

Self-organization in the English suffix system

**Mark Lindsay
Stony Brook University
Stony Brook, NY**

QMMMD

University of California, San Diego

15-16 January 2011

Introduction

"As many more individuals of each species are born than can possibly survive; and as, consequently, there is a frequently recurring struggle for existence, it follows that any being, if it vary... in any manner profitable to itself... will have a better chance of surviving, and thus be naturally selected."

— Charles Darwin, [The Origin of Species](#) (1859)

- ▶ Self-organization is a natural process
 - ▶ Ecosystems organize themselves
 - ▶ Language does too

Introduction

- ▶ Suffix productivity
- ▶ 3 major factors:
 - ▶ Randomness
 - ▶ Productive derivation
 - ▶ Intolerance of synonymy
- ▶ Leads to “survival of the fittest”

Source of randomness

- ▶ New word creation
- ▶ Borrowings from other languages
- ▶ Reanalysis
- ▶ Speech errors

Competition

- ▶ Synonymy is not tolerated
- ▶ Suffix must combine with new words
 - ▶ Or, it “dies off” (productively)
 - ▶ Why?
- ▶ An out-competed suffix must differentiate itself (somehow)

Means of Propagation

- ▶ Suffixes attach to words, creating new words
 - ▶ Maximize words to attach to
 - ▶ Minimize restrictions
- ▶ Exemplars extend/reinforce the pattern

Adaptation

- ▶ Co-existence by defining separate domains:
 - ▶ phonological (stress, syllables, prosody)
 - ▶ morphological (potentiation)
 - ▶ semantic (no longer synonymous)
 - ▶ pragmatic (formal/informal, technical)
 - ▶ etc.

Borrowed Suffixes

- ▶ Only looking at English suffixes that arise through borrowing
 - ▶ First: borrowing a lot of words
 - ▶ Then: productivity
 - ▶ Don't arise through grammaticalization
- ▶ All of these suffixes *must* have self-organized

Self-organization

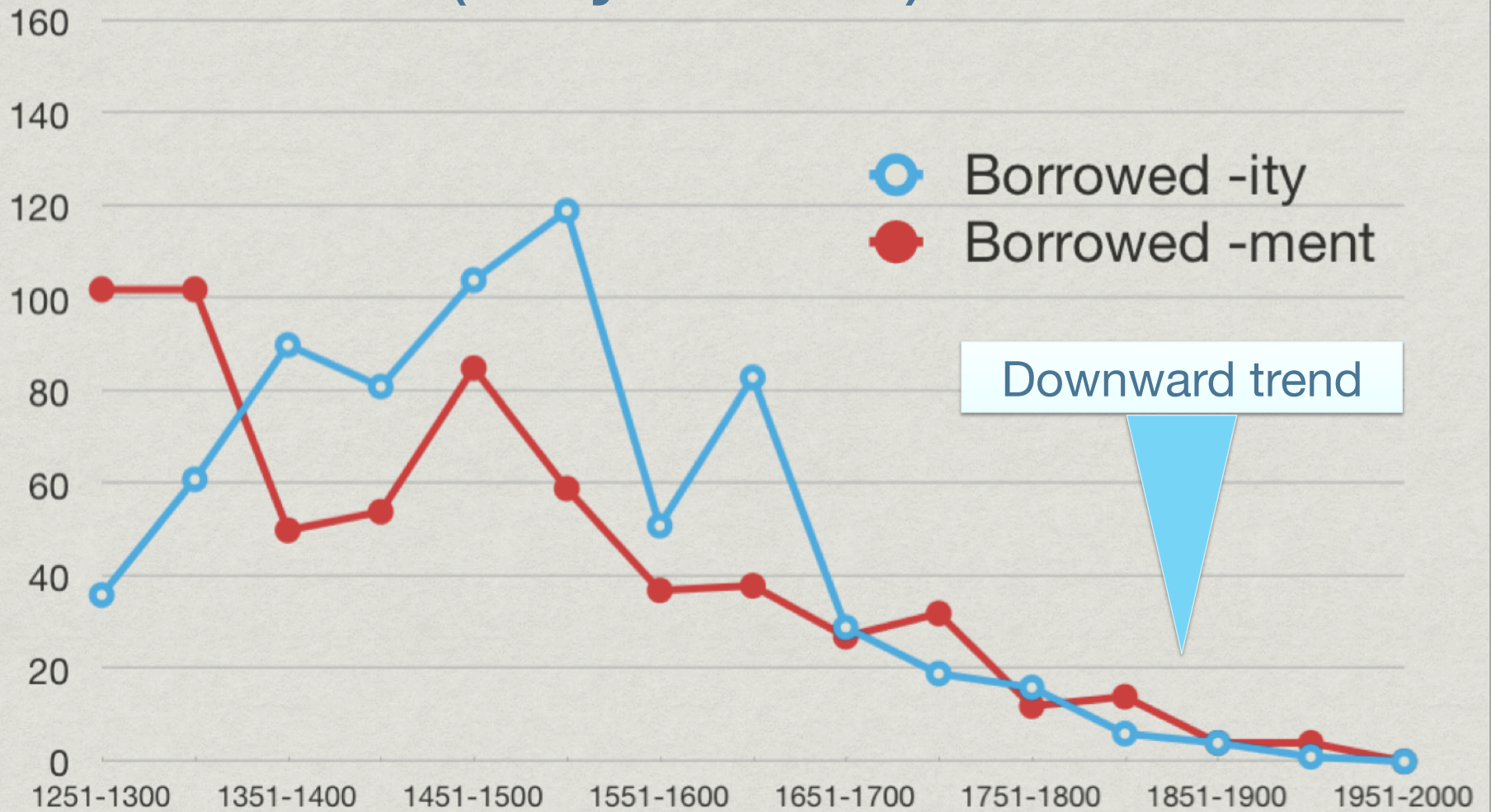
- ▶ Competition for productivity in language
- ▶ Google as a tool for investigating productivity
- ▶ English suffixes:
 - I. -ment, -ation and -ity** (diachronic)
 - II. -ic and -ical** (synchronic)
 - III. -ize and -ify** (synchronic)

-ment and -ity

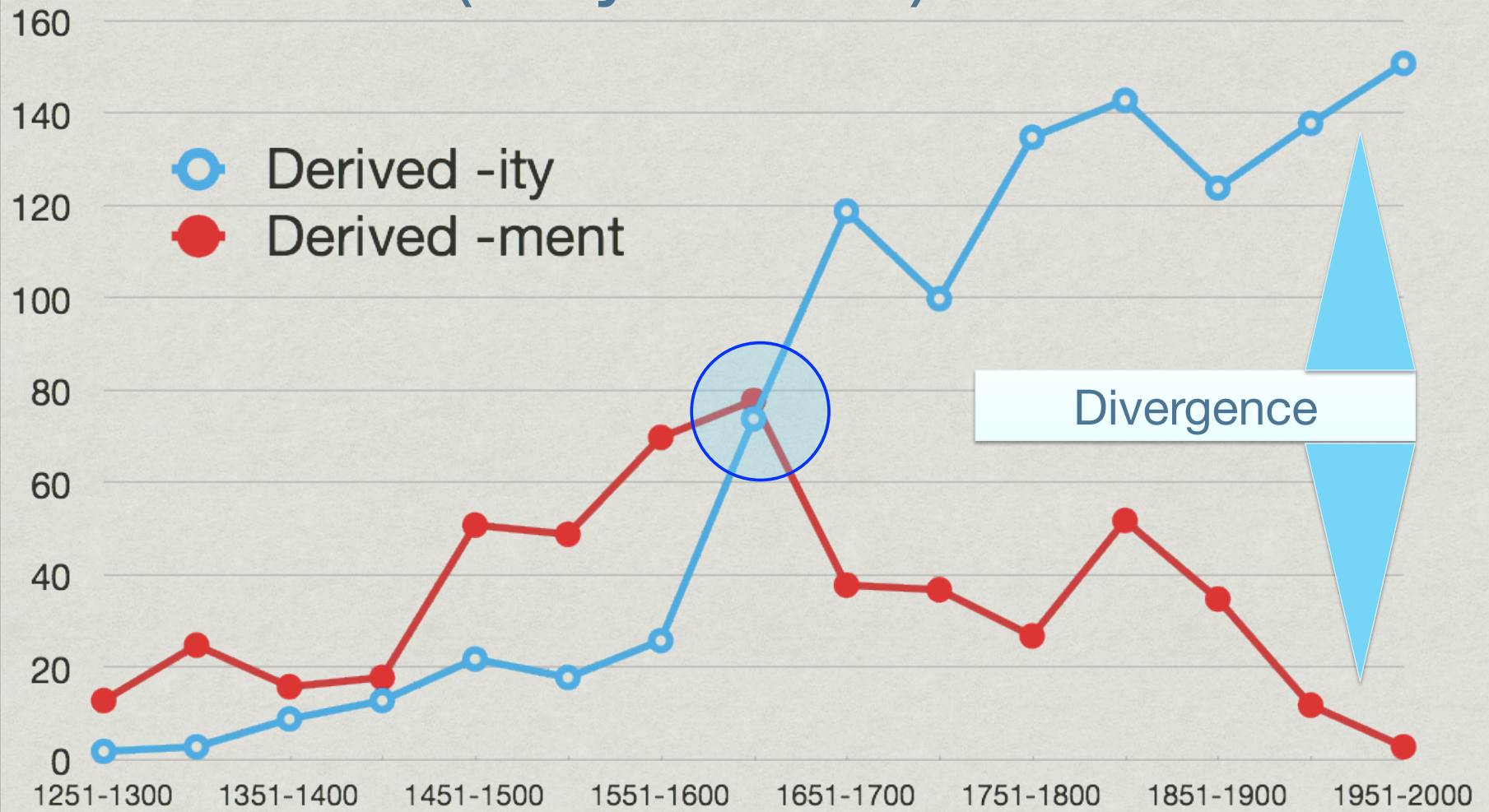
(Anshen & Aronoff 1999, Lindsay & Aronoff 2010)

- ▶ e.g. *deployment*, *amendment*, *concealment*
vulgarity, *overactivity*, *reality*
- ▶ Originally borrowed into English from French
- ▶ **Over time, borrowings decreased, while derivations increased**
- ▶ **-ity** remained productive, **-ment** did not
 - ▶ **-ment** is now (productively) “dead”

Borrowed forms since 1251 (adjusted)



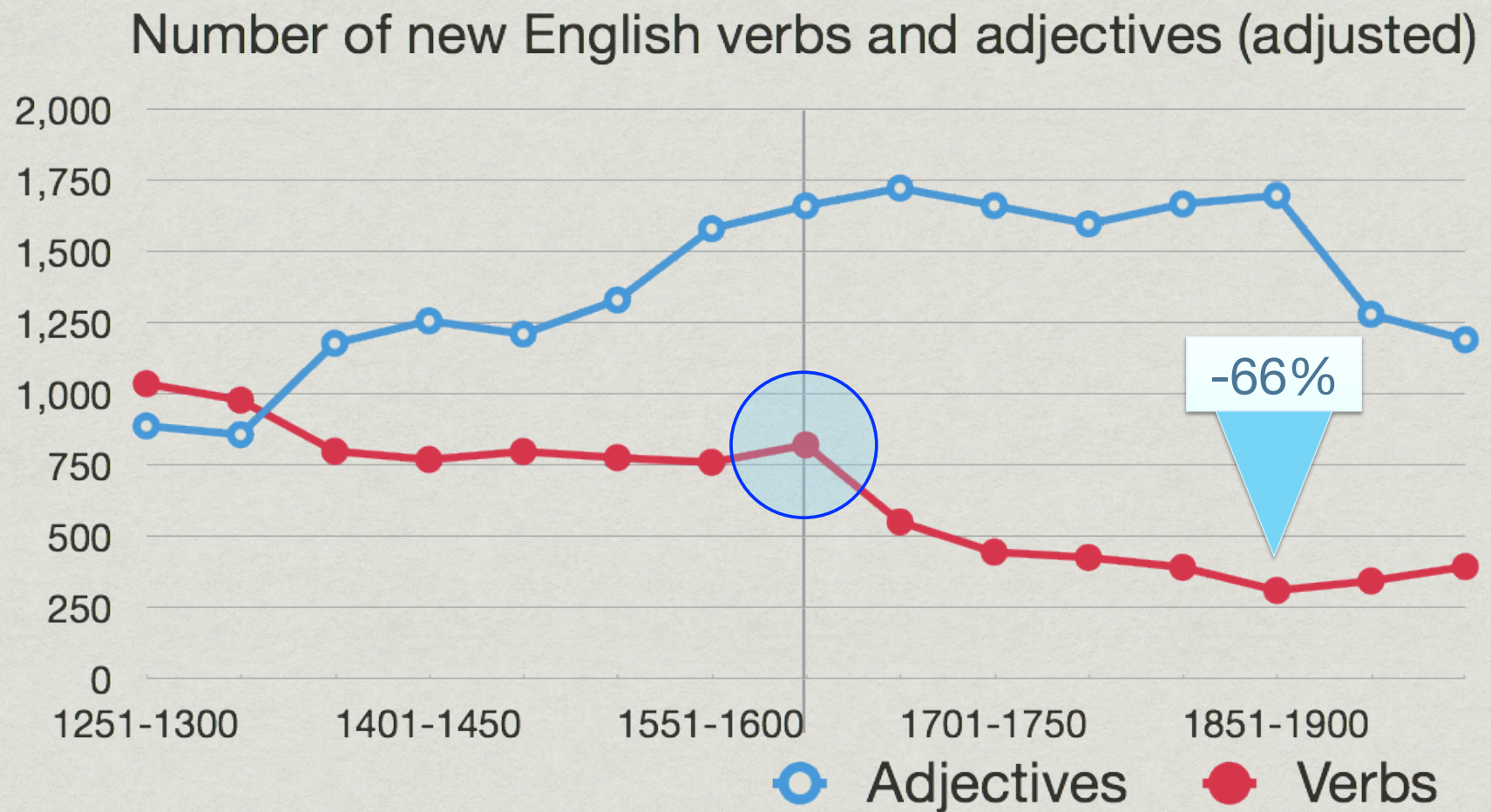
Derived forms since 1251 (adjusted)



What went wrong with **-ment**?

- ▶ New borrowings had decreased significantly
- ▶ **-ment** and **-ity** had different niches
 - ▶ **-ity** converts adjectives to nouns
e.g. *equal* → *equality*
 - ▶ **-ment** converts verbs to nouns
e.g. *punish* → *punishment*
- ▶ The number of new verbs decreased significantly at the same time as the decline of **-ment** (decline of 66% by the 1850s)

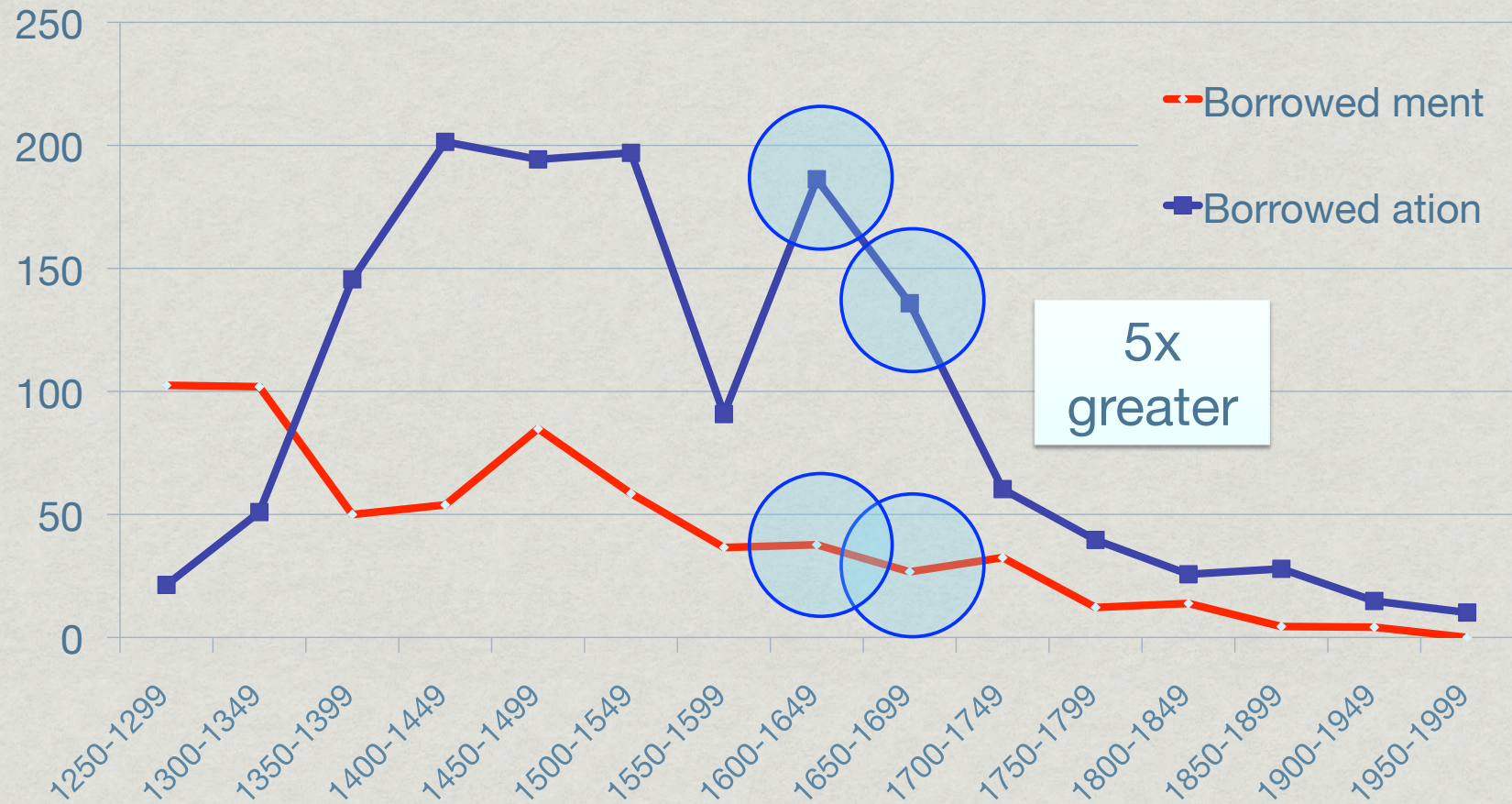
Different fates from changes in **habitat**



Competition: -ation

- ▶ e.g. **exportation**, **glorification**, **organization**
- ▶ Borrowed from French and Latin
- ▶ Same ultimate decrease in borrowings over time
- ▶ **-ation** outcompeted **-ment**
 - ▶ What gave **-ation** an edge?

-ment vs. -ation (adjusted)



-ation and -ment

- ▶ During critical time (1600s)
 - ▶ Far fewer new verbs
 - ▶ **5x** as many *-ation* borrowings vs. *-ment*
- ▶ *-ment* could not compete with *-ation* and could not survive on its own

Rival suffixes: ***-ic*** and ***-ical***

- ▶ Synonymous suffixes
 - ▶ **-ic**: came from Greek *-ikós*
 - ▶ **-ical**: *-al* from French
- ▶ Sometimes one suffix is favored over the other, sometimes both forms exist for one stem

Examples

▶ *Doublets:*

- ▶ ***symmetric*** vs. ***symmetrical***
- ▶ ***historic*** vs. ***historical***
- ▶ ***electric*** vs. ***electrical***

▶ *Clear preferences:*

- ▶ ***electronic*** > ***electronical***
- ▶ ***surgical*** > ***surgic***
- ▶ ***atomic*** > ***atomical***

Word-by-word

- ▶ Common doublets have developed distinct meanings:
 - ▶ **historical:** “of, pertaining to, treating, or characteristic of history or past events”*
 - ▶ **historic:** “well-known or important in history”*

*Definitions from Random House Dictionary (2010)

Distinctions of words, not of suffixes

- ▶ Difference in meaning between e.g. ***historic*** and ***historical*** cannot be generalized to the differences in other **-ic/-ical** pairs

~~historic : electric :: historical : electrical~~

On **-ic** vs. **-ical**:

“There was, at the beginning, indiscriminate coexistence of two synonymous adjectives. But language does not like to have two words for one and the same notion, and competition was bound to come.”

– Marchand (1969, p. 241-242)

-ic/-ical synonymy

- ▶ These suffixes are synonymous → they are in direct competition
- ▶ **Questions:**
 - ▶ Are both productive?
 - ▶ If so, which one is more productive?
 - ▶ How do they coexist?
- ▶ **A closer look – using Google**

Using Google

- ▶ Snapshot of productivity at a moment in time
- ▶ Large: virtually all words are measurable
- ▶ Free (working with Google)
- ▶ Lots of noise
- ▶ Searching is limited
- ▶ No part-of-speech-tagging

Using Google

1. Generate stems

- Use Webster's 2nd Edition as starting point
- Extract all words ending in **-ic** or **-ical**
- Strip off the suffixes to compile list of all unique stems

▶ Result: **11,966 words ending in -ic or -ical (or both)**

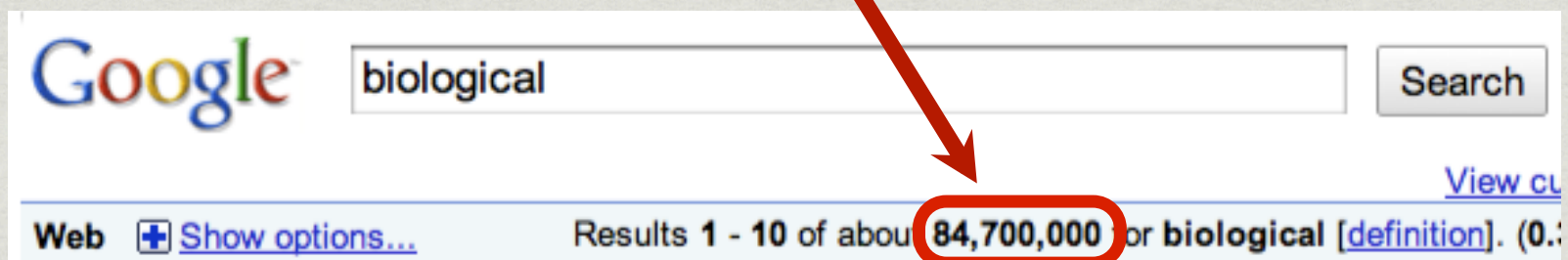
abenter- ablet-
abietin-
abiogenet-
abiolog- abiot-
abiotroph-
ablastem- ablept-
Abraham-
Abrahamit-
absinth-
absinthism-
absolutist-
absorptiometr-
abul-
abyssobenthon-
abyssopelag-
academ- acanon-
acantholog-
acarotox-
acatalect-
acatalept-
acatallact-
acatastat-
acategor-
acathol-

Using Google

2. Run automated Google queries for all stems

- *stem + ic* (e.g. *biologic*)
- *stem + ical* (e.g. *biological*)

THIS RESULT: Estimated Total Matches



Using Google Search API

Using Google

3. Store the results in a database and look at the general picture

DB

stem	-ic	-ical
materialist-	3010000	1470
mathemat-	1740000	48900000
mathet-	9030	6550
matriarch-	4990	5230
matr-	948000	22000
matroclin-	589	0
matronym-	16600	20
maxillozygomat-	373	0
maximist-	540	0
maz-	44900	541
mazolyt-	186	0
mazopath-	349	0
mechan-	43800000	139000000
mechanicochem-	0	1210
mechanist-	3390000	18900
mechanomorph-	2870	0
mechanotherapeut-	369	17
mecon-	17600	45
medall-	90600	0
median-	4120	1130
medianim-	433	0

Comparison

- ▶ Some stems had a significant number of Google hits for both forms:
 - ▶ *historic/historical*
- ▶ Others had only results for one:
 - ▶ *overenthusiastic* (**overenthusiastical*)
- ▶ *Most had non-zero results for both*
- ▶ **For 90% of pairs, Google hits differed by 1+ order of magnitude**
- ▶ **For each pair, word with most hits = “winner”**

Basic Results

suffix	-ic	-ical
winners	10,613	1,353
ratio	7.84	1

-ic is more productive than -ical (~8 to 1)

Morphological niche

- ▶ Why does **-ical** survive?
 - ▶ *Potentialization*: one affix is appended to another specific affix (Williams 1981)
 - ▶ e.g. **-ity** attaches to **-able**, while **-ness** does not fare well in that domain
 - ▶ Perhaps **-ical** is potentializing in some domain
- ▶ *Let's explore possible domains by looking at neighborhoods these suffixes encounter...*

Neighborhoods

- ▶ Sort all stems into right-to-left alphabetical neighborhoods of 1 to 5 letters, e.g.:

t set

a	u	t	o	m	a	t	-	i	c	(a	l)		
c	a	t	h	a	r	t	-	i	c	(a	l)		
f	u	t	u	r	i	s	t	-	i	c	(a	l)	
s	o	l	i	p	s	i	s	t	-	i	c	(a	l)
d	i	a	g	n	o	s	t	-	i	c	(a	l)	

neighborhood length 1

Neighborhoods

- ▶ Sort all stems into right-to-left alphabetical neighborhoods of 1 to 5 letters, e.g.:

st set

automat	-ic(al)
cathart	-ic(al)
futurist	-ic(al)
solipsist	-ic(al)
diagnost	-ic(al)

neighborhood length 2

Neighborhoods

- ▶ Sort all stems into right-to-left alphabetical neighborhoods of 1 to 5 letters, e.g.:

ist set

~~automat-ic(al)~~
~~cathart-ic(al)~~
futurist-ic(al)
solipsist-ic(al)
~~diagnost-ic(al)~~

neighborhood length 3

Neighborhoods

- ▶ Sort all stems into right-to-left alphabetical neighborhoods of 1 to 5 letters, e.g.:

rist set

~~automat-ic(al)~~
~~cathart-ic(al)~~
futurist-ic(al)
~~solipsist-ic(al)~~
~~diagnost-ic(al)~~

neighborhood length 4

Neighborhoods

- ▶ Sort all stems into right-to-left alphabetical neighborhoods of 1 to 5 letters, e.g.:

urist set

~~automat-ic(al)~~
~~cathart-ic(al)~~
futurist-ic(al)
~~solipsist-ic(al)~~
~~diagnost-ic(al)~~

neighborhood length 5

Neighborhoods

- ▶ Examples:
 - ▶ **-t(ic/ical)** — length 1 (4166 members)
 - ▶ **-mat(ic/ical)** — length 3 (399 members)
 - ▶ **-graph(ic/ical)** — length 5 (294 members)
- ▶ *Only* neighborhood favoring **-ical** with more than a handful of members:
 - olog(ical)**
(475 members: largest set of length 4)

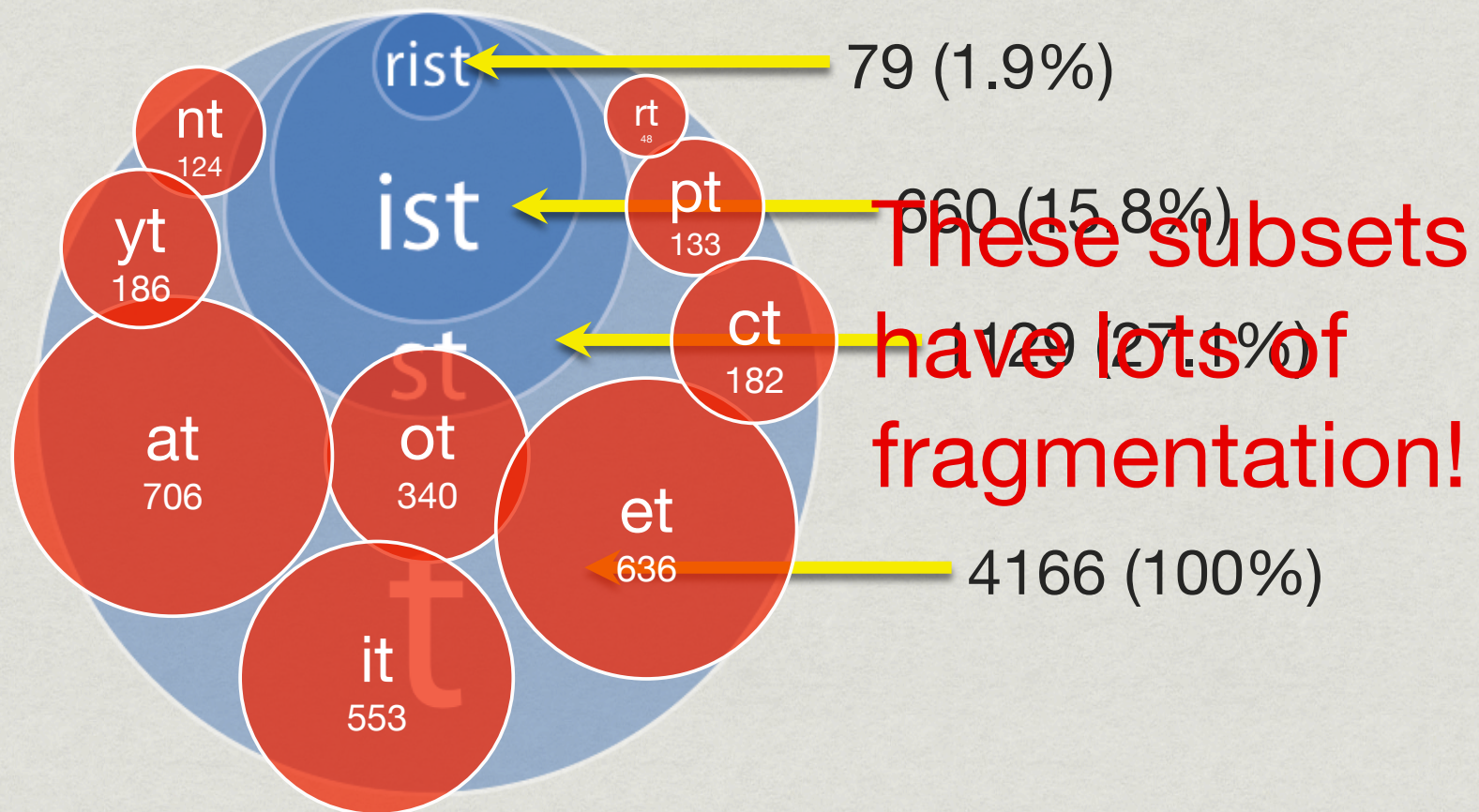
-olog(ical) subset

Ratio of -ic to -ical

suffix	-ic	-ical
-ological	1	8.30
average	7.84	1

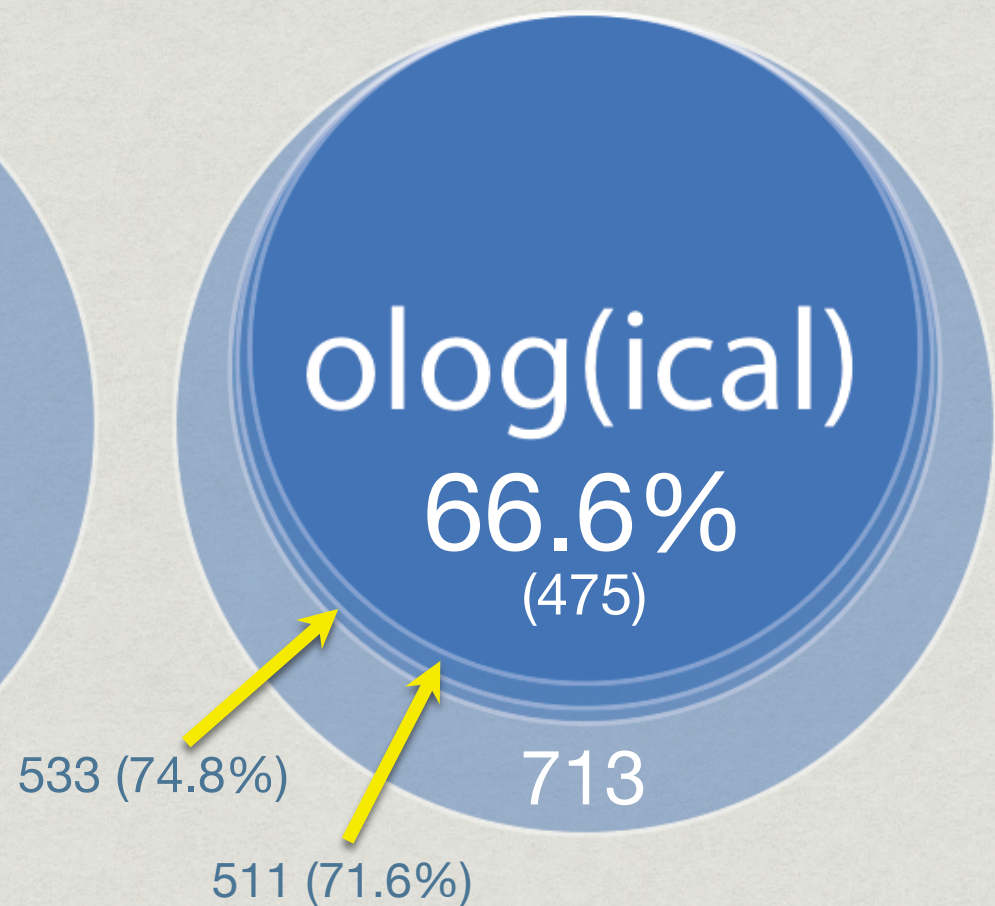
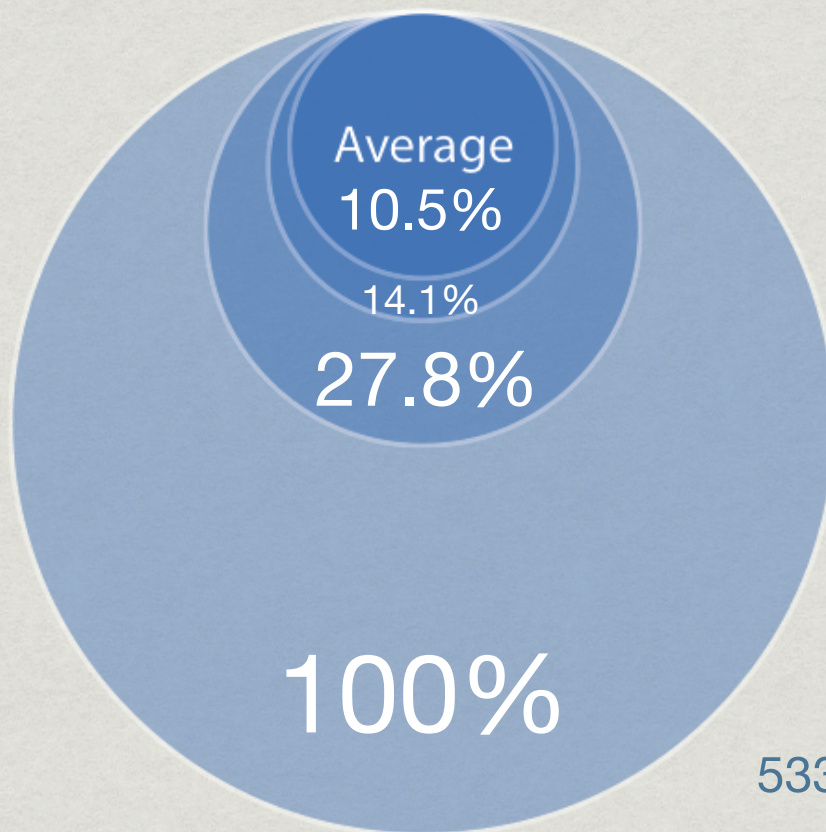
The ratio is reversed in the **-ological** set!

Sizing up the competition

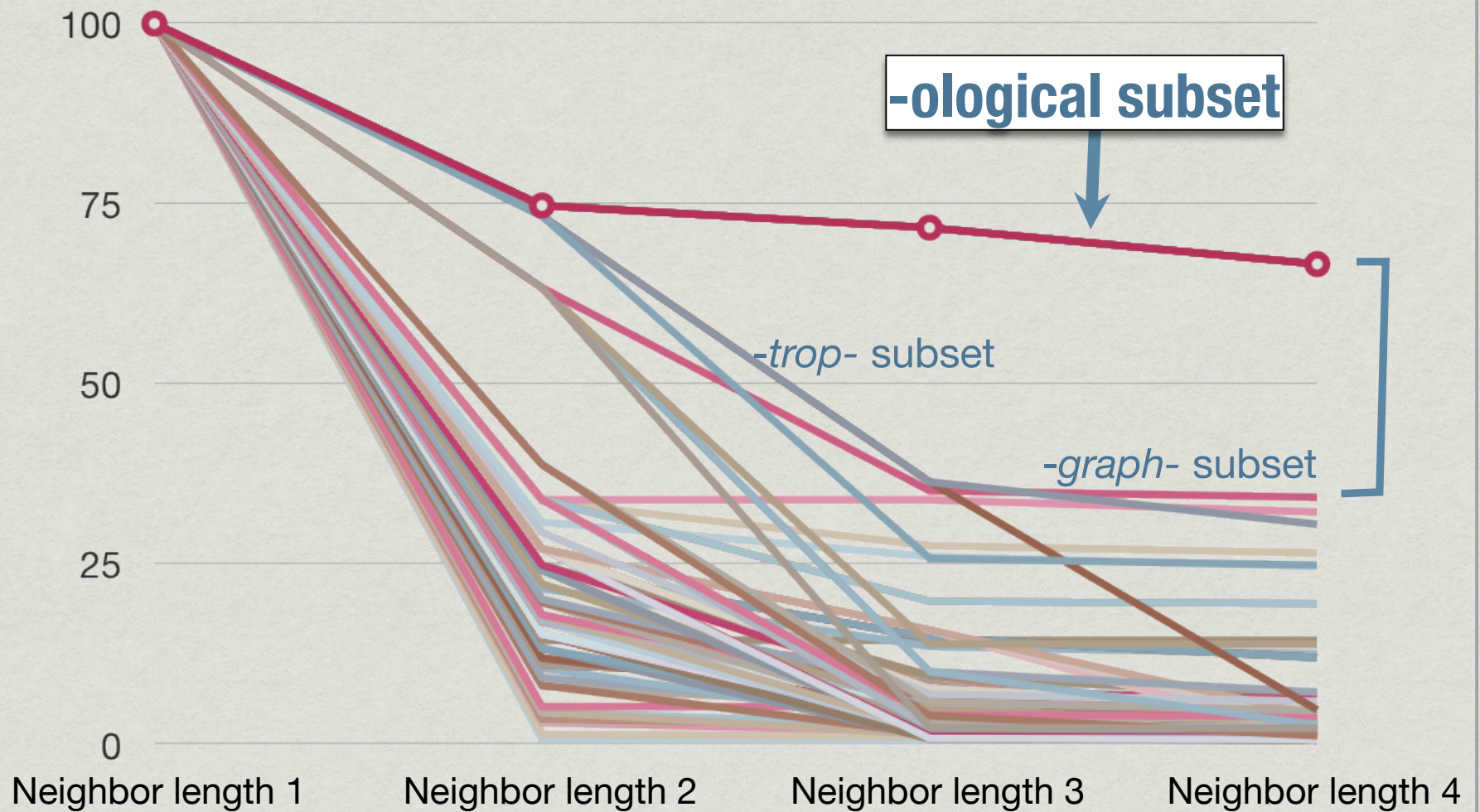


Average + olog(ical)

length 4:
66% of all -g stems =
uniformity & much less
competition!



-log(ical) subset is distinct



-ological niche

- ▶ **-ic** and **-ical** are synonymous competitors
- ▶ **-ic** has outcompeted overall
- ▶ **-ical** has survived in the **-ological** subsystem, because it found a niche:
 1. Sufficiently large set of words
 2. Far more uniformity than any other neighborhood (less fragmented)
- ▶ (Nothing special about **-olog-** itself - arbitrary)

-ize and -ify

- ▶ e.g. **theorize, stigmatize, agonize**
purify, fortify, stupefy
- ▶ Both from French; create verbs and add the meaning 'render, make, convert into', etc.
- ▶ Do these competing suffixes co-exist?
- ▶ Automated Google queries:
 - ▶ **stem + -ize/-izes/-izing/-ized** (also -ise)
 - ▶ **stem + -ify/-ifies/-ifying/-ified**

Google results

suffix	-ize	-(i)fy
winners	2217	419
ratio	5.29	1

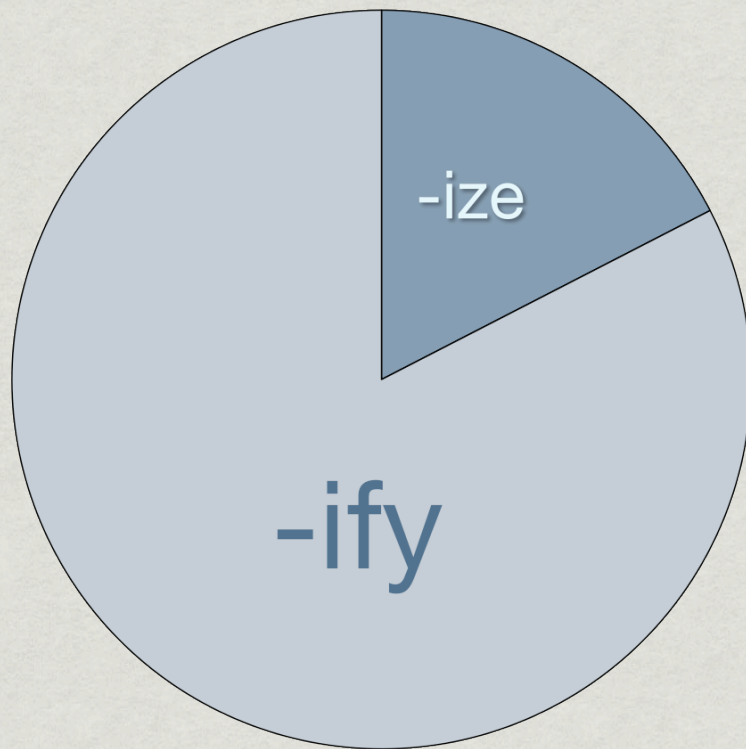
-ize is more productive than -ify (~5 to 1)

-ify: phonological niche

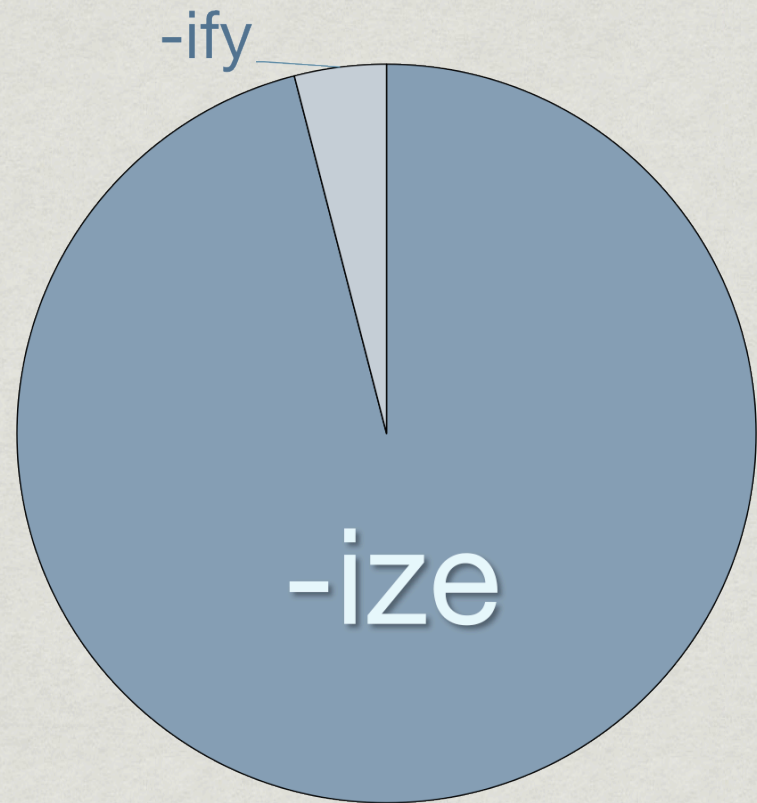
suffix	-ize	-(i)fy
monosyllabic stem	68	322
polysyllabic stem	2127	89

-ify is more productive in words with monosyllabic stems

-ize vs. -ify by stem syllables

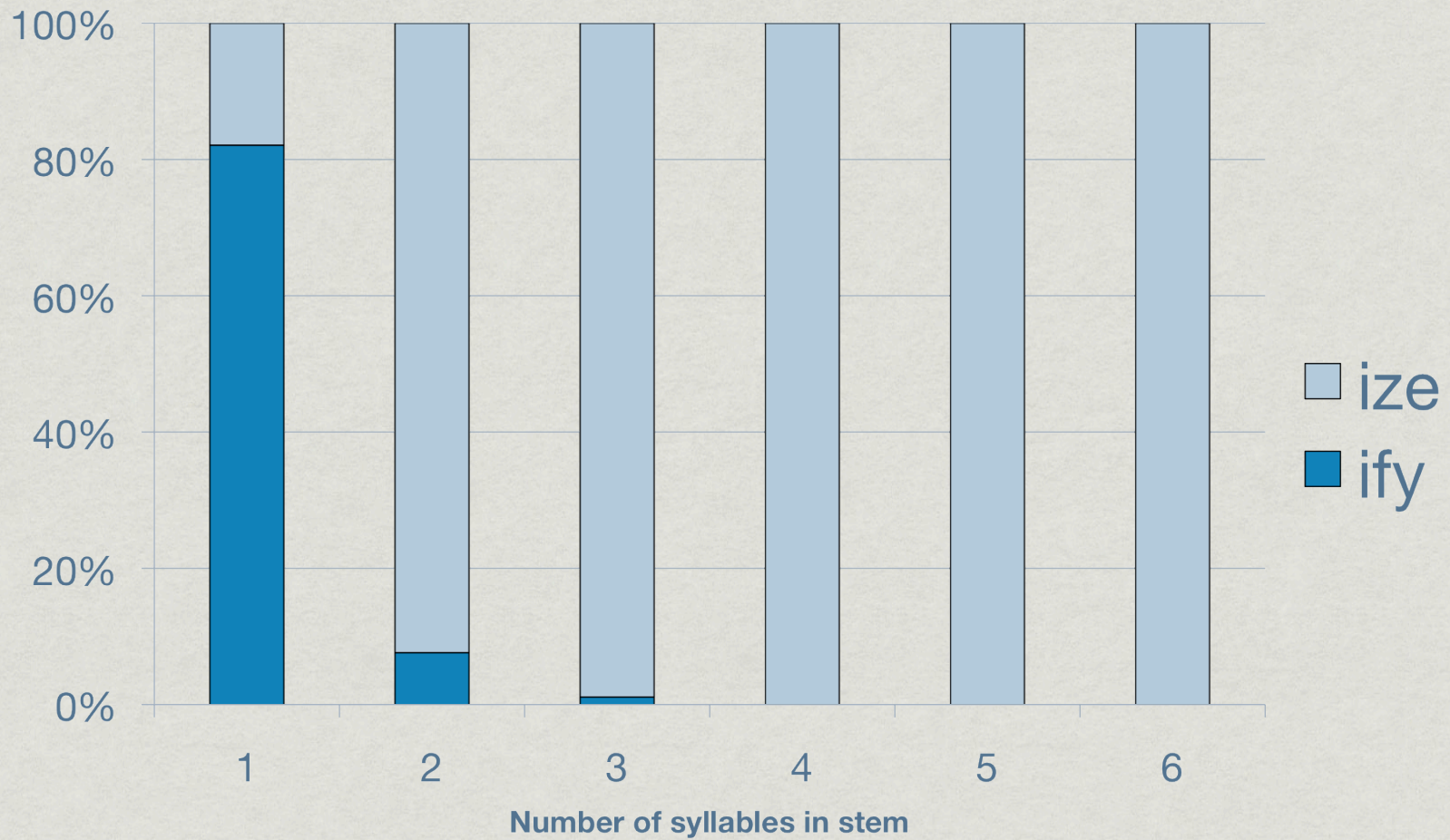


Monosyllabic Stems



Polysyllabic Stems

-ize vs. -ify by stem syllables



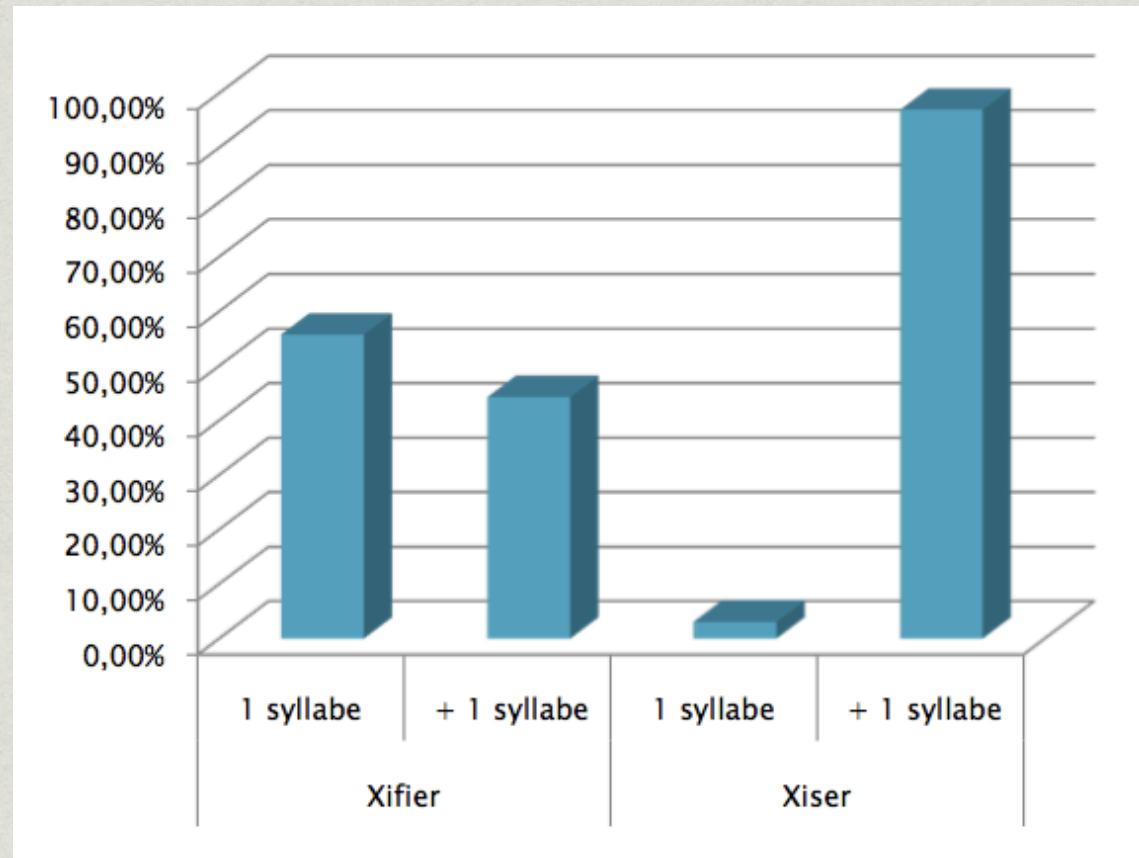
-ize and -ify coexist

- ▶ Suffix **-ize** is dominant overall
- ▶ **-ify** is dominant in words with monosyllabic stems
 - ▶ A large subset of words
 - ▶ Well-defined phonological niche

French -iser and -ifier

Lignon (2010)

- ▶ Monosyll.:
~55% -ifier
- ▶ Polysyll.:
~90% -iser
- ▶ Same
tendency



Summary

- Suffixes compete for productivity
- A suffix only remains productive with a source of productivity
- A suffix can prosper within a niche
 - Morphological (**-ical**)
 - Phonological (**-ify**)
- For certain problems, Google queries can provide insight into productivity

Further exploration

- **-ity vs. -ness**
- **-dom, -hood, and -ship**
 - **-tum, -heit, -schaft** (German)
 - **-dom, -heid, -schap** (Dutch)
- Suffix ordering
- Diachronic simulations

University Research Program for Google Search

- ▶ This research draws on data provided by the University Research Program for Google Search, a service provided by Google to promote a greater common understanding of the web.



Thank you

Selected References

Anshen, Frank & Mark Aronoff. 1999. Using dictionaries to study the mental lexicon. *Brain and Language* 68.10.

Aronoff, Mark. 1976. *Word formation in generative grammar* Cambridge, Mass.: MIT Press.
Aronoff, Mark. 1994. *Morphology by itself : stems and inflectional classes* Cambridge, Mass.: MIT Press.

Baayen, R. Harald. 1993. On frequency, transparency, and productivity. In Booij, G. and J. van Marle (eds.) *Yearbook of morphology 1992*. Dordrecht: Kluwer. 181-208.

Marchand, Hans. 1969. *The categories and types of present-day English word-formation. A synchronic-diachronic approach*. München: Beck.

Plag, Ingo. 1999. *Morphological productivity : structural constraints in English derivation* Berlin ; New York: Mouton de Gruyter.

Williams, Edwin. 1981. Argument structure and Morphology. *Linguistic Review*, 1:81-114.