

English lexicology

Lecture 2. The International Structure of Words

- 2.1 Defining the word
- 2.2 Morphemes
- 2.3 morphs
- 2.4 Roots, bases, stems
- 2.5 Affixes
- 2.6 The analysis of words into morphs and morphemes
- 2.7 Allomorphs and morphemic rules

2.1 Defining the Word

Figure 1



There are **formal criteria** for wordhood which all speakers use:

1. **Orthographic**: a word is what occurs between spaces in writing.
2. **Semantic**: a word has semantic coherence; it expresses a unified semantic concept.
3. **Phonological**:
 - a. potential pause: a word occurs between potential pauses in speaking. Though in normal speech, we generally do not pause, we may potentially pause between words, but not in the middle of words.
 - b. stress: a word spoken in isolation has one and only one primary stress.
4. **Morphological**: a word has an internal cohesion and is indivisible by other units; a word may be modified only externally by the addition of suffixes and prefixes.
5. **Grammatical**: words fall into particular classes.
6. **Syntactic**: a word has external distribution or mobility; it is moved as a unit, not in parts.

By **the criterion of orthography**, *supermarket* and *noteworthy* would be considered a single word, as would hyphenated forms such as *jack-of-all-trades*, *forget-me-not*, or *runner-up*, while phrases such as *travel agency*, *take out*, or *pins and needles* must be considered as multiple words, or phrases.

Yet by **the second criterion, semantic unity**, the words and the phrases all appear to be equally unified conceptually. The discrepancy is especially apparent if you compare *supermarket* with related concepts such as *toy store* or *grocery store*. In fact, **the conventions of spacing between words, as well as hyphenation practices, are often quite arbitrary in English.**

As well as being hyphenated, *forget-me-not*, *jack-of-all-trades*, and *runner-up* meet **the syntactic criterion** of wordhood: they are moved as a single unit.

However, they differ in respect to **the morphological criterion**; while *forget-me-not* always behaves as a single word, with external modification (*two forget-me-nots*, *forget-me-nots*), *runner-up* is inconsistent, behaving as a single word when made possessive (*runner-up's*), but as a phrase, that is, with internal modification, when pluralized (*runners-up*); *jack-of-all-trades* is similarly inconsistent.

The third criterion, a single primary stress, would seem to be the most reliable, but even here compound adjectives such as *noteworthy* pose a problem: they have two primary stresses and are phonologically phrases but **are treated orthographically, morphologically, and syntactically as single words**.

Phrasal verbs such as *try out* also present an interesting case. Though having many of the qualities of a phrase – internal modification occurs (*tried out*), material may intercede between the parts (*try out the car*, but also *try the car out*), and both *try* and *out* receive primary stress – **phrasal verbs seem to express a unified semantic notion**, the same as expressed in this case by the single word *test*.

Another difficulty when treating words is the term **word** itself, which may be used in a number of different ways:

1. It may refer to the word form, the physical unit or concrete realization, either the orthographical word (the written form) or the phonological word (the uttered or transcribed form).

2. It may refer to the lexeme, which is rather like a dictionary entry. A lexeme includes all inflected forms of a word. It is thus a kind of abstraction or class of forms and is indicated by small capitals, as in the following examples:

walk – walk, walks, walked, walking

run – run, runs, ran, running

sing – sing, sings, sang, sung, singing

The same word form may in fact represent different lexemes:

- a. A **homonym** is a single orthographic and phonological word standing for two lexemes, as *bear* is either the verb or the noun.
- b. A **homograph** is a single orthographic word (but separate phonological words) standing for two lexemes, as *lead* is either the noun /lɛd/ or the verb /lid/.
- c. A **homophone** is a single phonological word (but separate orthographical words) standing for two lexemes, as /mit/ is either the noun *meat* or the verb *meet*.

The same lexeme might also have quite distinct word forms, as in the case of the definite article *the*, represented by /ði/ or /ðə/, or the indefinite article *a/an*, represented by /ei/, /ə/, /ən/, or /æn/.

3. *Word* may also refer to a morphosyntactic word (or grammatical word).

A morphosyntactic word consists of a lexeme and associated grammatical meaning:

I take the garbage out every week. (take + present)

I took the garbage out yesterday. (take + past)

I have taken the garbage out already. (take + past participle). The different morphosyntactic words are represented by different word forms (take, took, taken).

But in

I put the garbage out every week. (put + present)

I put the garbage out yesterday. (put + past)

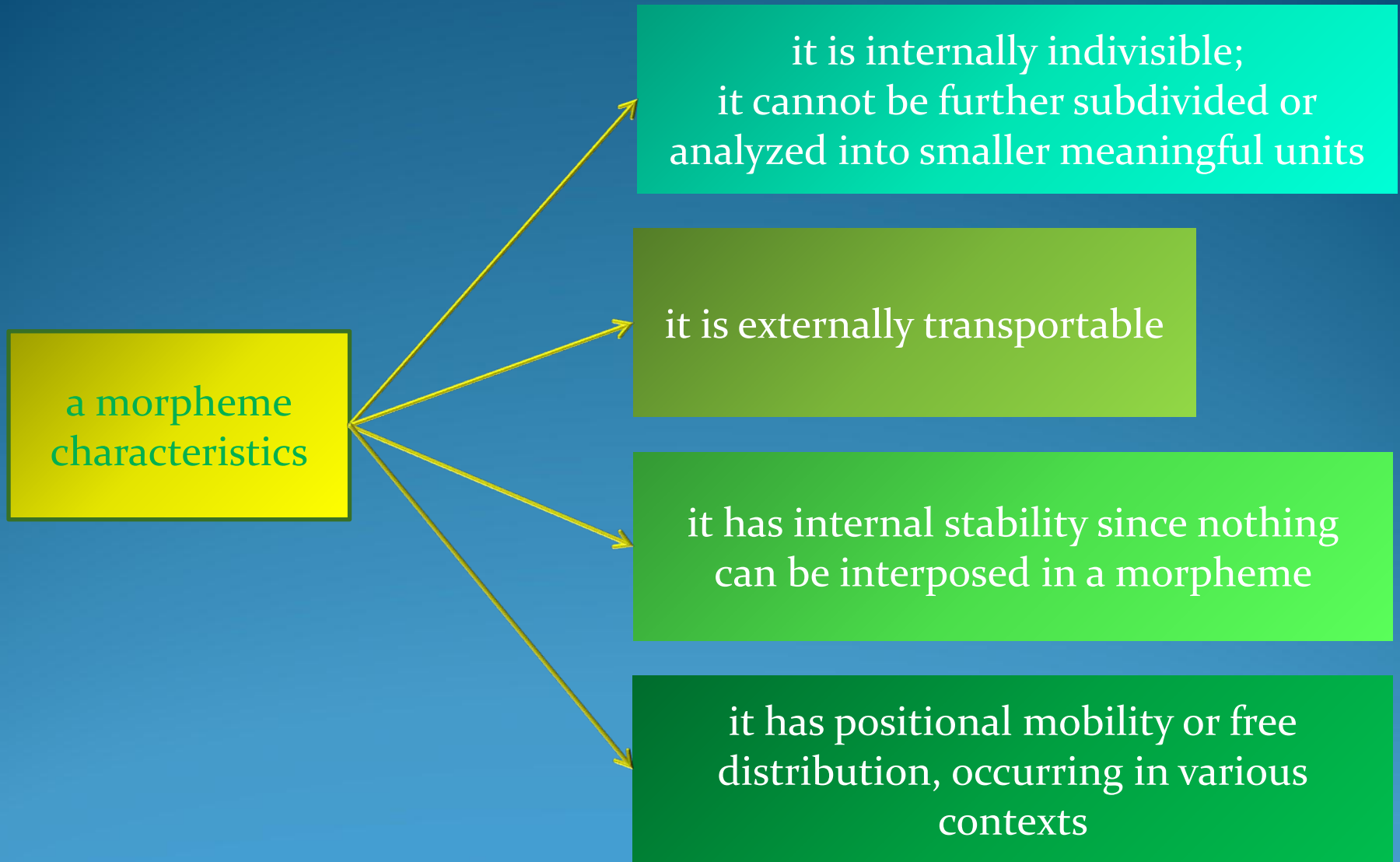
I have put the garbage out already. (put + past participle). The different morphosyntactic words are represented by the same word form (put).

2.2 Morphemes

A morpheme is the smallest meaningful unit in a language. The word *headphones* consists of the three morphemes *head*, *phone*, and *-s*; the word *ringleader* consists of three morphemes, *ring*, *lead*, and *-er*.

Some of these morphemes may stand alone as independent words (*head*, *phone*, *ring*, *lead*), others must always be attached to some other morpheme (*-er*, *-s*).

Morphemes are represented within curly braces { }.

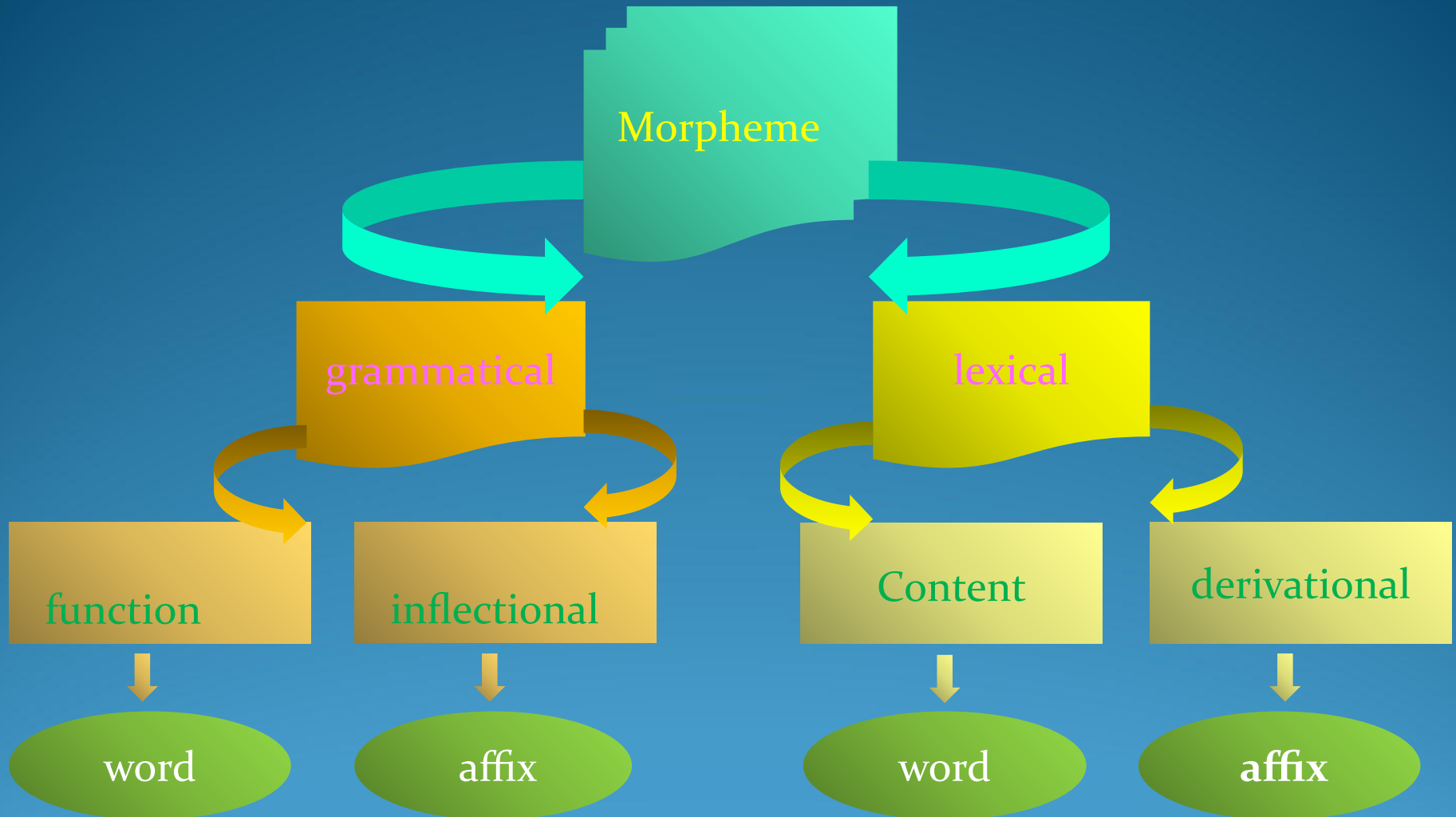


The morpheme refers to either a class of forms or an abstraction from the concrete forms of language.

For example, the feminine morpheme is an abstraction which can be realized in a number of different ways, by *-ess*, as in *actress*, or by a personal pronoun such as *she*.

Because morphemes are abstractions we place them within curly braces { } using capital letters for lexemes and descriptive designations for other types of morphemes. For *actress*, the morphemes are {ACTOR} and {f} (for {feminine}).

Figure 3. Types of Morphemes



Lexical morphemes express lexical, or dictionary, meaning. They can be categorized into the major lexical categories, or word classes: noun, verb, adjective, or adverb; these are frequently called “content words”.

They constitute open categories, to which new members can be added.

Lexical morphemes are generally independent words (free roots) or parts of words (derivational affixes and bound roots).

Grammatical morphemes express a limited number of very common meanings or express relations within the sentence.

They do not constitute open categories; they can be exhaustively listed. Their occurrence is (entirely) predictable by the grammar of the sentence because certain grammatical meanings are associated with certain lexical categories, for example, tense and voice with the verb, and number and gender with the noun.

Grammatical morphemes may be parts of words (inflectional affixes) or small but independent “function words” belonging to the minor word classes: preposition, article, demonstrative, conjunction, auxiliary, and so on, e.g. *of, the, that, and, may*.

2.3. Morphs

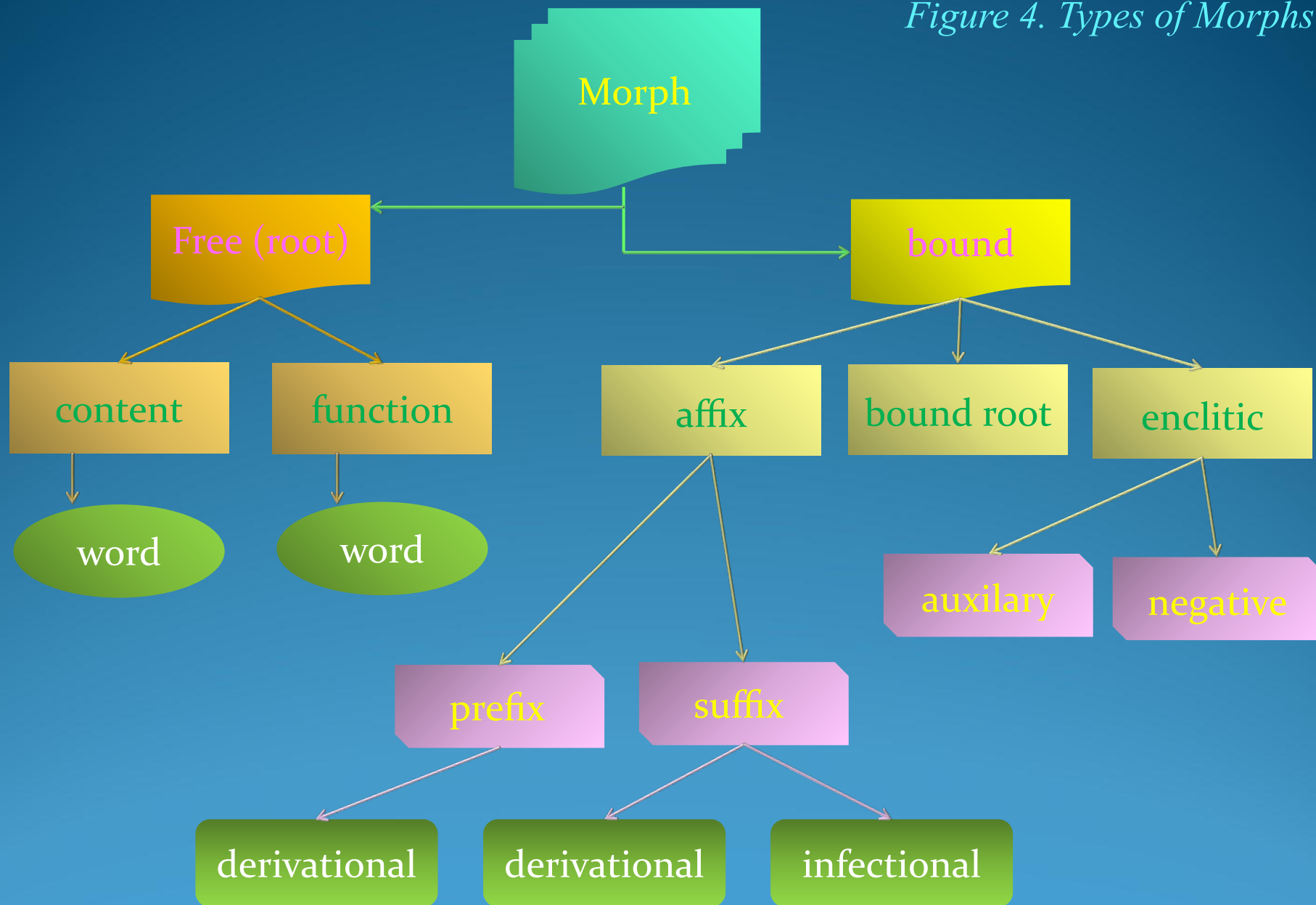
The level of the morph is the concrete realization of a morpheme, or the actual segment of a word as it is spoken or pronounced. Morphs are represented by phonetic forms.

We must introduce the concept of the morph distinct from the morpheme because sometimes although we know that a morpheme exists, it has no concrete realization (i.e, it is silent and has no spoken or written form). In such cases, we speak of a zero morph, one which has no phonetic or overt realization.

There is no equivalent on the level of the phoneme. For example, plural *fish* consists of the morphemes {fish} + {pl}, but the plural morpheme has no concrete realization (i.e. the singular and plural forms of *fish* are both pronounced /fɪʃ/).

Another example of a zero morph is the past tense of *let*; although the past consists of the morpheme {let} plus the morpheme {past}, the past tense morpheme has no concrete expression (i.e. the present and past forms of *let* are both pronounced /lɛt/).

Figure 4. Types of Morphs



A free morph may stand alone as a word, while a bound morph may not; it must always be attached to another morph.

A free morph is always a root. That is, it carries the principal lexical or grammatical meaning. It occupies the position where there is greatest potential for substitution; it may attach to other free or bound morphemes.

2.4 Roots, Bases, Stems

A **root** is often distinguished both from a **base** (a root plus associated derivational affixes, to which derivational affixes are added) and from a **stem** (a root plus associated derivational affixes, to which inflectional affixes are added). Thus, in the word *engagement*, *gage* is the root, *engage* is the base, and *engagement* is the stem.

unavoidably *overgrewn* *altruistic*
decoration *provocative* *disheartened*
reclassify *heterosexuality* *upbringing* *realization*

Roots are also occasionally bound morphs (called bound roots). Bound roots are often foreign borrowings that were free in the source language, but not free in English:

-vert convert, revert, subvert, introvert, pervert

-mit transmit, commit, remit, admit, omit, submit

-ceive conceive, perceive, receive, deceive

-fer transfer, refer, prefer, defer, confer

However -vert, -mit, -ceive, and -fer cannot stand alone as independent words, and we would also find it very difficult to state the meaning of any of these roots, unless we know Latin, from which these words derive: -vert is from Latin *vertere* meaning ‘to turn, -mit is from Latin *mittere* meaning ‘to send’, -ceive is from Latin *capere* meaning ‘to seize’, and -fer is from Latin *ferre* meaning ‘to bring’.

Bound roots may also be native English, as with *-kempt* (< *unkempt*) and *-couth* (< *uncouth*), where the positive form no longer exists.

You could say that the bound roots have a meaning only if you know their history, or etymology. For this reason, they have been termed etymemes.

2.5. Affixes

Unlike a root, **an affix does not carry the core meaning**. It is always **bound to a root**. It occupies a position where there is limited potential for substitution; that is, **a particular affix will attach to only certain roots**.

English has two kinds of affixes, **prefixes, which attach to the beginnings of roots**, and **suffixes, which attach to the end of roots**.

Some languages regularly use **infixes, which are inserted in the middle of words**.

In Modern English, infixes are used only for humorous purposes, as in *im-bloody-possible* or *abso-blooming-lutely*.

While it might initially be tempting to analyze the vowel alternation indicating plural (as in *man, men*) or past tense (as in *sing, sang*) in Modern English as **a kind of infix, the vowels are not added or inserted but actually replace the existing vowels**.

Affixes may be of two types, derivational or inflectional, which have very different characteristics.

A derivational affix in English is either a prefix or a suffix. There may be more than one derivational affix per word. A particular derivational affix may attach to only a limited number of roots; which roots it attaches to is not predictable by rule, but highly idiosyncratic and must be learned.

A derivational affix has one of two functions:

- 1) to convert one part of speech to another (in which case, it is class changing) and/or
- 2) to change the meaning of the root (in which case, it is class maintaining).

Such affixes function, then, in word formation and are important in the creation of new lexemes in the language. They always precede an inflectional affix.

An inflectional affix in English is always a suffix.

A particular inflectional affix attaches to all (or most) members of a certain word class.

The function of inflectional affixes is to indicate grammatical meaning, such as tense or number.

Because grammatical meaning is relevant outside the word, to the grammar of the entire sentence, inflectional affixes always occur last, following the root and any derivational affixes, which are central to the meaning or class of the root.

Table 1. Derivational vs. Inflectional Affixes in English

Derivational affixes	Inflectional affixes
<ul style="list-style-type: none"> ❖ either prefixes or suffixes ❖ optionally more than one per word ❖ attach idiosyncratically to only a limited number of roots ❖ have two functions <ol style="list-style-type: none"> 1) to convert one part of speech to another 2) to change the meaning of the root ❖ precede the inflectional suffix 	<ul style="list-style-type: none"> ❖ only suffixes ❖ only one per word ❖ attach to all (or most) members of a word class ❖ have one function <ol style="list-style-type: none"> 1. to indicate grammatical meaning ❖ follow derivational suffix(es)

A distinction can be made between **productive inflections**, which would attach to any new word entering the language to express a particular grammatical category, and **nonproductive**, or **remnant**, inflections, which are found on select members of a class, but would never be added to a new word.

There are only **eight productive inflections in Modern English**, as shown in Table 2.

Some examples of nonproductive inflections are the plural vowel alternation *tooth-teeth*; the *-most* superlative of *foremost*; the *-en* past participle of *write-written*; or the past tense vowel alternation of *ring-rang*.

Table 2. The Productive Inflections of Modern English

plural number	-s	Noun
possessive case	-s	
present (nonpast) tense, 3 rd . p. sg.	-s	Verb
past tense	-ed	
past participle	-ed	
present participle	-ing	
comparative degree	-er	Adjective
superlative degree	-est	

An enclitic is a kind of contraction, a bound form which derives from an independent word and must be attached to the preceding word.

In English, we have two kinds of enclitics: contracted auxiliaries, which are attached to the preceding subject, and the negative contraction -n't, which is attached to the preceding auxiliary.

Certain auxiliaries (e.g. *may, can, must, should, might, was*) cannot be contracted, while the contraction of *not* produces marginally acceptable forms in some cases (e.g. *?mayn't, ?mightn't*) or unacceptable forms in other cases (e.g. **am't*).

Contracted auxiliaries . Contracted negatives -n't

will, shall > 'll *won't, ?shalln't*

would, had > 'd *wouldn't, hadn't*

is, has > 's *isn't, hasn't*

have > 've *haven't*

am > 'm **am't (ain't)*

are > 're *aren't*

*was > *'s* *wasn't*

Words are analyzed morphologically with the same terminology used to describe different sentence types:

- 1) a **simple word** has one free root, e.g. *hand*;
- 2) a **complex word** has a free root and one or more bound morphs e.g. *unhand, handy, handful*, or it has two or more bound morphs, e.g. *transference, reception, conversion*
- 3) a **compound word** has two free roots, e.g. *handbook, handrail, handgun*; and
- 4) a **compound-complex** word has two free roots and associated bound morphs, e.g. *handwriting, handicraft*.

2.6 The analysis of words into morphs and morphemes

The importance of the distinction between morph and morpheme is that there is not always a one-to-one correspondence between morph and morpheme, and morphemes can combine or be realized in a number of different ways.

We can analyze words in two different ways:

- 1) into morphs following formal or structural divisions, or
- 2) into morphemes, recognizing the abstract units of meaning present.

Table 3. The two analyses of each of the words

Morphs		Morphemes
<i>Writers</i>	3 morphs <i>writ/er/s</i>	3 morphemes {write} + {-er} + {pl}
<i>authors</i>	2 morphs <i>author/s</i>	2 morphemes {author} + {pl}
<i>mice</i>	1 morph <i>mice</i>	2 morphemes {mouse} + {pl}
<i>fish</i>	1 morph <i>fish</i>	2 morphemes {fish} + {pl}
<i>children</i>	2 morphs <i>child/ren</i>	2 morphemes {child} + {pl}
<i>teeth</i>	1 morph <i>teeth</i>	2 morphemes {tooth} + {pl}
<i>mans</i>	2 morphs <i>man/s</i>	2 morphemes {man} + {poss}
<i>mens</i>	2 morphs <i>men/s</i>	3 morphemes {man} + {pl} + {poss}

Inflectional morphemes can often be realized by a number of different forms, or the same form may denote a number of different inflectional morphemes.

Therefore, it is usual to use descriptive designations for inflectional morphemes, such as {pl} (rather than {-s} or {-es}) for the plural marker and {poss} (rather than {-s}) for the possessive marker.

A noun such as *sheep* raises a difficulty for morphemic analysis, since it is either singular or plural. Should we postulate two morphemic analyses?

{sheep} + {pl}

{sheep} + {sg}

Let us look at how **morphological and morphemic analysis** works in **adjectives**:

Morphs

Morphemes

smaller 2 morphs *small/er*

2 morphemes {small} + {compr}

smallest 2 morphs *small/est*

2 morphemes {small} + {supl}

better 1 morph *better*

2 morphemes {good} + {compr}

best 1 morph *best*

2 morphemes {good} + {supl}

For **verbs**, the two analyses work as follows:

	Morphs	Morphemes
<i>worked</i>	2 morphs <i>work/ed</i>	2 morphemes {work} + {past}
<i>wrote</i>	1 morph <i>wrote</i>	2 morphemes {work} + {pstprt}
<i>written</i>	1 morph <i>written</i>	2 morphemes {write} + {past}
<i>working</i>	2 morphs <i>work/ing</i>	2 morphemes {write} + {pstprt}
<i>put</i>	1 morph <i>put</i>	2 morphemes {work} + {prsprt}
		3 morphemes {work} + {gerund} + {sg}
		2 morphemes {put} + {past}
		2 morphemes {PUT} + {pstprt}

Here, pstprt = past participle,
prsprt = present participle

Note that we have to analyze *-ing* verbal forms not only as present participles, but also as “gerunds”. Since gerunds are functioning as nouns, they may sometimes be pluralized, e.g.:

readings 3 morphs *read/ing/s* 3 morphemes {read} + {gerund} + {pl}

We need to postulate a morpheme {pres}, which is never realized, **to account coherently for the distinction past versus present:**

work 1 morph *work* 2 morphemes {work} + {pres}

write 1 morph *write* 2 morphemes {write} + {pres}

The morphemic analysis of pronouns is somewhat more complicated:

	Morphs	Morphemes
<i>we</i>	1 morph <i>we</i>	3 morphemes {1st p} + {pl} + {nomn}
<i>him</i>	1 morph <i>him</i>	4 morphemes {3rd p} + {sg} + {m} + {obj}
<i>its</i>	2 morphs <i>it/s</i>	4 morphemes {3rd p} + {sg} + {n} + {poss}

Here, nomn = nominative case and obj = objective case

Morphemes combine and are realized by one of four morphological realization rules:

1) **agglutinative rule**: two morphemes are realized by morphs which remain distinct and are simply “glued” together, e.g. {writer} + {pl} > *writers*

2) **fusional rule**: two morphemes are realized by morphs which do not remain distinct but are fused together, e.g. {tooth} + {pl} > *teeth*

3) **null realization rule**: a morpheme is never realized as a morph in any word of the relevant class, e.g. {sg} on nouns, which never has concrete realization in English.

4) **zero rule**: a morpheme is realized as a zero morph in particular members of a word class, e.g. {sheep} + {pl} > *sheep*. Note that in most other members of the class noun, {pl} has concrete realization as *-s*.

agglutinative	{work} + {past} > <i>worked</i>
fusional	{write} + {past} > <i>wrote</i>
null	{work} + {pres} > <i>work</i>
zero	{put} + {past} > <i>put</i> , {put} + {pstprt} > <i>put</i>
fusional & agglutinative	{man} + {pl} + {poss} > <i>men's</i>

Distinguishing between the concept of a **null rule** and a **zero rule** can be difficult.

Remember that **in the case of a null rule, the morpheme is never concretely realized**. For example, {pres} on verbs is always unmarked. No verb has an overt marker of the present.

In contrast, **when a morpheme is usually concretely realized, but is not realized on certain words, then we have a zero rule**. For example {pl} on nouns is typically expressed by *-s*, but on a noun such as *deer*, it is not marked and hence a zero rule.

2.7 Allomorphs and morphemic rules

Morphemes have predictable variants called allomorphs. Allomorphs are the members of the class, morpheme, or the phonetic realizations of the abstraction, morpheme.

Allomorphs are semantically similar and in complementary distribution. They needn't be phonologically similar, however.

Allomorphs are predicted, or “conditioned”, in one of three ways:

- 1) the appearance of a particular allomorph is predictable from the phonetic environment, hence **phonologically conditioned**;
- 2) the appearance is unpredictable phonologically but is determined by the grammar of the language, hence **grammatically conditioned**; or
- 3) the allomorphs are used interchangeably in all environments, hence in **free variation**.

Table 4. Regular Plural Formation in Nouns

A	B	C	
bushes /ʃ/	maps /p/	knobs /b/	rays /ei/
buses /s/	cats /t/	rods /d/	sofas /ə/
mazes /z/	racks /k/	logs /g/	toys /ɔi/
judges /dʒ/	ropes /p/	seals /l/	keys /i/
matches /tʃ/	laughs /f/	mirrors /r/	news /iu/
boxes /s/	paths /θ/	pans /n/	lathes /ð/
garages /ʒ/		tombs /m/	coves /v/
rouges /ʒ/		rings /ŋ/	

Table 4 gives the forms of noun plural in English that are phonologically conditioned, but certain noun plurals are **grammatically conditioned**:

Ø	<i>fish, sheep, deer</i>
vowel alternation	<i>mice, lice, geese</i>
-en	<i>children, brethren, oxen</i> foreign plurals
foreign plurals	
-a	<i>phenomena, data, criteria</i>
-i	<i>stimuli, foci</i>
-ae	<i>alumnae, formulae</i>
-ices	<i>indices, appendices</i>
-es	<i>bases, axes</i>
-im	<i>kibbutzim, cherubim</i>

These endings are not productive: they are either linguistic fossils (remnant forms from an earlier stage of English) or **foreign borrowings**. Note that if a noun such as *mouse* took a productive ending, it would be the [əz] allomorph, *child* would take /z/, and *tooth* would take /s/.

Let's look at one set of forms that does not seem to follow the morphemic rule for plural allomorphs given above. We would expect the plural allomorph of words ending in /f/ (a voiceless non-sibilant consonant) to be /s/, as in the following words:

belief – beliefs;

chief – chiefs;

proof – proofs;

safe – safes

However, what we find in the following set of forms is not /s/, but instead the plural allomorph /z/, with a simultaneous voicing of the final root consonant:

wolf – wolves *leaf – leaves*

knife – knives *loaf – loaves*

sheaf – sheaves *wife – wives*

elf – elves *life – lives*

shelf – shelves *calf – calves*

thief – thieves *self – selves*

In some cases, we also find **variation between the phonologically expected and unexpected forms**:

wharf – wharfs/wharves *dwarf – dwarfs/dwarves*

hoof – hoofs/hooves *scarf – scarfs/scarves*

A similar irregularity appears in the following words ending in /s/; the expected /əz/ allomorph is found, but there is also voicing of the final root /s/:

house – houses *blouse – blouses*

We could have a morphological realization rule which changes final voiceless fricatives to voiced fricatives when {pl} is added. However, such a rule would have to apply generally to all roots ending in voiceless fricatives, and it does not.

Instead, we say that **there are two predictable variants of the root, what is called root allomorphy**. The two allomorphs of the root are grammatically conditioned, by the presence of either a following {sg} and {pl} morpheme. The rule for *leaf/leaves* is as follows:

{lif} → [liv]/ – {pl}

[lif]/ elsewhere

Note that “elsewhere” would include the environment before both {sg} and {poss}. Hence, **this form has the widest distribution**. Actually, the –{pl} environment is too restricted since we also have voicing when a verb is formed from the noun (for example, *to shelve, to calve, to halve*).

A similar kind of root allomorphy is thus seen in cases of shifts from noun to verb where

- (a) the nominal forms have /s/ and the verbal forms have /z/, or
- (b) the nominal forms have /θ/ while the verbal forms have /ð/

Table 5. Root Allomorphy

a. N: /s/	V: /z/	b. N: /θ/	V: /ð/
house	to house	bath	to bathe
blouse	to blouse	cloth	to clothe
use	to use	breath	to breathe
excuse	to excuse	mouth	to mouthe
advice	to advise	teeth	to teethe
abuse	to abuse	wreath	to wreathe

Bound roots may show root allomorphy; for example, *-cept* is a predictable variant of *-ceive* before *-ion*, as in *conception*, *perception*, *reception*, and *deception*.

Generally, English is not rich in allomorphy.

