What is a word?

Outline

1. **What is a word?**

2. **Morphemes**

3. **Derivation and inflection**

4. **Morphological typology**

5. **Innovation**

- We begin with a familiar object: the written word.
- Zipf's Law.
- What is the basis for our dividing the stream of sounds (or letters) as we do?
- The lexeme: how well do we know how to organize a dictionary of our language?
- Morphology is the (study of the) structure of words: internal and external (=relational) structure
- Internal analysis of words
- External analysis: inflectional morphology and derivational morphology
- Productivity

Rank	OE Corpus lemmas	word forms	Tom Sawyer	Warren Commission
1	the	the	the	the
2	be	of	and	of
3	to	and	a	to
4	of	a	to	and
5	and	to	of	in
6	a	in	he	a
7	in	is	was	that
8	that	you	it	he
9	have	that	in	was
10	I	it	that	his
11	it	he	his	on
12	for	was	I	Oswald
13	not	for	you	had
14	on	on	Tom	at
15	with	are	with	for

Television/movies	
you	1,222,421
I	1,052,546
to	823,661
the	770,161
a	563,578
and	480,214
that	413,389
it	388,320
of	332,038
me	312,326

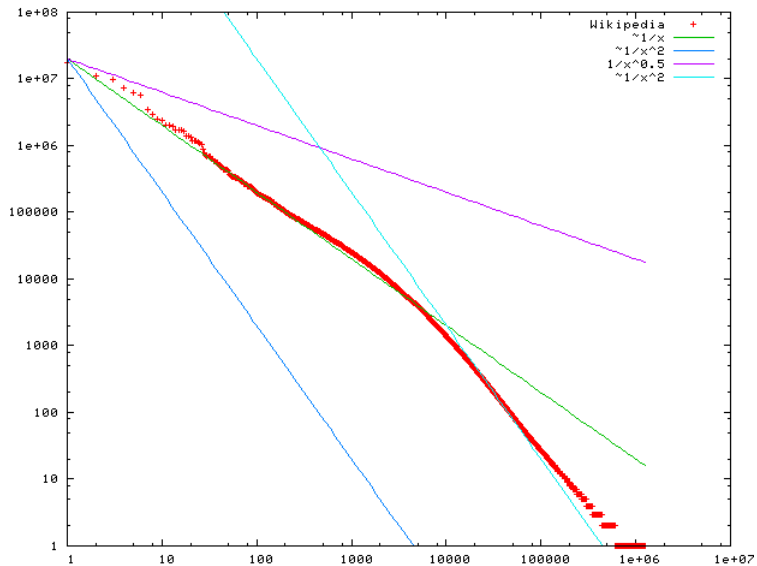| Project Gutenberg	
the	56,271,872
of	33,950,064
and	29,944,184
to	25,956,096
in	17,420,636
I	11,764,797
that	11,073,318
was	10,078,245
his	8,799,755
he	8,397,205
it	8,058,110
with	7,725,512
is	7,557,477
for	7,097,981
as	7,037,543
had	6,139,336

The product of a word's rank and its frequency is a constant across a large corpus.

$rank(w) \times freq(w) = Z$

where Z is a constant for a given language.

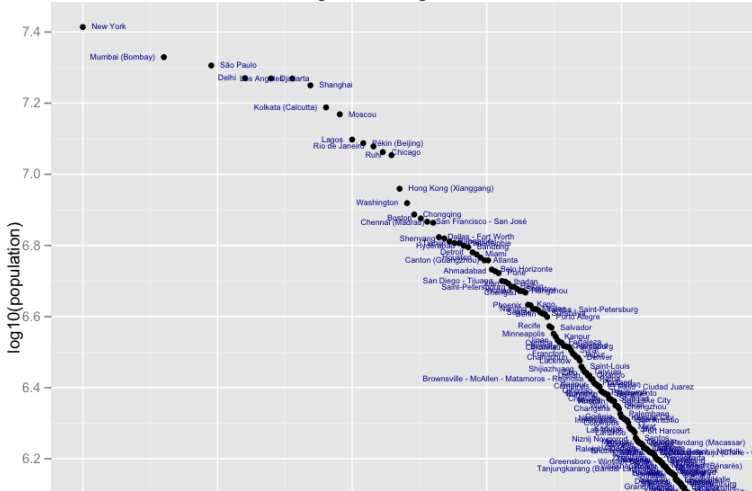From AI and Social Science website, Brendan O'Connor.

world city populations for 8 countries
log-size vs log-rank

Idea the First: Orthography

A first try:

Definition

A word is a written sequence which has a white space at each end but no white space in the middle.

- Problem the First: this begs the question.

Begging the question

Idea the First: Orthography

Definition

A word is a written sequence which has a white space at each end but no white space in the middle.

- Problem the First: this begs the question.

- Problem the Second: not everyone uses spaces.

Before there were spaces

There continue to not be spaces

Bangkok

From Wikipedia, the free encyclopedia

This article **needs additional citations** for **verification**.
Please help improve this article by adding reliable references. Unsourced material may be challenged

Bangkok is the capital, largest urban area and primary city of Thailand. Known in Thai as *Krung Thep Mahanakhon* (Thai: กรุงเทพมหานคร, pronounced [krūŋtʰêːp máhǎːnákʰɔ̄ːn] (listen)), or กรุงเทพฯ *Krung Thep* (listen) (help-info), meaning "City of the Deity") for short, it was a small trading post at the mouth of the Chao Phraya River during the Ayutthaya Kingdom. It came to the forefront of Siam when it was given the status as the capital city in 1768 after the burning of Ayutthaya. However, the current Rattanakosin Kingdom did not begin until 1782 when the capital was moved across the river by Rama I after the death of King Taksin. The Rattanakosin capital is now more formally called "Phra Nakhon" (Thai: พระนคร), pertaining to the ancient boundaries in the metropolis' core and the name Bangkok now incorporates the urban build-up since the 18th century which has its own public administration and governor.

There continue to not be spaces

บทความ | อภิปราย | แก้ไข | ประวัติ

กรุงเทพมหานคร

พิกัดภูมิศาสตร์:

จากวิกิพีเดีย สารานุกรมเสรี
(เปลี่ยนทางมาจาก Bangkok)

บทความนี้เกี่ยวกับเมืองหลวง สำหรับองค์กรปกครองส่วนท้องถิ่นรูปแบบพิเศษ ดูที่ กรุงเทพมหานคร (องค์กรปกครอ

กรุงเทพมหานคร เป็นเมืองหลวง และเป็นเมืองที่มีประชากรมากที่สุดในประเทศไทย[3] รวมทั้งเป็นศูนย์กลางการ
ปกครอง การศึกษา การคมนาคมขนส่ง การเงินการธนาคาร การพาณิชย์ การสื่อสาร และความเจริญก้าวหน้าด้านอื่น ๆ
ของประเทศไทย นอกจากนี้ยังเป็นเมืองที่มีชื่อยาวที่สุดในโลกอีกด้วย มีแม่น้ำสำคัญคือ แม่น้ำเจ้าพระยา ไหลผ่าน ทำให้
แบ่งเมืองออกเป็น 2 ฝั่ง คือ ฝั่งพระนคร และ ฝั่งธนบุรี (เดิมฝั่งตะวันตกของแม่น้ำเป็นที่ตั้งของ กรุงธนบุรี ซึ่งต่อมาภายหลังได้
รวมเข้าเป็นส่วนหนึ่งของกรุงเทพมหานคร) โดยกรุงเทพมหานครมีพื้นที่ทั้งหมด 1,568.737 ตารางกิโลเมตร พิกัดทาง
ภูมิศาสตร์คือ ละติจูด 13°45' เหนือ ลองจิจูด 100°31' ตะวันออก[4]

กรุงเทพมหานครเป็นเขตปกครองพิเศษของประเทศไทย โดยมิได้มีสถานะเป็นจังหวัด ซึ่งคำว่า กรุงเทพมหานคร
นั้น ยังใช้เป็นคำเรียกสำนักงานปกครองส่วนท้องถิ่นของกรุงเทพมหานครอีกด้วย ปัจจุบันกรุงเทพมหานครใช้วิธีการเลือก
ตั้งผู้บริหารแบบ การเลือกตั้งผู้บริหารท้องถิ่นโดยตรง

There continue to not be spaces

簡介

芝加哥大學

芝加哥大学的校训译成中文是「益智厚生」，意思是「提升知识，以充实人生。」（拉丁文原文：Crescat scientia vita excolatur. 英译：Let knowledge grow from more to more; and so be human life enriched）。

理念與課程

芝大对教育观念的「宏观」与实验精神，奠定了它在美国教育史上的重要地位。而它在学术研究上的地位与贡献，也同样值得称道。

芝加哥大学的人类学、天文学、地球科学、经济学、地理学、历史学、语言学、数学、物理学、统计学、商学、社会学、神学等学科专业在美国具有较强的学术实力。经济学、社会学系实力尤其超强，"芝加哥学派"赫赫有名。

Idea the First: Orthography

1. A word is a written sequence which has a white space at each end but no white space in the middle.

- Problem the First: this begs the question.

- Problem the Second: not everyone uses spaces.

- Problem the Third: compounds.

Google

11. Hyphenated Terms

Because some people spell hyphenated words with a hyphen and others with a space, Google searches for variations on any hyphenated terms.

When Google encounters a hyphen (–) in a query term, e.g., [part-time], it searches for:

- the term with the hyphen, e.g., part-time
- the term without the hyphen, e.g., parttime
- the term with the hyphen replaced by a space, e.g., part time

- [part-time] matches "part-time," "part time," and "parttime"
- [part time] matches "part-time" and "part time", but
- ["part time"] (with quotes) is better for space-separated words

- "land holder" (incl. *land-holder*): 91,300 hits

- "landholder": 486,000 hits

- Is one in more dictionaries than the other? Does it matter?

Compounds

die Herzkreislaufwiederbelebung

Herz 'heart'
Kreis 'circle'
laufen 'to run' (> *Lauf* 'a run, circuit')
wieder 'again'
beleben 'to revive, animate' (> *Belebung* 'revival')

Compounds

**das Rindfleischetikettierungs-
überwachungsaufgabenübertragungsgesetz**

Rind 'beef, cattle'
Fleisch 'meat, flesh'
Etikett 'label'
Etikettierung 'labelling'
über 'over'
wach 'awake'

Überwachung 'control, monitoring'
Aufgaben 'responsibilities'
tragen 'to carry'
Übertragung 'transfer'
Gesetz 'law'

Is the corresponding English form a word, or a phrase?

Compounds and worse

French (Romance, Indo-European)

porte-	clé
carry	key
key holder	

je-	m'en-	fout-	isme
I	myself-of-it	shove	nominalization
	I-don't-care-ism		

He had one of those I-know-something-you-don't-know looks.

Lexical entries

Here's another tack:

Definition

A word is what precedes a dictionary entry.

Dictionary forms

- aka **citation form** or **lemma**

- Computational linguists use **lemma** to mean something a bit broader.

What's in the dictionary?

bi•as (bī′əs) *n.* 1. A line cutting diagonally across the grain of fabric. 2. Preference or inclination that inhibits impartiality; prejudice. *–adv.* On a diagonal; aslant. *–v.* **–ased** or **–assed, –as•ing** or **as•sing.** To cause to have a bias; prejudice. [< OFr. *biais*, oblique.

Dictionary forms

- Verbs tend to have just one dictionary entry

- Ditto singular and plural forms of nouns

- Forms of a lexeme share (some common element of) meaning.

So can we just use the dictionary?

- Not everything is in the dictionary

- Dictionaries are made by graduate students in linguistics and their supervisors, after reading a lot of older dictionaries.

- People love to make up new words

- This process is called *neologism*

- ...which itself only recently became a word:

Thomas Jefferson, amateur philologist
Letter to John Adams, 1820

"But if Dictionaries are to be the Arbiters of language, in which of them shall we find *neologism*? No matter. It is a good word, well sounding, obvious, and expresses an idea which would otherwise require circumlocution...

I am a friend to *neology*. It is the only way to give to a language copiousness and euphony. Without it we should still be held to the vocabulary of Alfred or of Ulphilas; and held to their state of science also: for I am sure they had no words which could have conveyed the ideas of oxygen, cotyledons, zoophytes, magnetism, electricity, hyaline, and thousands of others expressing ideas not then existing, nor of possible communication in the state of their language."

Thomas Jefferson, amateur philologist
Letter to John Adams, 1820

"What a language has the French become since the date of their revolution, by the free introduction of new words! The most copious and eloquent in the living world; and equal to the Greek, had not that been regularly modifiable almost ad infinitum. Their rule was that whenever their language furnished or adopted a root, all its branches, in every part of speech were legitimated by giving them their appropriate terminations: *adelphos* "brother", *adelphe* "sister", *adelphidion* "little brother", *adelphotes* "brotherly affection", *adelphixis* "brotherhood", *adelphidoys* "nephew", *adelphikos* "brotherly," adj., *adelphizo* "to adopt as a brother", *adelphikos* "brotherly," adv. And this should be the law of every language."

Thomas Jefferson, amateur philologist
Letter to John Adams, 1820

"Thus, having adopted the adjective *fraternal*, it is a root, which should legitimate *fraternity*, *fraternation*, *fraternisation*, *fraternism*, *to fraternate*, *fraternise*, *fraternally*. And give the word *neologism* to our language, as a root, and it should give us its fellow substantives, *neology*, *neologist*, *neologisation*; its adjectives *neologous*, *neological*, *neologistical*, its verb *neologise*, and adverb *neologically*."

Thomas Jefferson, amateur philologist
Letter to John Adams, 1820

"**Dictionaries are but the depositories of words already legitimated by usage.** Society is the work-shop in which new ones are elaborated. When an individual uses a new word, if illformed it is rejected in society, if wellformed, adopted, and, after due time, laid up in the depository of dictionaries. And if, in this process of sound neologisation, our transatlantic brethren shall not choose to accompany us, we may furnish, after the Ionians, a second example of a colonial dialect improving on its primitive."

What's right about deferring to the dictionary?

This inclination recognizes implicitly that in most languages, words cluster into groups that only deserve to be mentioned once, as a clan, in the dictionary.

The words in the clan divide up their usage among themsleves in a very regimented way: *the lexeme*.

More problems with dictionaries

		avoir (to have)	*speak* (fut.)
	1st	ai	parlerai
Sg	2nd	as	parleras
	3rd	a	parlera
	1st	avons	parlerons
Pl	2nd	avez	parlerez
	3rd	ont	parleront

The future forms all derived from an earlier stage of the language in which one said something like, "to speak have" to mark the future. These forms have now become suffixes.

Semantic domains

Yet another try:

Definition

A word is the smallest unit of meaning.

- The phoneme is too small...
    - Any phoneme can appear in a word with any meaning

- The sentence is too big....
    - Sentences are semantically compositional (more on this later)

- Maybe the word is just right?

Idioms

What about idioms?

- kick the bucket: this has nothing to do with kicking buckets.

- *It was the bucket that Uncle Goober kicked last Thursday.*

- tighten your belt: this sort of has something to do with tightening of belts, but not really...

- *After years of wretched excess, the studios' free-spending fat cats are only now learning to tighten their Gucci belts.*

Morphemes

Definition

The **morpheme** is the smallest unit of language that carries meaning or function.

The meaning of words is often predictable: *it is composed in a regular way of the meaning of its parts.* We will be discussing morphemes a lot. We will even consider the case where we want to posit morphemes independent of whether we can assign meaning to htem. But for now: morphemes carry meaning in a regular fashion. So words cannot be characterized as the smallest meaning-bearing units.

Morphemes

one	two	three	more than three
and			
couple	couple+s		
hunt	hunt+er	hunt+er+s	
act	act+ive	act+iv+ate	re+act+ive+ate

Pop quiz 1: spot the morpheme

Identify the morphemes in the following English words:

fly	reuse	spiteful	preplan
desks	triumphed	suite	optionality
untie	unkempt	fastest	prettier
tree	justly	deform	mistreat
dislike	payment	disobey	premature
receive	permit	submit	directors

Free vs. bound morphemes

Free morphemes can occur as separate words: *car, yes*

Bound morphemes cannot occur 'on their own':

dog	-s	dogs
de-	toxify	detoxify
create	-tion	creation
cran-	berry	cranberry

*I saw three s dog.
*John ate two apples and four crans.
*Alice had to de the water toxify.

Free vs. bound morphemes

What's free in one language may be bound in another.

In Hare (Athabaskan), body parts must be possessed:

without a possessor		with a possessor	
*fí	'head'	sefí	'my head'
*bé	'belly'	nebé	'my belly'
*dzé	'heart'	ʔedzé	'someone's heart/a heart'

Form a review of *Solar* (2010), by Ian McEwan:
Michael is a 50-something former Nobel laureate, resting on his fleshy laurels from twenty-two years ago, where he stood on the shoulders of Einstein and proposed a scientific "Conflation Theory" that was trailblazing at the time. Now, he tours around the globe giving lectures and consults for a large fee, and he sits idly as a member of a board at a center for renewable energy in the UK. His main pursuit is women, and he pursues them with -aholic depravity. As the novel opens, his fifth marriage is falling apart due to his infidelities. But this time, his wife got the last word by having some side dishes for herself and leaving him labeled as the cuckold.

Why is this funny?

IT had been a rough day, so when I walked into the party I was very chalant, despite my efforts to appear gruntled and consolate.

I was furling my wieldy umbrella for the coat check when I saw her standing alone in a corner. She was a descript person, a woman in a state of total array. Her hair was kempt, her clothing shevelled, and she moved in a gainly way.

I wanted desperately to meet her, but I knew I'd have to make bones about it, since I was travelling cognito. Beknownst to me, the hostess, whom I could see both hide and hair of, was very proper, so it would be skin off my nose if anything bad happened. And even though I had only swerving loyalty to her, my manners couldn't be peccable. Only toward and heard-of behavior would do.

Fortunately, the embarrassment that my maculate appearance might cause was evitable. There *were* two ways about it, but the chances that someone as flappable as I would be ept enough to become persona grata or a sung hero were slim. I was, after all, something to sneeze at, someone you could easily hold a candle to, someone who usually aroused bridled passion.

So I decided not to risk it. But then, all at once, for some apparent reason, she looked in my direction and smiled in a way that I could make heads or tails of.

Backformation

If you deal with employees, this blog is for you. As managers, executives, in-house counsel, and HR people, you know all about disgruntled employees. They cost employers billions of dollars each year in lawsuits, attorneys' fees, lost productivity, and wasted time. Here we discuss how to keep employees **gruntled**. Employer advocate and counsel Jay Shepherd leads the discussion.

```
http://www.gruntledemployees.com
```

Content vs. function morphemes

Content morphemes are not tied to grammatical function

- They can denote **things**: *sand, truck, Leslie*

- They can denote **actions**: *throw, kiss, live*

- They can denote **states**: *green, vile, open*

- You can make up new ones: *Smurf, nuke, grok, hoinh*

Content vs. function morphemes

Function morphemes tend to be tied to specific uses

- Prepositions: *to, by, with*

- Pronouns: *he, she, it*

- Articles: *a, the*

- Affixes: *-ness, sub-, -s*

Generally, you can't make up new ones (ever tried?)

Pop quiz 2: Farsi

xar means 'buy' and *-id* indicates past tense.

xaridam	'I bought'
xaridi	'you bought'
xarid	'(he) bought'
naxaridam	'I did not buy'
mixarid	'(he) was buying'
mixaridid	'you (pl.) were buying'

I
you (pl.)
not
you (sg.)
was/were + ing

How do you say...

- I was buying.
- You (sg.) did not buy.
- You (sg.) were buying.

Allomorphy

It turns out morphemes don't always map onto phonemes in the same way. For instance:

an	orange	*a*	basketball
an	accent	*a*	cash register
an	idea	*a*	hospital
an	eel	*a*	girl

- What is the generalization governing the behavior of the indefinite article in English?

- Is it based on pronunciation, or spelling?

- Other examples?

Hungarian

		ad give	**fej** milk	**öl** kill	**hoz** bring	**vés** chisel	**főz** cook
Sg	1st	adom	fejem	ölöm	hozom	vésem	főzöm
	2nd	adod	fejed	ölöd	hozod	vésed	főzöd
	3rd	adja	feji	öli	hozza	vési	főzi
Pl	1st	adjuk	fejjük	öljük	hozzuk	véssük	főzzük
	2nd	adjátok	fejitek	ölitek	hozzátok	vésitek	főzitek
	3rd	adják	fejik	ölik	hozzák	vésik	főzik

Hungarian: present definite

		ad give	**fej** milk	**öl** kill	**hoz** bring	**vés** chisel	**főz** cook
Sg	1st	adom	fejem	ölöm	hozom	vésem	főzöm
	2nd	adod	fejed	ölöd	hozod	vésed	főzöd
	3rd	adja	feji	öli	hozza	vési	főzi
Pl	1st	adjuk	fejjük	öljük	hozzuk	véssük	főzzük
	2nd	adjátok	fejitek	ölitek	hozzátok	vésitek	főzitek
	3rd	adják	fejik	ölik	hozzák	vésik	főzik

'I give', 'you milk', 's/he kills', 'we bring', 'they cook'...

Hungarian

		pad bench	**fej** head	**sör** beer	**bot** stick
Sg	1st	padom	fejem	söröm	botom
	2nd	padod	fejed	söröd	botod
	3rd	padja	feje	söre	botja
Pl	1st	padunk	fejünk	sörünk	botunk
	2nd	padotok	fejetek	sörötök	bototok
	3rd	padjuk	fejük	sörük	botjuk

Hungarian: nominal possessive

		pad bench	**fej** head	**sör** beer	**bot** stick
Sg	1st	padom	fejem	söröm	botom
	2nd	padod	fejed	söröd	botod
	3rd	padja	feje	söre	botja
Pl	1st	padunk	fejünk	sörünk	botunk
	2nd	padotok	fejetek	sörötök	bototok
	3rd	padjuk	fejük	sörük	botjuk

'my book', 'your head', 'his/her beer', 'our stick'...

Analyzing word structure

Simple words are unanalyzable: *the, trim, nine*

Complex words consist of a **root** plus one or more **affixes**
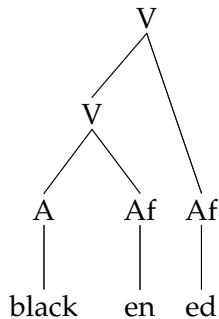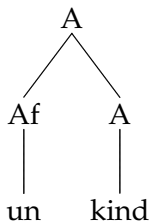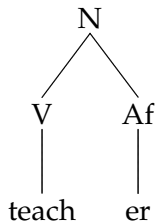
German superfix

gloss	infinitive	past participle
film	film-en	ge-film-t
ask	frag-en	ge-frag-t
praise	lob-en	ge-lob-t
show	zeig-en	ge-zeig-t

Roots and affixes

Roots tend to belong to a **lexical category**: noun, verb, preposition, adverb, etc.

Affixes include both **prefixes** and **suffixes**

prefixes	*suffixes*
de-active	faith-*ful*
re-play	govern-*ment*
il-legal	hunt-*er*
in-accurate	kind-*ness*

Different strokes

In Hungarian, 'prepositions' are either case endings...

case	form	gloss
NOMINATIVE	ház	'house'
INESSIVE	házban	'in a house'
ILLATIVE	házba	'into a house'
ADESSIVE	háznál	'at a house'
ELATIVE	házból	'out of a house'
ALLATIVE	házhóz	'to a house'

For different folks

...or **postpositions**:

elő**tt**	in front of
*a ház el**ő**tt*	'in front of the house'
mögött	behind
a ház mögött	'behind the house'
között	between, among
a fák között	'among the trees'
alatt	beneath
a kert alatt	'beneath the garden'

More postpositions

Hindi/Urdu uses postpositions almost exclusively:

(1) māliyāt kā mahekma
     finance of department
     'Department of Finance'

(2) śāhid   nē   apnī vālida kō tār   bhējā
     Shahid AGENT POS mother to telegram send
     'Shahid sent a telegram to his mother.'

(3) laṛkā dōstōṁ   kē sāth khēl rahā hai
     boy   friend.OBL.PL with   play is-doing
     'The boy is playing with friends.'

Infixes

Affixes may also occur *within* another morpheme

This is known as **infixation**, and such affixes are called **infixes**

Infixation in Tagalog:

base		*infix*	
bili	'buy'	b-**in**-ili	'bought'
basa	'read'	b-**in**-asa	'read'
sulat	'write'	s-**in**-ulat	'wrote'

Infixes

Nominalizing *-ni-* in Leti (Austronesian; data from Blevins 1999):

base		*infixed form*	
kaati	'to carve'	k-**ni**-aati	'carving'
polu	'to call'	p-**ni**-olu	'act of calling, call'
kini	'to kiss'	k-**n**-ini	'act of kissing, kiss'
tutu	'to support'	t-**n**-utu	'act of supporting, support'

Is there infixation in English?

Expletive infixation

English expletive infixation (McCarthy 1982):

togéther	to-**bloody**-gether
enóugh	e-**bloody**-nough
Kalamazóo	Kalama-**goddam**-zoo
absolútely	abso-**goddam**-lutely
fantástic	fan-**friggin**-tastic
unbelíevable	un-**friggin**-believable

Note that this has something to do with phonology...

Homeric infixation (Yu 2004)

'A hundred bucks? For a comic book? Who drew it, Micha-ma-langelo?'

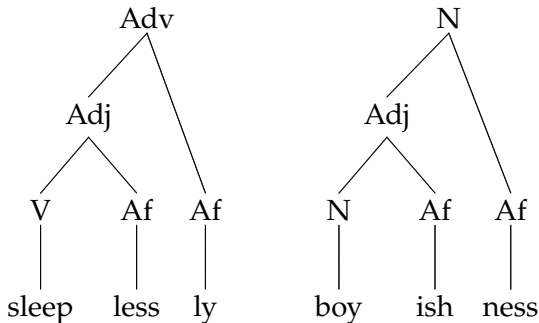|	
oboe	oba-**ma**-boe
opus	opa-**ma**-pus
tuba	tuba-**ma**-ba
violin	vio-**ma**-lin
Alabama	Ala-**ma**-bama
educate	edu-**ma**-cate
complicated	compli-**ma**-cated

(cf. 'diddly' infixation: Elfner & Wimpner 2008)

More English infixation?

What about words like *sleeplessly*, *boyishness*?
What is the status of *-less-*, *-ish-*?

The relationship between words and morphemes

- A **morpheme** is the minimal unit of meaning
  - *bound morphemes* cannot occur in isolation: *cran-, -ness*
  - *free morphemes* can stand alone: *blue, teach*
  - *allomorphs* are functionally equivalent forms of the same morpheme used in distinct contexts: *a, an*

- A **word** is a minimal free form that can occur in isolation and/or whose position with respect to neighboring elements is not entirely fixed
  - *simple words* cannot be broken down further: *the, kayak*
  - *complex words* have multiple parts: *himself = him + self*

The relationship between words and morphemes

Can a word be a morpheme?

word (=morpheme)	category
think	verb
true	adjective
succotash	noun
gosh!	interjection
under	preposition
that	conjunction
often	adverb

The relationship between words and morphemes

Can a morpheme *not* be a word?

morpheme	category
un-	prefix
dis-	prefix
-ness	suffix (nominalizing)
-s	suffix (pluralizing)
kempt	bound morph (*unkempt*)
cran-	bound morph

The relationship between words and morphemes

Can a word be a syllable?

word	*category*
car	noun
work	verb
in	preposition
whoops!	interjection

The relationship between words and morphemes

Are there morphemes that aren't syllables?

| morpheme	category
under	preposition ($> \sigma$)
spider	noun ($> \sigma$)
-s	'plural' ($< \sigma$)

The relationship between words and morphemes

Are there syllables that aren't morphemes?

word	*syllables*
kayak	kai.jæk
broccoli	bɹɔ.kə.li, bɹɔk.li
angle	æŋ.gl
jungle	jʌŋ.gl

Putting morphemes together

- In English (and Indo-European languages), morphemes are *largely* put together in a sequence, or **concatenated**

- This is not the only way to do things

- Semitic languages (especially) are **non-concatenative**: they form words using a special type of infixation sometimes called **templatic morphology**

Putting morphemes together

affix

k   i   t   a   b

root ('write')
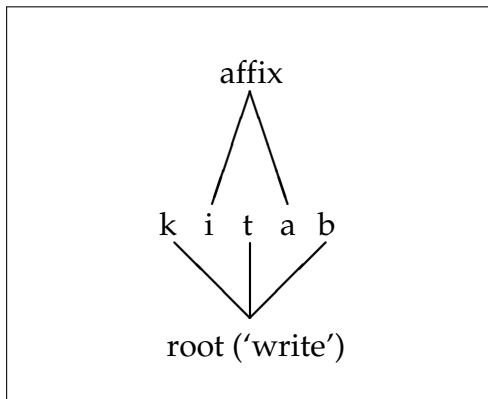
ka**t**a**b**a   **k**u**t**i**b**   a**kt**u**b**
'wrote'   'has been written'   'am writing'

Derivational versus inflectional morphology

Linguists struggle to find a simple account of the difference between *inflectional* morphology and *derivational* morphology.

But anyone who studies the relations between words of a language knows that this distinction is real.

Content words (nouns, verbs, adjectives) very often are organized into families of words organized along a set of (nearly) independent dimensions, as far as linguistic function is concerned; we call these *paradigms*.

Present tense of the verb *aller* 'to go'

person	singular	plural
1st	vais	allons
2nd	vas	allez
3rd	va	vont

3 dimensions...

Present tense of the verb *aller* 'to go'

person	singular	plural
1st	vais	allons
2nd	vas	allez
3rd	va	vont
imperfect of verb *aller*		
1st	allais	allion
2nd	allais	alliez
3rd	allait	allaient

Inflectional morphology

Inflectional affixes indicate *grammatical information* (tense, number, person...) but don't change category.

When we linguists say that, we mean that our statements about syntax will remain the same regardless of which particular choice of the inflectional paradigms is used in a sentence. We find the verb in the same place in a French sentence regardless of what person, number, or aspect it is in; this is what we mean when we say that inflectional morphology does not change part of speech.

form 1	*form 2*	*lexical category*
dog	*dogs*	noun
run	*runs*	verb
tall	*taller*	adjective
old	*oldest*	adjective

Inflectional morphology

English has pretty anemic inflectional morphology:

affix	function	example
-s	3rd sing. pres.	*He talks.*
-ed	past tense	*He talked.*
-ing	progressive	*He's talking.*
-ed, -en	past participle	*He talked.*
-s	plural (of noun)	*The cats are sleeping.*
-'s	possessive	*The cat's food is ready.*
-er	comparative	*John is taller than Mary.*
-est	superlative	*John is Bob's oldest son.*

Inflectional morphology

There are also a few other ways to mark inflectional contrast:

- **ablaut** (quasi-predictable vowel change)
- **suppletion** (substituting one form for another)

present	past
go	*went*
am	*was*
come	*came*
fall	*fell*
eat	*ate*

singular	plural
foot	*feet*
moose	*moose*

When a rule of syntax (grammar) involves agreement between two words in a sentence for some grammatical feature, the realization of that feature by each of the words is part of inflectional morphology. Examples: subject-verb agreement; noun-adjective agreement.

gender	Number:singular	Number:plural
masculine	pequeño	pequeños
feminine	pequeña	pequeñas

$$\text{pequeñ} \left\{ \begin{matrix} o \\ a \end{matrix} \right\} \left\{ \begin{matrix} \emptyset \\ s \end{matrix} \right\}$$

French adjectives

gender	masc. sg.	masc.plural	fem. sg.	fem. plural
small	pəti	pəti	pətit	pətit
large	grã	grã	grãd	grãd
normal	normal	normo	normal	normal
green	ver	ver	vert	vert
red	ʁuʒ	ʁuʒ	ʁuʒ	ʁuʒ
good	bõ	bõ	bɔn	bɔn
gray	gʁi	gʁi	gʁiz	gʁiz
long	lõ	lõ	lõg	lõg
hot	ʃo	ʃo	ʃod	ʃod
white	blã	blã	blãʃ	blãʃ
fresh	fʁe	fʁe	fʁɛʃ	fʁɛʃ
false	fo	fo	fos	fos

Subtractive morphology in the formation of the masculine form.

Estonian (Blevins 2006)

	SING(ULAR)	PLUR(AL)
NOM(INATIVE)	—	—d
GEN(ITIVE)	—	—de/—te
PART(ITIVE)	—/—t/—d	—sid/—id
STEM PART PLUR (PART2)		—
SHORT ILLA SG (ILLA2)	—	
ILLA(TIVE)	—sse	
INES(SIVE)	—s	
ELA(TIVE)	—st	
ALLA(TIVE)	—le	
ADES(SIVE)	—l	
ABLA(TIVE)	—lt	
TRANS(LATIVE)	—ks	
TERM(INATIVE)	—ni	
ESS(IVE)	—na	
ABES(SIVE)	—ta	
COM(ITATIVE)	—ga	

Table 1. Declensional endings in Estonian

Estonian (Blevins 2006)

	Disyllabic		Quadrisyllabic		Quantitative Grade		Qualitative Grade	
	SING	PLUR	SING	PLUR	SING	PLUR	SING	PLUR
NOM	maja	*majad*	koridor	*koridorid*	`tool	*toolid*	rida	`read
GEN	*maja*	majade	*koridori*	koridoride	tooli	`toolide	`rea	ridade
PART	maja	(majasid)	koridori	(koridorisid)	`tooli	(`toolisid)	rida	(ridasid)
PART2		maju		koridore		`toole		ridu
ILLA2	`majja		koridori		`tooli		`ritta	
ILLA	*majasse*	majadesse	*koridorisse*	koridorisse	*toolisse*	`toolidesse	`reasse	ridadesse
INES	*majas*	majades	*koridoris*	koridoris	*toolis*	`toolides	`reas	ridades
ELA	*majast*	majadest	*koridorist*	koridorist	*toolist*	`toolidest	`reast	ridadest
ALLA	*majale*	majadele	*koridorile*	koridorile	*toolile*	`toolidele	`reale	ridadele
ADES	*majal*	majadel	*koridoril*	koridoril	*toolil*	`toolidel	`real	ridadele
ABLA	*majalt*	majadelt	*koridorilt*	koridorilt	*toolilt*	`toolidelt	`realt	ridadelt
TRANS	*majaks*	majadeks	*koridoriks*	koridoriks	*tooliks*	`toolideks	`reaks	ridadeks
TERM	*majani*	majadeni	*koridorini*	koridorini	*toolini*	`toolideni	`reani	ridadeni
ESS	*majana*	majadena	*koridorina*	koridorina	*toolina*	`toolidena	`reana	ridadena
ABES	*majata*	majadeta	*koridorita*	koridorita	*toolita*	`toolideta	`reata	ridadeta
COM	*majaga*	majadega	*koridoriga*	koridoriga	*tooliga*	`toolidega	`reaga	ridadega
	'house' (3)		'corridor' (6)		'chair' (20)		'row' (16)	

Table 12. Exemplary first declension paradigms

Derivational morphology

Derivational morphology deals with the relationship between separate lexemes, typically in different parts of speech, and typically in a way that is both semantically irregular, and limited in its range of application.

As a consequence, a derivationally derived form can always be replaced by a non-derived form — something that is rarely the case for inflectionally complex forms.

verbal base	derived noun
sell	sell-er
write	writ-er
teach	teach-er
sing	sing-er
discover	discover-er

How many derivational affixes does English have? ...

English derivational affixes

affix	root	derived form
-ation	is added to a verb	to give a noun
	finalize	finalization
	confirm	confirmation
un-	is added to a verb	to give a verb
	tie	untie
	wind	unwind
un-	is added to an adjective	to give an adjective
	happy	unhappy
	wise	unwise
-al	is added to a noun	to give an adjective
	institution	institutional
	universe	universal

Idiosyncracy in the derivational lexicon

–ist	–ism	–ize
baptist	baptism	baptize
exorcist	exorcism	exorcize
terrorist	terrorism	terrorize

Idiosyncracy in the derivational lexicon

–ist	–ism	–ize
baptist	baptism	baptize
exorcist	exorcism	exorcize
terrorist	terrorism	terrorize
violinist	*violinism	*violinize

Idiosyncracy in the derivational lexicon

–ist	–ism	–ize
baptist	baptism	baptize
exorcist	exorcism	exorcize
terrorist	terrorism	terrorize
violinist	*violinism	*violinize
organist	!organism	!organize
publicist	*publicism	publicize

Idiosyncracy in the derivational lexicon

–ist	–ism	–ize
baptist	baptism	baptize
exorcist	exorcism	exorcize
terrorist	terrorism	terrorize
violinist	*violinism	*violinize
organist	!organism	!organize
publicist	*publicism	publicize
womanist	*womanism	womanize
materialist	materialism	!materialize

Idiosyncracy in the derivational lexicon

–ist	–ism	–ize
baptist	baptism	baptize
exorcist	exorcism	exorcize
terrorist	terrorism	terrorize
violinist	*violinism	*violinize
organist	!organism	!organize
publicist	*publicism	publicize
womanist	*womanism	womanize
materialist	materialism	!materialize
atheist (1568)	aetheism	!atheize
linguist (1588)	*linguism	*linguize

Idiosyncracy in the derivational lexicon

–ist	–ism	–ize
baptist	baptism	baptize
exorcist	exorcism	exorcize
terrorist	terrorism	terrorize
violinist	*violinism	*violinize
organist	!organism	!organize
publicist	*publicism	publicize
womanist	*womanism	womanize
materialist	materialism	!materialize
atheist (1568)	aetheism	!atheize
linguist (1588)	*linguism	*linguize
humanist (1589)	humanism	?humanize
rationalist (1627)	rationalism	!rationalize

More idiosyncracy in the derivational lexicon

–ese	Bhutanese, Chinese, Vietnamese Japanese, Lebanese, Maltese Taiwanese, Portuguese
–an	African, Alaskan, American Angolan, Cuban, Jamaican Mexican, Nicaraguan
–ian	Argentinian, Armenian, Canadian Ethiopian, Bolivian, Serbian Jordanian, Palestinian
–ish	Scottish, British, Flemish Swedish, Polish, Danish Irish
–i	Iraqi, Israeli, Pakistani
–?	French, German, Greek, Thai

What's going on here?

- *Brazil::Brazilians, Australia::Australians...*

- So why not *East Timor::East Timorians*?

- Why *not* regularize?

Pop quiz 3: inflectional vs. derivation

Are the following affixes inflectional or derivational?

sof**ten**	*(Heating the cheese will soften it.)*
pollu**tion**	*(The pollution of the forest was tragic.)*
read**ing**	*(I am reading a book.)*
read**ing**	*(The reading of the poem was beautifully done.)*
king**dom**	*(The knight rode across the kingdom.)*
happi**er**	*(My friend is happier than I am.)*

(are both the *ing*s the same? are there others?)

Interesting excercise for you

1. Estimate the number of words you know.
2. Specifically:
   - Estimate the number of words in your active vocabulary (those words or word forms you have used or would use in spoken or written language), and
   - Estimate the number of words in your passive vocabulary (those words whose meanings you recognize and would understand when they are used in an appropriate context, but which you have never used yourself).

Interesting exercise (cont.)

1. Method:
   Take an unabridged dictionary of English. Open to a random
   page: count the number of entries (lemmata) on that page. Ask
   yourself for each entry whether you 1. would understand the
   word if you heard it in the context of a sentence (call the total such
   forms on the page p), and 2. whether you have ever yourself used
   the word in speech or writing (=a). Multiply p and a separately by
   (1) the number of pages in the dictionary that have entries (n_e) and
   then by (2) the total number of claimed entries in the dictionary
   [found on the cover or in the introduction] (n_c) divided by the
   total number of entries on the page you looked at (n_p). I.e.,

   $$A = p \cdot n_e \quad C = p \cdot (n_c/n_p)$$
   $$B = a \cdot n_e \quad D = a \cdot (n_c/n_p)$$

Interesting exercise (cont.)

1. Repeat, but now counting every form of every word on the page (for nouns: plural and singular; for adjectives: comparative, superlative; for verbs: all parts); also use any sublemmata (typically found when affixation gives rise to new forms that are entirely derivative or transparent). Feel free to note if the dictionary does not indicate all the possible forms (especially for derivational affixation).

2. Take results with large grain of salt.

What is the meaning of an affix?

un-	*untie*	con-	*constitution*
	unshackle		*confess*
	unharness		*connect*
	unhappy		*contract*
	untimely		*contend*
	unthinkable		*conspire*
	unmentionable		*complete*

outshoot	outthink	outtrade	outfox
outrun	outdo	outsmart	outpatient

Some *out*-s from Merriam-Webster

outachieve	outact	outbargain	outbid
outbitch	outbluff	outbox	outbrag
outbrawl	outbulk	outbuy	outcatch
outcharge	outclimb	outcoach	outcompete
outdance	outdazzle	outdebate	outdeliver
outdesign	outdrag	outdress	outdrink
outdrive	outduel	outearn	outeat
outfight	outfigure	outfish	outfly
outfumble	outgain	outglitter	outgross
outhit	outhomer	outhunt	outhustle
outjump	outkick	outkill	outlast
outleap	outlearn	outman	outmaneuver
outmanipulate	outmarch	outmuscle	outorganize
outpass	outperform	outpitch	outplay

out-

Compare with:

outage	outback	outboard	outbreak
outcast	outcome	outcry	outdoors
outgrow	outlandish	outlet	outnumber?
outpour	output	outright	outstanding
outward			

Morphological typology

The standard terminology goes back to Edward Sapir *Language* (1921): In an **isolating** language, words are (all) composed of single morphemes: low morpheme/word ratio.

In a **synthetic** language, words tend to consist of multiple morphemes.

A distinction is often drawn between **(in)flectional** and **agglutinative** systems. This distinction deals with inflectional systems. Compare French with an imaginary agglutinative language:

	French		Agglu	
person	singular	plural	singular	plural
1st	vais	allons	skip	skips
2nd	vas	allez	skipo	skipso
3rd	va	vont	skipum	skipsum

One word/one morpheme?

bàn	*ghế*	*bàn ghế*
table	chair	'furniture'
xe	*đạp*	*xe đạp*
vehicle	kick	'bicycle'
trả	*lời*	*trả lời*
to return	word	'answer'

Polysynthetic languages

- Greenlandic (and other Eskimo-Aleut languages) make extensive use of **postbases**: multiple, recursively addable derivational suffixes

- Given a root, you can create a 'word' of alarming length

- Thus the potential number of 'words' is arguably limitless...

Greenlandic *qani-* 'snow'

quanik	'snowflake'
qaniit	'snow (in the air)'
qinoq	'slush snow'
qannerpoq	'it's snowing'
qanimavoq	'shivers'
qanipalaat	'clumps of falling snow'

Greenlandic *oqaq-* 'tongue'

oqarpoq	'says'
oqaaseq	'word'
oqaluppoq	'speaks'
oqaasilerisoq	'linguist'
oqaasilerissutit	'grammar'
oqaluttualiortoq	'author'
oqaatiginerluppaa	'speaks badly about him'

New words

- Back to Jefferson: neologism

- Another way that morphology is markedly different from other areas of grammar

- (Seemingly) arbitrary innovations aren't allowed in syntax, or phonology...

- But people add new words all the time (and change the meaning of existing words)

climb down

- Can Santa really **climb down** the chimney?

`http://christmas.howstuffworks.com/traditions/santa-chimney.htm`

- You'll find a ladder to **climb down** into the ship's hold.

`http://www.gamespot.com/features/2871423/p-9.html`

- Boys **climbing down** to the water

`http://www.eveandersson.com/photo-display/large/france/south/`
`cassis-calanque-de-port-miou-boys.html`

silly

1a. Deserving of pity, compassion, or sympathy.

- **1556** in W. H. Turner *Select Rec. Oxford* 246
  'The fire raging upon the silly Carcase.'

1b. Helpless, defenceless; esp. of women and children. *Obs.*

- **1591** SHAKES. *Two Gent.* IV. i. 72
  'Prouided that you do no outrages On silly women, or poore passengers.'

1c. Of animals, esp. as a (poetic) epithet of sheep.

- **1866** M. ARNOLD *Thyrsis* v,
  'He could not keep..Here with the shepherds and the silly sheep.'

silly

2d. Scanty, sorry, meagre, poor. *Obs*.

- **1767** SIR R. COLVILLE in Dossie *Mem. Agric.* (1768) I. 412
  'Marsh land, of a light, silly, hungry soil.'

3b. Of humble rank or state; lowly. *Obs.*

- **1577-87** HOLINSHED *Chron.* II. 96
  'He was shot thorough with an arrow amongst his men by a sillie footman.'

5a. Lacking in judgement or common sense; foolish.

- **1840** DICKENS *Barn. Rudge* iii,
  "Heaven help this silly fellow,' murmured the perplexed locksmith.'

1991
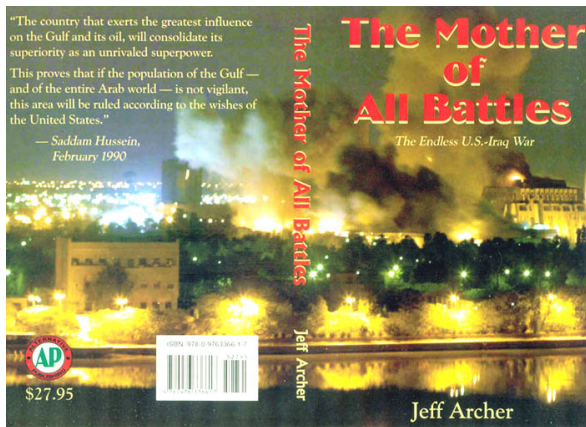
mother of all (not actually a word, right...)

1992
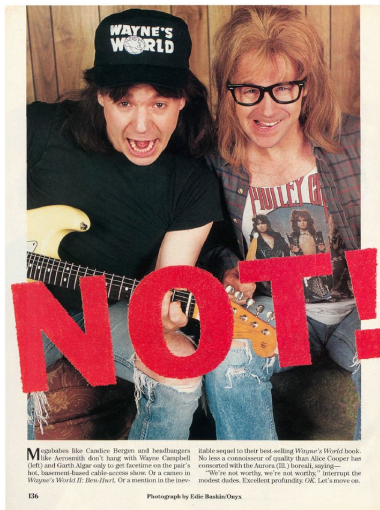
not (as in 'just kidding'. What was the innovation here?)
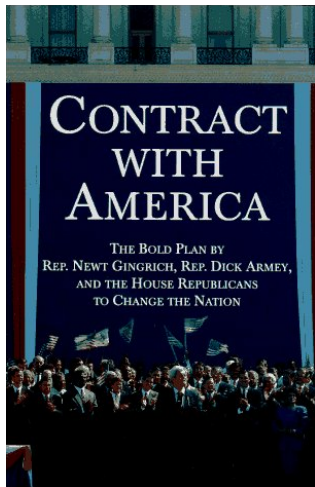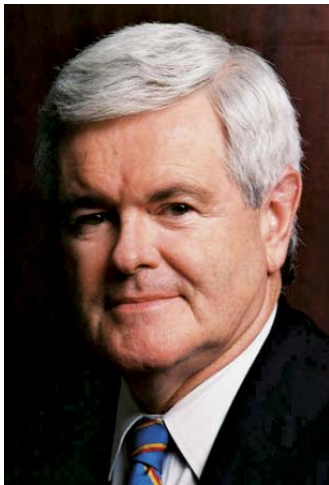
1993

information superhighway

1994

cyber, morph (as in 'to change form')
This spawned all sorts of children...

1995

(to) newt

('act aggressively as a newcomer')

1998

e-
(*i-* hadn't been invented yet...)

2003

metrosexual

- *n.* a fashion-conscious heterosexual male, or, as coiner Mark Simpson put it, a man who "has clearly taken himself as his own love object."

- a *portmanteau* (or *blend*):
  - *smog*
  - *brunch*
  - *spork*
  - *Brangelina*
  - ...etc.

2005

1. truthiness (noun)

1 : "truth that comes from the gut, not books" (Stephen Colbert, Comedy Central's "The Colbert Report," October 2005)

2 : "the quality of preferring concepts or facts one wishes to be true, rather than concepts or facts known to be true" (American Dialect Society, January 2006)

truthiness

2009

n. a short message sent via the Twitter.com service

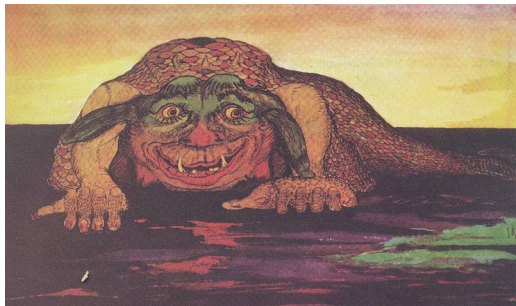*v.* the act of sending such a message.

tweet

Word of the Decade

google

('to search the internet'. Cf. *Xerox*, *Kleenex*, etc.)

A century earlier...

The sun is setting –
Can't you hear
A *something* in the distance
                    Howl!!?
I wonder if it's –
Yes!! it *is*
That horrid Google
On the prowl!!!

V. C. Vickers, *The Google Book* (1913)

`http://blogoscoped.com/googlebook/`