Statistical information encoded in English spelling

Anastasia Ulicheva [1], Hannah Harvey [1], Mark Aronoff [2], and Kathleen Rastle [1]

1 Royal Holloway, University of London, UK
2 Stony Brook University, US

Grant Number: ES/N016440/1

Background

Sound-to-spelling correspondences in English are inconsistent: compare spellings of the final sound /-əs/ in FAMOUS, SOLACE, ATLAS, CYPRESS.

Is this irregularity functional?

Source for systematicity	Example
Some sound-spelling correspondences are predictable	/b/ spelled "b"
Meaning	HEAL is related to HEALTH
Etymology	'deca' in DECADE, DECIMAL
Morphology	-ED for past tense (cf. KICKED, ROAMED, BATTED)

Berg and Aronoff (2017) found that there are systematic relationships between spellings of four English suffixes –OUS, –IC, –AL, –Y and grammatical class.

For example, -/əs/ is spelled as –OUS in adjectives, and as something else in other classes.

		OUS spelling	Other spelling
Number of words (and example)	adjectives	346 *Marvellous*	6 *Citrus*
	NOT adjectives	0 --	314 *Cactus*

We asked: how common is this systematicity between spelling and grammatical class in English? If it is common, then do people extract this statistical information as they become literate and do they use it do support their reading and spelling?

Goals of this study

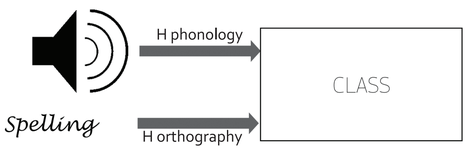
1. Is regularity between spelling and grammatical class true of English suffixes in general?
   ⇨ Study 1: Computational linguistic analysis
2. Are people sensitive to this regularity?
   ⇨ Study 2: Explicit judgment experiment
   ⇨ Study 3: Spelling experiment
3. What does the degree of sensitivity depend on?
   ⇨ Analysis of individual differences

Study 1: Large-scale linguistic analysis

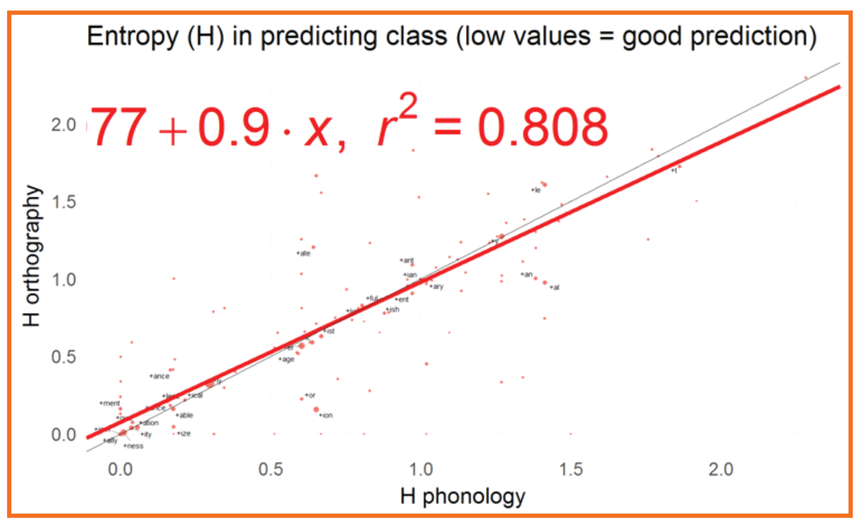
Question
Is systematicity between spelling and class true of English derivation in general?

Idea
Spelling disambiguates grammatical class
- Is there a dependency between spelling and class?
- Is this dependency stronger than that between phonology and class?

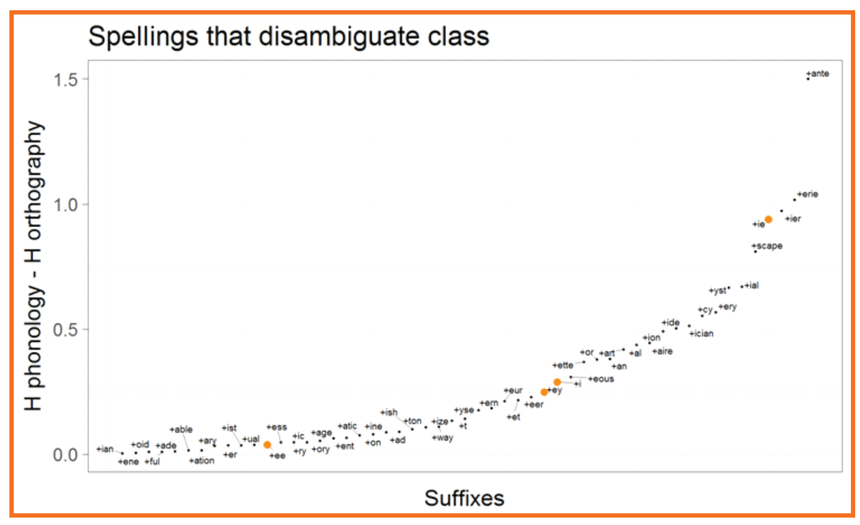
Method
- 159 suffixes extracted from CELEX
- for each suffix its entropy (H) for class was calculated
- Entropy is a measure of prediction precision



Results



Orthography predicts class better than phonology does



$$H = \Sigma[-p_i \times \log_2(p_i)]$$

where p_i is the proportion of words belonging to a given grammatical class.

Low H means that the prediction of class is good
High H means that the prediction of class is poor

Example
The sound /i/ is most frequently spelled as "Y" e.g. BUSY. But there are other spellings:
<ie>: calorie
<ee>: employee
<i>: Israeli
<ey>: alley
It turns out that each of these alternative spellings is used to denote nouns.

Conclusions
- Spelling provides additional information about grammatical class
- This is true of English derivation in general

Study 2: Explicit judgement experiment

Question Are people sensitive to regularities between spelling and class?

Idea
- We manipulate spellings of nonwords
- Does this manipulation influence people's decisions about which grammatical class these nonwords may belong to?

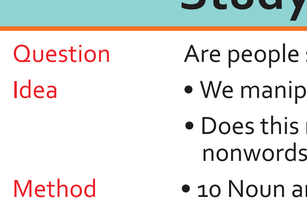
Method
- 10 Noun and 10 Adjective suffixes that strongly predict class
  - Nouns: NESS, ITY, MENT, AGE, LET, IST, AN, ER, EE, ENCE
  - Adjectives: ABLE, OUS, LESS, ICAL, LIKE, Y, IC, IVE, ISH, ATIC
- Joined them with CVC non-existing stems

JIXLET
Noun or Adjective?

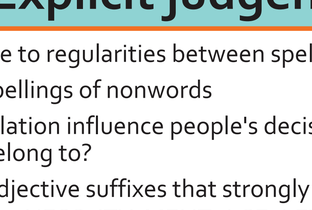
- We explained to people what nouns are and what adjectives are
  - NOUN is a person, animal, place, thing, or idea: For example, AUNT, CAT, FOREST, CUP, LOVE.
  - ADJECTIVE is an attribute of a noun: For example, SWEET, RED, SIMPLE
- 46 participants

Analysis and Results Mixed logistic regression with subjects and suffixes as random effects ($z = -4.18$, $p < 0.0001$)

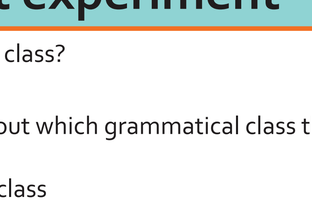
Main Result



"Good" suffixes



"Bad" suffixes



Conclusions People have explicit awareness of systematicities between spelling and class

Study 3: Spelling experiment

Question Are people sensitive to regularities between spelling and class?

Idea
- Nonwords are placed into different sentence frames
- Does context influence people's spellings?

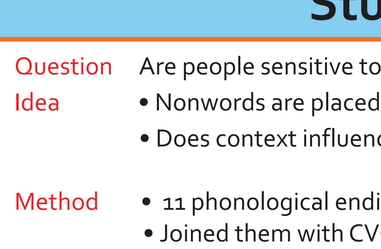
Method
- 11 phonological endings that can be spelled differently
- Joined them with CVC non-existing stems
- 66 nonword recordings
- Biasing sentence contexts
- 29 participants

[sedʒnɪs]

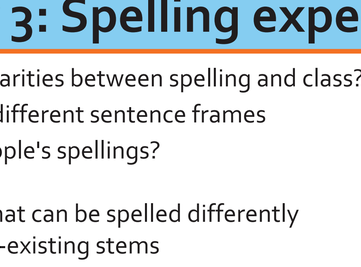
Context	Sentence (beg.)	Target	Sentence (end.)	Spelling
CON	The presentation recognised the impressive		of the protestors.	-NESS
INC	The mourners began to sadly		as the coffin disappeared.	???

Analysis and Results Mixed logistic regression with subjects and suffixes as random effects ($z = 4.84$, $p < 0.0001$).

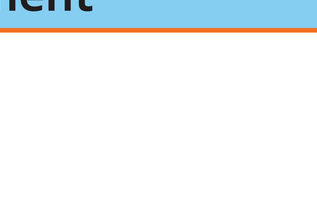
Each sound elicited a variety of spellings.
Our focus is on the target spelling.



Main Result

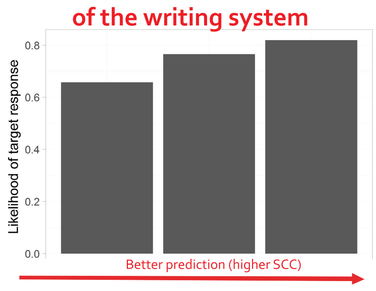


Variability across suffixes



Conclusions People exploit their knowledge of regularities between class and spelling to indicate class

Suffix behaviour mirrors statistics of the writing system



Why are there differences across suffixes?

Spelling to class consistency (SCC)

$$SCC = \frac{N_{+spelling+class}}{N_{+spelling}}$$

Where $N_{+spelling}$ is the number of words with a given spelling, $N_{+spelling+class}$ is the number of words with a given spelling that belong to a given class.

	SCC	CSC
OUS	0.99 OUS-words are adjectives	0.81 Alternative spelling for /əs/-adjectives: (e.g. emeritus)
Y	0.58 Y-words are adjectives, but also nouns/verbs	0.95 Alternative spelling for /i/-adjectives: rare

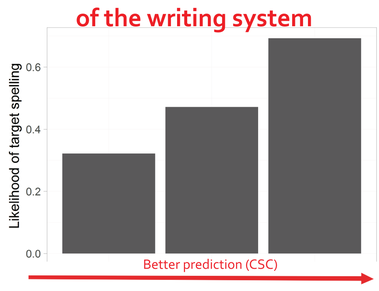
Why are there differences across suffixes?

Class to spelling consistency (CSC)

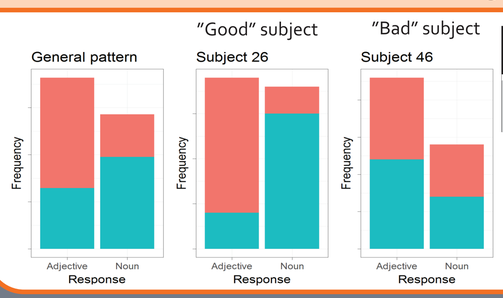
$$CSC = \frac{N_{+spelling+pronunciation+class}}{N_{+pronunciation+class}}$$

Where $N_{+spelling+pronunciation+class}$ is the number of words with a given spelling and pronunciation that belong to a given class, $N_{+pronunciation+class}$ is the number of words with a given pronunciation that belong to a given class.

Suffix behaviour mirrors statistics of the writing system



Differences across subjects: Explicit judgement



	General spelling ability
Explicit awareness of spelling-class regularities	0.52, $p < 0.001$

People with explicit awareness of spelling-class regularities are better spellers.

General conclusions

- Regularities between spelling and grammatical class are ubiquitous.
- People are sensitive to these regularities. They extract statistical information about grammatical class from print without any formal instruction and exploit it when dealing with written language.
- People`s behaviour mirrors the statistics of the writing system: We are better at picking up and using the information about spellings that disambiguate grammatical class.

Thanks to:

Rebecca Crowley
Nardeen Massoud

Ana.Ulicheva@rhul.ac.uk