

10. Historical Sources of English Word Formation

The Structure of Words (Ling 350)

Prof. Pentangelo

Fall 2025

Quiz #8

1. x
2. x

First, let's finish up chapter 8, Productivity

1. Semantic blocking
2. Measuring productivity: the significance of neologisms

Semantic Blocking

- An otherwise grammatical word-formation process may be blocked by the preexistence of a lexeme with the same meaning.

curious_A → curiosity_N

~~glorious_A → gloriosity_N~~

Process blocked by existence of *glory*_N

~~famous_A → famosity_N~~

Process blocked by existence of *fame*_N

- There are exceptions though. Think about HATE_N, which has largely displaced HATRED_N.
- And some speakers say CONVERSATE_V, despite preexisting CONVERSE_V.

Now do activity 2 on last week's sheet.

Measuring Productivity

- It can be extremely hard to quantify how productive one morphological process is vs. another.
- Probably the best way to do this involves working with **corpora**. (N.B. *corpora* is plural; *corpus* is singular.)
- What are corpora?
- What is a **neologism**?

“For a process to be productive, in one sense, it should be a process that can be used to form brand new lexemes, or neologisms.” (103)

- What are **hapax legomena**?

“words that [...] appear only once in the whole corpus” (103)

Measuring Productivity

- One study used the Cobuild corpus to assess the productivity of *-ity* vs. *-ness*.
- Think back to our second meeting: what are **types** vs. **tokens**?
- In the Cobuild corpus:

“... the number of word-types exhibiting *-ity* (roughly 400) is not greatly less than the number exhibiting *-ness* (roughly 500). However, most of the *-ity* words are of common occurrence (more technically, their token-frequency is high), while many more of the types exhibiting *-ness* have low token frequency, including hapax legomena. That is, although by one measure *-ness* seems to be not much more productive than *-ity* is, it is far more likely than *-ity* to be used in the creation of neologisms.” (104)

Neologisms and semantic blocking

- Often, a neologism won't be created—or just won't catch on—if it is blocked by a preexisting word with the same meaning.
- I'll list adjectives (X). Tell me the noun that means, roughly, “the state of being X.”

“It's good to be smart. _____ is a virtue.”

smart

curious

glorious

stupid

red

deadly

quick

green

active

Japanese

shiny

attractive

tall

healthy

supportive

- We can actually query a corpus ourselves to see what processes are more or less productive.
- Let's learn how to use english-corpora.org.

English-corpora.org

- The Corpus of Contemporary American English (COCA) has 1 billion entries, up to 2019.
- The NOW Corpus (News on the Web) has 23.5 billion entries, up to yesterday.
- Go to english-corpora.org and make an account.
- How to form a query:

* You're limited to 20 queries in a 24-hour period.

Query	Results
stupid	All results for just “stupid”
stupid*	All results that start with “stupid,” including “stupid,” “stupidity,” “stupidness,” and “stupider”
sill*_n	All nouns that start with “sill,” including “sill” and “silliness” but excluding “silly”

* is a wildcard. It can stand for any number of letters.

_n says you only want nouns, _v verbs, _j adjectives, _r adverbs.

Note: It's not perfectly tagged. Some words are misclassified, so even if you search just for nouns, some adjectives will slip through.

-maxxing

- The journal *American Speech* regularly publishes an “Among the New Words” roundup of new words. The .pdf of the most recent one is on the homepage of our class site. Access it.
- One thing they mention is the suffix -maxxing. What does this mean?
- What are some words the article lists under this entry?

looksmaxxing

jawmaxxing

testosterone-maxxing

in-flights-maxxing

Other one-offs found in the wild are *nostalgia-maxxing*, *buzzword-maxxing*, *vocabulary-maxxing*, *malaise-maxxing*, *Debussy-maxxing*, *aura-maxxing*, *racism-maxxing*, *nap-maxxing*, and *follow-up-maxxing*. Now that's some strong suffix-maxxing. [Mark Peters]

“one offs” = hapax legomena

-maxxing

- Which corpus could we use to explore this term's productivity?
- Let's use the NOW Corpus.
- What query can we construct that would include “looksmaxxing” and “nap-maxxing” as well as “looksmaxxed” or “nap-maxxes,” but (hopefully) not TJMaxx?

`*maxx*_v`

The * takes the place of any number of letters

The _v says we only want verbs

- Before conducting the search, click OPTIONS and set # HITS to 1000.
- So what results do we get?
- Are there any hapax legomena in the results?

Is *-maxxing*
productive?

Activity!

- Working in groups, come up with a neologism you've heard that involves either compounding (e.g. *brainrot*) or derivation (like *-maxxing*).
- (To explore *brainrot*-style compounds, you might search for “N rot” or “*rot”.) “N rot” will return any strings of a noun plus rot, with a space in between.
- Using one of the corpora on English-corpora.org, try to evaluate how productive the process you've chosen is.
- You'll report back on your results to the class.

Inflection

- In the history of English, we used to have *way more* inflectional morphology than we do today. English inflection today is strangely minimal.
- In Old English, nouns had one of three genders (masc., neut., fem.). They also were marked for class (nom., acc., gen., dat.). The way that these were marked differed for singular and plural nouns. This was mostly done with suffixes.
- Adjectives had to agree with nouns (like in Spanish or Italian today).
- Verbs were also a lot richer, being conjugated for person in present as well as in past tenses.
- This mostly fell away as a result of sound changes and language contact.

Inflection

- Some things, like words and derivational affixes, are pretty easily borrowed. Inflectional morphology is not.
- However, we do often borrow nouns in singular and plural forms.

<i>Source language</i>	<i>Singular</i>	<i>Plural</i>
Greek	phenomenon	phenomena
	schema	schemata
Latin	cactus	cacti
	formula	formulae
	datum	data
Hebrew	cherub	cherubim
	kibbutz	kibbutzim

English's Origins (briefly)

- English is a Germanic language, like Dutch, German, Swedish, Afrikaans, Danish, etc.
- English didn't come from German. English and German are sister languages.
- Germanic languages came from Proto-Germanic, an ancient language which has been reconstructed with more or less confidence.
- Proto-Germanic came from Proto-Indo-European, another ancient reconstructed language. Many other languages—and thus language families—came from PIE as well.



Simplified Indo-European family tree by Fortson

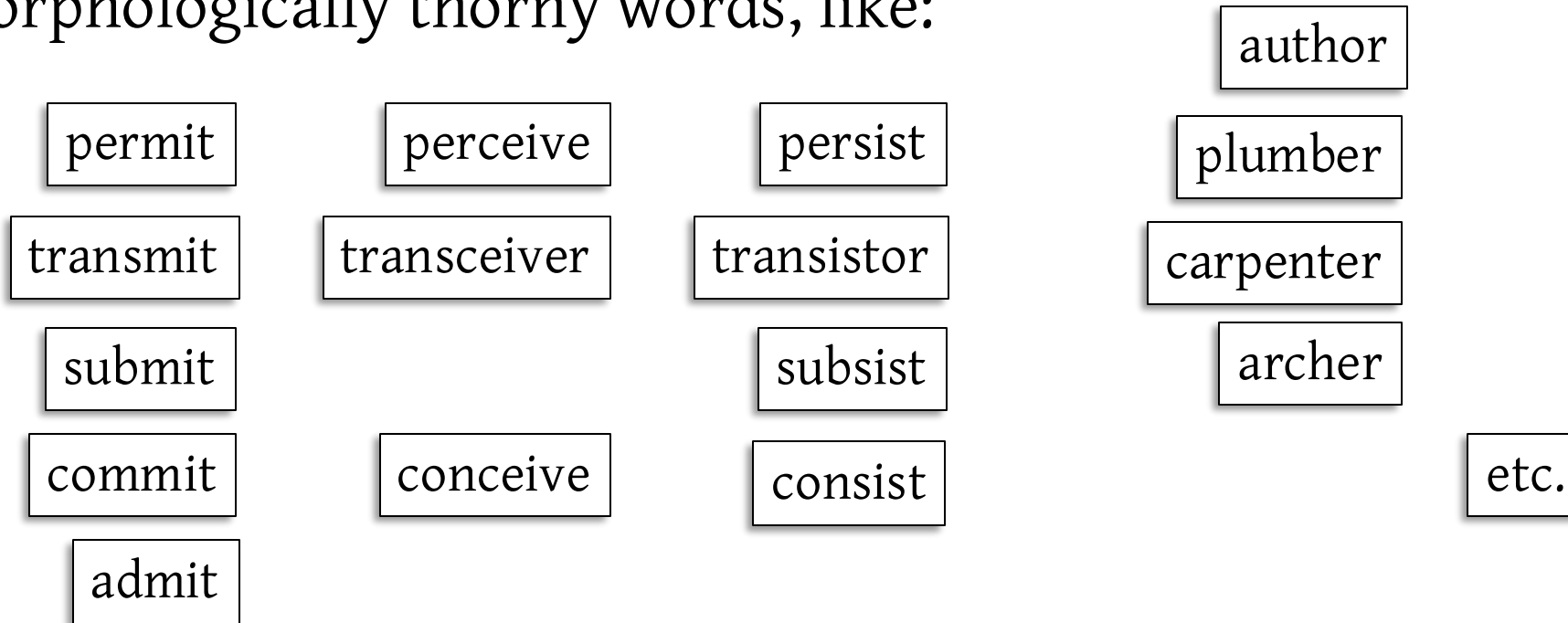
Historical Sources of English Word Formation

- So English is a Germanic language, but approximately 2/3 of its vocabulary comes from Romance languages, particularly French and Latin.
- Why?
- Also, a lot of Greek borrowings entered during the Early Modern era, around 1500 AD.

“The variety of the sources that have contributed to the vocabulary of English accounts for the existence of many pairs or groups of words that are descended, in whole or in part, from the same ancestral morpheme in the extinct **Proto-Indo-European** language from which Greek and the Romance and Germanic languages are descended – words that are thus **cognate**, in the terminology of historical linguistics – but that have reached their contemporary English form by a variety of routes, through direct inheritance or through borrowing at different times.” (109)

Bound Roots

- We have this massive scale of borrowing to thank for many of our morphologically thorny words, like:



In Proto-Indo-European, the word was **kért*.

This became the native English word *heart*.

In Latin, **kért* became *cord-* (a bound root).

From here, we get *cordial*.

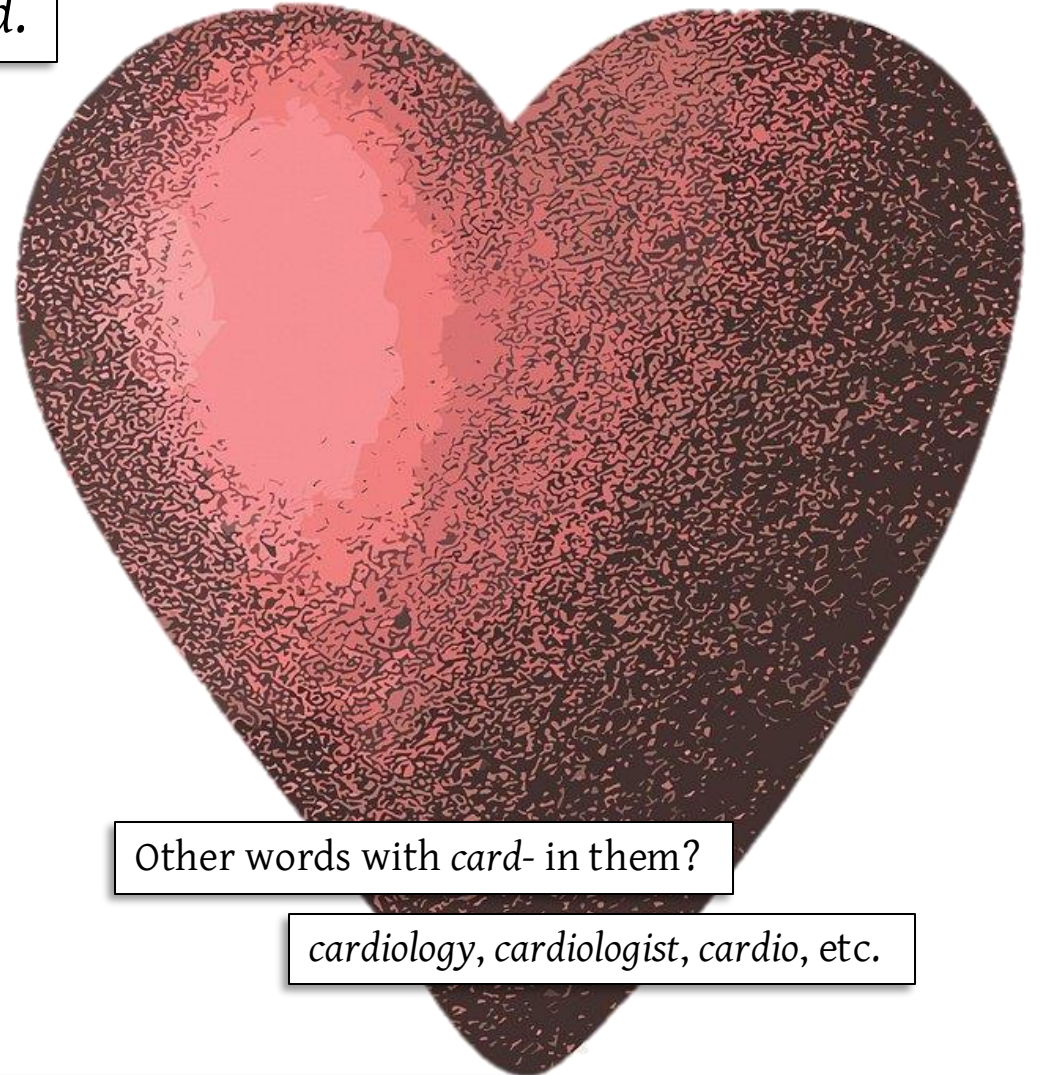
In French, Latin *cord-* became *cœur*.

From this, we get *courage*.

In Greek, **kért* became *kard-* (another bound root).

From the Greek *kardiakós*, we get *cardiac*.

It's significant that the borrowed roots are all **bound**, while the native English root is **free**. This happens a lot throughout English vocabulary.



Other words with *card-* in them?

cardiology, cardiologist, cardio, etc.

- What are the Greek-derived words with the following meanings? Identify the morphemes and their meanings. Do you know other words with these morphemes in them?

study of life

flying dinosaur with wing membrane connected to an elongated finger

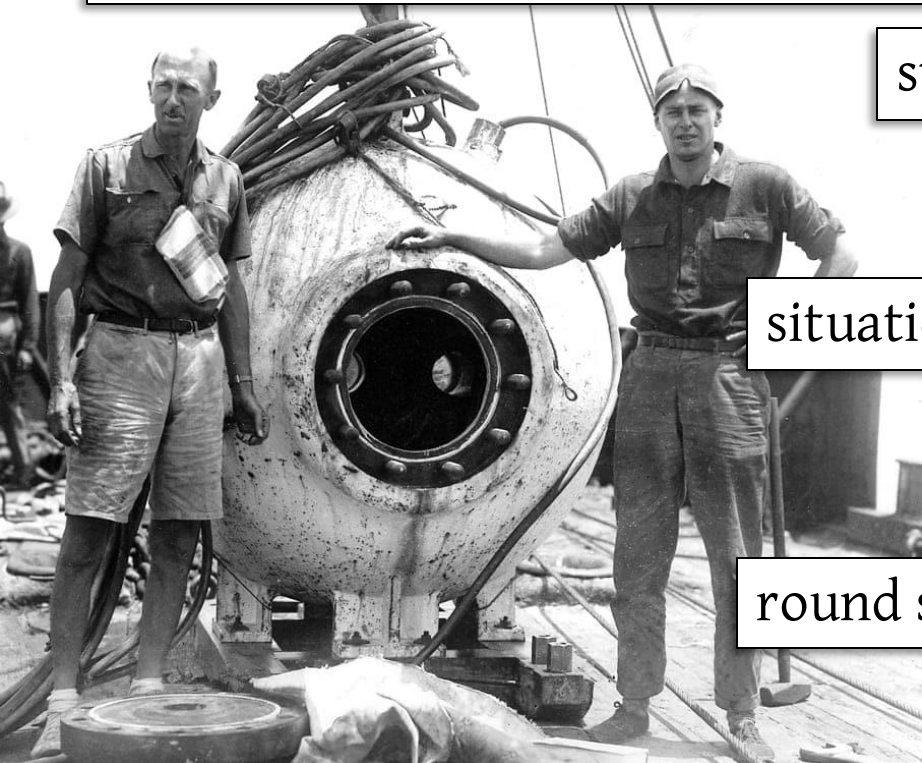
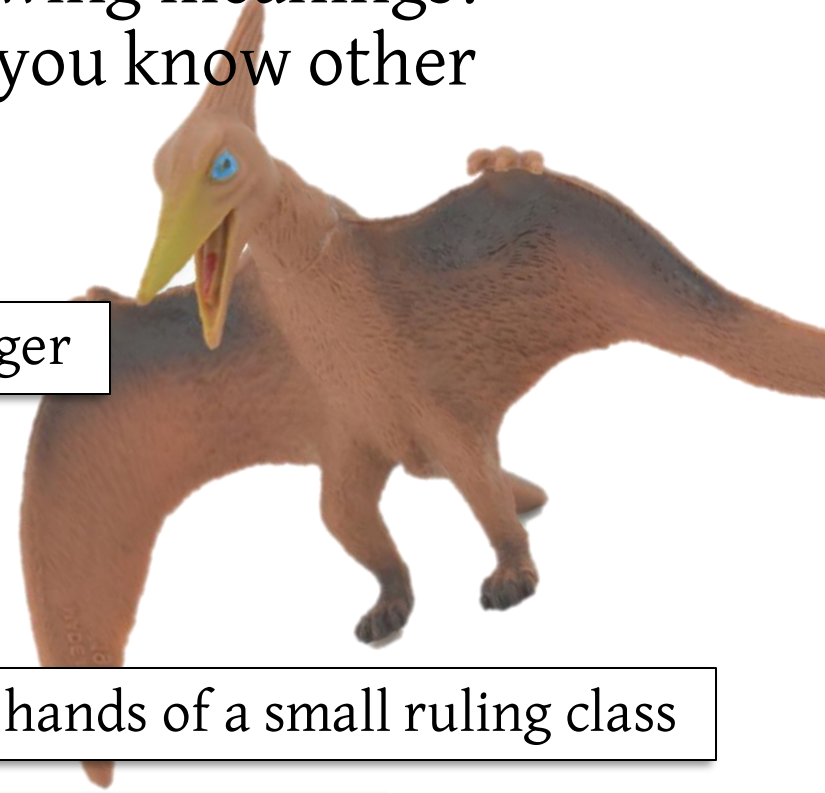
study of skin diseases

red blood cell

situation where political power is in the hands of a small ruling class

members of that class

round submarine vessel for exploring the depths of the ocean



Activity!

- In groups, do activity 1 on your sheets.
- Now do activity 2.
- Now 3.
- Now 4.
- Now 5.

Derivation

Germanic

-ish
-ed
-en
-er
-hood
-ie (*as in* doggie)
-let
-ship
-y (*as in* misty)

Romance or Greek

-((a)t)ion
-(i)an
-(i)fy
-al
-ance, -ence
-ar
-ent, -ant
-ess
-ette
-ine
-ise
-ism
-ist
-ment
de-
dis-

Which ones take mostly free bases? Which ones bound?

Germanic takes mostly free; Romance or Greek mostly bound.

Again, this is a generalization —
it does not apply in all cases.

Does this tally with your findings?

For next time

- We're done with Carstairs-McCarthy's book! Now read chapter 7 in Bauer.
- This chapter analyzes (and defines) morphemes in a somewhat different way from Carstairs-McCarthy. Pay attention, in your reading, to how Bauer's treatment of the morpheme differs from the one we've used so far.
- There will be a quiz on this reading at the start of class.