the result of a historical mix-and-match between *three* unrelated verbs. The present tense forms *am, are,* and *is* come from a verb stem *es-*, which meant 'to be' all the way back to proto-Indo-European. The past tense forms *was* and *were* come from a stem *wes-* that originally meant 'remain, stay, continue to be'. Those two verbs collapsed into one, using *es-* forms for the present and *wes-* forms for the past. Later, around 1200 AD, the infinitive and participle forms of a third verb, *béo-n, '*to become', were co-opted to serve as the infinitive and participle forms for the *am-was* verb. In fact, in some dialects of English, *be* made a bid at taking over the whole paradigm: there were forms like *he beeth* and *thou beest* in the south. However, around 1500 the *am/are/is* group had solidified their hold on the standard dialect of English, and they've been part of the standard verb ever since.

The learnability of such forms is a deeper problem. One of the main clues that a word-learning child has that a novel concept is being discussed is whether or not the concept has a name that's different than any of the other words that he's learned so far. So, for instance, when you show a child two items, one of which he knows a word for and the other of which he doesn't, and ask him to hand you the "timp", he's likely to assume that "timp" is a word that names the unfamiliar object. He's using a heuristic—"new word form, therefore new meaning"—as a guiding principle in guessing new word meanings. Another way of putting this principle is to say that children assume that there are no true synonyms.

If children generally operate on that principle, then suppletive forms ought to be particularly difficult to learn. Imagine a child who knows *go*, but not *went*. He'll hear the new word, *went*, in some past tense context, and make a guess at its meaning. But, since he's using the new form/new meaning heuristic, he'll crucially assume that it can't mean exactly the same thing as *go*, which he already knows. In order to learn the true connection between *go* and *went*, the child has to notice that the form *goed* never shows up. Everywhere the child might expect to hear *goed*, he hears *went* instead. To get the connection between *go* and *went*, he must

notice the *absence* of an otherwise expected form—he has to learn from *negative evidence*. Theorists have long supposed that learning from negative evidence is close to impossible. Nonetheless, although suppletion is rare, and it tends to occur only in very frequent words, it does crop up repeatedly. It should become an important topic of language acquisition research in the near future.

5.6 Keeping Irregulars: An overview

One moral to our story so far is that 'irregulars' really are irregular. While there might be a phonological or a semantic clue which reminds a speaker of English that this word *might* be a member of an irregular class, the only sure way to know if a given root is morphologically irregular is to see it used in context, and notice and remember its behavior. For every kind of irregular pattern, there are exceptions: *wolf/wolves* but *gulf/gulfs*; **a rice* but *a bean*, **billiard/billiards* but *pool/*pools*.

Why do we retain irregular forms? Why don't we just forget all these tricky homosemes and do everything regularly? Instead of *many children,* we'd have *many childs*; rather than *feel/felt* we'd have *feel/feeled*, and instead of *stupidity* we'd have *stupidness*.

Indeed, when children are learning a language, one of the most interesting things they do is *overgeneralize*—apply a regular morpheme to a stem that normally selects for an irregular homoseme. Children at a certain stage of language acquisition will say *falled* instead of *fell*, *feets* instead of *feet*, and *sheeps* instead of *sheep*. After they've heard the irregular often enough, they'll memorize it, and stop overgeneralizing.

An advantage to memorizing some forms is that you can often produce them faster than you could if you had to perform the extra operation of separately looking up and adding an affix. For words that you need to access often—highly frequent words—a memorized form can speed up language processing. For highly frequent words, too, there'll be more opportunities for a learner to hear the irregular form, and learn that this word is different from most. Consequently, words tend to retain their irregular forms better the more frequent they are.

Let's take our stems that call for the irregular *-t* homoseme of the past tense, *mean, feel, learn, burn, dream, spell, spoil, smell,* and *spill*. We

noted above that some of them seem firmly attached to their allomorph, like *mean* and *feel*, some of them seem to be alternating with a regular form, like *learnt/learned* and *dreamt/dreamed*, and some are starting to sound downright archaic in the irregular, like *spoilt* and *smelt*. If the frequency hypothesis is correct, the more frequent a word is, the more likely it is to retain its irregular form, because of the greater opportunity that learners will have to memorize it.

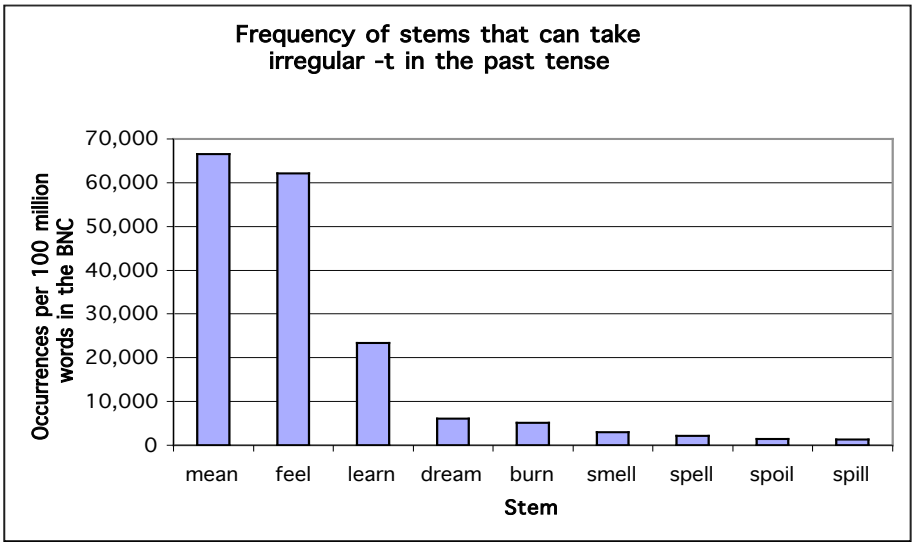
Looking in a list of the most frequent 8000 words in the British National Corpus, a collection of modern British English texts, we see that these verbs are ordered as follows, from most frequent to least frequent.

(88)

Verb	*# of occurrences in BNC*	*Frequency rank*
mean	66,556	134
feel	62,185	148
learn	23,394	432
dream	6,050	1580
burn	5,091	1829
smell	3,037	2680
spell	2,181	3300
spoil	1,455	4373
spill	1,296	4697

A bar chart of these numbers is given in Fig. 1., so you can see the frequencies graphically. The ranking correlates fairly well with my intuitions about how 'natural' the *-t* past tense is with each of these stems.

Fig. 1



Retention of an irregular form is also helped if a word has other characteristics that can help make memorization and retrieval of the irregular form faster than application of the regular affix. Any of the mnemonic clues that we considered above could help. For instance, if a certain irregular form is associated with several stems that have a similar phonological shape, the ease of access of that irregular might be increased—the frequency of all the similar-sounding stems could 'add up' to quite a large number, even if the frequency of any one form is low. Similarly, if a certain irregular form is associated with a particular general kind of meaning, the same effect could occur: the meaning association could help you zero in on the memorized listed form more quickly than one might expect given the basic frequency of the word.

To explain the importance of frequency for these and other phenomena, psycholinguists propose that listemes are sorted in the mind in order of frequency. The idea is that your mental lexicon is organized so that you can get to more frequently used listemes quickly. When you attempt to produce an inflected word, there's a sort of competition going on. You can't spend forever looking around for the right form in your mental lexicon. If you can sort through your lists and come up with an irregular form within a certain time window, then you produce an irregular. On the other hand, if the irregular listeme is farther down the

frequency list, so that it takes more than your allowed time window to find it, you will give up on the hunt and just go ahead and produce the default, regular form. A diagram of this word production model is given below:

(89) a. Think of a meaning to convey:

Speaker A	Speaker B
HE MEAN+PastTense IT	HE SPOIL+PastTense IT

b. Start looking for the right listemes. You're on the clock!

A finds:	B finds:
...	.....
/mɛn/ +	/spɔjl/ +
...	.....
/t/	nothing!

c.

A produces	B produces
meant	*spoiled*
(as soon as she finds it)	(after the clock runs out)

This predicts that irregular forms should vary in how fast they are produced according to how frequent they are, but that regular forms should be produced at the same speed no matter what their frequency is. When people are tested on how fast they are able to determine whether a particular form is a word or not[77], this prediction seems to be borne out.

5.7 Productivity, blocking and Bushisms

Some of the affixes we've been considering seem to hardly be 'alive' in English anymore. Affixes like *-ship* (as in *friendship, guardianship, kingship, partnership*) or *-ary* (as in *visionary, missionary,*

[77] This is called a 'lexical decision task'. The subject is asked to hit a button 'yes' when they're shown a word that exists in English, like *meant*, and hit a button 'no' when they're shown a word that doesn't exist (like *pimble*). Their reaction times on 'yes' are measured to see if there's any pattern among words in how quickly they can decide.

secretary, adversary) seem to occur primarily as part of a few, nonvarying words; it's not so often that someone will make up a noun ending in *-ship* or *-ary* out of thin air.

Regular affixes, though, are constantly used to form new words on the spot. Regular inflectional affixes, of course, are all used in this way, to inflect any word that comes along, whether it's made up or borrowed or whatever. Many derivational affixes are also used this way. President Bush used the suffixes *-ist* and *-ize* to make up *explorationist* and *securitize* on the spot. Perusing a few recent pages of the *New Yorker*, the *Tucson Weekly* and *The Nation*, I find the following nonce coinages, none of which are recognized by my spell-checker: *deroyalization, unfinishable, horkening, ginchy, non-city, Disneyfied,* and *regurgitant*. These words testify to the fact that that *-ation, un-*, *-able*, *-en*, *non-, -y, -ify* and *-ant* are all alive and well in the hands of professional creators of English prose.

Such affixes are termed *productive,* because, of course, they are used to produce new words on a regular basis. Productive affixes are the ones which, over time, squeeze out more infrequent irregulars. Irregular affixes are not productive—they only apply to a limited set of listed stems, and if you try to apply one to a form that's not on their list, you get something quite odd-sounding (consider *seem-semt,* built on the same principle as *dream/dreamt*).

Given the model of lexical production described above, when you hunt for a word form, and find an irregular in time, you won't produce the regular form, as in *mean/meant/*meaned*. In such cases, we can say the productive suffix is *blocked* by the irregular allomorph. When children overgeneralize and say *foots* instead of *feet*, or *mouses* instead of *mice*, it's because they don't know the irregular form well enough for it to block the regular one yet.

The principle of blocking can help us understand the funny-soundingness of a couple of the cases that we started with: George Bush's production of *analyzation* from *analyze* and *securitize* from *security*.

Here's a hypothesis about what happened during his production of these words. He wanted to convey the following meanings: a noun meaning the ACTION of ANALYZING and a verb meaning CAUSE to be SECURE. He rummaged through his listemes and found, close to the top, some frequent forms with the correct roots: *anal-yze* and *secur-ity*.

However, the former is a verb, and he needed the noun of action, and the latter is a noun, and he needed the causative verb form. He began the hunt for irregular forms with the right meaning, but ran out of time before he found the noun of action *analysis*, formed with the irregular nominalizing suffix *-sis* (like *diagnosis* and *hypnosis*), and before he found the causative verb *to secure*, formed with the null causative suffix *-Ø* (like *to open* or *to clear*). Instead, he used the appropriate productive Latinate suffixes with the right meanings, *-ation* and *-ize*, as suffixes on the more-frequent stems he had originally found, *analyze* and *security*.

There's no way to test this hypothesis directly, of course, but if lexical access of irregulars is determined by frequency, then we can at least test one prediction it makes: the noun *security* and the verb *analyze* should be more frequent than the irregular forms that failed to block Bush's overregularizations. For *secure/security* this is true: *security* is quite frequent—number 644 in the BNC top-8000 list—but the verb *to secure* is less frequent, ranked 1717. For this pair, then, our imagined sequence of events in Bush's language-generator is potentially plausible.

Unfortunately, the numbers for *analyze/analysis* go the wrong way: *analyze* is ranked 2166th of the BNC's 8000 most-frequent words, but the noun *analysis* is considerably more frequent, ranked 732. Can we come up with another idea to explain the failure of Bush's blocking mechanism for these?

Well, one thing that is immediately noticeable about *analysis* is that it sounds very different from *analyze*—the stress is in a different place, and consequently the vowels are reduced in a different pattern. *Analyze* is pronounced /ˈænəlajz/, while *analysis* is pronounced /əˈnæləsɪs/. Further, the nominalizing suffix *-sis*, although quite common in medical terminology, is relatively uncommon in everyday speech; only 8 of the 8000 most frequent words have it at all. Of these 8, there are only three that have causative verb forms in *-ize* besides *analysis*: *emphas-ize* from *empha-sis, hypothes-ize* from *hypothe-sis*, and *synthes-ize* from *synthe-sis*. For these three words, though, the change from the verbal *-ize* suffix to the nominal *-sis* suffix doesn't involve *any* change in pronunciation of the stem. In those words the *-sis* suffix is attached directly onto the same stem that *-ize* attaches to. In *analyze/analysis*,

however, the stem for *analyze* is *anal-* /ænəl/, while the stem for *analysis* is *analy-* , /ənælə/. Both the infrequency of the *sis* morpheme and the homosemic stem forms could have led Bush's word-analyzing machinery to conclude that *analysis* and *analyze* are not different forms of the same stem at all, but rather are separate listemes entirely. In that case, *analysis* would fail to block *analyzation* because it wouldn't even be in the competition for the noun-of-action form—and since *analyze* has the *-ize* suffix in it, *-ation* is in fact the only possible choice for the noun of action (see our lexical entry for *-ation* in section XX)

For English speakers whose lexical inventories do make the connection between *analyze* and *analysis*, and who are more familiar with the verb *to secure* than Bush is, Bush's failure to exhibit blocking in these cases sounds funny, like a child's failure to exhibit blocking with *foots* or *hitted*. It does, however, illustrate the fact that the order of listemes in an English speaker's mental inventory, as well as the particular set of listemes in there, will vary from person to person, depending on how much exposure to each listeme they have had, and on whether their word-analysis machinery has identified particular pieces as being related or not.

We've learned a lot about word *forms*. What about word *meanings*? We turn to this important topic in the next chapter.

XX to come: references, further reading, exercises.

Exercise on vowel change in old English
Exercise on finding the source of an irregular pattern applied to a regular verb.