

A **morpheme** is a grammatical unit composed of a phonological UR and a meaning. The progressive marker *-ing* in English is a morpheme: it has the UR /ɪŋ/ and the meaning PROGRESSIVE. Meanings are indicated either in single quotes (usually for substantive meanings, like nouns and verbs) or in small capitals (usually for grammatical functions, like plurality or tense).

Many morphemes have different **allomorphs**, pronunciations in particular environments. The past tense marker *-ed* in English has three allomorphs: [d], [t], and [əd]. Phonology intersects with morphology when determining what phonological conditions predictably trigger allomorphs.

A **root** or **stem** is a morpheme that determines the category and primary meaning of the word. Roots are usually nouns, verbs, adjectives, and prepositions. Various forms of a root can be made by adding **affixes**. Affixes attached before the root are called **prefixes** (like *un-* in English), while affixes attached after the root are called **suffixes** (like *-s* in English). Some languages like Tagalog also have **infixes**, which are affixes that are attached inside other morphemes: [bili] is the root meaning 'buy', but the infinitive 'to buy' is [bumili]. English also has infixation, as in *abso-bloody-lutely* and Homer Simpson's *edu-ma-cation* and *saxo-ma-phone*. Roots may also be combined together by **compounding**, as in English *blackboard*, *doghouse*, and *overcook*.

An affix is **productive** if it can be used with (almost) any root of the same type, especially new words. The nominalizer *-ness* in English is productive. Compare *-ness* to *-ity*, which has the same meaning, but is unproductive: *vain-ness*, *plain-ness*, *Jane-ness*; *vanity*, **planity*, **Janity*.

An affix is usually classified as **inflectional** if it does not change the category of the root (like noun pluralization, subject-verb agreement, verb tense, etc.), and as **derivational** if it does (like *-ness*, *-ish*, and *-ize*). Inflectional affixes are generally productive, while derivational affixes may or may not be productive.

Languages may be classified by their type of morphology. A language is **analytic** if it has no affixes (i.e., every word is one morpheme big). In Mandarin Chinese, [wɔ məŋ tən tʃin] means 'we are playing the piano (literally: 1PERS PL play piano)', while [wɔ məŋ tən tʃin lə] means 'we played the piano (literally: 1PERS PL play piano PAST)'. A **synthetic** language is not analytic; some words are composed of multiple morphemes. For example, in contrast to Mandarin, English indicates past tense with an affix rather than an independent word. Purely analytic languages are rare, so most languages are synthetic to some extent.

Synthetic languages may be further classified based on how meanings are combined in morphemes. In an **agglutinating** language, a given meaning is consistently expressed by one morpheme, in every environment. In Finnish, agreement on the verb for a first-person singular subject is always marked by [-n], regardless of the verb's tense. In a **fusional** language, however, meanings may combine in different ways, seemingly at random. In Russian, first-person singular agreement on a verb is marked by [-jo] in the present tense and [-l] in the past tense.

Finally, a **polysynthetic** language allows productive compounding. For example, in Sora, the single word [ɲɛnədʒdʒadarsɪəm] means 'I will not receive cooked rice from your hands', where [dʒa] means 'receive', [dar] means 'cooked rice', and [si] means 'hand'; all three of these are clearly roots (verbs and nouns). Compare Sora's polysynthetic compounding to English compounds: *blackboard* versus *whiteboard* and **greenboard*; *steakhouse* versus *doghouse* and **cowhouse*; *overcook* versus *overlook* and **overdie*. These English compounds are not productive and do not have consistent meanings. In Sora, however, you could substitute any noun for [dar] and get an analogous, compositional meaning.

Swahili verbs

Swahili is a Southern Bantoid language of the Niger-Congo family, spoken in East Africa. Note: no data is given for second-person plural (2PL) subjects or objects, so all instances of ‘you’ are second-person singular (2SG).

nitakupenda	‘I will like you’	atanipiga	‘s/he will beat me’
nitampenda	‘I will like him/her’	atakupiga	‘s/he will beat you’
nitawapenda	‘I will like them’	atampiga	‘s/he will beat him/her’
utanipenda	‘you will like me’	alinipiga	‘s/he beat me’
utampenda	‘you will like him/her’	alikipiga	‘s/he beat you’
atanipenda	‘s/he will like me’	alimpiga	‘s/he beat him/her’
atakupenda	‘s/he will like you’	amenipiga	‘s/he has beaten me’
atampenda	‘s/he will like him/her’	amekupiga	‘s/he has beaten you’
atatupenda	‘s/he will like us’	amempiga	‘s/he has beaten him/her’
atawapenda	‘s/he will like them’	ananipiga	‘s/he is beating me’
tutampenda	‘we will like him/her’	anakupiga	‘s/he is beating you’
watampenda	‘they will like him/her’	anampiga	‘s/he is beating him/her’
tulikulipa	‘we paid you’	unamsumbua	‘you are annoying him/her’
wametulipa	‘they have paid us’	atakusumbua	‘s/he will annoy you’

(Tserdanelis and Wong 2004)

Ístmo Zapoteco possessed nouns

Ístmo Zapoteco (a.k.a. Isthmus Zapotec) is a Zapotecan language of the Oto-Manguean family, spoken in Mexico.

<i>root</i>	<i>3SG POSS</i>	<i>2PL POSS</i>	<i>gloss</i>
palu	spalube	spalulu	‘stick’
ku:ba	sku:babe	sku:balu	‘dough’
tapa	stapabe	stapalu	‘four’
geta	sketabe	sketalu	‘tortilla’
bere	sperebe	sperelu	‘chicken’
do?o	sto?obe	sto?olu	‘rope’

(Tserdanelis and Wong 2004)

Tashlhiyt Berber

Tashlhiyt is a Berber language of the Afro-Asiatic family, spoken in Morocco. Tashlhiyt has two prefixes of interest, each with two forms: a reflexive [m-/n-] and an agentive [am-/an-].

mxazar	‘scowl at oneself’	nfara	‘disentangle oneself’
msaggal	‘look for oneself’	nhaffam	‘be shy to oneself’
m ^ɣ awar	‘ask oneself for advice’	nxalaf	‘place oneself crosswise’
mɣla	‘lose oneself’	nkaddab	‘consider oneself a liar’
amlas	‘one who shears’	anr ^ɣ mi	‘one who is tired’
amkrz	‘plower’	anbur	‘one who remains celibate’
amagur	‘one who remains’	and ^ɣ fur	‘follower’
amzug	‘one who absconds’	an ^ɣ azum	‘one who is fast’

(Alderete 1997)

Indonesian nasals (expanded from Data Handout #4)

Indonesian is a Malayo-Polynesian language of the Austronesian family, spoken in Indonesia.

bantu	məmbantu	'help'	pukul	məmukul	'hit'
dəjar	məndəjar	'hear'	tulis	mənulis	'write'
d̥ʒahit	mənd̥ʒahit	'sew'	t̥ʃatat	mənt̥ʃatat	'note down'
gambar	məŋgambar	'draw a picture'	kirim	məŋirim	'send'
hituŋ	məŋhituŋ	'count'	lempar	mələmpar	'throw'
ambil	məŋambil	'take'	rasa	mərasa	'feel'
isi	məŋisi	'fill up'	wakil	məwakili	'represent'
undang	məŋundang	'invite'	jakin	məjakini	'convince'
masak	məmasak	'cook'			
nikah	mənikah	'marry'			
ŋatʃo	məŋatʃo	'chat'			
ŋaŋi	məŋaŋi	'sing'			

(Halle and Clements 1983)

Rendaku, Lyman's Law, and ga-gyō bionka in Tokyo Japanese

Japanese is a language of the Japanese-Ryukyuan family (perhaps part of a larger linguistic group including Korean and/or the Altaic family), spoken in Japan. In compounds words made of *yamato* (native Japanese) vocabulary, a rule called *rendaku* applies. (Note that due to the peculiarities of Japanese phonology, [b] and [h] are treated as a voiced/voiceless pair. Similarly, but slightly less strangely, [d̥ʒ] and [ʃ] are also such a pair.)

kami	'paper'	ori-gami	'folding paper'
sumo:	'sumo'	o:-zumo:	'sumo tournament'
tera	'temple'	jama-dera	'mountain temple'
hana	'nose'	mizu-bana	'runny nose'
hata	'side'	kawa-bata	'river bank'

(Ito and Mester 1986)

However, *rendaku* is blocked in some compounds by Lyman's Law.

kaze	'wind'	kami-kaze	'divine wind'	*kami-gaze
tabi	'tabi (socks)'	jiro-tabi	'white tabi'	*jiro-dabi
ʃizuka	'tranquil'	mono-ʃizuka	'tranquility'	*mono-d̥ʒizuka
hadaka	'naked'	maru-hadaka	'completely naked'	*maru-badaka

(Ito and Mester 1986)

Further complicating matters is a rule in the Tokyo dialect called *ga-gyō bionka*, in which /g/ is lenited through nasalization to [ŋ].

geta	'clogs'	*ŋeta	toŋe	'thorn'	*toge
giri	'debt'	*ŋiri	kaŋo	'basket'	*kago
gutʃi	'complain'	*ŋutʃi	sasaŋeru	'give'	*sasageru

(Ito and Mester 1996)

In the Tokyo dialect, *ga-gyō bionka* causes /ori-gami/ 'folding paper' to be pronounced as [ori-ŋami], instead of *[ori-gami].

In addition, *ga-gyō bionka* alters the blocking effect of Lyman's Law, causing /bara-toge/ 'rose thorn' to be pronounced [bara-toŋe], instead of expected *[bara-doŋe] (cf. [mizu-bana] runny nose, in which Lyman's Law and *rendaku* interact normally).

Yowlumne Yokuts verbs

Yokuts is a Yok-Utian language of the Penutian family, spoken in California. Yowlumne Yokuts (a.k.a. Yawelmani Yokuts) is the dialect that was spoken in the southeastern San Joaquin Valley but became extinct in 1960 upon the death of its last speaker, Wahumcha. Some data here has been reconstructed based on regularity with other forms.

	INDIRECT			INDIRECT			
FUTURE	NONFUTURE	NONFUTURE	GERUND	DUBITATIVE	IMPERATIVE	IMPERATIVE	gloss
xaten	xathin	xatsithin	xatmi	xatal	xatka	xatsitka	'eat'
maxen	maxhin	maxsithin	maxmi	maxal	maxka	maxsitka	'procure'
xilen	xilhin	xilsithin	xilmi	xilal	xilka	xilsitka	'tangle'
boken	bokhin	boksithin	bokmi	bokol	bokko	boksitka	'find'
koʔen	koʔhin	koʔsithin	koʔmi	koʔol	koʔko	koʔsitka	'throw'
dubon	dubhun	dubsuthun	dubmu	dubal	dubka	dubsutka	'lead by hand'
tulon	tulhun	tulsuthun	tulmu	tulal	tulka	tulsutka	'burn'
la:nen	lanhin	lansithin	lanmi	la:nal	lanka	lansitka	'hear'
me:ken	mekhin	meksithin	mekmi	me:kal	mekka	meksitka	'swallow'
do:sen	doshin	dossithin	dosmi	do:sol	dosko	dossitka	'report'
wo:nen	wonhin	wonsithin	wonmi	wo:mol	wonko	wonsitka	'hide'
paʔten	paʔithin	paʔitsithin	paʔitmi	paʔtal	paʔitka	paʔitsitka	'fight'
lihmen	lihimhin	lihimsithin	lihimmi	lihmal	lihimka	lihimsitka	'run'
logwen	logiwhin	logiwsithin	logiwmi	logwol	logiwka	logiwsitka	'pulverize'
ʔugnon	ʔugunhun	ʔugunsuthun	ʔugunmu	ʔugnal	ʔugunka	ʔugunsutka	'drink'
ʔamlen	ʔa:milhin	ʔa:milsithin	ʔa:milmi	ʔamlal	ʔa:milka	ʔa:milsitka	'help'
salken	sa:likhin	sa:liksithin	sa:likmi	salkal	sa:likka	sa:liksitka	'wake up'
mo:jnen	mo:jinhin	mo:jinsithin	mo:jinmi	mo:jnol	mo:jinka	mo:jinsitka	'get tired'
co:mon	comhun	comsuthun	commu	co:mal	comka	comsutka	'destroy'
woʔjon	wo:ʔujhun	wo:ʔujsthun	wo:ʔujmu	woʔjal	wo:ʔujka	wo:ʔujstka	'fall asleep'

(Newman 1944)