

# What makes a writing system optimal?

Kathy Rastle

Royal Holloway, University of London



ROYAL  
HOLLOWAY  
UNIVERSITY  
OF LONDON

@kathy\_rastle  
www.rastlelab.com

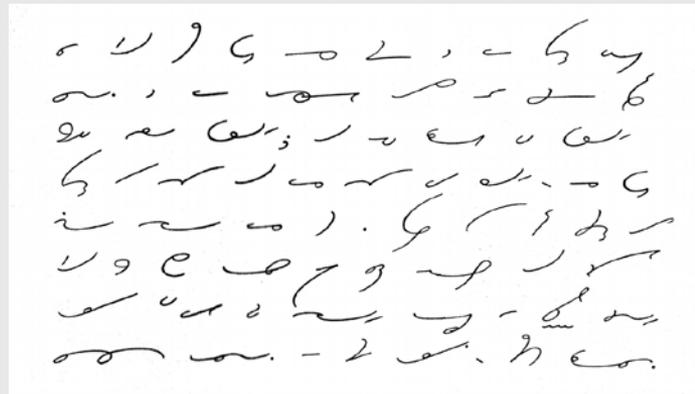
LEVERHULME  
TRUST \_\_\_\_\_

# Writing systems



ROYAL  
HOLLOWAY  
UNIVERSITY  
OF LONDON

가 ga	나 na	다 da	라 ra	마 ma	바 ba	사 sa	아 a
가 gya	냐 nya	다 dya	랴 rya	마 mya	바 bya	샤 sya	야 ya
거 geo	너 neo	더 deo	러 reo	머 meo	버 beo	서 seo	어 eo
겨 gyeo	녀 nyeo	더 dyeo	려 ryeo	머 myeo	버 byeo	셔 syeo	여 yeo
고 go	노 no	도 do	로 ro	모 mo	보 bo	소 so	오 o
교 gyo	뇨 nyo	도 dyo	료 ryo	묘 myo	뇨 byo	쇼 sya	요 yo



zhège chéngshì wánquán biàn le nǐ bùzhīdào  
这个城市完全变了。你不知道  
wèntí yě kěnéng shì yīnwèi rénmen zhījiān de  
问题，也可能是因为人们之间的  
chéngshì hé yǐqián wánquán bù yíyàng le dào  
城市和以前完全不一样了。到  
shù yě dōu sǐ le yǐqián zài zhèlǐ shēnghuó  
，树也都死了，以前在这里生活  
shēnghuó zàidìxià qí tā rén xiàng nǐ yíyàng zhī  
生活在地下。其他人像你一样住  
shǎo yìxiē hái kěyǐ hūxī dào gānjìng de kōng  
少一些，还可以呼吸到干净的空

Upper case

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Lower case

a b c d e f g h i j k l m n o p q r s t u v w x y z



# What is writing?



Writing is a representation of spoken language; “rendering spoken language into a form that can be reconstructed by other humans separated by space or time.”

But writing isn't the same as spoken language.

# What is writing?



## CHAPTER I

### *I am born*

Whether I shall turn out to be the hero<sup>1</sup> of my own life, or whether that station will be held by anybody else, these pages must show. To begin my life with the beginning of my life, I record that I was born (as I have been informed and believe) on a Friday, at twelve o'clock at night. It was remarked that the clock began to strike, and I began to cry, simultaneously.

In consideration of the day and hour of my birth, it was declared by the nurse, and by some sage women in the neighbourhood who had taken a lively interest in me several months before there was any possibility of our becoming personally acquainted, first, that I was destined to be unlucky in life; and secondly, that I was privileged

*Charles Dickens*

- Writing uses richer vocabulary and more complex syntax than spoken language.
- Writing is highly impoverished relative to spoken language: no prosody, gesture, audio-visual information, background / environmental context.
- Is there pressure for writing to 'depart' from spoken language to offer additional information that supports rapid access to *language through vision*?

Seit August gilt Deutschland aus britischer Sicht als grünes, also sicheres Reiseland. Diese Meldung löste große Erleichterung bei meiner Familie aus. Meine Tochter, die an einer schottischen Universität studiert und uns zur Weihnachtszeit in Deutschland besucht hatte, konnte bisher wegen diverser Lockdowns und verschärfter Einreisebestimmungen nicht dorthin zurückkehren. Monatelang war sie in Deutschland gestrandet und musste ihr Studium aus der Ferne fortführen.

Die Ausbreitung der Delta-Variante und die Aussicht, nach der Einreise tagelang in einem teuren Quarantäne-Hotel verbringen zu müssen, hatten eine frühere Reise nach Großbritannien verhindert.

**I'm Nobody! Who are you?**

**Are you - Nobody - too?**

**Then there's a pair of us!**

**Don't tell! they'd advertise - you know!**

**How dreary - to be - Somebody!**

**How public - like a Frog -**

**To tell one's name - the livelong June -**

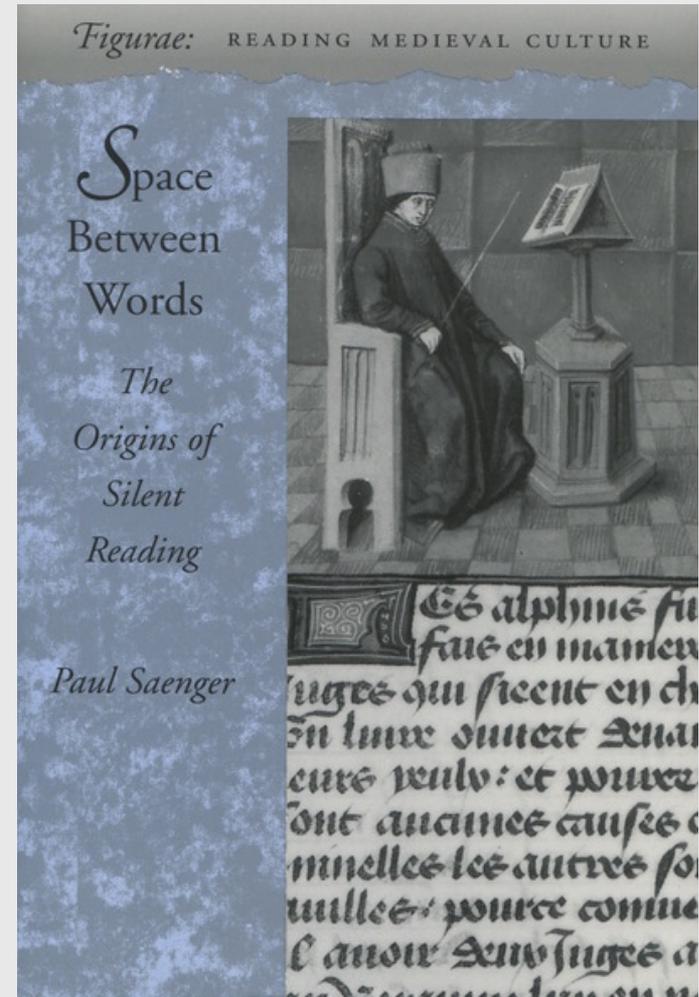
**To an admiring Bog!**

*Emily Dickinson*

# Word spacing



- Word spacing breaks the link between spoken and written language; yet it supports skilled, silent reading .
- The emergence of spacing between words co-occurred with the development of silent reading.
- Was this a causal relationship? If so, what is the mechanism of change?



# Is writing “optimal”?



“Every language gets the writing system it deserves.”

(Halliday; Mattingly; Frost, Seidenberg)

“Every writing system ... is a living, breathing organism that must adapt to the ever-changing needs of its users, their culture, and the technology of communication.” (Share, 2012)

# Is writing “optimal”?



“Every language gets the writing system it deserves.”

(Halliday; Mattingly; Frost, Seidenberg)

“Every writing system ... is a living, breathing organism that must adapt to the ever-changing needs of its users, their culture, and the technology of communication.” (Share, 2012)

“Writing systems evolve to provide *optimal* information by weighting the need for maximal cues about the spoken words and their specific meanings while using minimal orthographic load.” (Frost, 2012)

# Is writing “optimal”?



“Every language gets the writing system it deserves.”

(Halliday; Mattingly; Frost, Seidenberg)

“Every writing system ... is a living, breathing organism that must adapt to the ever-changing needs of its users, their culture, and the technology of communication.” (Share, 2012)

“Writing systems evolve to provide *optimal* information by weighting the need for maximal cues about the spoken words and their specific meanings while using minimal orthographic load.” (Frost, 2012)

“The writing systems that have survived support comprehension about equally well. Reading comprehension is a constant that is maintained via trade-offs between orthographic complexity ... and spoken language complexity.” (Seidenberg, 2011)



# Orthographic depth in alphabetic systems



Shallow

Deep



**Serbian**

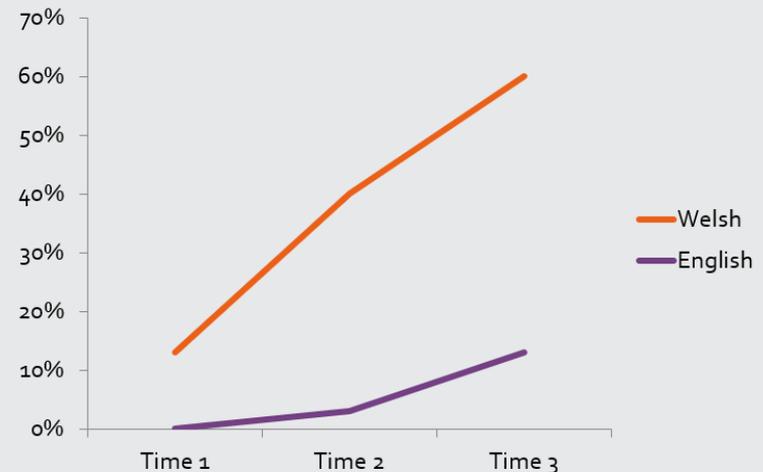
one symbol-one sound  
no exceptions

**French, Italian, Dutch...**

**English**

one symbol-to-many sounds  
many symbols-to-one sound  
many exceptions

Language	Reading Accuracy
Austrian	97%
German	98%
Norwegian	92%
Icelandic	94%
Swedish	95%
Dutch	95%
Danish	71%
English	34%



# English spelling not optimal for learning



09:08 Sun 18 Apr

85%

THE SUNDAY TIMES NEWS | NEWS

## New spelling rools to make English more predictable for pupils

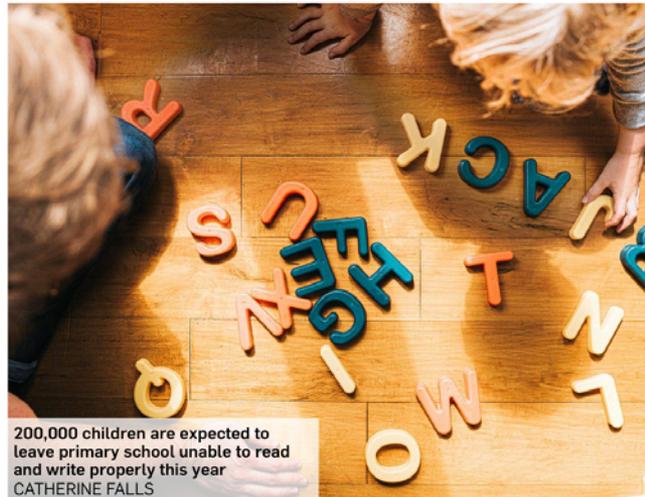
Sian Griffiths, Education Editor

“We shall fight on the beaches, we shall fight on the landing grounds, we shall fight in the feelds and in the streets, we shall fight in the hills; we shall never surrender...”

This is how Churchill’s famous speech should be spelt, according to reformers who have voted on a new system after deliberating for nearly three years.

The new system, Traditional Spelling Revised (TSR), will now be promoted in the hope that if it gains support “it will eventually become the new norm, thereby accelerating access to literacy,” the International English Spelling Congress said.

The congress was set up by The English Spelling Society, a UK-based group that has campaigned to “repair our broken spelling” for 113 years, and the American Literacy Council. The society argues that English



spelling has been so “chopped and changed” by invaders, scribes and printers over the centuries that the result is “crazy”.

Owing to its irregularity, English spelling takes up to three years longer to master at

primary school than other languages, according to the society. This year 200,000 children are expected to leave primary school unable to read and write properly, many because they have never mastered spelling.

Stephen Linstead, the former civil servant who designed the new system, says it is “a handicap that can stay with them for life”.

TSR changes up to 18 per cent of words in the language: wash becomes wosh, love is luv and educate becomes eddicate. Linstead said the result is as predictable as modern French spelling and will ease the burden of tens of thousands of children.

The system was selected from a shortlist of six put forward by an expert commission. Among other simplifications it eliminates silent letters such as the w in wrong and the g in gnash and cuts irregular spellings to just 60.

Jack Bovill, chairman of the English Spelling Society, said: “Parents on the front line teaching their children over the pandemic at home will have become familiar with how difficult it can be to teach English spelling.

“The highly irregular English spelling system has significant economic and social costs compared with those of other languages. The search has gone on for many years to find

spelling-sound predictability -> ease of learning

# English spelling not optimal for learning



## Initial Teaching Alphabet (ITA)

### Consonants

b c d f g h j k l m n

b k d f g h dz k l m n  
bib cake dad fife gag hat judge kick lull mime noon

ŋ p r s z t v w y z ʒ

ŋ p r s z t v w j z ʒ  
sing pipe roar sauce is tot valve will yes zoo vision

### Joined consonants

ch sh th th wh a e i o u ω

tʃ ʃ θ ð m æ ε ɪ ɒ ʌ ʊ  
church shush thin then whale at egg in odd up book

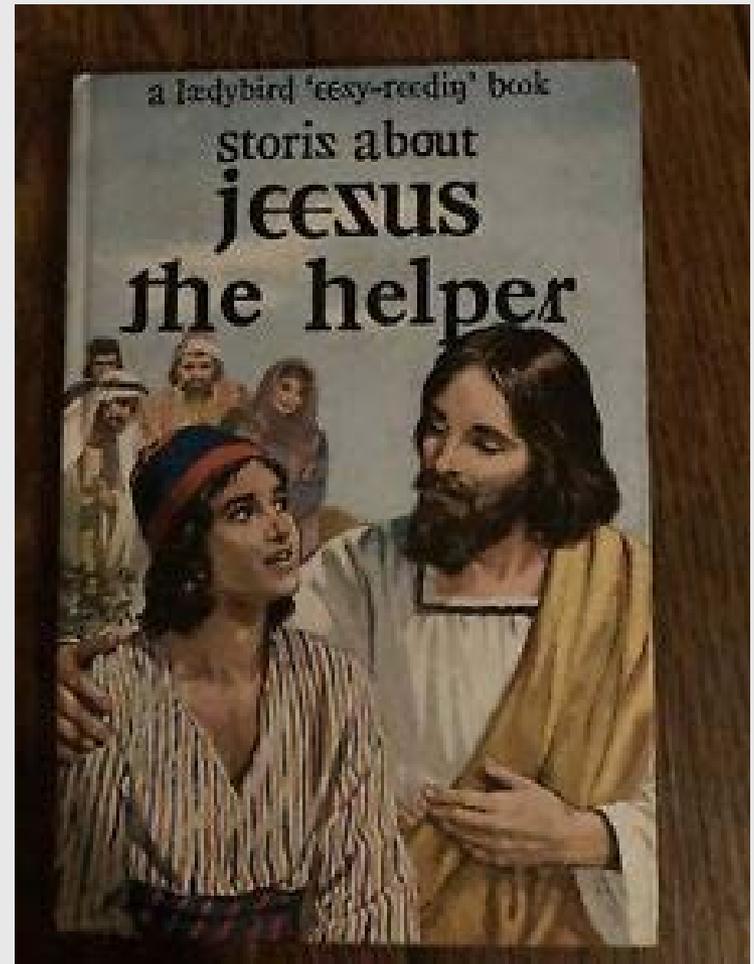
### Short vowels

### Long vowels / diphthongs

ɑ æ au ɛ ɛ ω ue ie oi ou

ɑː eɪ ɔː iː oʊ uː juː aɪ oɪ aʊ  
father ape all eat oak ooze use ice oil owl

Also, *ŋ* is used following a vowel letter to write the sound in "earn" etc



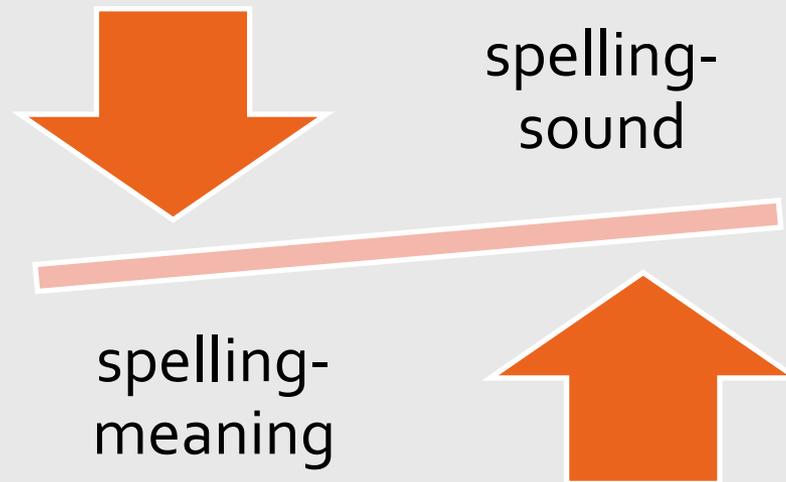
# Would reform optimize English spelling?



Ana Ulicheva

We don't know

Why does spelling-sound inconsistency exist?

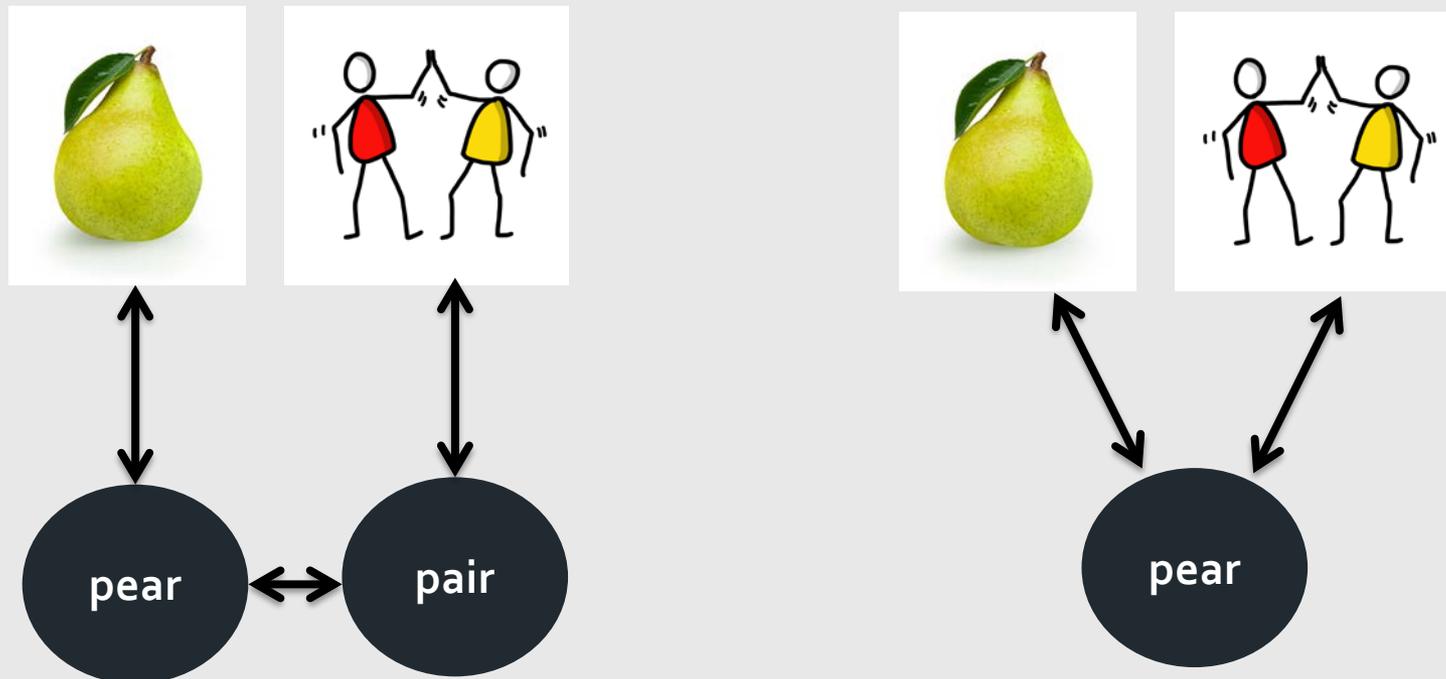


Would simplification lead to loss of meaningful information (thus hindering comprehension)?

# Would reform optimize English spelling?



Simplification might make it easier to learn but harder to understand



# Meaningful morphemic information



cooked, planned, needed

–ED means past tense

curious, atlas, citrus

–OUS means 'adjective'

- Letter clusters in English become “reserved” for communicating meaning; ubiquitous phenomenon (*Berg & Aronoff, 2017; Ulicheva et al., 2018*)
- This information is not present in spoken language; and communicating it requires the spelling system to sacrifice spelling-sound transparency.

# Meaningful morphemic information



cooked, planned, needed

cookt, pland, needed

–ED means past tense

curious, atlas, citrus

curius, atlus, citrus

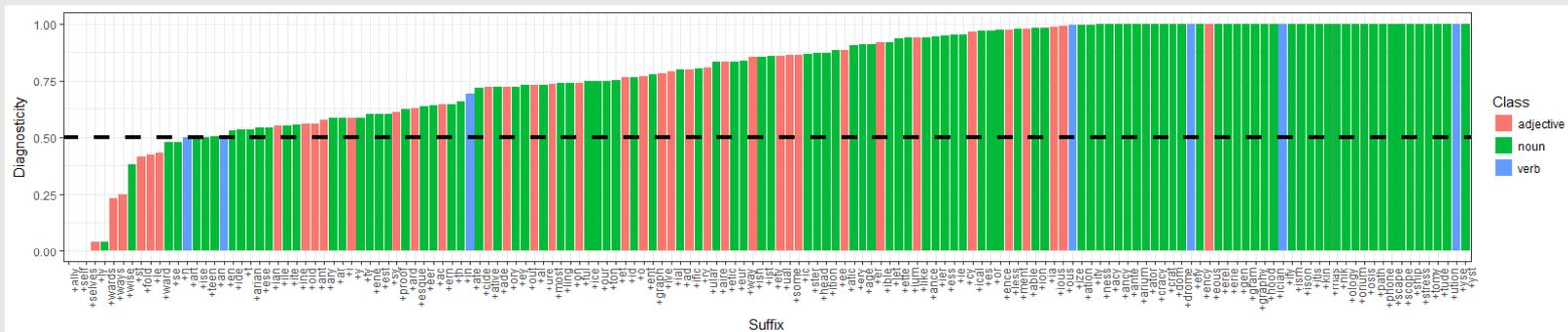
–OUS means 'adjective'

- Letter clusters in English become “reserved” for communicating meaning; ubiquitous phenomenon (*Berg & Aronoff, 2017; Ulicheva et al., 2018*)
- This information is not present in spoken language; and communicating it requires the spelling system to sacrifice spelling-sound transparency.
- This information would be lost with a more transparent spelling system.
- Is English spelling in a sweet spot between learnability and comprehension?

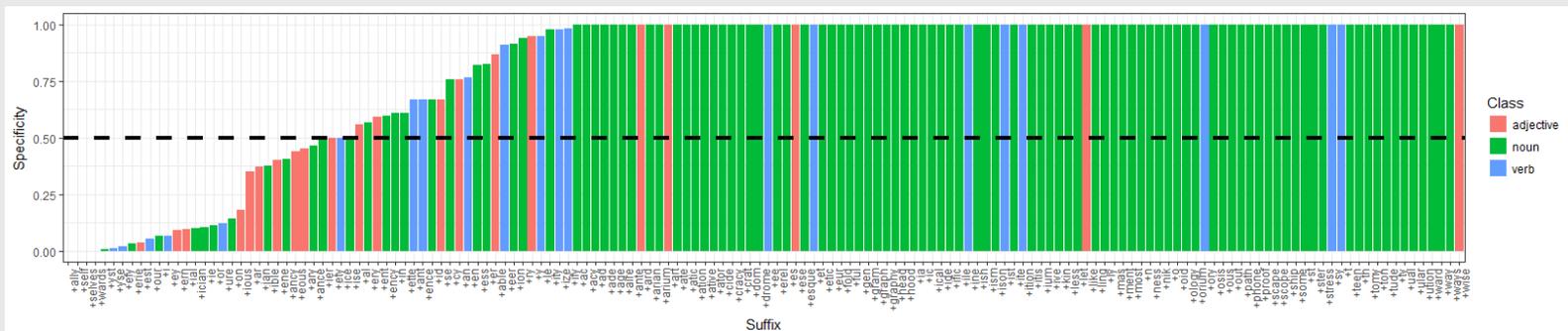


## Meaningful information is *highly visible* in English spelling

Most English suffix spellings are highly diagnostic of grammatical category



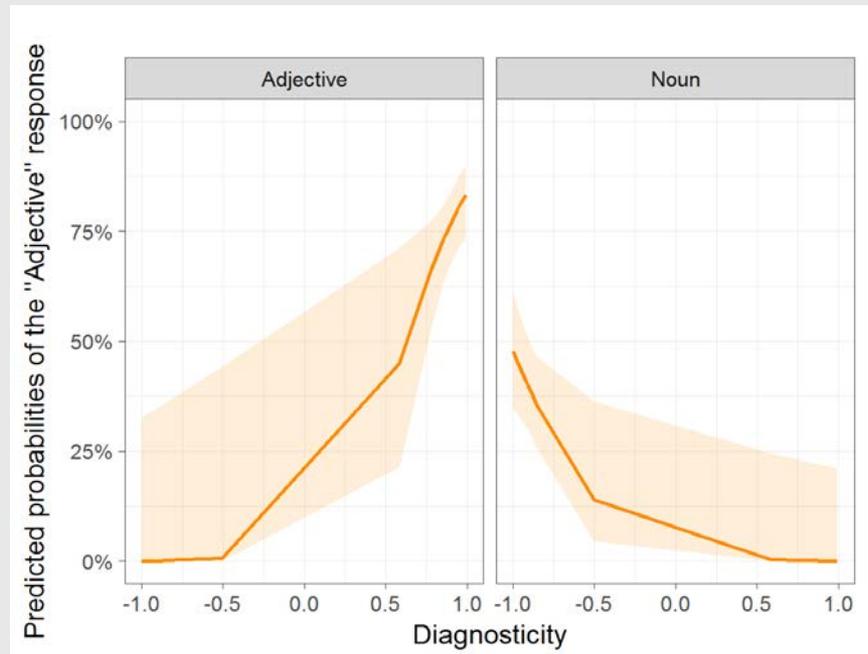
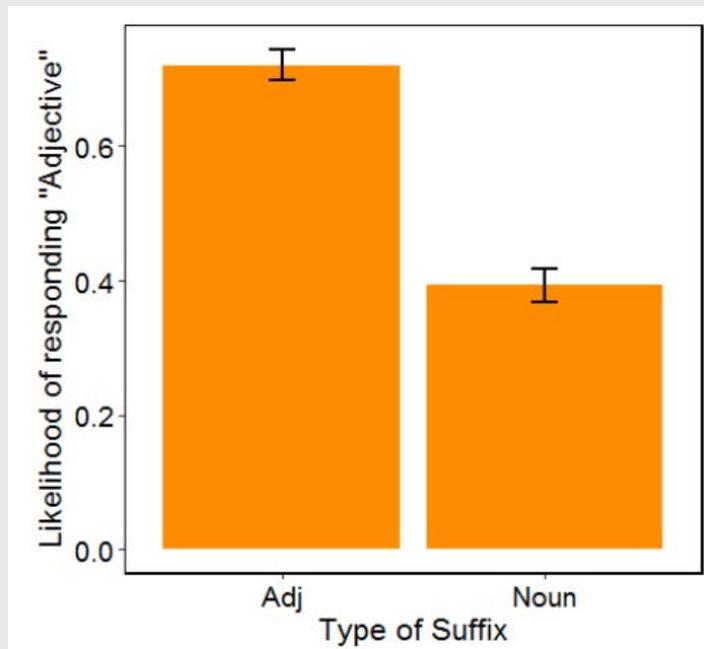
Most English suffix spellings provide the *only* means of expressing a sound sequence for a particular grammatical category



# Meaningful morphemic information



Is it an adjective or noun?  
DOMOUS, JIXLET, TERISH, RABNESS ...

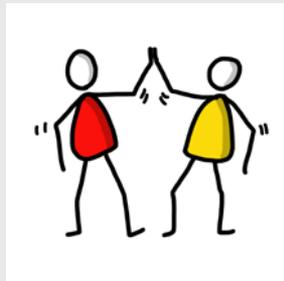


- Explicit knowledge of object / property status, linked to strength of cue
- Knowledge superior for adults with higher vocabulary & spelling
- Similar effects in spelling and in eye-tracking experiments



**Levenstein distance** between two words is the minimum number of single-character edits (insertions, deletions or substitutions) required to change one word into the other.

/pɛər/ – /pɛər/



pear – pair

**more dense: Levenstein distance 0**

**less dense: Levenstein distance 2**

# Estimating lexicon density



For spoken and written monosyllables separately

1. Compare each word  
to every other word

word<sub>1</sub>-word<sub>2</sub>  
word<sub>1</sub>-word<sub>3</sub>  
word<sub>1</sub>-word<sub>4</sub>  
...

6,448 word pairs per word

2. Calculate Levenstein  
distance for each pair

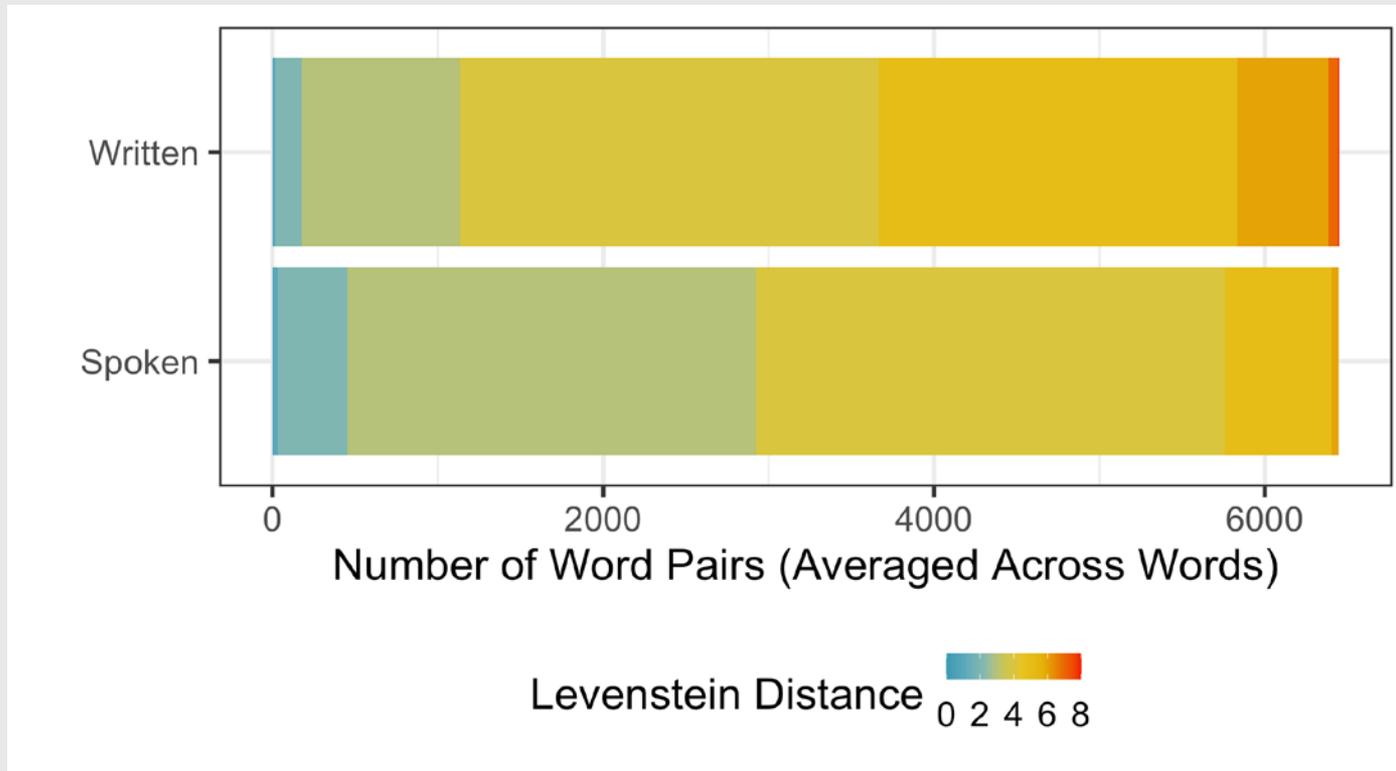
0  
1  
6  
...

6,448 distance values

Characterise  
frequency  
distribution  
(how many pairs  
at each distance)

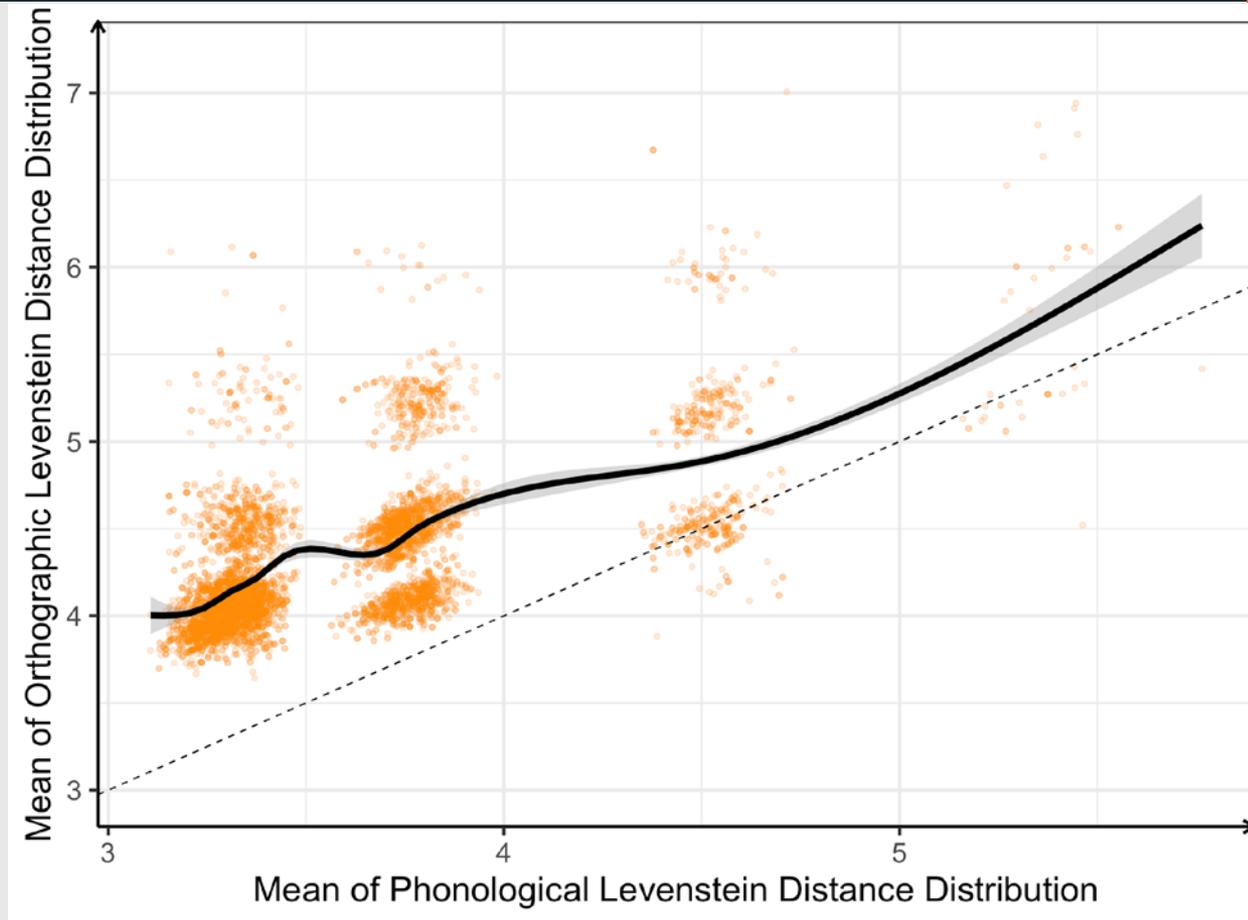
**More close neighbours = denser lexicon**

# English lexicon density



Written lexicon is SPARSER than spoken lexicon  
(note bias against this given fewer letters than phonemes)

# English lexicon density



Written lexicon is sparser

Writing pulls phonologically similar words further apart

# What happens when we reform the spelling?



## List of phoneme-grapheme correspondences (PGCs)

ɔʒ → J  
ɔʒ → DGE  
ɔʒ → GE  
...



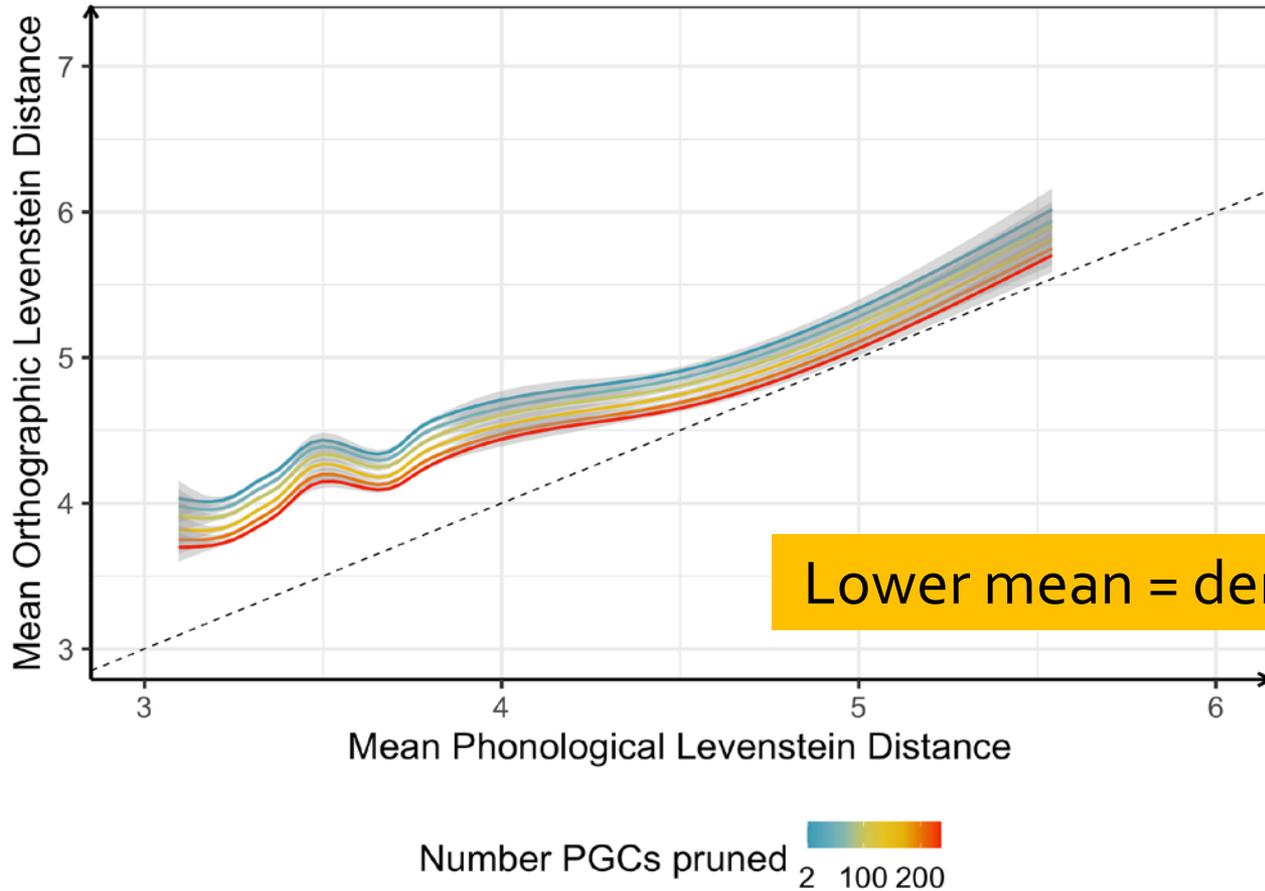
## Lexicon

JAM  
LOJ  
BINJ  
...

We progressively prune more & more inconsistent PGCs

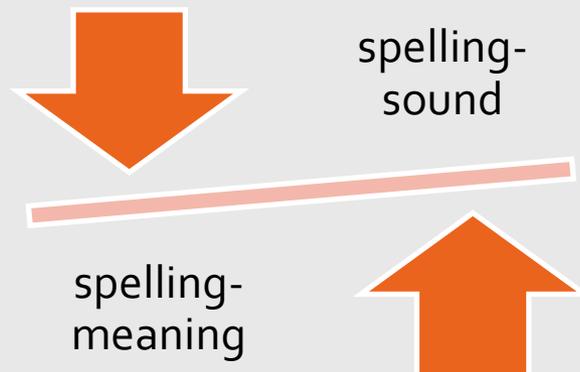
Measure impact on lexicon density

# What happens when we reform the spelling?

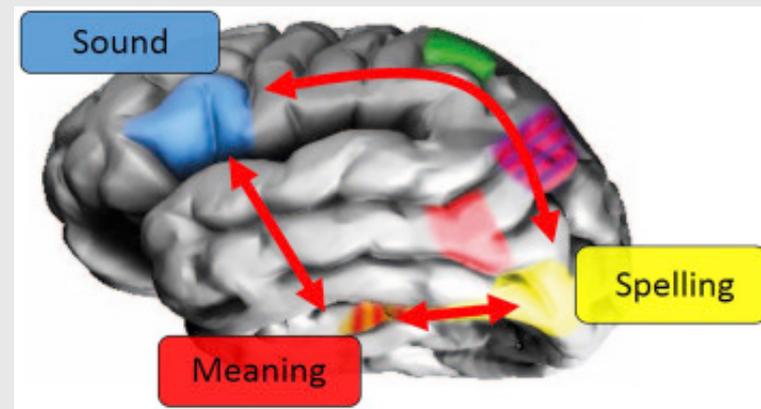
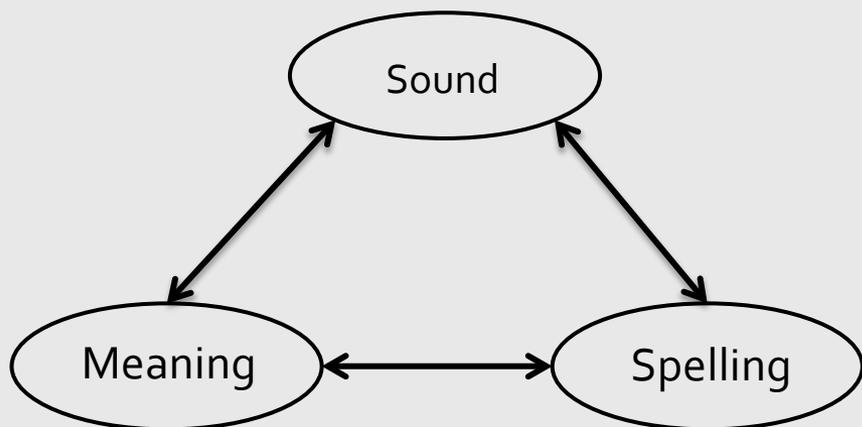


Simpler spelling → less disambiguation

# Optimal system occupies a sweet spot (?)

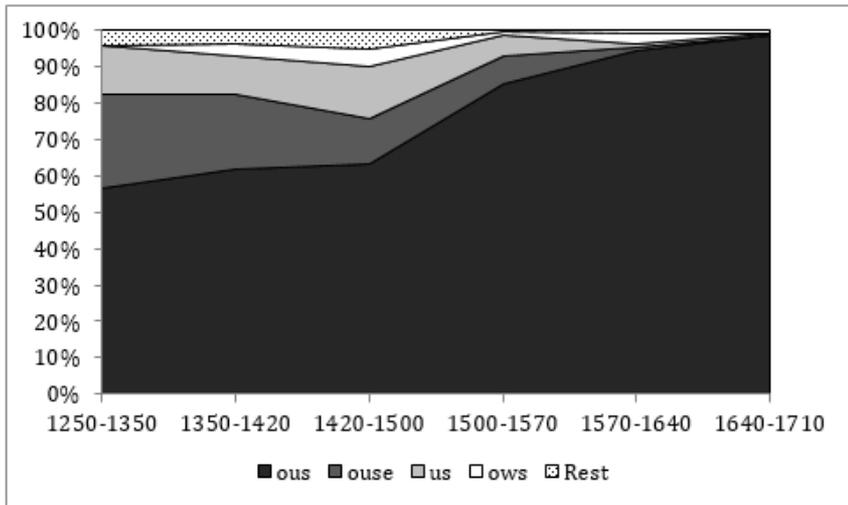


Learnability  
versus  
Comprehension



Intuitions about what makes an optimal writing system for a given language may be wrong.

# How do writing systems become optimal?



*Berg & Aronoff, 2017*

- Some (anecdotal) evidence of change in a direction that increases the specificity of English spelling.
- Much more systematic research needed to infer diachronic change that is relevant to a theory based on psychological pressures.
- Moreover, any theory of the “evolution” of writing requires a mechanism: variation, selection, propagation. (*Heyes, 2018*)
- Theoretical claims e.g. “self organizing”, “living, breathing” are hand-waving without a mechanism.

# Conclusions



- Writing is not the same thing as spoken language.
- Our intuitions about what makes a writing system optimal may be wrong.
- The field has focused on learnability; representing the sounds of spoken language faithfully.
- But comprehension also matters; reasonable evidence that English spelling trades learnability for disambiguation.
- How is this tradeoff expressed in other languages that vary in their inherent ambiguity?
- Substantial need in this field for more precise theorizing about diachronic change and its underpinning mechanisms.



Ana Ulicheva, Jo Taylor, David Crepaldi, Jakke Tamminen, Betty Mousikou, Clare Lally, Sam McCormick, Maria Ktori, Nicky Dawson, Adam Jowett, Benedetta Cevoli, Chloe Newbury, Becky Crowley, Oxana Grosseck



Mark Aronoff, Matt Davis, Aureliu Lavric, Michal Ben-Shachar, Maya Yablonski, Chang Lee



**The Leverhulme Trust**



Thank you!



ROYAL  
HOLLOWAY  
UNIVERSITY  
OF LONDON