Spare us the surprise

The interplay of paradigmatic predictability and frequency

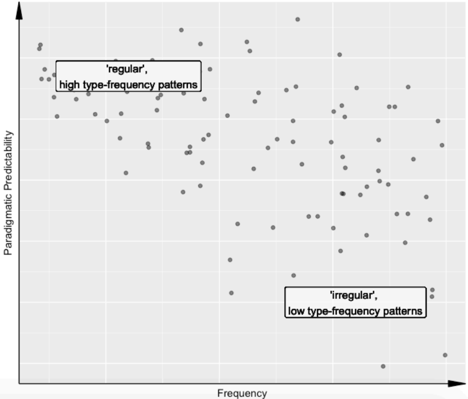
Maria Copot Olivier Bonami

ISMo 2021

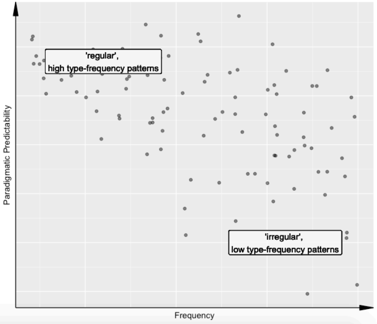
Université de Paris, LLF

Introduction

- A known inverse link between **frequency and paradigmatic predictability** of a word form (Wu, Cotterell & O'Donnell, 2019; Marcus et al. 1992; Bybee, 1985):
  - Paradigmatically unpredictable word forms (suppletives/irregulars) tend to be frequent
  - Infrequent lexemes tend to have predictable word forms

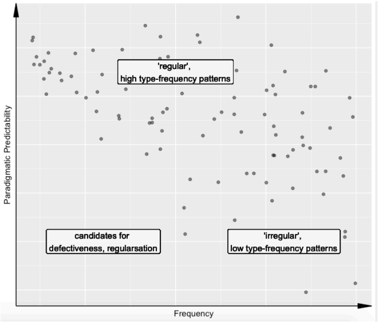


Uncertainty, frequency and memory



- The more high frequency a word form, the more it can afford to be paradigmatically unpredictable.
  - The unpredictable word form can be well anchored in memory thanks to its frequency

Uncertainty, frequency and memory



- If a paradigmatically unpredictable word is infrequent/in an infrequent context...
    - Regularisation (Eng. *helped* ← *holp*) (Lieberman et al. 2007)
    - Avoidance (*forego* → *foregoed?/forewent?*) (Albright, 2003; Sims, 2015)
    - If a whole context is infrequent and a locus of low predictability, it may drop out of use (It. *passato remoto*)

Summing up

- A negative correlation between frequency and predictability
- For an unpredictable form to survive, it must be frequently attested
- Words can't afford to be both syntagmatically and paradigmatically unpredictable (Filipović Đurđević & Milin, 2019)
    - Frequent words are an expected way to continue a sentence (= syntagmatically more predictable), so they can tolerate paradigmatic uncertainty.

The effect of paradigmatic predictability on speaker production

- When producing a sentence, we incrementally have to find words that
    1. are inflectionally appropriate (e.g. have the correct agreement, are the correct part of speech: *she eats/*eat dinner*)
    2. are an appropriate lexical choice (contribute the intended lexical semantics: *she eats/*coagulates dinner*)

The effect of paradigmatic predictability on speaker production

- At each word boundary, the inflectional requirements are often clear
  - *"You should beware of the dog!" - "Indeed, yesterday, I ____ (it)!"* needs a past tense form.
- Several appropriate lexemes (HEED, BEWARE OF, WATCH OUT FOR). Some parameters for the choice:
  - Overall strength of the lexeme's mental representation (a function of recency, frequency in input, salience given context...)
  - The ease of accessibility of the necessary form of the lexeme (a function of predictability)

The plan

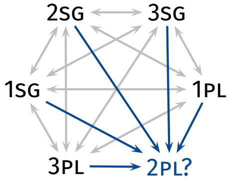
- How does paradigmatic predictability impact token frequency?
- The hypothesis:
    - at parity of lexeme frequency, less paradigmatically predictable words will be used less frequently.
    - The more frequent a lexeme, the less predictability will matter for frequency of use (frequent words need to be retrieved from memory rather than actively predicted)

Aspects of form predictability

- Several aspects of form predictability may be relevant to token frequency, e.g.:
  1. Local entropy: the uncertainty surrounding how to fill a given cell
     - PRS *fling* → PST $\begin{cases} \textit{flung?} \\ \textit{flang?} \\ \textit{flinged?} \end{cases}$
  2. Surprisal: the predictability of the particular form actually filling the cell
     - PRS *fling* → PST *flung*
- Following a corpus study, we conclude that the measure relevant for written production is surprisal How to operationalise?

Form predictability as average surprisal i

- Need to measure:
  - given **knowledge of the rest of the paradigm**...
  - how **confident** should a speaker be that they are **producing the right form in the necessary cell?**
- This is clearly a variant of the **Paradigm Cell Filling Problem** (Ackerman, Blevins & Malouf, 2009; Ackerman & Malouf, 2013).



- We rely on a purely **word-based** approach to the PCFP of Bonami & Beniamine (2016)

Operationalising surprisal

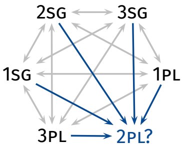
- Beniamine's (2018) Qumin package was used for all computations.
- Intuitively: **conditional probability of output form given the phonological shape of the input form**
- Surprisal is computed over **pairs of cells** ($C1 \rightarrow C2$). For a given form pair...
  - Find all patterns compatible with the input form
  - $x = \frac{\textit{type freq. of instantiated pattern}}{\textit{type freq. of all applicable patterns}}$
  - Turn it into bits: $-log_2(x)$

$$\text{1PL IND PRES } \textit{sortons} \rightarrow \text{PST PART} \begin{cases} \textit{sorté?} \\ \textit{sortu?} \\ \textit{sorti?} \end{cases}$$

PATTERN	PATTERN TYPE FREQUENCY	SURPRISAL
Xons ~Xé	most lexemes	0.06
Xons ~Xu	∼ 15 lexemes	4.7
Xons ~Xi	**∼ 5 lexemes**	**7.2**

Average surprisal

- Average over predictor cells c to get an overall estimation of how surprising c' is given the rest of the paradigm.



  - Ideally, this should be weighted by cell frequency.
    - But we do not have quality estimations of cell frequency, because of pervasive syncretism.
    - For lack of a better solution we use unweighted frequency.

Methodology

- We set out to confirm that paradigmatic surprisal has a negative effect on token frequency throughout the lexicon.
- And that the effect is reversed for high-frequency lexemes.
- Case study: French verbal cells

Methodology

- For the items within each cell, we constructed a model of the shape
  - *token frequency ∼ surprisal + lexeme frequency + surprisal:lexeme frequency*
- The value of surprisal we employ is the average surprisal of the given form based on each of the other forms in the paradigm.
- Lemma frequency is included as a control variable (= familiarity)
- The interaction: test the intuition that for high values of lemma frequency, surprisal matters less (words with a strong representation in memory don't need to be predicted)
- Separate bayesian poisson regressions with weakly-informative priors were fitted to the data in each cell.

Methodology

- Resources used:
  - Frequency counts: FrCoW (Schäfer & Bildhauer, 2016) for token and lemma counts.
  - Paradigms & excluding homographs: GLàFF (Hatout, Sajous & Calderone, 2014)
  - Surprisal: values computed using Qumin (Beniamine, 2018) on the Flexique verb dataset (Bonami, Caron & Plancq, 2014)

Data selection

- Which cells in the paradigm of French verbs can we work with?
- Working with our dataset, we exclude...

Finite forms

	1SG	2SG	3SG	1PL	2PL	3PL
IND.PRS	2	3	183	2	5	14
IND.IPFV	0	0	5083	10	10	5076
IND.PST	4484	4448	4694	5116	5116	5101
FUT	5211	5207	5213	5190	5212	5221
SBJV.PRS	0	250	2	8	7	13
SBJV.IPFV	4701	4725	5119	4726	4738	4740
COND	0	0	5220	5212	5212	5215
IMP	—	0	—	2	2	—

Nonfinite forms

INF	PRS.PTCP	PST.PTCP			
		M.SG	F.SG	M.PL	F.PL
5006	4311	3935	3055	2903	3199

Number of verbs from Flexique with no homograph documented in the GLÀFF, by paradigm cell

Data selection

- Which cells in the paradigm of French verbs can we work with?
- Working with our dataset, we exclude...
    - cells with high numbers of homographs according to the GLÀFF;

Finite forms

	1SG	2SG	3SG	1PL	2PL	3PL
IND.PRS	2	3	183	2	5	14
IND.IPFV	0	0	5083	10	10	5076
IND.PST	4484	4448	4694	5116	5116	5101
FUT	5211	5207	5213	5190	5212	5221
SBJV.PRS	0	250	2	8	7	13
SBJV.IPFV	4701	4725	5119	4726	4738	4740
COND	0	0	5220	5212	5212	5215
IMP	—	0	—	2	2	—

Nonfinite forms

INF	PRS.PTCP	PST.PTCP			
		M.SG	F.SG	M.PL	F.PL
5006	4311	3935	3055	2903	3199

Number of verbs from Flexique with no homograph documented in the GLÀFF, by paradigm cell

Data selection

- Which cells in the paradigm of French verbs can we work with?
- Working with our dataset, we exclude…
  - cells with high numbers of homographs according to the GLÀFF;
  - cells out of current usage (i.e. most attestations are likely to be archaic);

Finite forms

	1SG	2SG	3SG	1PL	2PL	3PL
IND.PRS	2	3	183	2	5	14
IND.IPFV	0	0	5083	10	10	5076
IND.PST	4484	4448	4694	5116	5116	5101
FUT	5211	5207	5213	5190	5212	5221
SBJV.PRS	0	250	2	8	7	13
SBJV.IPFV	4701	4725	5119	4726	4738	4740
COND	0	0	5220	5212	5212	5215
IMP	—	0	—	2	2	—

Nonfinite forms

INF	PRS.PTCP	PST.PTCP			
		M.SG	F.SG	M.PL	F.PL
5006	4311	3935	3055	2903	3199

Number of verbs from Flexique with no homograph documented in the GLÀFF, by paradigm cell

Data selection

- Which cells in the paradigm of French verbs can we work with?
- Working with our dataset, we exclude...
  - cells with high numbers of homographs according to the GLÀFF;
  - cells out of current usage (i.e. most attestations are likely to be archaic);
  - past participle cells, for which tagging is inherently unreliable.

Finite forms

	1SG	2SG	3SG	1PL	2PL	3PL
IND.PRS	2	3	183	2	5	14
IND.IPFV	0	0	5083	10	10	5076
IND.PST	4484	4448	4694	5116	5116	5101
FUT	5211	5207	5213	5190	5212	5221
SBJV.PRS	0	250	2	8	7	13
SBJV.IPFV	4701	4725	5119	4726	4738	4740
COND	0	0	5220	5212	5212	5215
IMP	—	0	—	2	2	—

Nonfinite forms

INF	PRS.PTCP	PST.PTCP			
		M.SG	F.SG	M.PL	F.PL
5006	4311	3935	3055	2903	3199

Number of verbs from Flexique with no homograph documented in the GLÀFF, by paradigm cell

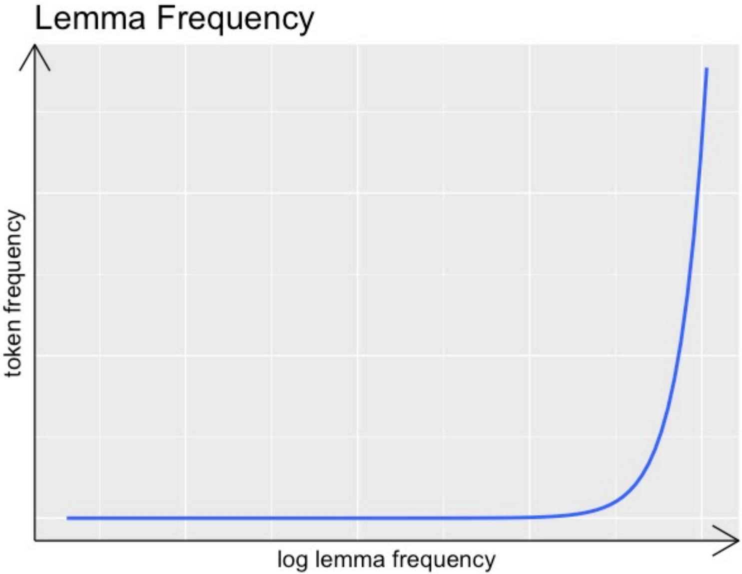
Properties of the selected cells

- The selected cells correspond to 3 areas of high interpredictability.

	FUT.1SG	FUT.2SG	FUT.3SG	FUT.1PL	FUT.2PL	FUT.3PL	COND.3SG	COND.1PL	COND.2PL	COND.3PL	IPFV.3SG	IPFV.3PL	PRS.PTCP	INF
FUT.1SG	0	0	0	0	0	0	0	0	0	0	0.24	0.24	0.24	0.23
FUT.2SG	0	0	0	0	0	0	0	0	0	0	0.24	0.24	0.24	0.23
FUT.3SG	0	0	0	0	0	0	0	0	0	0	0.24	0.24	0.24	0.23
FUT.1PL	0	0	0	0	0	0	0	0	0	0	0.24	0.24	0.24	0.23
FUT.2PL	0	0	0	0	0	0	0	0	0	0	0.24	0.24	0.24	0.23
FUT.3PL	0	0	0	0	0	0	0	0	0	0	0.24	0.24	0.24	0.23
COND.3SG	0	0	0	0	0	0	0	0	0	0	0.24	0.24	0.24	0.23
COND.1PL	0	0	0	0	0	0	0	0	0	0	0.24	0.24	0.24	0.23
COND.2PL	0	0	0	0	0	0	0	0	0	0	0.24	0.24	0.24	0.23
COND.3PL	0	0	0	0	0	0	0	0	0	0	0.24	0.24	0.24	0.23
IPFV.3SG	0.35	0.34	0.34	0.34	0.35	0.34	0.35	0.33	0.33	0.35	0	0	0.0004	0.34
IPFV.3PL	0.35	0.34	0.34	0.34	0.35	0.34	0.35	0.33	0.33	0.35	0	0	0.0004	0.33
PRS.PTCP	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.33	0.33	0.34	0	0	0	0.32
INF	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.076	0.077	0.074	0

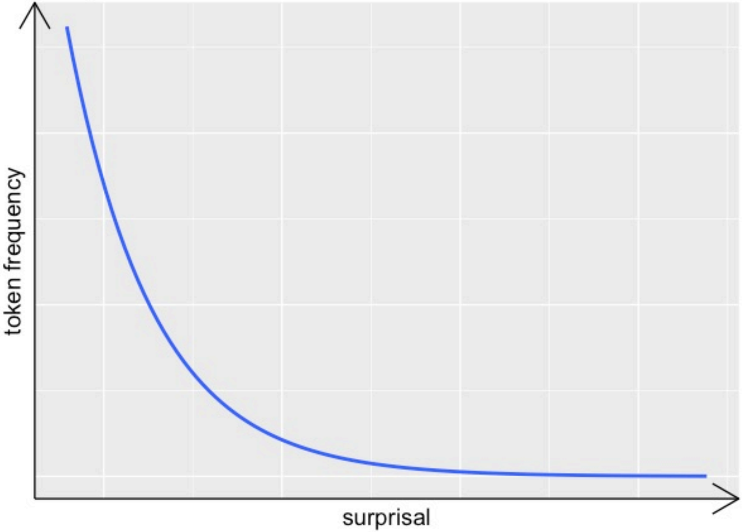
Implicative entropy (Bonami & Beniamine, 2016) between selected cells

Predictions

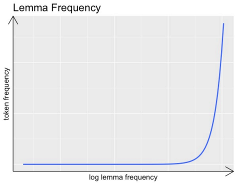


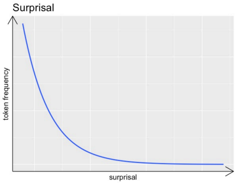
Predictions

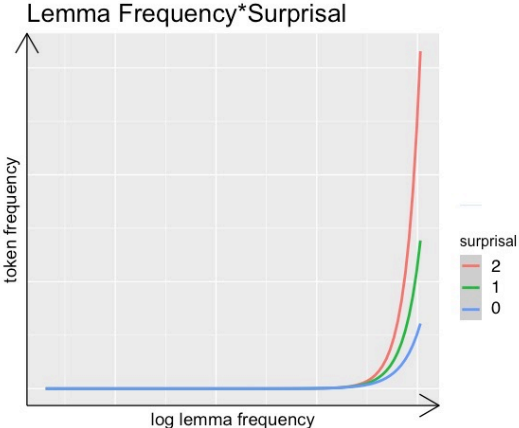
Surprisal



Predictions







Results

- Lemma frequency has a uniform positive effect on token frequency in all cells.
- Surprisal had a negative effect in 12/14 cells, an effect indistiguishable from 0 in 1/14, and an unexpected positive effect in 1/14.
- The interaction between surprisal and lemma frequency had a positive coefficient in 11/14 cells and an effect indistiguishable from 0 in 1/14. 2/14 have unexpected negative coefficients.
- Overall, 11/14 cells behaved exactly as predicted, two behaved counter to expectations and one showed non-significant impact for surprisal and surprisal:lemma

Model Output - Coefficients

Cell	Lemma freq.	Surprisal	Interaction
FUT.1SG	0.9935	−0.3783	0.0675
FUT.2SG	1.0771	−0.2306	0.0447
FUT.3SG	1.1764	−0.0261	0.0073
FUT.1PL	0.9693	−0.1932	0.0415
FUT.2PL	1.1072	−0.3368	0.0647
FUT.3PL	1.1466	−0.0040	0.0088
COND.3SG	1.2509	−1.0392	0.1835
COND.1PL	1.2544	−1.7739	0.2876
COND.2PL	1.2583	−2.7622	0.4486
COND.3PL	1.2312	−1.3889	0.2404
IPFV.3SG	1.1707	−0.0441	−0.0010
IPFV.3PL	0.9352	−0.5588	0.0959
PRS.PTCP	0.5916	0.0545	0.0053
INF	0.9438	0.0620	−0.0089

Unexpected coefficient sign (red highlight)

95% Credible interval overlaps with zero (blue highlight)

Outlier Cells

- **Cells that didn't conform to predictions:** infinitive, imperfect 3sg, present participle.
- These are by far the three most frequent cells in the dataset.
- Hypothesis: the effect of surprisal is therefore nullified at the level of the whole cell (same mechanism for frequent lexemes)
    - while the coefficients for surprisal and the interaction have unexpected monotonicity, their value is much smaller compared to other cells, and very close to 0 (for pres. part. it is indistinguisheable from 0)

Discussion

- Overall, token frequency is negatively impacted by paradigmatic form predictability.
    - The pattern is reversed for items of high lemma frequency.
    - High frequency lexemes are more familiar to speakers, so the predictability of their word forms matters less for access/usage
- The method performs well on 11/14 cells, and the exceptions exist for principled reasons.
    - Showcases the importance of paradigmatic information in predicting frequency.
    - Frequent contexts and lexemes diminish the importance of paradigmatic predictability.

What next?

- Obtaining a good estimate of cell frequency (existing resources yield poor estimates, especially for the person dimension)
  - It would allow a weighed average of surprisal to be used
  - It would help interpret outlying results.
- Currently exploring the Italian verbal system with the same method (less homography)
  - Results going in roughly the same direction, some kinks to iron out
- Testing the general effect of surprisal psycholinguistically.
  - Speakers appear sensitive to paradigmatic surprisal between individual nonwords.

Ackerman, Farrell, James P. Blevins & Robert Malouf (2009). “Parts and wholes: implicative patterns in inflectional paradigms”. **in**: *Analogy in Grammar*. **byeditor**James P. Blevins & Juliette Blevins. Oxford: Oxford University Press, **pages** 54–82 (**backrefpage** 12).

Ackerman, Farrell & Robert Malouf (2013). “Morphological organization: the low conditional entropy conjecture”. **in**: *Language* 89, **pages** 429–464 (**backrefpage** 12).

Albright, Adam (2003). “A quantitative study of Spanish paradigm gaps”. **in**: *WCCFL22 Proceedings*. **byeditor**G. Garding & M. Tsujimura. Sommervile, MA: Cascadilla Press, **pages** 1–14 (**backrefpage** 6).

Beniamine, Sacha (2018). “Typologie quantitative des systèmes de classes flexionnelles”. phdthesis. Université Paris Diderot (**backrefpages** 13, 17).

Bonami, Olivier & S. Beniamine (2016). “Joint predictiveness in inflectional paradigms”. **in**: *Word Structure* 9.2, **pages** 156–182 (**backrefpages** 12, 22).

Bonami, Olivier, Gauthier Caron & Clément Plancq (2014). “Construction d’un lexique flexionnel phonétisé libre du français”. **in**: *Actes du quatrième Congrès Mondial de Linguistique Française*. **byeditor**Franck Neveu et al. **pages** 2583–2596 (**backrefpage** 17).

Bybee, Joan L. (1985). "Morphology. A study of the relation between meaning and form.". in: *Typological Studies in Language* 9.2, pages 493–496 (backrefpage 4).

Filipović Đurđević, Dušica & Petar Milin (2019). "Information and learning in processing adjective inflection". in: *Cortex* 116. Structure in words: the present and future of morphological processing in a multidisciplinary perspective, pages 209–227. ISSN: 0010-9452. URL: https://www.sciencedirect.com/science/article/pii/S0010945218302375 (backrefpage 7).

Lieberman, Erez et al. (2007). "Quantifying the evolutionary dynamics of language". in: *Nature* 449.7163, pages 713–716. ISSN: 1476-4687. URL: https://doi.org/10.1038/nature06137 (backrefpage 6).

Marcus, Gary F. et al. (1992). "Overregularization in Language Acquisition". in: *Monographs of the Society for Research in Child Development* 57.4, pages i–178. ISSN: 0037976X, 15405834. URL: http://www.jstor.org/stable/1166115 (backrefpage 4).

Sims, Andrea (2015). *Inflectional defectiveness*. Cambridge: Cambridge University Press (backrefpage 6).

Wu, Shijie, Ryan Cotterell & Timothy J. O'Donnell (2019). "Morphological Irregularity Correlates with Frequency". in: *CoRR* abs/1906.11483. arXiv: 1906.11483. URL: http://arxiv.org/abs/1906.11483 (backrefpage 4).

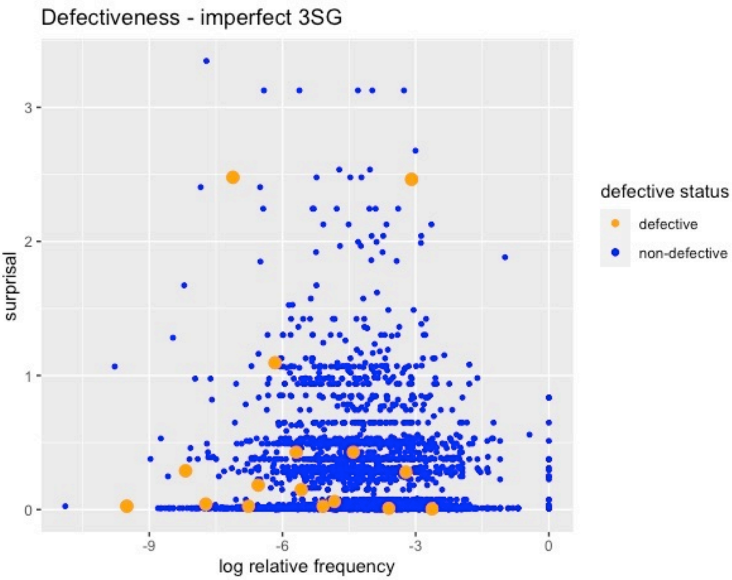
Annex

Model Output - Coefficients & Cell Frequency

Cell	Lemma freq.	Surprisal	Interaction	Cell Frequency	Freq. Rank
FUT.1SG	0.9935	-0.3783	0.0675	1345435	10
FUT.2SG	1.0771	-0.2306	0.0447	303754	13
FUT.3SG	1.1764	-0.0261	0.0073	6575463	5
FUT.1PL	0.9693	-0.1932	0.0415	789963	11
FUT.2PL	1.1072	-0.3368	0.0647	1506039	9
FUT.3PL	1.1466	-0.004	0.0088	4069211	6
COND.3SG	1.2509	-1.0392	0.1835	7394571	4
COND.1PL	1.2544	-1.7739	0.2876	255317	14
COND.2PL	1.2583	-2.7622	0.4486	365173	12
COND.3PL	1.2312	-1.3889	0.2404	1848943	8
IPFV.3SG	1.1707	-0.0441	-0.001	19020206	2
IPFV.3PL	0.9352	-0.5588	0.0959	3726892	7
PRS.PTCP	0.5916	0.0545[1]	0.0053[2]	14297764	3
INF	0.9438	0.062	-0.0089	112986370	1

[1] 95% Credible interval overlaps with zero.

Illustrating Defectiveness



Surprisal in detail i

- For each pair of cells (c, c') in the paradigm:
  1. Assign each pair to an alternation pattern, optimizing alignments between pairs of words.

Lexeme	PRS.3PL	PRS.2PL
CROIRE	kʁwa	kwaje
BAVER	bav	bave
LEVER	lɛv	ləve
MENER	mɛn	məne
PEINER	pɛn	pɛne
MORDRE	mɔʁd	mɔʁde

⇒

Lexeme	PRS.3PL	PRS.2PL	Alternation
BAVER	bav	bave	$\pi_1 : \_ \rightleftharpoons \_\mathrm{e} / X^{+}C\_\#$
PEINER	pɛn	pɛne	
MORDRE	mɔʁd	mɔʁde	
LEVER	lɛv	ləve	$\pi_2 : \_\ \_ \rightleftharpoons \_\mathrm{ə}\_\mathrm{e} / X^{+}\_C\_\#$
MENER	mɛn	məne	
CROIRE	kʁwa	kwaje	$\pi_3 : \_ \rightleftharpoons \_\mathrm{je} / X^{+}\mathrm{wa}\_\#$

Surprisal in detail ii

Lexeme	PRS.3PL	PRS.2PL	Alternation
BAVER	bav	bave	
PEINER	pɛn	pɛne	$\pi_1 : \_ \rightleftharpoons \_\mathrm{e} / X^{+}C\_\#$
MORDRE	mɔʁd	mɔʁde	
LEVER	lɛv	ləve	$\pi_2 : \_\ \_ \rightleftharpoons \_\mathrm{ə}\_\mathrm{e} / X^{+}\_C\_\#$
MENER	mɛn	məne	
CROIRE	kʁwa	kwaje	$\pi_3 : \_ \rightleftharpoons \_\mathrm{je} / X^{+}\mathrm{wa}\_\#$

2. Classify predictor cell shapes on the basis of which patterns they are compatible with.

Lexeme	PRS.3PL	PRS.2PL	π_1	π_2	π_3	Predictor shape
BAVER	bav	bave	✓			κ_1
MORDRE	mɔʁd	mɔʁde	✓			
PEINER	pɛn	pɛne	✓	✓		
LEVER	lɛv	ləve	✓	✓		κ_2
MENER	mɛn	məne	✓	✓		
CROIRE	kʁwa	kwaje			✓	κ_3

⇒ Puts words from predictor cell c into classes $\kappa_1, \ldots, \kappa_m$ that share phonological properties relevant for determining what happens in cell c'.

Surprisal in detail iii

Lexeme	PRS.3PL	PRS.2PL	π_1	π_2	π_3	Predictor shape
BAVER	bav	bave	✓			κ_1
MORDRE	mɔʁd	mɔʁde	✓			
PEINER	pɛn	pɛne	✓	✓		κ_2
LEVER	lɛv	ləve	✓	✓		
MENER	mɛn	məne	✓	✓		
CROIRE	kʁwa	kwaje			✓	κ_3

3. Compute the surprisal of the form found in cell c' given the form found in cell c:

$$S = -\log_2 P(\pi_i \mid \kappa_j)$$

Lexeme	PRS.3PL	PRS.2PL	Pattern	Class	p	S
BAVER	bav	bave	π_1	κ_1	1	0
MORDRE	mɔʁd	mɔʁde	π_1	κ_1	1	0
PEINER	pɛn	pɛne	π_1	κ_2	$1/3$	1.585
LEVER	lɛv	ləve	π_2	κ_2	$2/3$	0.585
MENER	mɛn	məne	π_2	κ_2	$2/3$	0.585
CROIRE	kʁwa	kwaje	π_3	κ_3	1	0