

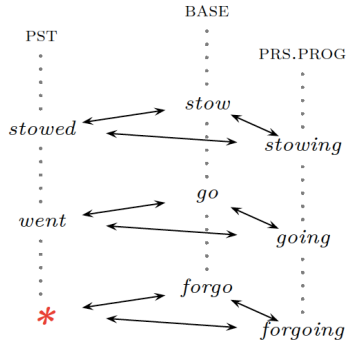
Reconciling the contradictory phenomenology of defectiveness

Maria Copot

Université Paris Cité, LLF, CNRS

Introduction

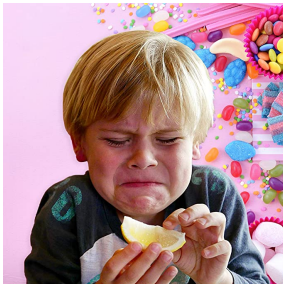
Defectiveness



- A paradigm cell remaining **unfilled despite expectations** (Sims, 2015)

A psychoanalytic parenthesis - a felt sense of defectiveness

A psychoanalytic parenthesis - a felt sense of defectiveness



Disgust Icky, ugly, unfortunate, “eugh”, “why are you doing this to me”

Defectiveness has certain **negative qualia** for speakers.



One expects these qualia to have an **impact on the use** of defective forms.



Defective forms should **not be used** in language.

The situation with usage

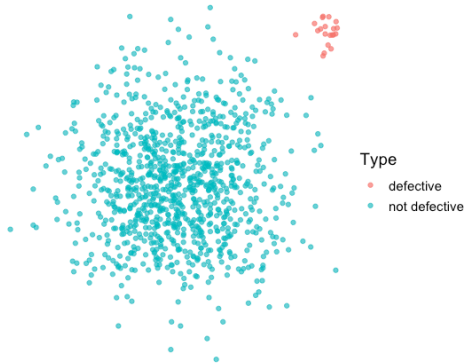
- PhD thesis originally on defectiveness.

The situation with usage

- PhD thesis originally on defectiveness.
- Understand the **processing of defective words**.

The situation with usage

- PhD thesis originally on defectiveness.
- Understand the **processing of defective words**.
 - Step zero: create a method to **find defective words** in corpora, as basis for experimental items.



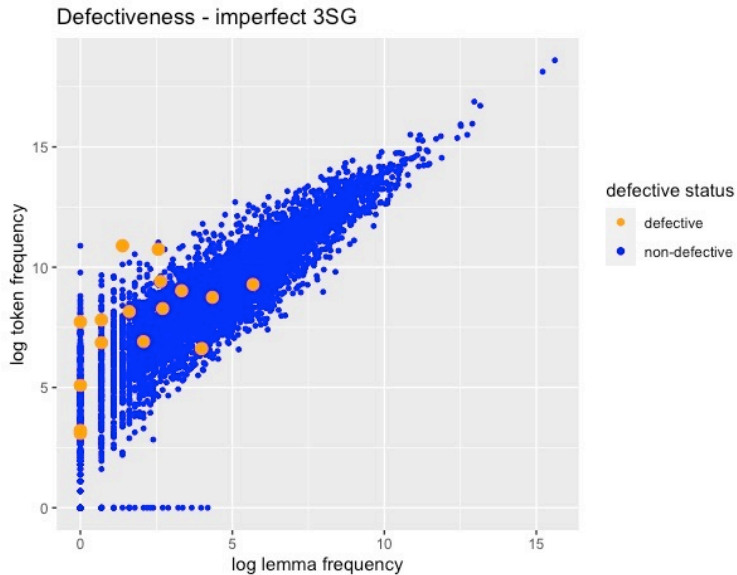
The situation with usage - taxonicity

- Every quantitative study on defectiveness finds that defective words **as a group** are different **on average** along relevant dimensions from non-defective words.
- Finding minimal overlap between defective and non-defective words hasn't been explored to my knowledge.

The situation with usage

- Step zero: create a method to **find defective words** in French and Russian corpora, as basis for experimental items.
- If defectiveness is the **unexpected absence** of a word form to fill a cell, then it should manifest as a lexeme being **much less frequent than expected** in a particular cell.

Must meet the data where it is



Finding defectiveness in corpora - a cursed endeavour

- Repeatedly **unsuccessful**, took $\frac{1}{3}$ of my PhD, using ~ 5 methods on multiple parts of speech in both French and Russian.
 - **Frequency analysis**: defective words as less frequent than expected
 - ...compared to various reference classes
 - **Statistical modeling**: defective words as large negative residuals in models attempting to predict token frequency from
 - frequency of cells belonging to the same lexeme
 - frequency of cells of related lexeme
 - distributional semantics
 - information theory

- Speakers are **using defective words** in corpora and on the internet
 - Not just metalinguistic attestations
 - Not just irony
 - Not just L2 speakers

For this to happen, at least one of these must be true

- **Dictionaries** are wrong about what words are defective
- **Defectiveness** felt sense does not cash out in diminished usage

- **Social factors** have a crucial role in explaining the origins and manifestation of **defective felt sense**.

- **Social factors** have a crucial role in explaining the origins and manifestation of **defective felt sense**.
 - Question the assumption that defectiveness “is in the grammar”.

- **Social factors** have a crucial role in explaining the origins and manifestation of **defective felt sense**.
 - Question the assumption that defectiveness “is in the grammar”.
 - Explain **contradictory results** in the behavioural literature on defectiveness.

- **Social factors** have a crucial role in explaining the origins and manifestation of **defective felt sense**.
 - Question the assumption that defectiveness “is in the grammar”.
 - Explain **contradictory results** in the behavioural literature on defectiveness.
 - Help shed light on the **ontology** of defectiveness and how it relates to other phenomena.

What follows was done in
collaboration with Andrea Sims

Defectiveness in French verbs: a sociomorphological phenomenon?

A typology of French verbal defectiveness

Based on what dictionaries deem defective, Boyé and Cabredo Hofherr (2010) propose a typology

A typology of French verbal defectiveness

Based on what dictionaries deem defective, Boyé and Cabredo Hofherr (2010) propose a typology

1. **Form indeterminacy**: speakers don't have a plausible form for the defective paradigmatic slot.
 - APPAROIR 'to appear' is only used in the IND.PRS.3SG. Hard to predict rest of paradigm.

A typology of French verbal defectiveness

Based on what dictionaries deem defective, Boyé and Cabredo Hofherr (2010) propose a typology

1. **Form indeterminacy**: speakers don't have a plausible form for the defective paradigmatic slot.
 - APPAROIR 'to appear' is only used in the IND.PRS.3SG. Hard to predict rest of paradigm.
2. **Form conflict**: plausible forms exist but they are rejected for independent reasons.
 - NEIGER 'to snow' - **je neige* 'I snow', defective for semantic reasons

A typology of French verbal defectiveness

Based on what dictionaries deem defective, Boyé and Cabredo Hofherr (2010) propose a typology

1. **Form indeterminacy**: speakers don't have a plausible form for the defective paradigmatic slot.
 - APPAROIR 'to appear' is only used in the IND.PRS.3SG. Hard to predict rest of paradigm.
2. **Form conflict**: plausible forms exist but they are rejected for independent reasons.
 - NEIGER 'to snow' - **je neige* 'I snow', defective for semantic reasons
3. **Form gaps**: plausible stems are not used without any synchronic motivation.
 - CLORE 'to close'. *Vous *closez* has no reason not to exist, and is the obviously correct form, yet it is deemed defective.

French speakers have defective felt sense

le verbe clore 17 views



Ampac

to

8 Feb 2001, 18:17:26

Présent

Je clos

Tu clos

Il clot

Nous clo ... ?

Vous clo ?

Ils closent

Impératif

Clos

Clo ???

Clo ???

Merci de me donner toutes les formes de "clore", on a cherche au boulot mais on n'a pas trouve

Ampac C.

Yet the forms of CLORE are **not used any less frequently** than one might expect in corpora. Why?

The status of defective words - highlighting important pieces of the puzzle

1. In corpora, **defective forms are used** as if they were non-defective.

The status of defective words - highlighting important pieces of the puzzle

1. In corpora, **defective forms are used** as if they were non-defective.
2. Speakers do not like **defective forms**, but do not mind other words that have the same structural properties
 - **overabundant** forms (Bermel, Knittle, and Russell, 2018)
 - (defective-resembling) **nonwords** (Albright, 2003)

The status of defective words - highlighting important pieces of the puzzle

1. In corpora, **defective forms are used** as if they were non-defective.
2. Speakers do not like **defective forms**, but do not mind other words that have the same structural properties
 - **overabundant** forms (Bermel, Knittle, and Russell, 2018)
 - (defective-resembling) **nonwords** (Albright, 2003)
3. Although we are making progress in individuating **structural factors** correlated with defectiveness (Baerman and Corbett, 2010; Fábregas, 2018; Sims, 2023), these are **not sufficient nor necessary**

Explanatory accounts of defectiveness - the role of social variables

- Theories of the **synchronic** nature of defectiveness focus on providing mechanisms for **defectiveness to fall out of the grammar**. (Albright, 2003; Sims, 2006, 2015, 2023)

Explanatory accounts of defectiveness - the role of social variables

- Theories of the **synchronic** nature of defectiveness focus on providing mechanisms for **defectiveness to fall out of the grammar**. (Albright, 2003; Sims, 2006, 2015, 2023)
- **Diachronic** theories of defectiveness do acknowledge the role of **social factors** (Baerman, 2008, 2011; Broadbent, 2009; Gilliéron, 1919).
 - e.g. Baerman (2011) notes that CLORE owes its defectiveness to having fallen out of use and then having been resurrected by the Académie française in certain cells.

Explanatory accounts of defectiveness - the role of social variables

- Theories of the **synchronic** nature of defectiveness focus on providing mechanisms for **defectiveness to fall out of the grammar**. (Albright, 2003; Sims, 2006, 2015, 2023)
- **Diachronic** theories of defectiveness do acknowledge the role of **social factors** (Baerman, 2008, 2011; Broadbent, 2009; Gilliéron, 1919).
 - e.g. Baerman (2011) notes that CLORE owes its defectiveness to having fallen out of use and then having been resurrected by the Académie française in certain cells.
- The **synchronic** role of social factors remains underexplored.

The research question

- We propose that attending to **system-external factors** resolves the **paradox** of how defectiveness manifests.
- The research question: what is the role of **linguistic prescriptiveness** in accounting for speakers' **felt sense of defectiveness**?

Why should linguists care about prescriptivism?

- To many, speakers' and societies' beliefs about how language **should be** are irrelevant to studying the way that language **is**.

Why should linguists care about prescriptivism?

- To many, speakers' and societies' beliefs about how language **should be** are irrelevant to studying the way that language **is**.
- **Prescriptivism** is simultaneously seen as
 1. irrelevant to the study of language - “linguists should be descriptivists, not prescriptivists”
 2. to be countered- “leave your language alone”

Why should linguists care about prescriptivism?

- Attitudes like “leave your language alone” and “prescriptivism is harmful” rest on the **belief that prescriptivism can actually affect language use**
 - If prescriptivism had no consequences, we wouldn't spend any effort combating it.

Why should linguists care about prescriptivism?

- Attitudes like “leave your language alone” and “prescriptivism is harmful” rest on the **belief that prescriptivism can actually affect language use**
 - If prescriptivism had no consequences, we wouldn't spend any effort combating it.
- Historically, plenty of **examples**:
 - Diminished usage of anglicisms in formal French.
 - *Une *application/candidature* ‘a job application’
 - Changes resulting from hypercorrection
 - e.g. *between you and I, octopi*

Why should linguists care about prescriptivism?

- Attitudes like “leave your language alone” and “prescriptivism is harmful” rest on the **belief that prescriptivism can actually affect language use**
 - If prescriptivism had no consequences, we wouldn't spend any effort combating it.
- Historically, plenty of **examples**:
 - Diminished usage of anglicisms in formal French.
 - *Une *application/candidature* ‘a job application’
 - Changes resulting from hypercorrection
 - e.g. *between you and I, octopi*
 - Increasing adoption of *they* as a gender-neutral 3SG pronoun in English.

Prescriptivism - and more generally speakers' **metalinguistic awareness** - should be taken into account when looking at **patterns of usage**, like any other sociolinguistic factor.

- High **metalinguistic awareness** of language
 - Tradition of *remarqueurs*, columns about language in popular media (Ayres-Bennett, 1994, 2006).
 - Lots of schooling in grammar and literature.
- **Language planning**
 - Low tolerance for local languages and varieties.
 - Attempts to exclude all foreign borrowings and loanwords.

Prescriptivism in France - Académie Française

Since 1635, its mission is to “keep the French language **pure and elegant**, and fit for discussing the arts and sciences”.

(to everyone’s big surprise, barely any linguists involved)



The link between prescriptivism and defectiveness

Vogel (2019)'s paradox of grammatical taboos

1. A taboo in a language L can only hold over a construction C, if C exists. Thus, C must be part of Ls language system.
2. Because of the taboo over C, speakers of L who conform to the taboo nevertheless believe that C should not and therefore does not belong to L.

- **Standardising culture:** one and only one way to speak correctly
 1. **Implicit avoidance:**
 - there is a right way to express a message →
 - speakers uncomfortable with morphological indeterminacy →
 - **avoid form** rather than risk choosing wrong one.

- **Standardising culture:** one and only one way to speak correctly
 1. **Implicit avoidance:**
 - there is a right way to express a message →
 - speakers uncomfortable with morphological indeterminacy →
 - **avoid form** rather than risk choosing wrong one.
 2. **Explicit avoidance:** such cases may be codified as defective in grammars, yielding to the **acquisition of the gap as explicit linguistic knowledge** in a prescriptively authoritative text.
 - But words may be coded as defective for reasons other than indeterminacy → CLORE 'to close', resurrected.

Prescriptivism interacts with and substantiates **metalinguistic knowledge** of defectiveness.

Predictions of a prescriptiveness account

- **Speakers vary** in whether they deem a word defective
 - Depends on their **personal inclinations** towards prescriptivism.
 - Depends on whether they have **explicit metalinguistic knowledge** of the word's defective status.

Predictions of a prescriptiveness account

- **Speakers vary** in whether they deem a word defective
 - Depends on their **personal inclinations** towards prescriptivism.
 - Depends on whether they have **explicit metalinguistic knowledge** of the word's defective status.
- **Defective words vary** in the extent to which they are deemed defective
 - Some words are **more central** to prescriptivist discourse - a function of frequency

- **Dictionaries disagree**, reflecting variation in conservatism and usage.
 - Core group is defective in all dictionaries
 - FALLOIR 'to have to', QUÉRIR 'to seek'
 - Some lexemes are defective in some dictionaries but not in others
 - BRAIRE 'to bray' is defective in Le Robert but not Larousse
 - Even if the same lexeme is defective in two dictionaries, different cells might be listed
 - OCCIRE 'to kill' is defective for its IND.PRS, IND.IPFV, IND.PST in both Le Robert and Larousse, only Larousse lists it as defective in its IND.FUT.

Methodology

- **French** speakers
 - Well-documented list of defective forms
 - Strong prescriptive culture

- **French** speakers
 - Well-documented list of defective forms
 - Strong prescriptive culture
- **Acceptability judgement task**

Aujourd'hui j'ai raton laveur mes amis

Est-ce que cet usage du mot est correct?

Pas du tout correct  Parfaitement correct

Task Conditions

Normative Judgement Could you find this usage in a dictionary? Would a teacher mark it as correct?

+normative, +formal

Aujourd'hui j'ai raton laveur mes amis

Est-ce que cet usage du mot est correct?

Pas du tout correct  Parfaitement correct

Possibility Judgement Could you hear this usage from friends hanging out at a bar, or students after school?

-normative, -formal

Aujourd'hui j'ai raton laveur mes amis

Est-ce que cet usage du mot est possible?

Totalement impossible  Tout à fait possible

- **Defective** verb forms marked as defective in at least two French dictionaries.
 - Removed lexemes that were marked as formal or archaic or register-restricted (e.g. legal)
 - (most defective lexemes are either explicitly marked as formal or connotated as such - CLORE VS FERMER 'to close')

- **Defective** verb forms marked as defective in at least two French dictionaries.
 - Removed lexemes that were marked as formal or archaic or register-restricted (e.g. legal)
 - (most defective lexemes are either explicitly marked as formal or connotated as such - CLORE VS FERMER 'to close')
- **Slang**: grammatical taboos - informal French words
- **Subject agreement errors**: ungrammatical - the verb featured an incorrect agreement marker¹

¹not homophonous with the correct option

If prescriptivism is what causes defectiveness, we expect defective words to...

If prescriptivism is what causes defectiveness, we expect defective words to...

- Be rated **higher in the possibility task** than the normative task
 - They are **used** in the language but are **stigmatised**
 - Similar to slang in this respect

Hypotheses

If prescriptivism is what causes defectiveness, we expect defective words to...

- Be rated **higher in the possibility task** than the normative task
 - They are **used** in the language but are **stigmatised**
 - Similar to slang in this respect
- Be rated **more variably** than the other two item conditions
 - Depending on how strong the **prescriptive pressure against using a given lexeme** is
 - Depending on the individual's **level of agreement with prescriptive norms and their knowledge of them.**

Procedure

80 participants from Prolific.co



Administered a **prescriptiveness questionnaire** (aimed at three types of prescriptivism) and collected demographic info



Assigned to a **task condition**, given instructions for the kind of judgement required



9 items for each of the three item conditions in a randomised order, no distractors



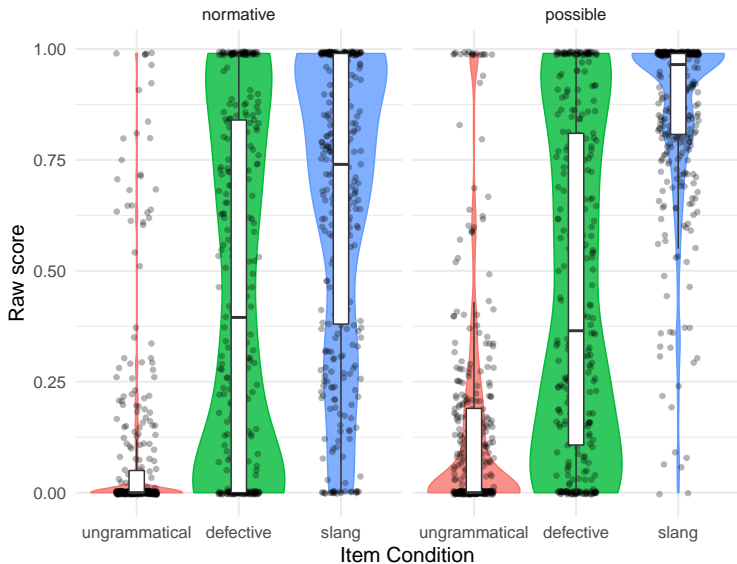
Verification of **lexeme knowledge**

A bayesian **zero-and-one-inflated beta regression** was fitted to participant judgements.

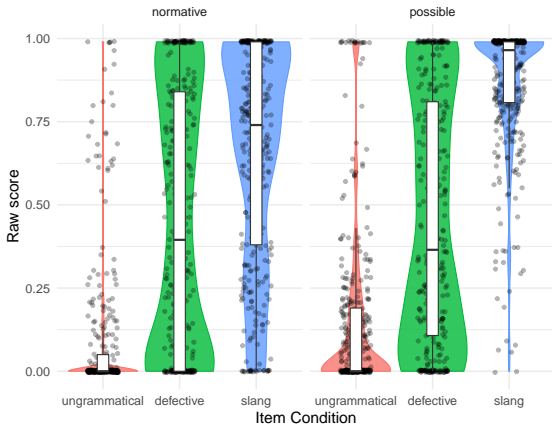
```
judgement ~  
item_condition * task_condition * frequency * prescriptivism +  
(item_condition * frequency | participant) +  
(task_condition * prescriptivism | item)
```

Results

Task and item condition - raw data

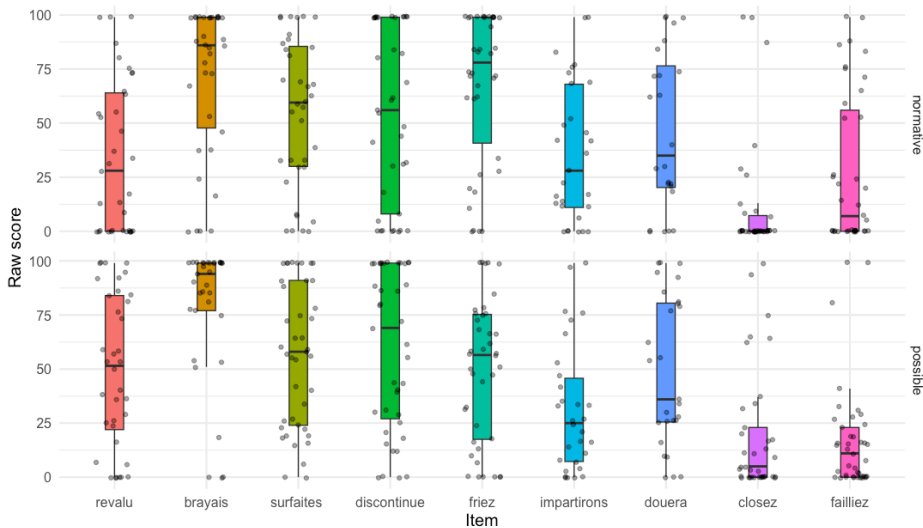


Task and item condition - raw data

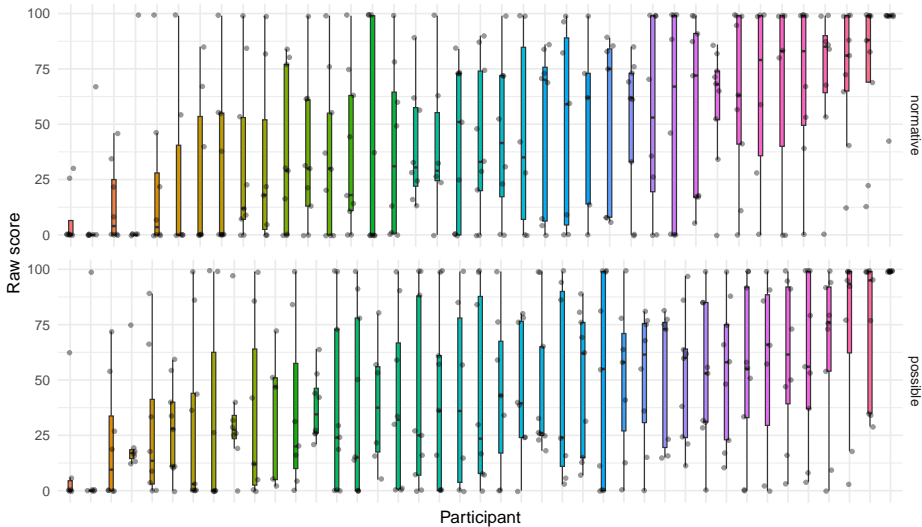


	ungrammatical	defective	slang
median $\hat{\sigma}$ by participant	0.08	0.36	0.11
median $\hat{\sigma}$ by item	0.23	0.34	0.26

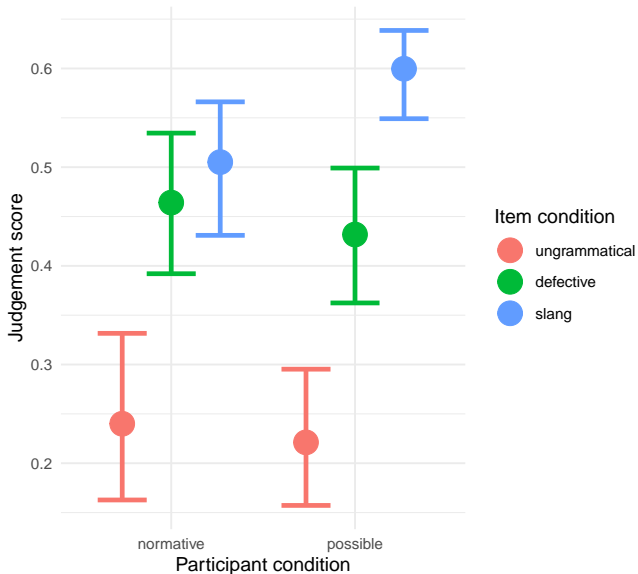
Variability for defective words by item and task



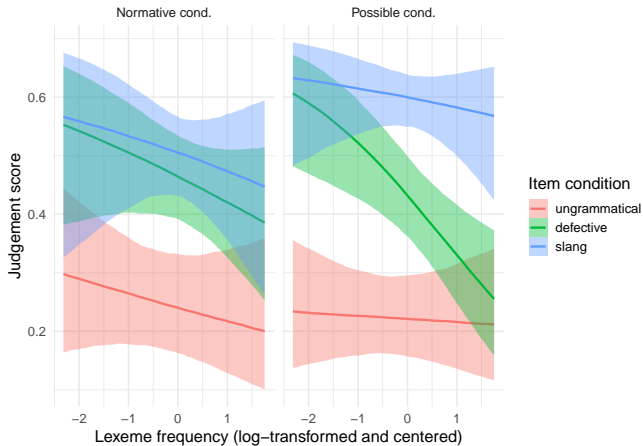
Variability for defective words by participant and task



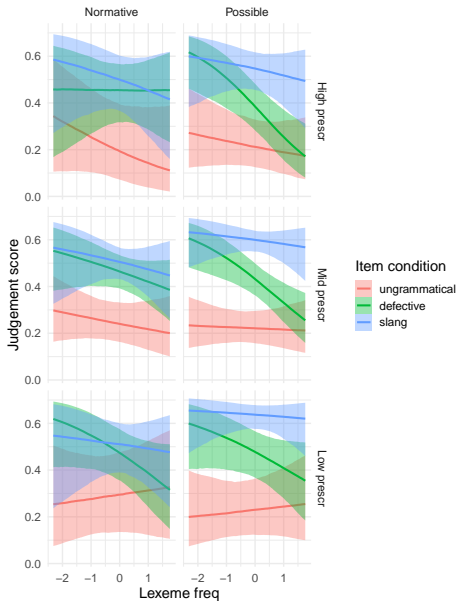
Conditional plots - item x task



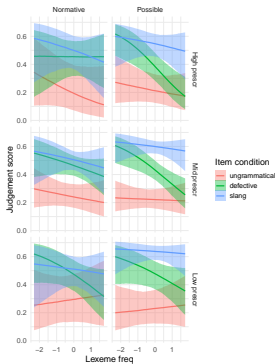
Conditional plots - item x task x frequency



Conditional plots - item x task x frequency x prescriptivism

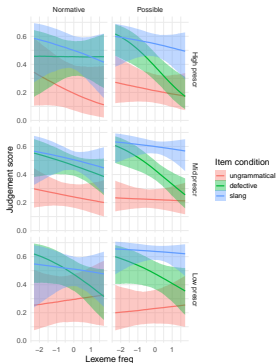


The effect of prescriptivism - inaccurate predictions



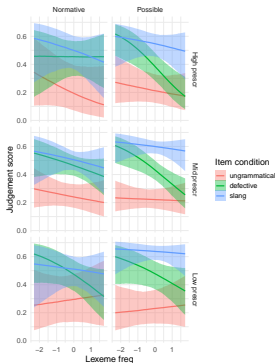
- Prediction: in **normative** task, **defective words** rated more harshly by **more prescriptive people**, moderated by **frequency**.

The effect of prescriptivism - inaccurate predictions



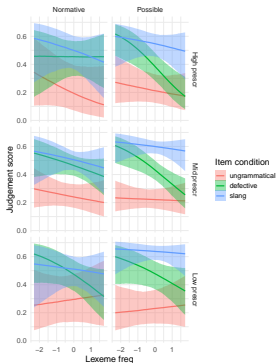
- Prediction: in **normative** task, **defective words** rated more harshly by **more prescriptive people**, moderated by **frequency**.
 - Instead: high prescriptiveness people in normative task rate **lower on average**, but have a **flatter slope for frequency**. Why?

The effect of prescriptivism - inaccurate predictions



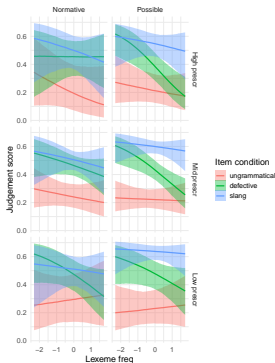
- Prediction: in **normative** task, **defective words** rated more harshly by **more prescriptive people**, moderated by **frequency**.
 - Instead: high prescriptiveness people in normative task rate **lower on average**, but have a **flatter slope** for frequency. Why?
 - Treating all defective words as **equally bad**.

The effect of prescriptivism - inaccurate predictions



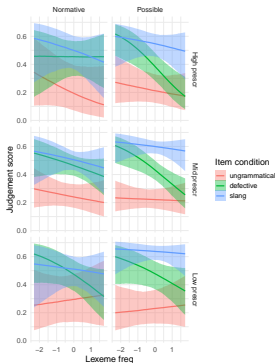
- Prediction: **frequency** shouldn't matter at all in the **possibility** task - instead, strong negative slope

The effect of prescriptivism - inaccurate predictions



- Prediction: **frequency** shouldn't matter at all in the **possibility** task - instead, strong negative slope
 - **Formal lexemes** are disproportionately the target of prescriptivism (no slang word is defective...)

The effect of prescriptivism - inaccurate predictions



- Prediction: **frequency** shouldn't matter at all in the **possibility** task - instead, strong negative slope
 - **Formal lexemes** are disproportionately the target of prescriptivism (no slang word is defective...)
 - **Register clash** - unlikely to hear formal lexemes between friends at the bar.

Discussion

- Defective words elicit much more **variable responses** than other item types.
 - Different speakers count different words as defective.

Core findings

- Defective words elicit much more **variable responses** than other item types.
 - Different speakers count different words as defective.
- A defective word's **lexeme frequency** matters a lot for its acceptability (not true for other item types)
 - Normative task: words that are known to be defective are rated worse
 - Possibility task: words that are known to be formal are rated worse

Core findings

- Defective words elicit much more **variable responses** than other item types.
 - Different speakers count different words as defective.
- A defective word's **lexeme frequency** matters a lot for its acceptability (not true for other item types)
 - Normative task: words that are known to be defective are rated worse
 - Possibility task: words that are known to be formal are rated worse
- **Participant prescriptivism** has an important role in the normative task.
 - More prescriptive participants rate defective words worse on average, and treat them as a unified group regardless of frequency.

Trusting dictionaries for morphological investigation

- Do so with **care**.
- We already know this is unwise for phenomena subject to **individual variation** and **sociolinguistic conditioning**.
 - But I suspect many phenomena to which this applies are still flying under the radar.
 - Underestimating the importance of this could mislead attempts to theorise about linguistic phenomena.

Started with a puzzle. Both of these are true:

1. Defective words have frequency profiles similar to nondefective words in corpora.
2. (Some) speakers dislike (some) defective words (to a greater or lesser extent).

1. **Aggregating over the variation** in items and speakers might be enough to yield attested frequency patterns for defective words in corpora.
 - I think prescriptivism is behind this variation, but it doesn't have to be.
2. If we concede prescriptivism has a role, defective words may be used when **social sanctioning risk** is low
 - Internet: perceived low risk (generally)
 - Experiments: perceived high risk

A cognitive mechanism?

The **Negative Feedback Cycle** (Kapatsinski, 2022): cognitive mechanism that might explain a disconnect between production (corpora) and comprehension (acceptability judgement) of defective words.

First stages of processing about **generating options**, later stages about **suppressing suboptimal options** (message, communicative context etc)

- **Low vs high stakes** : determines threshold of uncertainty for engaging filter
 - Low stakes: defective forms more likely

A cognitive mechanism?

The **Negative Feedback Cycle** (Kapatsinski, 2022): cognitive mechanism that might explain a disconnect between production (corpora) and comprehension (acceptability judgement) of defective words.

First stages of processing about **generating options**, later stages about **suppressing suboptimal options** (message, communicative context etc)

- **Low vs high stakes** : determines threshold of uncertainty for engaging filter
 - Low stakes: defective forms more likely
- **Production vs comprehension**: former happens fast, latter can take indefinite time.
 - Production: defective forms more likely

Disgust-oriented qualia for **defectiveness** in French very similar to seeing someone make a social faux pas.

Different than qualia for **ungrammaticality**, which are **confusion-oriented**.

- The experiment's results are **suggestive** of prescriptivism playing an important role in accounting for defectiveness, but not decisive. Unavoidable because
 - Defective lexemes in French have **inherently formal** connotations.
 - The task conditions differed in both **normativity and formality**.
- A follow-up study:
 - Compare the outcome of the same structural conditions in **±prescriptive language planning** approaches.
 - Joint work with Mari Aigro, Virve-Anneli Vihman, Andrea Sims.

- **Defectiveness** in French is often linked to **uncertainty or indeterminacy**.

- **Defectiveness** in French is often linked to **uncertainty or indeterminacy**.
- The same structural conditions lead to **overabundance in Estonian**.

- **Defectiveness** in French is often linked to **uncertainty or indeterminacy**.
- The same structural conditions lead to **overabundance in Estonian**.
- French has a language planning culture that is **antivariationist**, while Estonian **embraces variation**.

- **Defectiveness** in French is often linked to **uncertainty or indeterminacy**.
- The same structural conditions lead to **overabundance in Estonian**.
- French has a language planning culture that is **antivariationist**, while Estonian **embraces variation**.
- Is defectiveness what happens when **grammatical uncertainty meets antivariationist pressures**?

- **Defectiveness** in French is often linked to **uncertainty or indeterminacy**.
- The same structural conditions lead to **overabundance in Estonian**.
- French has a language planning culture that is **antivariationist**, while Estonian **embraces variation**.
- Is defectiveness what happens when **grammatical uncertainty meets antivariationist pressures**?
- An experiment coming soon!

Thank you for not having **forgoed* this talk!

References



Albright, Adam (2003). “A quantitative study of Spanish paradigm gaps”. In: *West Coast Conference on Formal Linguistics 22 proceedings*. Ed. by Gina Garding and Mimu Tsujimura. Somerville, MA: Cascadilla Press, pp. 1–14 (cit. on pp. 26–31).



Ayala, Alexandra Lilia (2022). *A method for modelling nominal defectiveness in an Icelandic corpus*. ms., University of British Columbia (cit. on p. 131).



Ayres-Bennett, Wendy (1994). “Les ailes du temps et la plume du «remarqueur»: La tradition puriste au XIXe siècle”. In: *Romantisme* 24.86, pp. 33–46 (cit. on p. 39).







— (2006). “Reading the *Remarqueurs*: Changing perceptions of ‘classic’ texts”. In: *Historiographia Linguistica* 3 (33), pp. 263–302 (cit. on p. 39).









Baerman, Matthew (2008). “Historical observations on defectiveness: The first singular non-past”. In: *Russian Linguistics* 32, pp. 81–97 (cit. on pp. 29–31).

Bibliography

-  Baerman, Matthew (2011). “Defectiveness and homophony avoidance”. In: *Journal of Linguistics* 47:1, pp. 1–29 (cit. on pp. 29–31).
-  Baerman, Matthew and Greville G. Corbett (2010). “Introduction: Defectiveness: Typology and diachrony”. In: *Defective paradigms: Missing forms and what they tell us*. Ed. by Matthew Baerman, Greville G. Corbett, and Dunstan Brown. London: British Academy Press, in coordination with Oxford University Press, pp. 1–18 (cit. on pp. 26–28).
-  Bermel, Neil, Luděk Knittle, and Jean Russell (2018). “Frequency data from corpora partially explain native-speaker ratings and choices in overabundant paradigm cells”. In: *Corpus Linguistics and Linguistic Theory* 14, pp. 197–231 (cit. on pp. 26–28).
-  Boyé, Gilles and Patricia Cabredo Hofherr (2010). “Defectiveness as stem suppletion in French and Spanish verbs”. In: *Defective paradigms: Missing forms and what they tell us*. Ed. by Matthew Baerman, Greville G. Corbett, and Dunstan Brown. London: Oxford University Press, in coordination with British Academy Press, pp. 35–52 (cit. on pp. 21–24).

Bibliography

-  Broadbent, Judith M. (2009). “The *amn’t gap: The view from West Yorkshire”. In: *Journal of Linguistics* 45, pp. 251–284 (cit. on pp. 29–31).
-  Cameron, Deborah (1995). *Verbal hygiene*. London: Routledge (cit. on p. 100).
-  Curzan, Anne (2014). *Fixing English: Prescriptivism and language history*. Cambridge: Cambridge University Press (cit. on pp. 100, 101).
-  Daland, Robert, Andrea D. Sims, and Janet Pierrehumbert (June 2007). “Much ado about nothing: A social network model of Russian paradigmatic gaps”. In: *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*. Prague: Association for Computational Linguistics, pp. 936–943. URL: <https://aclanthology.org/P07-1118> (cit. on p. 103).
-  Fábregas, Antonio (2018). “Defectiveness in morphology”. In: *Oxford research encyclopedia of linguistics*. Ed. by Mark Aronoff. Oxford University Press (cit. on pp. 26–28).
-  Gilliéron, Jules (1919). *La faillite de l’étymologie phonétique: Résumé de conférences faites a l’école pratique des hautes études*. Neuveville: Beerstecher (cit. on pp. 29–31).

Bibliography

-  İleri, Muhammed and Ömer Demirok (to appear). “A paradigm gap in Turkish”. In: *Proceedings of the 7th Workshop on Turkic and Languages in Contact with Turkic* (cit. on p. 104).
-  Kapatsinski, Vsevolod (2022). “Morphology in a Parallel, Distributed, Interactive Architecture of Language Production”. In: *Frontiers in Artificial Intelligence* 5, Article 803259 (cit. on pp. 81, 82).
-  Nikolaev, Alexandre and Neil Bermel (2022). “Explaining uncertainty and defectivity of inflectional paradigms”. In: *Cognitive Linguistics* 33 (3), pp. 585–621 (cit. on p. 104).
-  Pertsova, Katya (2016). “Transderivational relations and paradigm gaps in Russian verbs”. In: *Glossa* 1 (1), pp. 1–13 (cit. on p. 104).
-  Pertsova, Katya and Julia Kuznetsova (2015). “Experimental evidence for lexical conservatism in Russian: Defective verbs revisited”. In: *Formal Approaches to Slavic Linguistics 24 proceedings*. Ed. by Yohei Oseki, Masha Esipova, and Stephanie Harves. Ann Arbor, MI: Michigan Slavic Publications, pp. 418–437 (cit. on p. 104).

Bibliography



Sims, Andrea D. (2006). “Minding the gaps: Inflectional defectiveness in a paradigmatic theory”. PhD thesis. Columbus, OH: The Ohio State University (cit. on pp. 29–31, 104).



— (2015). *Inflectional defectiveness*. Cambridge: Cambridge University Press (cit. on pp. 3, 29–31, 103, 104).



— (2023). “Defectiveness”. In: *The Wiley Blackwell Companion to Morphology*. Ed. by Peter Ackema et al. Malden, MA: Wiley-Blackwell (cit. on pp. 26–31).



Vogel, Ralf (2019). “Grammatical taboos: An investigation on the impact of prescription in acceptability judgement experiments”. In: *Zeitschrift für Sprachwissenschaft* 38 (1), pp. 37–79 (cit. on p. 42).

Appendix

Sitting with the qualia

English Yesterday, she [**forgo**] the entrée so she could have dessert.

French Nous [**clôre**] la porte.

Italian Il sole ha [**splendere**] tutto il giorno.

Spanish Presidente, [**abolir**] esa ley!

Russian Мне нужно пять [**кочерга**].

The extension of defectiveness

- Phenomena falling under the umbrella label of defectiveness vary in
 - Their diachronic pathway (lexeme once had full paradigm vs lexeme never had full paradigm)
 - The part of the linguistic system responsible (syntax, semantics, morphology, phonology, mystery)
 - Whether the issue is with the word's form, meaning or both
 - ...several other dimensions
- Wise to treat them as a natural class?

Establishing scope

- Because of
 1. the wide variety of things we mean by “defectiveness”
 2. the fact that the empirical work I will present is on French verbsit would be unwise to claim that the findings apply to all defectiveness
- Goal: highlight the role of social factors in inducing a felt sense of defectiveness

Defining prescriptivism

- Curzan (2014): practices of language regulation
 - created, maintained and enforced by institutions
 - reproduced and perpetuated by ideologically aligned speakers
- Cameron (1995): a sociolinguistic practice that is
 - elitist
 - authority-/institution-based
 - often, though not always, conservative

Types of prescriptivism (Curzan, 2014)

- **Standardising**

- Deciding what is part of the standard language (right) and what is outside of it (wrong).
- There is only one way to speak correctly.
- “it’s wrong to...”

- **Stylistic:**

- Deciding what counts as beautiful uses of the language.
- More variation-friendly.
- “it’s ugly to...”

- **Restorative:**

- Keep the language pure.
- “The language was better when... let’s go back to that.”

- **Politically responsive:**

- Promoting inclusive/politically expedient language.

Prescriptivism and defectiveness

- **Standardising culture**: one and only one way to speak correctly
 - There is a **right way to express a message** → speakers uncomfortable with morphological indeterminacy → **avoid form rather than risk choosing wrong one**.
 - Such cases may be **codified** as defective in grammars, yielding to the **acquisition of the gap as explicit linguistic knowledge** in a prescriptively authoritative text.
 - Indeterminacy is then not the only thing that leads to defectiveness, if for restorative prescriptivism reasons a form has a gap, then it will be learned as such.
- Prescriptivism interacts with and substantiates metalinguistic knowledge of defectiveness whatever the reason for the prescriptive force.

The relationship between defectiveness, lexeme frequency and prescriptivism

- Frequency has a much stronger effect on defective items than prescriptivism. Compatible with two causal structures.

[t]0.4

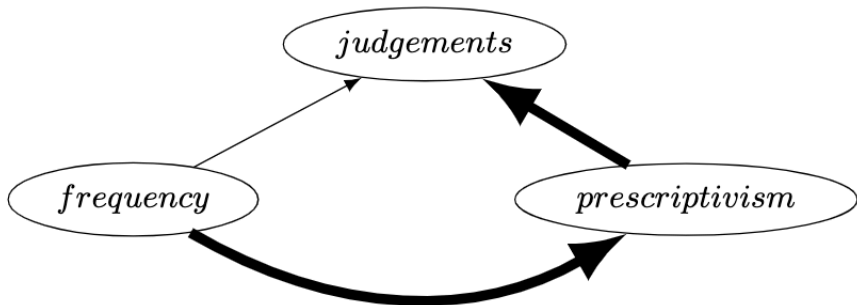


Figure 1

The negative effect of frequency

- Most previous research found a positive effect: less frequent lexeme → more uncertainty, feels 'worse'
 - **albright2003**; Sims (2006), Pertsova and Kuznetsova (2015), and Pertsova (2016): participant more confident producing form for high-freq lexemes (Spanish, Russian)
 - Sims (2015): gaps in Greek disproportionately affect low-freq lexemes
 - Nikolaev and Bermel (2022): less frequent defective lexemes more likely to be avoided.
- Not accidental: visible in the raw data, present in all internal replications. Not the only ones to find a negative effect of frequency on the judgement of defective words (İleri and Demirok, to appear, for Turkish)
- Much is different: task, items, type of defectiveness. A metanalysis is necessary - at minimum, defectiveness and frequency have more complex interaction than previously thought.

Finding defectiveness in corpora

- “A lexeme lacking any acceptable form”
- Naïve prediction: all defective forms have a frequency of 0.
Important to ward off against
- False positives:
 - Because of Zipf, most words in a lexicon occur zero times in corpora, regardless of defective status.
- False negatives:
 - Metalinguistic mentions of defective forms
 - Legitimate uses of defective forms (...errors? non-native speakers?)

Empirically finding defective words

Finding defectiveness in corpora

- Slightly less naïve prediction: defective forms have a frequency that is lower than expected
- ...but how to set up what is expected?

Distinguishing expected from unexpected absence

- Unsurprising if *methylhexanified* occurs 0 times in a corpus.
- Comparison to the frequency of other forms of the same lexeme helps quantify whether we can be certain that low frequency is not accidental.

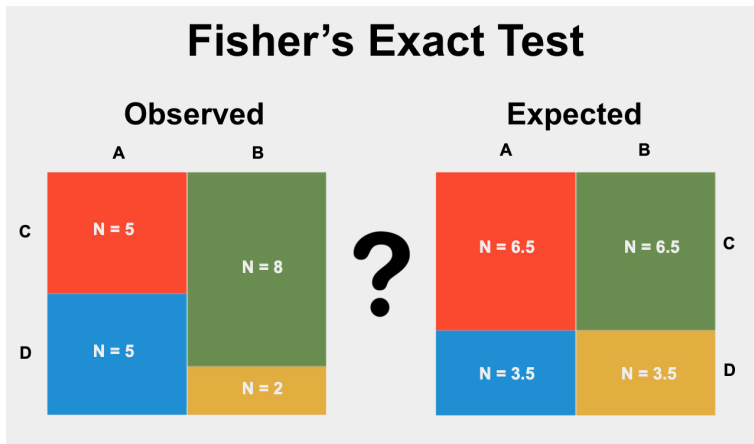
<i>methylhexanified</i>	0		<i>forgoed</i>	0
<i>methylhexanify</i>	2		<i>forgo</i>	76392

Comparing observed to expected frequency

- The overall frequency of the concept expressed by the lexeme is accounted for by comparison with another form of the same lexeme
- Must simultaneously compare with the frequency of other lexemes in the same cell, which serve as measure of what to expect.
- If inflection is a way of expressing the same meaning in different grammatical contexts, the ratio $\frac{\text{Cell A}}{\text{Cell B}}$ should be constant throughout the lexicon.
 - Lexemes with ratios below this trend = defective

Comparing observed to expected frequency

- Mathematically simplest way of performing these comparisons: Fisher's exact test.



Comparing observed to expected frequency

- Mathematically simplest way of performing these comparisons: Fisher's exact test.

<i>unearthed</i>	598		<i>forgoed</i>	0
<i>unearth</i>	6541		<i>forgo</i>	76392

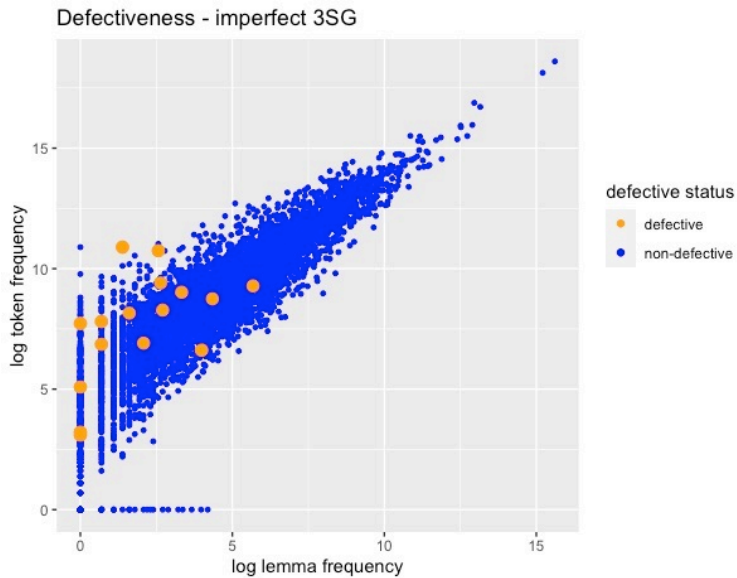
Frequency ratios - expectations

Defective words are expected to have a lower frequency ratio compared to other lexemes in the same cell.

$$\frac{\textit{forgoed}}{\text{FORGO}} < \textit{avg}\left(\frac{\textit{ate}}{\text{EAT}}, \frac{\textit{illuminated}}{\text{ILLUMINATE}, \dots}\right)$$

$$\frac{C1(\textit{defective})}{\text{DEFECTIVE}} < \textit{avg}\left(\frac{C1(\textit{nondefective})}{\text{NONDEFECTIVE}}\right)$$

Frequency ratios - the ground truth



Variations on the theme

- Attempted a few variations on this idea - similar results
 - Frequency of a reference form as the denominator.

IPFV.3SG VS INFINITIVE

- Frequency of a form with minimally different morphosyntactic properties as denominator.

IPFV.3SG VS IPFV.3PL

Defectiveness as large negative residuals

- Statistical modeling: learning how different variables relate to each other. Learned mappings can then be used to make predictions about specific data points.
- Simplest case: learn mappings between the frequency of a lexeme in the INFINITIVE and its frequency in the PST.PTCP

$$\text{freq}(\text{PST.PTCP}) \sim a + \text{freq}(\text{PRS.SIMPLE}) * b$$

- Can add more predictors, e.g.

$$\text{freq}(\text{PST.PTCP}) \sim a + \text{freq}(\text{PRS. SIMPLE}) * b + \text{freq}(\text{PRS.PROG}) * c$$

- Can allow for nonlinear relationship between predictors

Defectiveness as large negative residuals

- The mappings learned can be used to make predictions, e.g.

$$\begin{aligned} \text{freq(PST.PTCP)} &= a + \text{freq(PRS.SIMPLE)} * b + \text{freq(PRS.PROG)} * c \\ \text{freq(EATEN)} &= a + \text{freq(EAT)} * b + \text{freq(EATING)} * c \end{aligned}$$

- Residual: the difference between the predicted value and the observed value.
- Defectiveness: large negative residual? Lower frequency than expected given the behaviour of other lexemes.

Defectiveness as large negative residuals

- Played with a number of parameters
 - Variables: raw frequencies vs frequency ratios
 - Model structure: interaction terms between predictors
 - Distribution of predicted variable: poisson, negative binomial, gamma, normal, lognormal.
 - Mixture models: zero-inflation, hurdle (modeling zero frequencies as potentially generated by different means compared to non-zero frequencies)

Defectiveness as large negative residuals - outcome

- Fewer false positives than before, especially good at catching words that are defective for syntactic/semantic reasons (e.g. weather verbs), but still cannot dial into "morphological" defectives (of the FOREGO type).
- Defective words usually have negative residuals, but these are not particularly large, on par with many other non-defective words.

The missing ingredient: semantics?

- Perhaps not very informative to compare $\frac{\textit{foregoed}}{\textit{forego}}$ to all other lexemes.
- Some lexemes may have low frequency ratios in the cell for independent reasons:

	JOLI 'beautiful'		LESBIEN 'lesbian'
M.SG	joli	>	lesbien
F.SG	jolie	=	lesbienne

The missing ingredient: semantics?

- Restrict comparison to semantically similar lexemes.

$$\frac{\textit{foregoed}}{\textit{forego}} \text{ VS } \frac{\textit{avoided}}{\textit{avoid}}$$

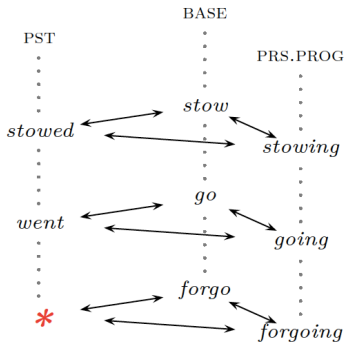
- Use word embeddings to define a comparison class. Convert the lexicon into a vector space and find lexemes with similar meaning to the lexeme of interest, based on the words they co-occur with.

Accounting for semantics

- Fisher test: only compare to lexemes with cosine similarity > 0.7 to target lexeme.
 - Words with too few near neighbours removed from the analysis.
- Modeling: include a dimensionally reduced vector as predictor, alongside lexeme frequency.
- Still a few false negatives, but mostly false positives.

Defectiveness and word form uncertainty

- Word form uncertainty is associated with "morphological defectiveness" (most saliently, Albright, 2003)



Defectiveness and word form uncertainty

- Word form uncertainty is associated with "morphological defectiveness" (most saliently, Albright, 2003)
- Exploit this fact to find defective forms?

Information theory and word form uncertainty

- Build a measure of how uncertain a given word form is, based on other members of its paradigm.
- Two aspects of form predictability may be relevant:
 1. How surprised are we to see that a particular pattern has applied?
 - SG *goose* → PL *geese*
 2. How uncertain are we about which pattern actually applies?
 - PRS *fling* → PST $\left\{ \begin{array}{l} \textit{flung?} \\ \textit{flang?} \\ \textit{flinged?} \end{array} \right.$

Information theory and word form uncertainty

Out of all the patterns compatible with the phonology of the form...

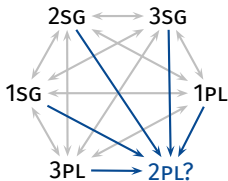
How surprised are we to see that pattern P applies?

surprisal

How uncertain are we about which of the possible patterns apply?

local entropy

A form's paradigmatic uncertainty is its average uncertainty when predicted from each of the other cells in its paradigm



Using word form uncertainty to find defectiveness

- Sorting low frequency items by uncertainty measures successfully captured a number of defective lexemes, but many nondefective lexemes also ranked highly
- Using uncertainty and lexeme frequency to predict token frequency and then selecting large negative residuals does not isolate defectives

Widening the search

- These methods were applied to French, attempting to hone in on the words deemed defective in Lexique (Bonami, Caron & Plancq, 2014) for verbs and nouns, and on those deemed defective in Bonami & Boyé (2003) for adjectives.
- They were also applied to Russian nouns, attempting to hone in on words deemed defective by the Zalizniak dictionary.

The taxonicity of defective words (in French)

- The methods employed all sought to find combinations of dimensions along which defective words formed a clearly separate extreme.
- Or at least have them all be close enough to one corner of the space
- Instead, defective words seem to pattern all over the place, and while there are correlations with factors, no combination of factors attempted is able to pick out defectives as a natural class.

The taxonicity of defective words (in French) - why?

- Things deemed defective in French dictionaries are indeed heterogeneous (Boyé & Cabredo Hoffher, 2008)
- Three main groups
 - Stem indeterminacy
 - Stem conflict
 - Stem gaps
 - (secret fourth group: defectiveness for reasons outside the morphology)

A paradox

How can both of these be true?

1. Speakers do not like defective forms
2. Defective forms are indistinguishable from non-defective forms in corpora.

The paradox - some hypotheses

- Maybe our methods were poor
 - But Ayala (2022) finds underattestation for Icelandic defective forms in corpora
 - Maybe interspeaker variation in what is defective is enough to hide underattestation in the aggregate
 - But why is there interspeaker variation on this?