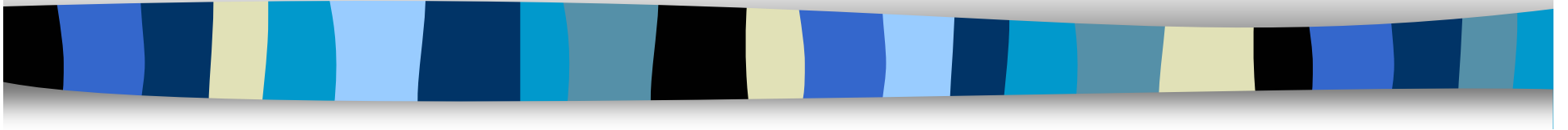


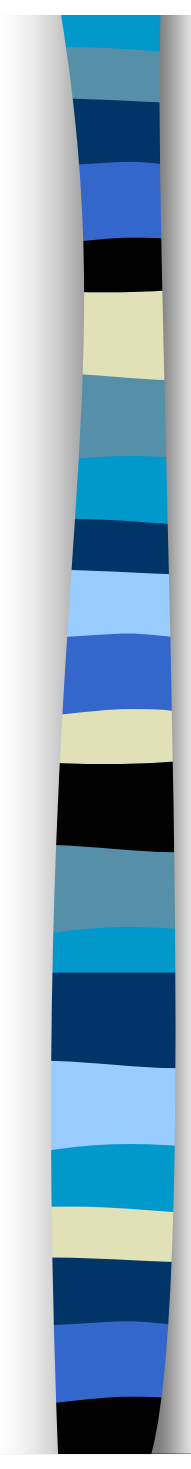
Introduction to Morphology:2





Words and morphology

- What's a word?



Word: a natural level of “chunking” in a language that’s about 1 to 5 segments in length, roughly, and its length is inversely related to its frequency.

It is perfectly clear in most languages that there exists a unit of organization at this level – that is, that words exist! But it is very hard to come up with a definition that works across all languages – or even perfectly in a single one!



Two more terms

Morphology: *the internal organization of words*, and the study of that organization.

Lexicon: the organized collection of words in a language. It has *organization* that is manifested in two ways: (a) redundancy – that's what we talked about in class last time; and (b) productivity: creating new words.



Breaking up sentences into words

- It is deceptively easy for us to break utterances into words in a language for which we already know the standard decisions. (Like for that sentence...)
- How were these decisions made? How will we make them, or justify them, in a new language? How do we justify *changes* in the decision as to what is a word in a language?



Sometimes it's the pronunciation

- *It's* from *it is*
 - I'm *gonna* (or *gunna*) do it this time.
- But this can be very deceptive. By the best linguistic analysis, *it's* and *we'll* are not *words* in the usual sense at all.



What's a word, phonologically?

- Some languages: one stress per word. This works, sometimes. Swahili is a pretty good case:
 - Mimi ni-na-sem-a Ki-Swahili kidogo tuu, with penultimate stress on each word.
- But it won't work in French, and it even in Spanish...



Spanish

- even in Spanish, you get one stress per word, but no stress on most “non-lexical” words:
- Sólo se que se fue. “I only know that she left”



Vowel harmony

In a number of languages, all vowels in a word agree in a specific vowel dimension (front/back, high/non-high)

- otta-a ‘s/he takes’
otta-vat ‘they take’
otta-vat-ko ‘do they take?’
- pitä-ä ‘s/he likes’
pitä-vät ‘they like’
pitä-vät-kö ‘do they like?’

But not in compounds, and not in most languages.



Sometimes it's non-compositional semantics...

- When the whole means something other than the “sum” of its parts (i.e., the normal composition of the meaning of its parts)
- This is a frequent characteristic of compounds: lighthouse,



Hyphens

- Lots of times we're really not sure – and we have that tricky old hyphen to deal with.
- *A come-hither look, a once-in-a-lifetime opportunity.* Are those words that the hyphen makes? I don't think so.



Sometimes a language marks its words clearly with phonology

- Demarcative stress: penultimate stress in Swahili:
 - Unaséma (you speak)
 - Unasemáje? (do you speak?)
- Vowel harmony:
 - All vowels in a word agree for frontness (or some other characteristics). This is relatively rare.



Syntax and morphology help define each other...

- In English, basic sentences are composed of a subject noun phrase followed by a verb phrase:
- John jumped.
- The big dog chewed on the bone.
- A baby who was lost cried in the playground.



Syntax does not look inside of words

- Words are the units whose placement in a sentence is explained by *syntax*; smaller units than these words (i.e., subword units) are explained by morphological structure.
- We're set, if the principles of syntactic analysis and morphological analysis are pretty different. That appears to be true.



Syntax defines units like *noun*

- The noun can be preceded by an article (determiner) like *the*, followed by one or more adjectives,
- And followed by a verb.

The dog barked.

- But the syntax says nothing about how *complex* that noun or adjective can be:

The antediluvian opera-singer



Some languages do give word demarcations

- Why do languages with Latin-based writing do this?
- One theory (Paul Saenger) is that pre-Latin alphabets that did not mark vowels did mark word boundaries; and the introduction of vowel indications made word-boundaries superfluous.

- 
- Spaces came in in the Western tradition in the 7th century in Irish texts of Latin.



Recent work on automatically detecting words

- Carl de Marcken (1996) *Unsupervised Language Acquisition*



OK! So even *justifying* where the word breaks are is hard.

- Suppose we could *do* that. Would we want to say that a word is anything that is found between word-breaks (i.e., spaces)?



Probably not.

- That's both too loose and too strong.
- Laurie Bauer's example: How many words are there in:
 - The cook was a good cook as cooks go, and as cooks go, he went.
- By one count, there are 15. But how many *different* words?



Probably not.

There is a sense in which cook and cooks are different forms of the same word

- That's both too loose and too strong.
- Laurie Bauer's example: How many words are there in:
 - The **cook** was a good **cook** as **cooks** go, and as **cooks** go, he went.
- By one count, there are 15. But how many *different* words?



Probably not.

There is a sense in which *go*
and *went* are different
forms of the same word

- That's both too loose and too strong.
- Laurie Bauer's example: How many words are there in:
 - The cook was a good cook as cooks **go**,
and as cooks **go**, he **went**.
- By one count, there are 15. But how many *different* words?



Summary so far:

- Defining a word as a sequence of letters between spaces (or something equivalent for spoken language, if there is something equivalent) is not good enough, because it *lumps* too much and it *splits* too much.
- It *lumps* words that are accidental homophones (*bank, counter, POLISH*) or different parts of speech (*hit*), and
- it *splits* different members of the same lexeme (*is, was, were, etc.*).



Some observations on **productivity** and new words

- -ize: winter-ize your car, automn-ize, *?fall-ize, *?spring-ize;
- work-aholic, coke-aholic, etc.
- nudnik, beatnik, cheapnik, jerknik.
- happiness, sadness, uncomfortableness, overbearingness, University of Chicago-ness, pissed-off-èd-ness



A morpheme

- Morphemes are to the morphs that are their allomorphs as phonemes are to the segments that are their allophones.
- If two morphs have the same meaning and/or grammatical function, and occur either in free variation or in complementary distribution, then they are allomorphs of the same morpheme.
- But otherwise, each morph realizes its own unique morpheme (*dog, cat, walk, ing, etc...*)



Allomorphs? conditioned by what?

- Sometimes conditioned by phonology:
 - an egg, an apple, an elephant, an igloo, an umbrella, an herb
 - a dog, a cat, a uniform, a hotel, a zebra
- Sometimes conditioned by grammatical information:
 - man/men, woman/women



Morphological effects

■ Affixation

– Suffixation

- Constitut-ion-al-ity
- Talo-i-ssa-an ‘in their houses’
 - House-plural-in-3rd-person-possessive

– Prefixation

- Dis-en-tangle
- Swahili: m-tu/wa-tu ‘person/people’
 - kitabu/vitabu ‘book/books’



■ Circumfixes

– German

– film-en

ge-film-t

‘to film’

– frag-en

ge-frag-t

‘to ask’

– lob-en

ge-lob-t

‘to praise’

– zeig-en

ge-zeig-t

‘to show’



infixes

■ Tagalog

- sulat ‘write’
- s-um-ulat ‘wrote’
- s-in-ulat ‘was written’



Template morphology ‘transfixes’

katab he wrote

jiktib he will write

maktuub written

maktaba bookshop

makaatib bookshops

mitaab book

maatib clerk

darashe studied

jidris he will study

madruus studied

madrasa school

madaaris schools

dars lesson

mudaris teacher



Reduplication: Maori

- amper ‘nearly’
- amperamper ‘very nearly’
- dik ‘thick’ dikdik ‘very thick’
- drie ‘three’ driedrie ‘three at a time’
- Tau ‘man’ tatau ‘men’
- Mero ‘boy’ bemero ‘boys’



Conversion

He walked round the car

She was looking round.

They sat at the round table.

As soon as I round the coner, I want you
to start running.

I always enjoy theater in the round.



Subtractive morphology: French masculine and feminine adjectives

mauvais	mauvaise	bad
heureux	heureuse	happy
grand	grande	big
long	longue	long
chaud	chaude	hot
vert	verte	green
froid	froide	cold
petit	petite	small
blanc	blanche	white
frais	fraiche	fresh, cool
faux	fausse	false



suppletion

Go/went

French:

Je vais (I go),

J' ir –ai (I will go)

J' all –ais (I was going)

All-er (to go); compare with donn-er,
where the stem remains the same
throughout.



Maybe a word is a minimal free form

Intuition: *What are those things in the tree?*
Birds.

Maybe a word is the smallest unit that can stand on its own as part of a discourse.

But not in French: *Qu'est-ce que c'est que ça?*
Des oiseaux.

Not: **Oiseaux.*



Besides, this definition cuts too sharp(ly)

- “jump” can stand on its own: does that mean that “jump” is a word in “I am jumping”? If not (and I think not), then it shows we can’t just take some sounds out of context, ask if they can be said as a stand-alone utterance, and conclude what their status was in the original utterance. (Which sounds to me like a pretty implausible thing to do anyway, once I put it like that.)



Laurie Bauer suggests 3 morphological considerations

- Words are syntactically mobile (but *phrases* are even more syntactically mobile, so this doesn't distinguish words from things bigger than words):
- When you have alternative word orders, it's generally words that are reordered, not subpieces of words. (Hopefully this can be made non-circular: we should speak not of "alternative word orders," but "alternative syntactic expressions of the same lexemes in syntactically related sentences.")

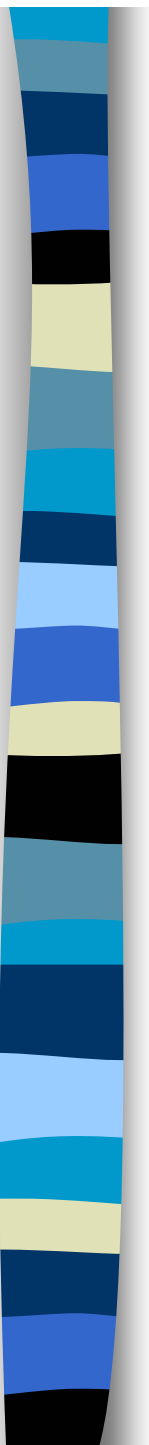
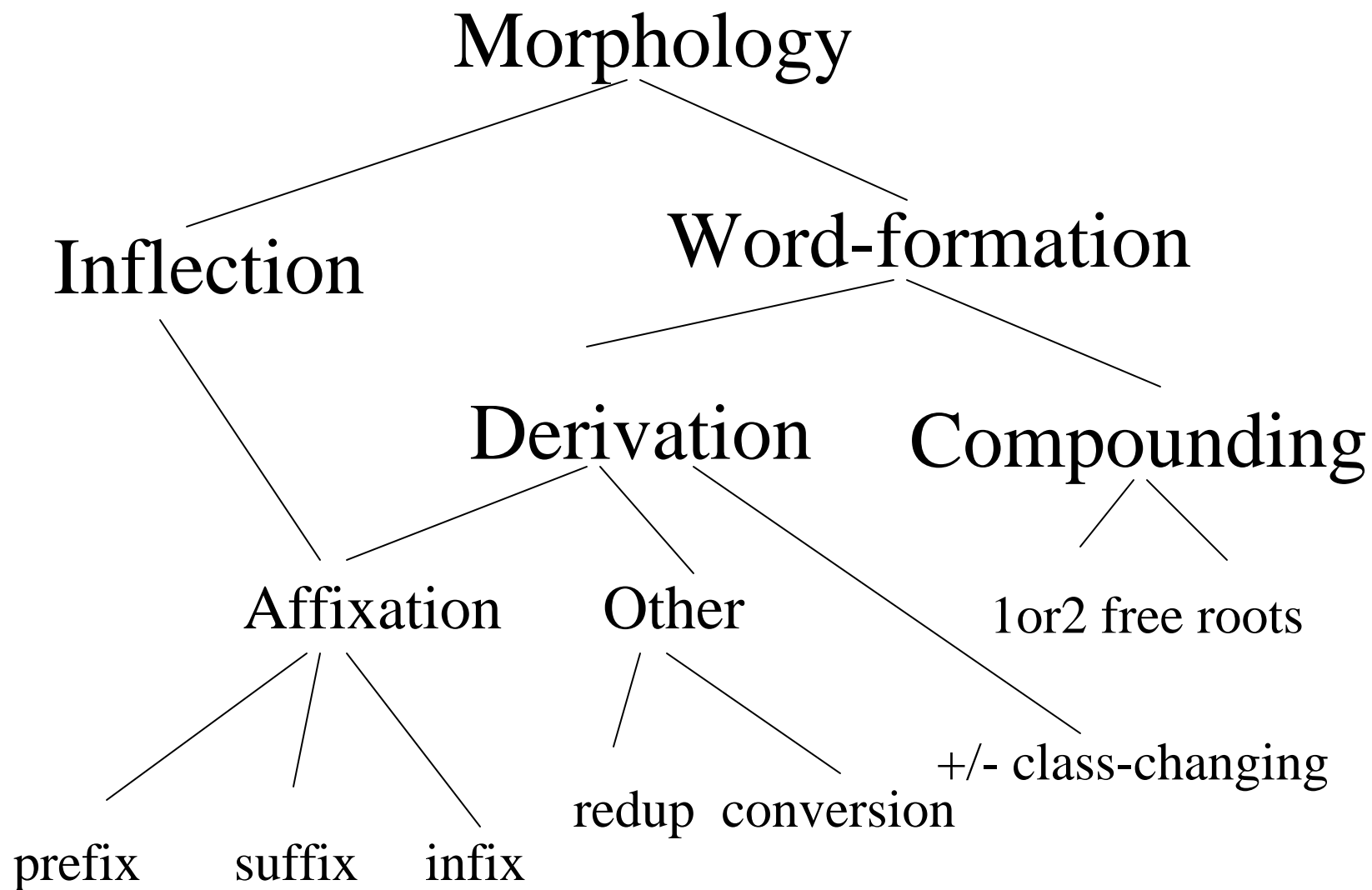


I love peaches. Peaches I love.

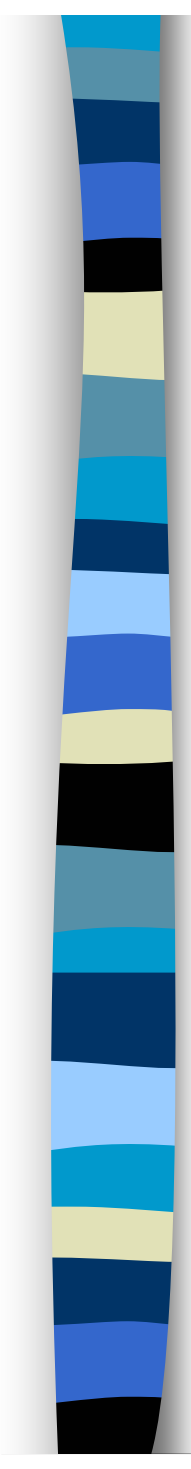
*Peach I love 's.

I love running. Running I really love.

*Run I really love ing.



inflection	derivation
produces (or relates) word forms of a single lexeme	produces new lexemes from old (or <i>relates</i> lexemes)
involves few variables of a closed system	may involve many variables (= semantic contributions) in an open system.
a lexeme is typically related to a lot of word-forms(e.g., inflected forms of verbs, nouns)	
low commutability within the sentence: you can't monkey around with inflectional morphology and still have a grammatical sentence	high commutability within the sentence: you can monkey around...
marks agreement	does not mark agreement
further from the root (the "deepest" morpheme) than derivational morphology	closer to the root than inflection
cannot be replaced by a single root form	often can be replaced by a single root form
no gaps	gaps in a paradigm, or just gaps
semantically regular	semantically irregular

- 
1. ninasema
 2. wunasema
 3. anasema
 4. ninaona
 5. ninamupika
 6. tunasema
 7. munasema
 8. wanasema
 9. ninapika
 10. ninaupika
 11. ninakupika
 12. ninawapika
 13. ananipika
 14. ananupika

Swahili (from Nida's workbook)



15. nilipika

16. nilimupika

17. nitakanupika

18. nitakapikiwa

19. wutakapikiwa

20. ninapikiwa

21. nilipikiwa

22. nilipikaka

23. wunapikizwa

24. wunanipikizwa

25. wutakanipikizwa

26. sitanupika

27. hatanupika

28. hatutanupika

29. hawatatupika