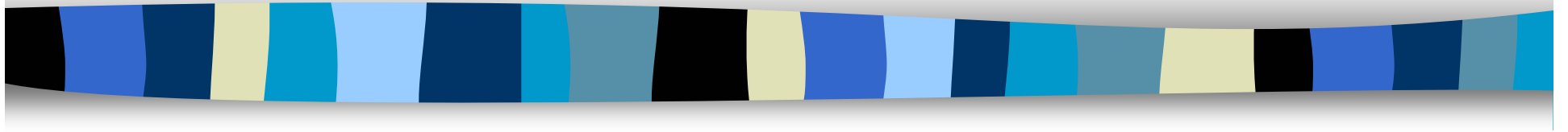


# Introduction to Morphology: 1



John Goldsmith



# Today

- We will make a number of assumptions to make the discussion easy: sweeping lots of problems under the rug.
- ***Later*** we will deal with more of the problems.



# 1. What is morphology?

Study of the internal structure of words:

- morph-ology word-s jump-ing
- Ni-li-mu-pik-a



# A list is not enough

An empirical fact:

AP newswire: mid-Feb – Dec 30 1988

Nearly 300,000 words.

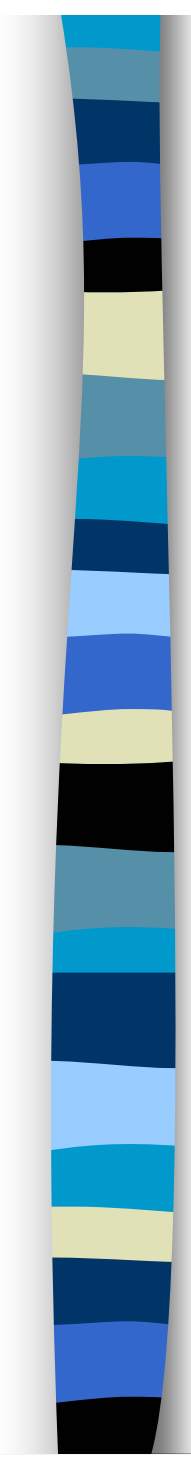
“New” words that appeared on Dec 31 1988:

compounds: prenatal-care, publicly-funded, channel-switching, owner-president, logic-loving, part-Vulcan, signal-emitting, landsite, government-aligned, armhole, signal-emitting...



...new words...

dumbbells, groveled, fuzzier, oxidized  
ex-presidency, puppetry, boulderlike,  
over-emphasized, hydrosulfite,  
outclassing, non-passengers, racialist,  
counterprograms, antiprejudice, re-  
unification, traumatological,  
refinancings, instrumenting, ex-critters,  
mega-lizard

- 
- ex-presidency: prefix *ex-*
  - boulder-like: suffix *-like*
  - over-emphasized: prefix *over-*
  - antiprejudice: prefix *anti*

This is often called the *OOV problem* (“out of vocabulary”) in computational linguistics



# Basic principles of morphology

- For some purposes, we need to think about phonemes, while for others it's more convenient to talk about letters.
- For our purposes, I'll talk about *letters* whenever we don't need to specifically focus on phonemes.



# Morpheme

- It is convenient to be able to talk about the *pieces* into which words may be broken, and linguists call these pieces *morphemes*: the smallest parts of a language that *can be regularly assigned a meaning*.



# Morph

- The realization (in phonemes or in letters) of a morpheme
- A morpheme can be realized as more than one more, depending on context (book-s, child-ren, ox-en) (a, an)
- A morph can realize more than one morpheme (plural -s, 3<sup>rd</sup> sg. -s)



# Morphemes

Uncontroversial morphemes:

door, dog, jump, -ing, -s, to

More controversial morphemes

sing/sang: s-ng + i/a

cut/cut: cut + PAST

# Classic distinctions in morphology:

## Analytic (isolating) languages:

- no morphology of derivational or inflectional sort.

## Synthetic languages:

- Agglutinative: 1 function per morpheme
- *Fusional* or *inflectional*: more than 1 function per morpheme

# Agglutinative: Finnish Nominal Declension

talo 'the-house'

kaup-pa 'the-shop'

talo-ni 'my house'

kaup-pa-ni 'my shop'

talo-ssa 'in the-house'

kaup-a-ssa 'in the-shop'

talo-ssa-ni 'in my house'

kaup-a-ssa-ni 'in my shop'

talo-i-ssa 'in the-houses'

kaup-o-i-ssa 'in the-shops'

talo-i-ssa-ni 'in my houses' kaup-o-i-ssa-ni 'in my shops'

# Fusional or inflectional: Latin

## Latin Declension of *hortus* 'garden'

	<b>Singular</b>	<b>Plural</b>
Nominative (Subject)	hort-us	hort-i
Genitive (of)	hort-i	hort-orum
Dative (for/to)	hort-o	hort-is
Accusative (Direct Obj)	hort-um	hort-us
Vocative (Call)	hort-e	hort-i
Ablative (from/with)	hort-o	hort-is



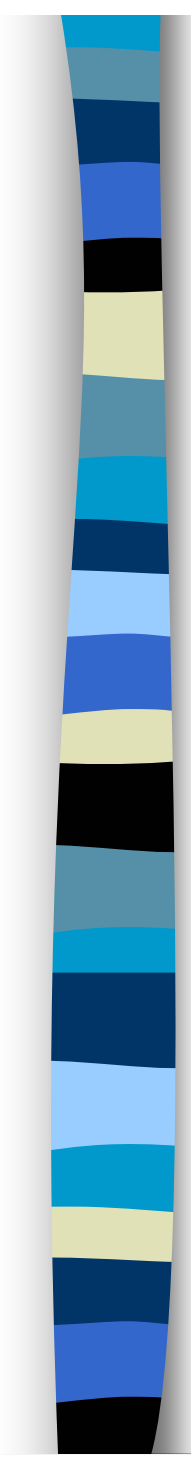
# Free and bound morphemes

- Free morphemes can form (free-standing) words; bound morphemes are only found in combination with other morphemes.
- Examples?



# Inflection and derivation

- We also speak of *inflectional morphology* and *derivational morphology*
- The core idea is this: *a word may take on a different form depending on the syntactic context in which it is found.* These differences constitute *inflectional morphology*.

- 
- A *lexeme* is the set of inflectionally related words.
  - Derivational morphology is that aspect of morphology which relates words from distinct lexemes.
  - Clear?
  - *We govern, they governed*: one lexeme
  - *Govern, government*: two lexemes.



Again:

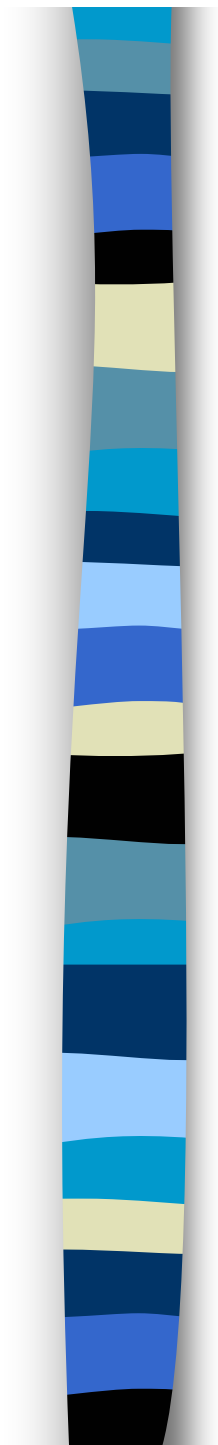
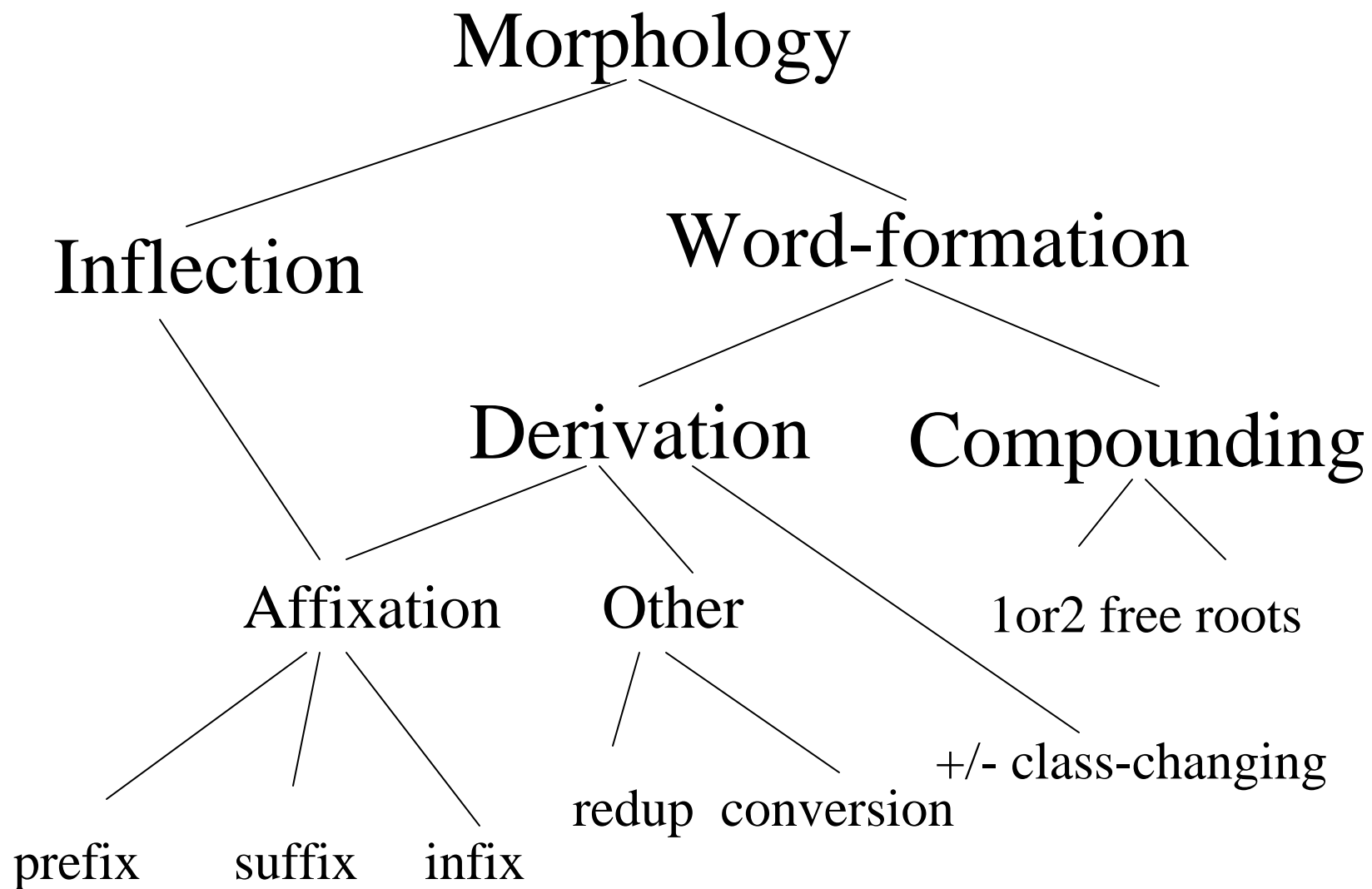
**Derivational morphology:** creates one lexeme from another

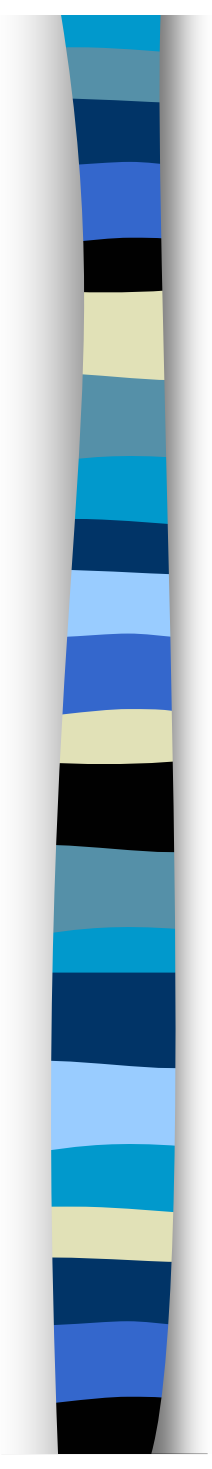
compute > computer > computerize > computerization

**Inflectional morphology:** creates the form of a lexeme that's right for a sentence:

the nominative singular form of a noun; or the past 3<sup>rd</sup> person singular form of a verb.

<b>inflection</b>	<b>derivation</b>
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)	
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



- 
- Word: an identifiable string of letters (or phonemes) *sing*
  - Word-form: a word with a specific set of syntactic and morphological features. The *sing* in *I sing* is 1<sup>st</sup> person sg, and is a distinct word-form from the *sing* in *you sing*.
  - Lexeme: a complete set of inflectionally related word-forms, including *sing*, *sings*, and *sang*
  - Lemma: a complete set of morphologically related lexemes: *sing*, *sings*, *song*, *sang*.



## A lexeme's stem

In many languages (unlike English), constellations of word-forms forming a lexeme demand the recognition of a basic *stem* which does not stand freely as a word:

*Italian*      *ragazzo, ragazza* (boy, girl)

*ragazzi, ragazze* (boys, girls)

**ragazz-**



# Compounds

Compounds are composed of 2 (or more) words or stems

Compounds: hot dog, White House, bookstore, cherry-covered



Languages vary in the amount of morphology they have and use

English has a lot of derivational morphology and relatively little inflectional morphology

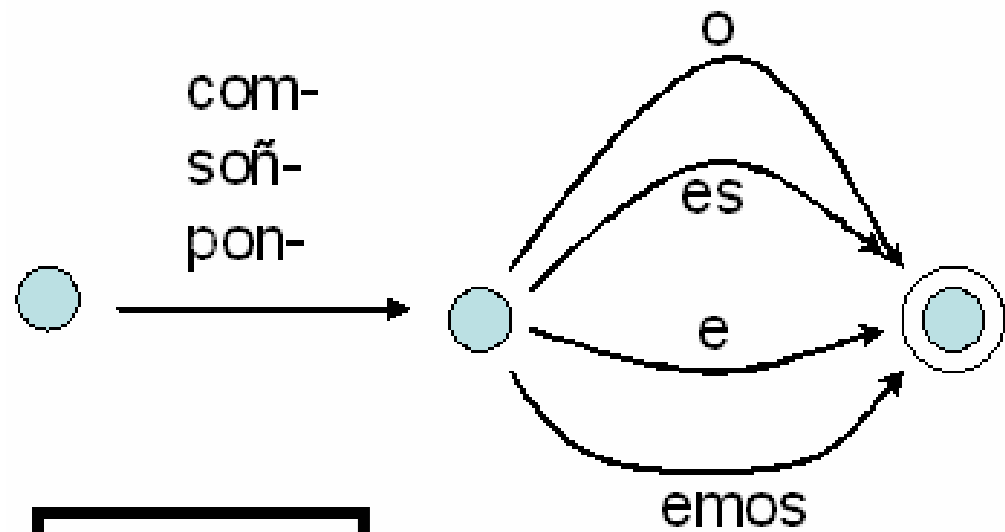
English verb's inflectional forms:  
bare stem, -s, -ed, -ing



# European languages

Not uncommon for a verb to have 30 to 50+ forms:  
marking tense, person and number of the subject

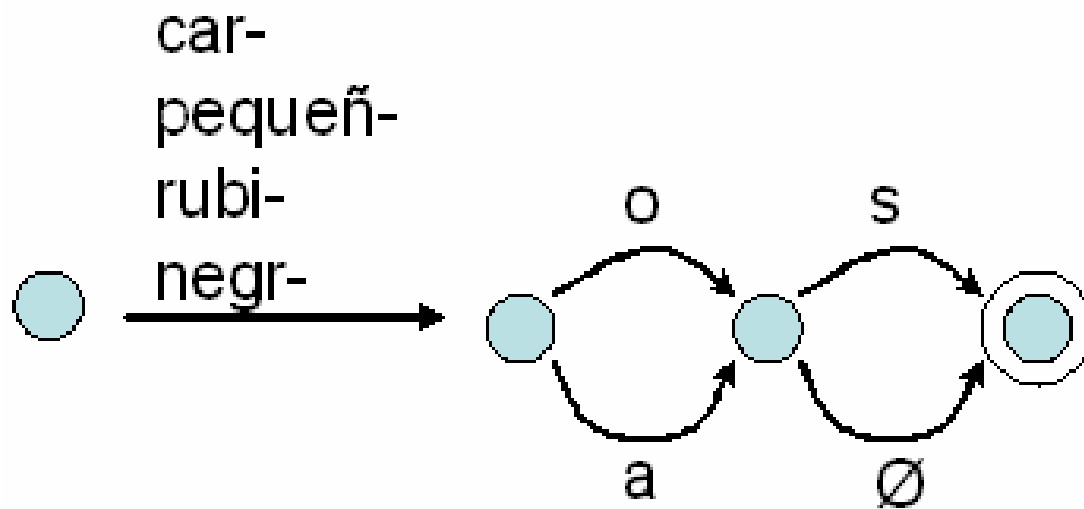
# Thinking about morphology like a computer scientist: “states” of a machine...



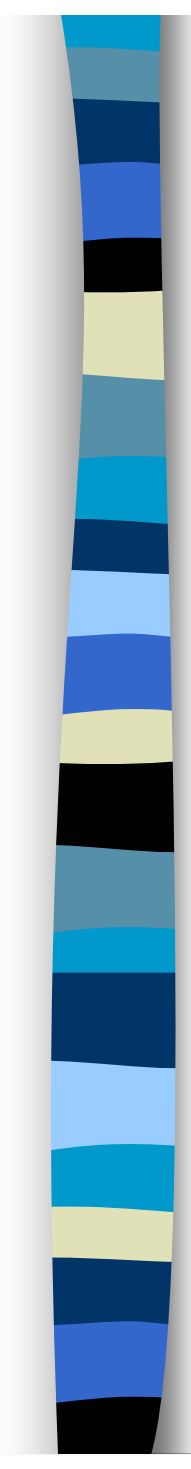
Spanish  
verb stems

a subset of  
verbal suffixes marking subject

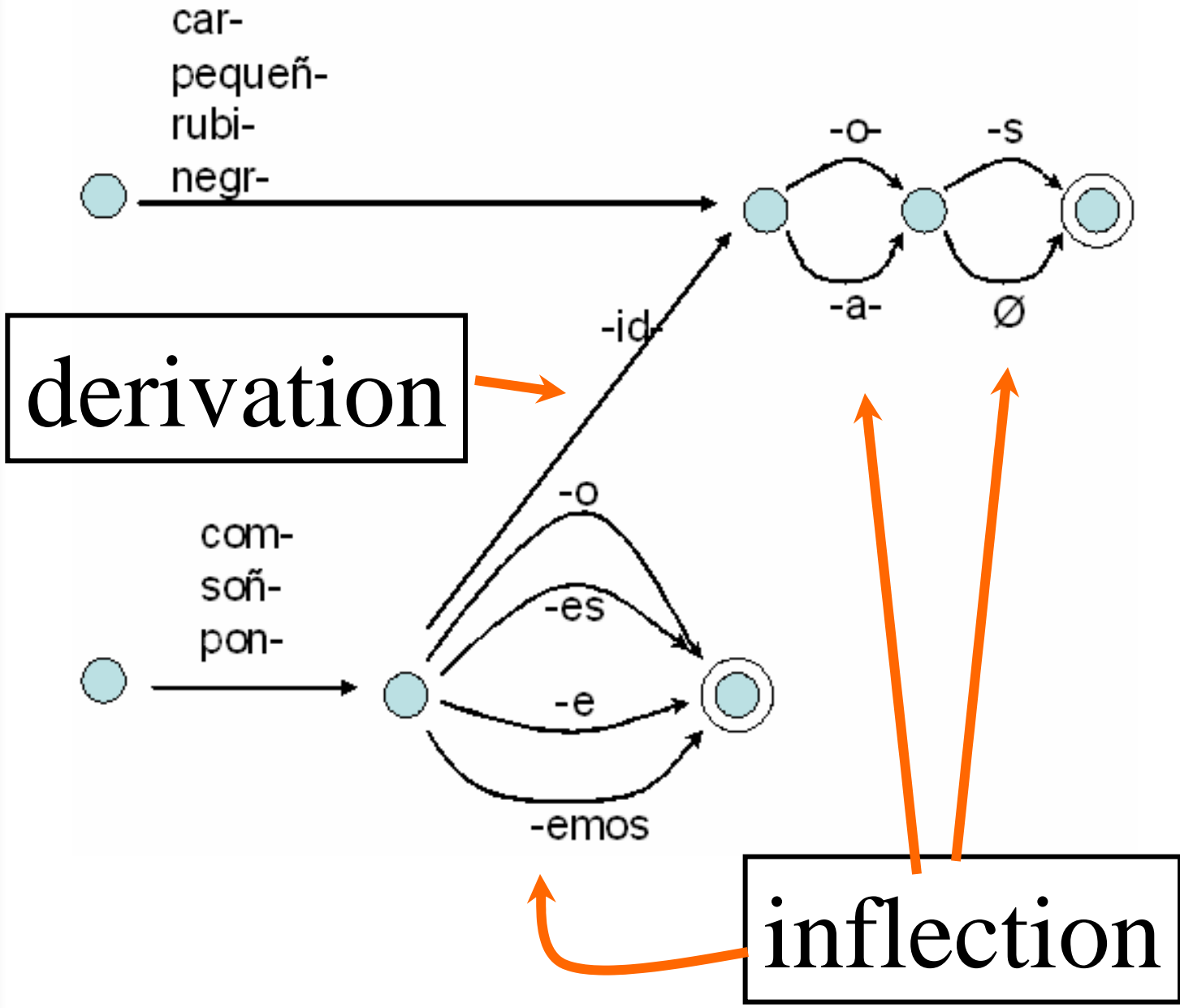
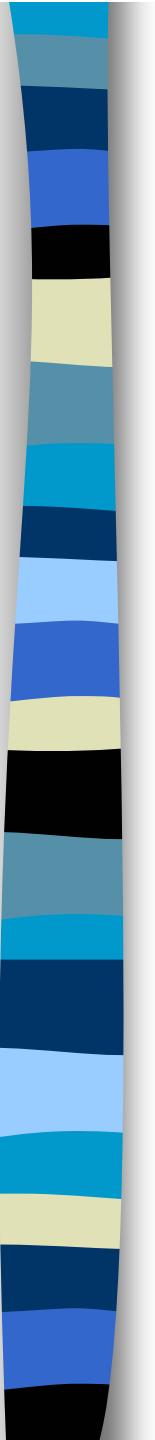
# Inflection



Spanish adjectival suffixes,  
inflection



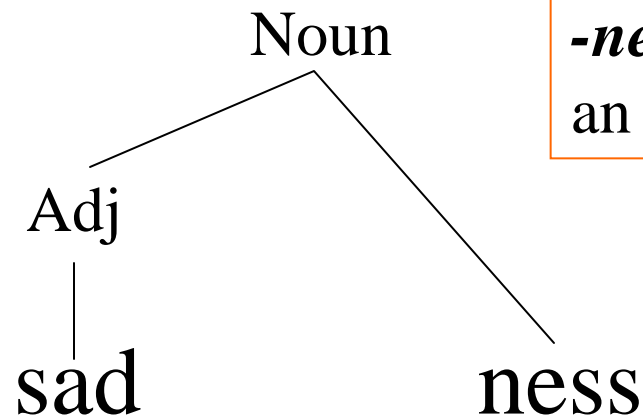
A way to use the “state machine”  
metaphor to think about the  
difference between inflectional and  
derivational morphology...



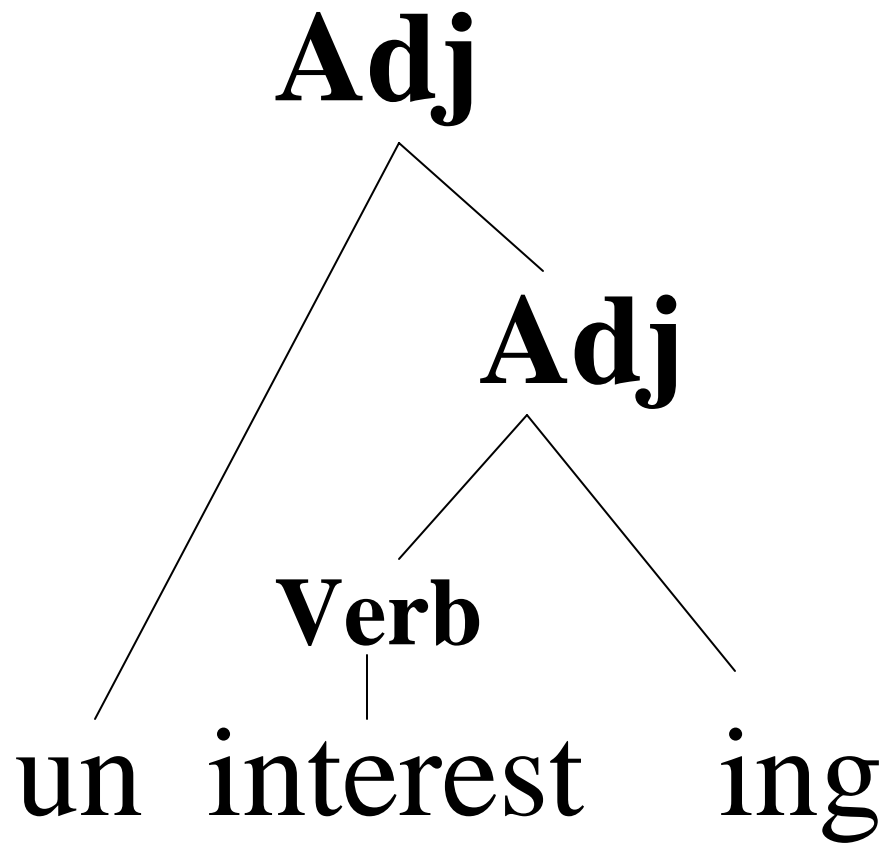
# Derivation

How *linguists* think about it...

Derivational morphology usually consists of adding a prefix or suffix to a base (= stem). The base has a lexical category (it is a noun, verb, adjective), and the suffix typically assigns a different category to the whole word.



***-ness***: suffix that takes an adjective, & makes a noun.



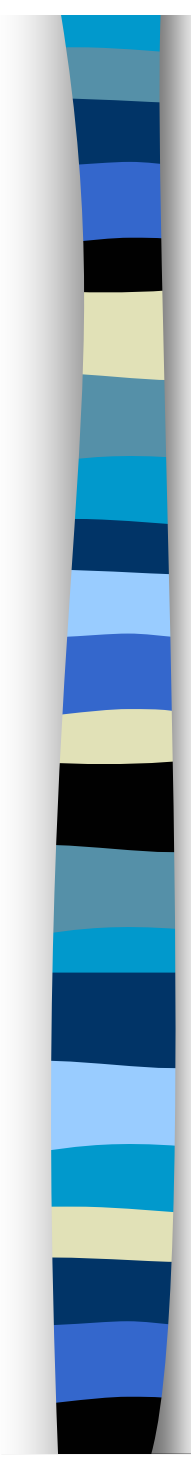


## Distinct from contractions...

English (and some other languages) permit the collapsing together of common words. In some extremely rare cases, only the collapsed form exists (English possessive **'s**).

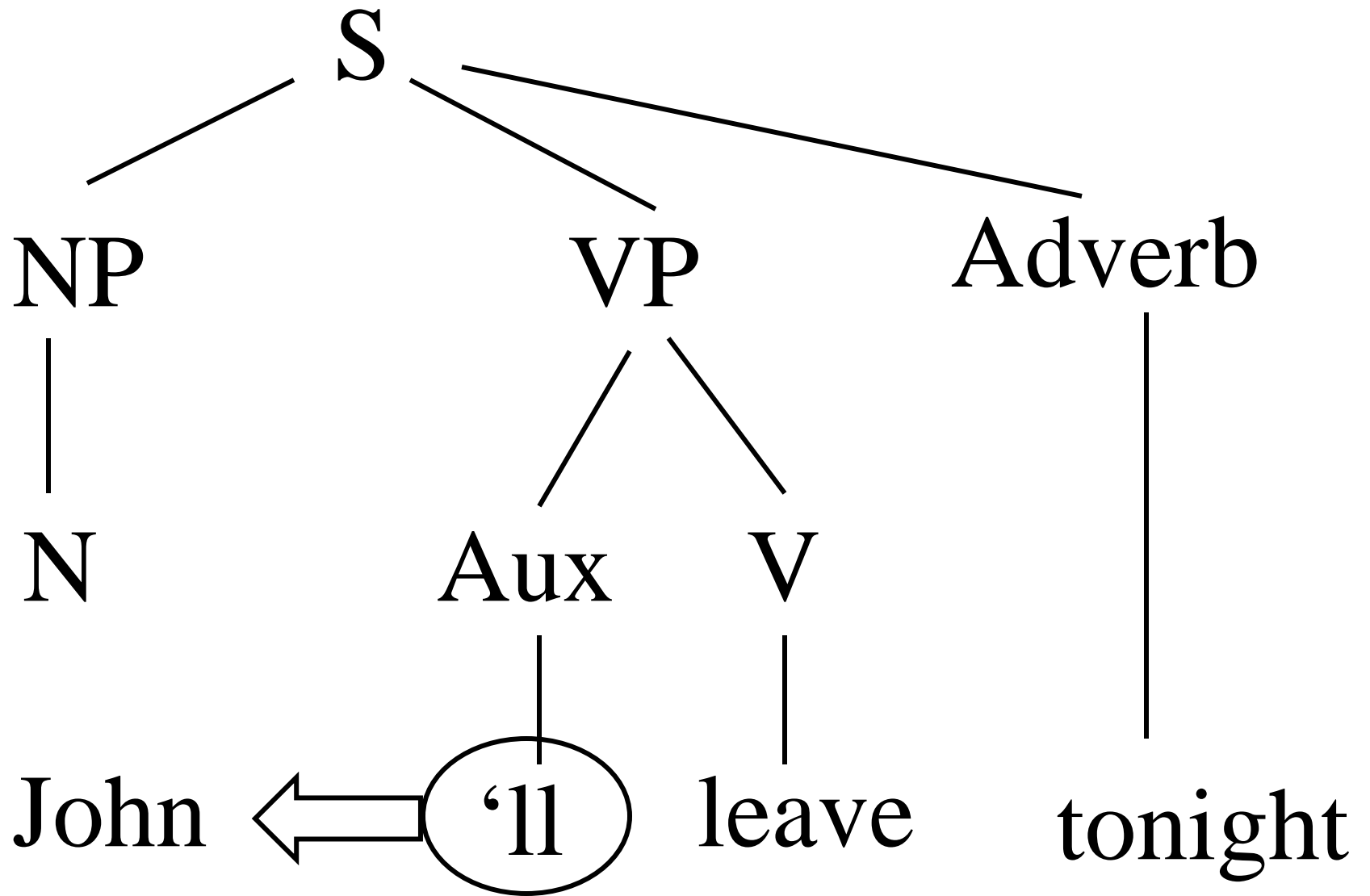
He will arrive tonight > he'll arrive...

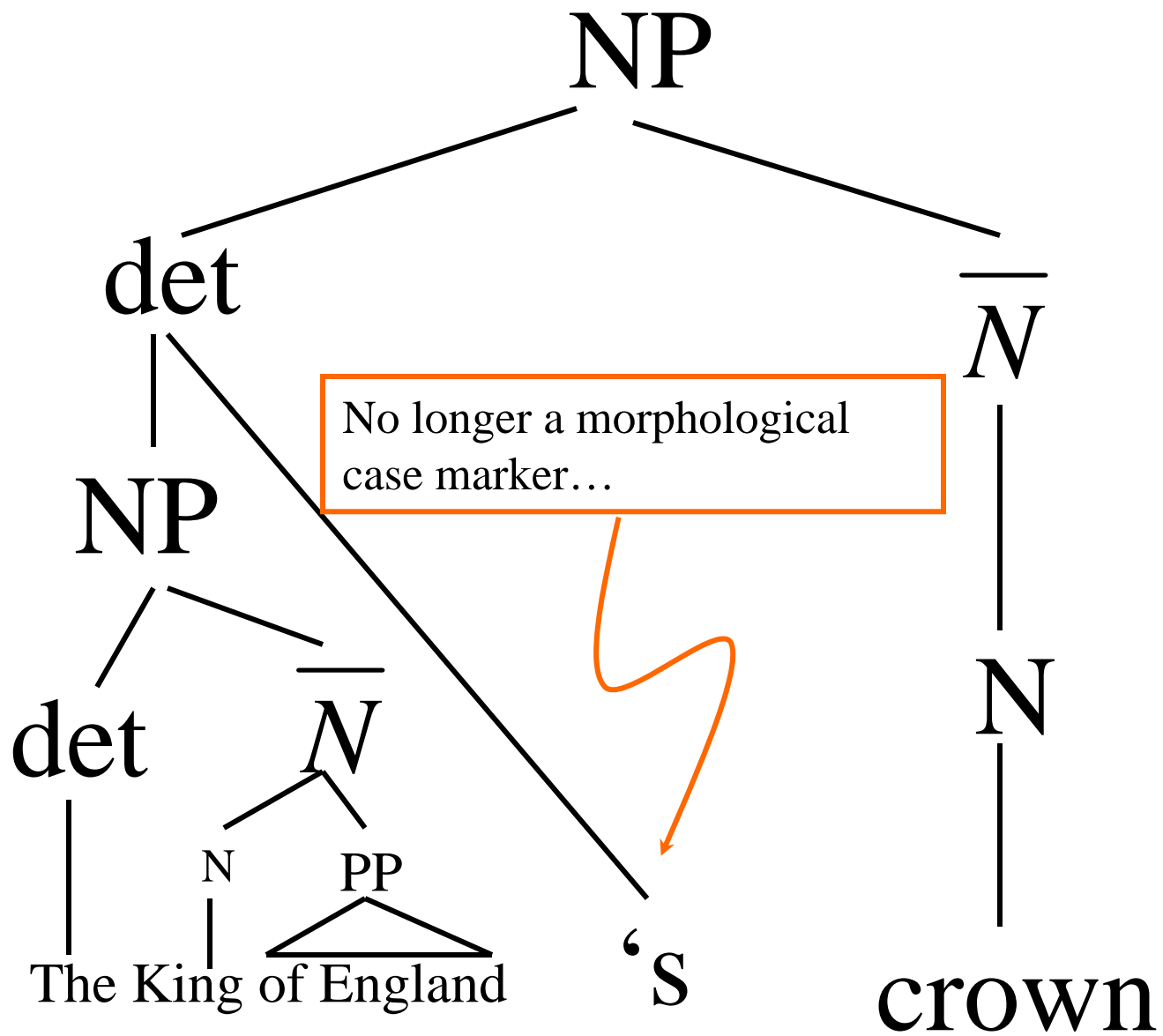
The [King of England]'**s** children



These are *different*, and are examples of what linguists call *clitics*

- From Greek *klitikos*, ***leaning***, from *klīnein*, *to lean*
- *Clitic* are *syntactically independent* objects that act as if they were morphologically or phonologically part of a neighboring word. English: *don't*, *I'll*, the king of England's crown.







# Some basics of English morphology

Inflectional morphology

**Nouns:** -NULL, -s, -'s

**Verbs:** -NULL, s, -ed, -ing

(so-called *weak verbs*)

**Strong verbs:** 3 major groups

- a. Internal verb change (sing/sang, drive/drove/driven, dive/dove)
- b. -t suffix, typically with vowel-shortening  
dream/dreamt, sleep/slept
- c. -aught replacement: catch, teach, seek,



# Derivational morphology in complex

This morphology creates new words, by adding prefixes or suffixes.

It is helpful to divide them into two groups, depending on whether they leave the pronunciation of the base unchanged or not.

There are, as always, some fuzzy cases.



### Level 1

ize, ization, al, ity, al, ic, al, ity,  
ion, y (nominaliz-ing),  
al, ate, ous, ive, ation

Can attach to non-word stems  
(fratern-al, paternal; parent-  
al)

Typically change stress and  
vowel quality of stem

### Level 2

Never precede Level 1  
suffixes

Never change stress  
pattern or vowel quality

Almost always attach to  
words that already exist

hood, ness, ly, s, ing, ish,  
ful, ly, ize, less, y (adj.)



## Combinations of Class 1,2

- Class 1 + Class 1: histor-ic-al, illumina-tion, indetermin-at-y;
- Class 1 + Class 2: fratern-al-ly, transform-ate-ion-less;
- Class 2 + Class 2: weight-less-ness
- ?? Class 2 + Class 1: \*weight-less-ity, fatal-ism-al

