# Form predictability and cell frequency

## A behavioural study

Maria Copot  Olivier Bonami

IMM2022

Université Paris Cité, LLF, CNRS

# Background

# The Paradigm Cell Filling Problem

How can we account for the ability of speakers to **produce and comprehend words** they've never been exposed to before?

# Two lines of research

- Psycholinguistic
  - Often not fully engaging with morphological theory
- Linguistic
  - Morphologically sophisticated but often not directly tested on behaviour

# Two lines of research

- Psycholinguistic
  - Often not fully engaging with morphological theory
- Linguistic
  - Morphologically sophisticated but often not directly tested on behaviour
- Straddling the boundary: the Minimal Generalisation Learner (Albright & Hayes, 2002)

# Investigating speaker behaviour on the PCFP

- Wug tests (Berko Gleason, 1958; Bybee, 1985; Prasada & Pinker, 1993; Albright & Hayes, 2003)

  *I pling every day. Yesterday I also* _____

  - Knowledge of system + form A = form B

# Investigating speaker behaviour on the PCFP

- Wug tests (Berko Gleason, 1958; Bybee, 1985; Prasada & Pinker, 1993; Albright & Hayes, 2003)

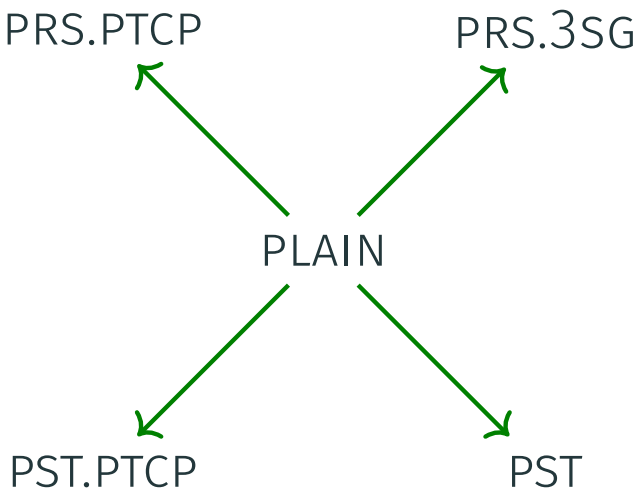
  *I pling every day. Yesterday I also* _____

  - Knowledge of system + form A = form B

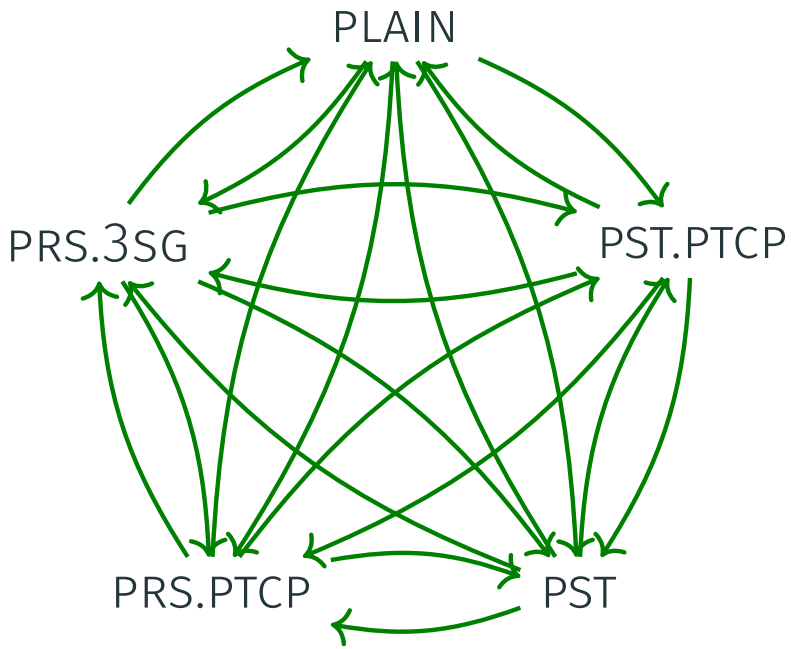
- Largest source of behavioural evidence on the PCFP: the infamous Past Tense Debate

# Blind spots of psycholinguistic approaches to the PCFP

- Speakers are predicting from the base outwards (and we presume to know what the base is)



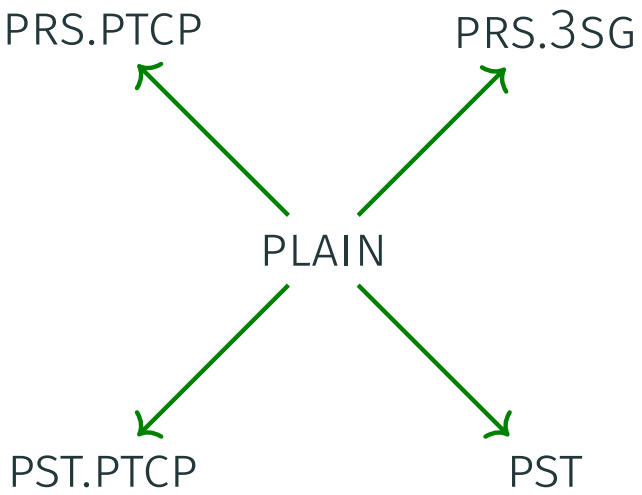

*Base-centric system*



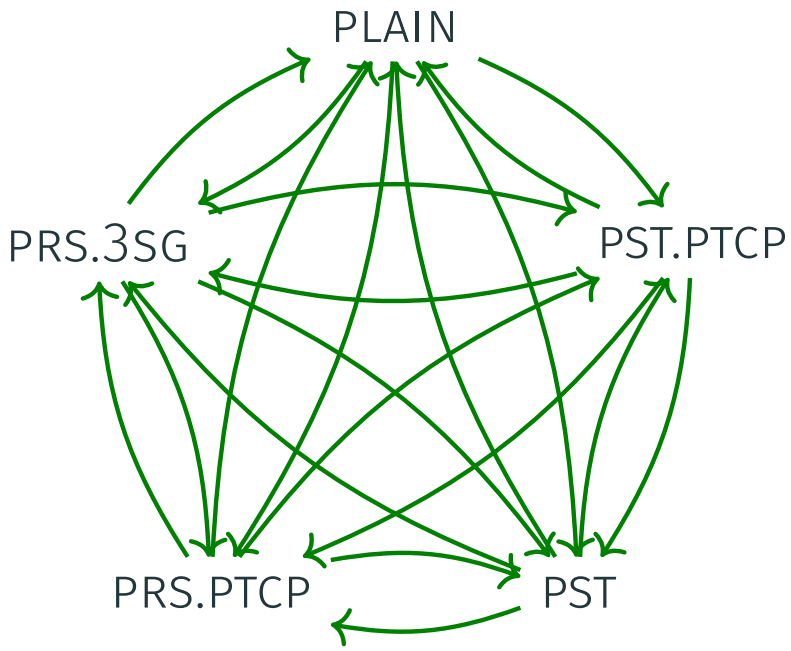

*Paradigmatic system*

# Blind spots of psycholinguistic approaches to the PCFP

- Speakers are predicting from the base outwards (and we presume to know what the base is)




*Base-centric system*




*Paradigmatic system*

- A focus on small, often dyadic, morphological systems
  - little insight about more complex cases

# Linguistic approaches to the PCFP

- Quantifying the information-theoretical difficulty of the PCFP
  - LOW CONDITIONAL ENTROPY CONJECTURE (Ackerman & Malouf, 2013): IT-quantified uncertainty in the system is low on average
    - Bonami & Beniamine, 2016; Sims & Parker, 2016; Beniamine, 2018; Guzmán Naranjo, 2020; Beniamine, Bonami, & Luís, 2021; Pellegrini, 2021; Wilmoth & Mansfield, 2021

# Linguistic approaches to the PCFP

- Quantifying the information-theoretical difficulty of the PCFP
  - LOW CONDITIONAL ENTROPY CONJECTURE (Ackerman & Malouf, 2013): IT-quantified uncertainty in the system is low on average
    - Bonami & Beniamine, 2016; Sims & Parker, 2016; Beniamine, 2018; Guzmán Naranjo, 2020; Beniamine, Bonami, & Luís, 2021; Pellegrini, 2021; Wilmoth & Mansfield, 2021
- Showing that the PCFP is not a learnability hazard
  - Artificial language experiments that the PCFP is learnable (Seyfarth, Ackerman & Malouf, 2014; Johnson et al. 2020)

# Linguistic approaches to the PCFP

- Quantifying the information-theoretical difficulty of the PCFP
  - LOW CONDITIONAL ENTROPY CONJECTURE (Ackerman & Malouf, 2013): IT-quantified uncertainty in the system is low on average
    - Bonami & Beniamine, 2016; Sims & Parker, 2016; Beniamine, 2018; Guzmán Naranjo, 2020; Beniamine, Bonami, & Luís, 2021; Pellegrini, 2021; Wilmoth & Mansfield, 2021
- Showing that the PCFP is not a learnability hazard
  - Artificial language experiments that the PCFP is learnable (Seyfarth, Ackerman & Malouf, 2014; Johnson et al. 2020)
  - Solving the PCFP as a byproduct of acquiring form-meaning mappings
    - Baayen (2011) & Baayen et al. (2019) with linear algebra, Malouf (2017) with deep learning. Ramscar (2021) suggests that the PCFP is not a natural learning task

# Linguistic approaches to the PCFP

- Quantifying the information-theoretical difficulty of the PCFP
  - LOW CONDITIONAL ENTROPY CONJECTURE (Ackerman & Malouf, 2013): IT-quantified uncertainty in the system is low on average
    - Bonami & Beniamine, 2016; Sims & Parker, 2016; Beniamine, 2018; Guzmán Naranjo, 2020; Beniamine, Bonami, & Luís, 2021; Pellegrini, 2021; Wilmoth & Mansfield, 2021
- Showing that the PCFP is not a learnability hazard
  - Artificial language experiments that the PCFP is learnable (Seyfarth, Ackerman & Malouf, 2014; Johnson et al. 2020)
  - Solving the PCFP as a byproduct of acquiring form-meaning mappings
    - Baayen (2011) & Baayen et al. (2019) with linear algebra, Malouf (2017) with deep learning. Ramscar (2021) suggests that the PCFP is not a natural learning task
- These approaches are paradigmatic
- The extent of human-based research is to show that the PCFP is learnable, not how humans behave in the face of the PCFP

# The goal

- Paradigmatically aware work on how speakers engage with the PCFP

# The Minimal Generalisation Learner

- Method to obtain mappings between the base and a cell of interest
- Quantifies how probable is an output given an input
- Both quantitative and behavioural evidence has been gathered thanks to it (Albright & Hayes, 2003; Albright & Hayes, 2002; Albright, 2003; Jun & Albright, 2016)

# The Minimal Generalisation Learner

1. Trained on pairs of forms belonging to two paradigm cells. MGL yields all possible mappings from the first form to the second

   [hæk] → [hækt]  ∅ → t/hæk_

   [dis] → [dist]  ∅ → t/dis_

# The Minimal Generalisation Learner

1. Trained on pairs of forms belonging to two paradigm cells. MGL yields all possible mappings from the first form to the second
   [hæk] → [hækt] ∅ → t/hæk_
   [dis] → [dist] ∅ → t/dis_
2. The obtained rules are compared to each other and are subsumed under generalised versions where possible
   ∅ → t/hæk_
   ∅ → t/dis_
   ∅ → t/[-voiced]_

# The Minimal Generalisation Learner

1. Trained on pairs of forms belonging to two paradigm cells. MGL yields all possible mappings from the first form to the second
   [hæk] → [hækt] ∅ → t/hæk_
   [dis] → [dist] ∅ → t/dis_
2. The obtained rules are compared to each other and are subsumed under generalised versions where possible
   ∅ → t/hæk_
   ∅ → t/dis_
   ∅ → t/[-voiced]_
3. Output: a set of rules with different degrees of specificity – a given input form will usually have more than one applicable rule.

# The Minimal Generalisation Learner

- For each rule, the model calculates its confidence score

$$\frac{\textit{affected lexemes}}{\textit{potentially affected lexemes}} - \textit{uncertainty penalty}$$

$$\propto P(\textit{target}|\textit{input})$$

- These scores have been repeatedly correlated with speaker behaviour (Albright & Hayes, 2003; Jun & Albright, 2016)

# Research questions

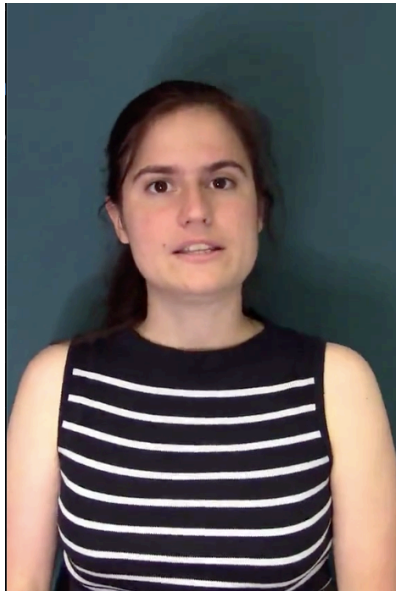
- Do individuals make use of individual paradigmatic implicative relationships in language use?
  - Is prediction from a base form only (as claimed by Jun & Albright (2016)) or does form predictability matter omnidirectionally?
  - Does familiarity with pattern distribution matter?
  - Does predictability truly operate paradigmatically? This is only testable with a larger paradigmatic system

# Methodology

# Acceptability judgement task

- A paradigmatic task
  - Comprehension parallel to wug task
  - Variables and items are constructed through a paradigmatically-informed perspective
- French verbal system (51 cells, rampant stem allomorphy)

# Acceptability judgement task



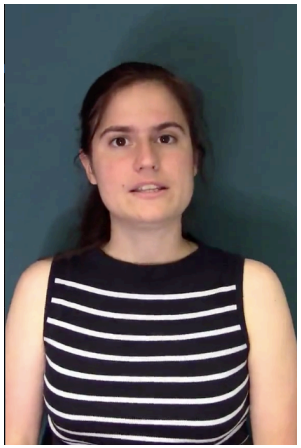
”Nous édrilons le quiz de culture générale presque tous les ans.
C’est Pierre qui l’a édrili l’anné dernière.”

*We* PRS.1PL *the pop culture quiz almost every year.*
*It’s Pierre who* PST.PTCP.M.SG *it this year*

---

Thanks to Julie Bauné and Alice Bruguier for the videos

# Acceptability judgement task



Est-ce que le deuxième mot sonne bien[1] en tant que mot inventé dans ce contexte ?

*Does the second word sound good in this context?*


[1]Participants were given instructions and examples about what it meant to ”sound good” - emphasis on relation, possibility, subjectivity, using the full scale

# The items - pseudolexemes

- Pseudolexemes based on the French verbal system
  - made with Wuggy (Keuleers & Brysbaert, 2010), to match phonology of items belonging to each inflectional class

# The items - cells

- The INFINITIVE is thought to be the base of the French verbal system
  - most frequent cell
  - citation form
  - on average, best predictor of the rest of the paradigm (Bonami & Boyé, 2014)
  - non-finite form and relatively unmarked
- Two cell pairs, bidirectional prediction
  - The cells differ in frequency, finiteness, morphosemantic properties and base status

| Predictor → Target |
|---|
| INFINITIVE → IMPERFECT INDICATIVE 2PL |
| IMPERFECT INDICATIVE 2PL → INFINITIVE |
| PAST PARTICIPLE MASC. SING. → PRESENT INDICATIVE 1PL |
| PRESENT INDICATIVE 1PL → PAST PARTICIPLE MASC. SING. |

# The items - levels of predictability

**Three possible levels of predictability** for each item. For any given item, the participant is shown just one of the versions below.

Nous **édrilons** le quiz de culture générale presque tous les ans.

C'est Pierre qui l'a { *édrilé* / *édrili* / *édrilu* } l'année dernière.

We **IND.PRS.1PL** the general culture quiz almost every year.

Pierre { **PST.PTCP.M.SG-1** / **PST.PTCP.M.SG-2** / **PST.PTCP.M.SG-3** } it last year.

# Hypothesis

1. The more predictable the second form is from the first, the better it will be rated.
    - If speakers use the distributional information inherent in the implicative relationships set up by the paradigm, this will hold true...
        - For all directions of prediction
        - For all cell pairs

# Hypothesis

1. The more predictable the second form is from the first, the better it will be rated.
    - If speakers use the distributional information inherent in the implicative relationships set up by the paradigm, this will hold true...
        - For all directions of prediction
        - For all cell pairs
2. We expect variability between cells.
    - The cell pairs chosen differ in
        - Whether they involve the supposed base
        - Their frequency
        - Their morphosemantic properties

# Quantifying predictability

- To quantify the predictability of the second form conditional on the first, we use the MGL scores.
  - The MGL was trained on the cell pairs of interest from data in Flexique (Bonami, Caron & Plancq, 2014)
  - We then input the pseudolexeme pairs to obtain confidence scores for pairing

# Quantifying predictability

- To quantify the predictability of the second form conditional on the first, we use the MGL scores.
  - The MGL was trained on the cell pairs of interest from data in Flexique (Bonami, Caron & Plancq, 2014)
  - We then input the pseudolexeme pairs to obtain confidence scores for pairing
- The analysis was attempted with several operationalisations of predictability
  - Surprisal (Beniamine, 2018)
  - Neural network (Calderone, Hathout & Bonami, 2021)
  - Transforming MGL scores into
    - Probabilities (softmax of different temperatures)
    - Log odds
- The results remain the same. We choose to report MGL-based results for continuity with previous work.

# Phonological well-formedness judgements

- Must partial out the naturalness of the pseudolexeme
- A different set of participants was asked to provide phonological well-formedness judgements on the target forms.
- 20 well-formedness judgements for each target form, averaged into a phonological well-formedness score for the word

# Analysis

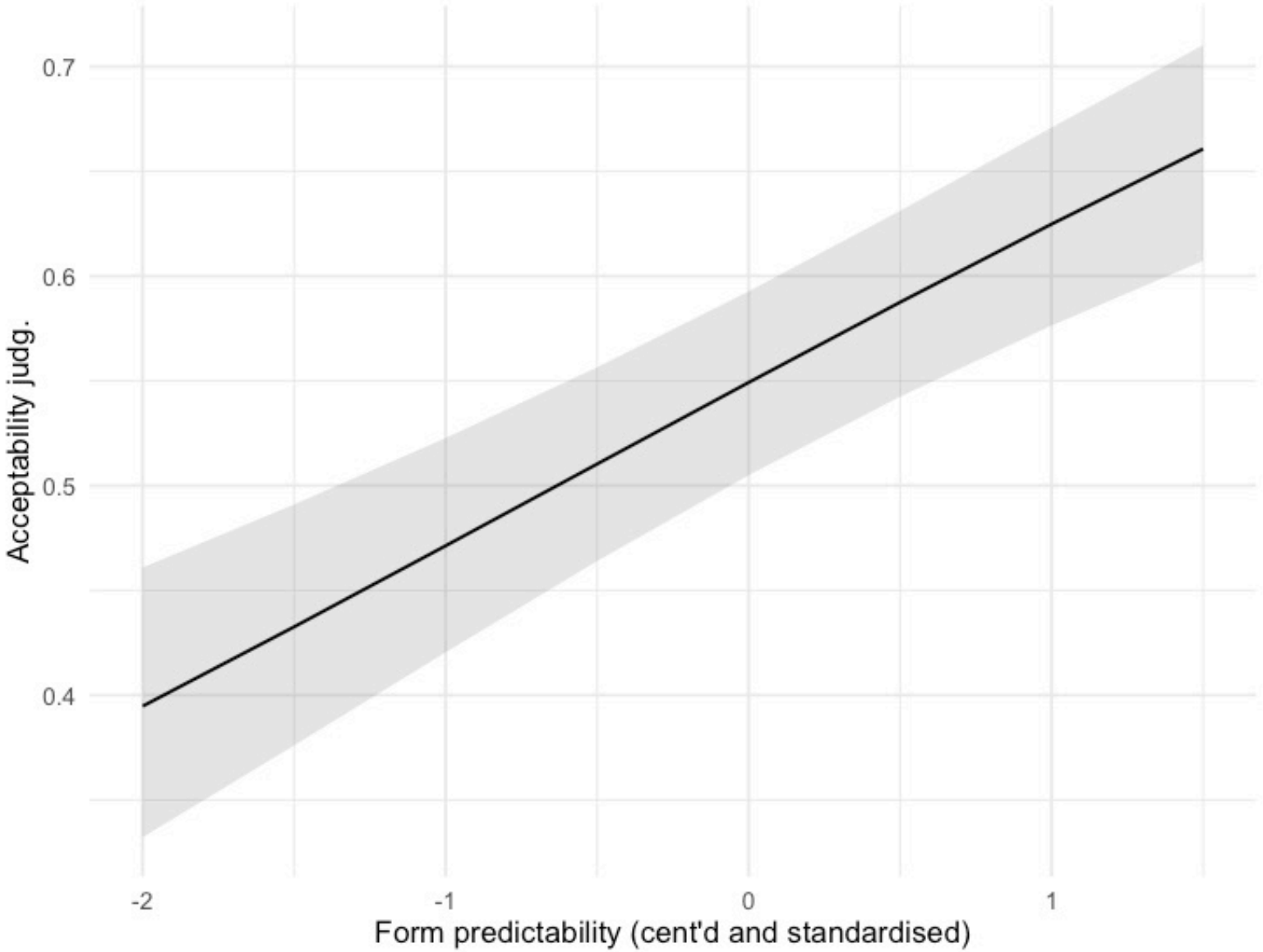
Predict acceptability judgement of the target form from...

- MGL predictability score of the target form given the predictor
- well-formedness judgement
- directed cell pair

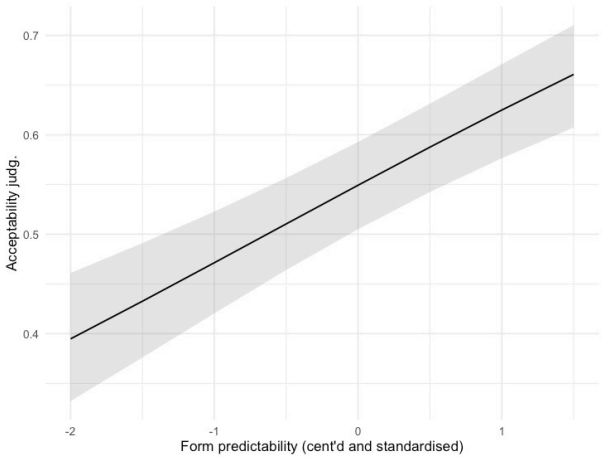
A maximal model with random intercepts for item and participant fitting a beta distribution.

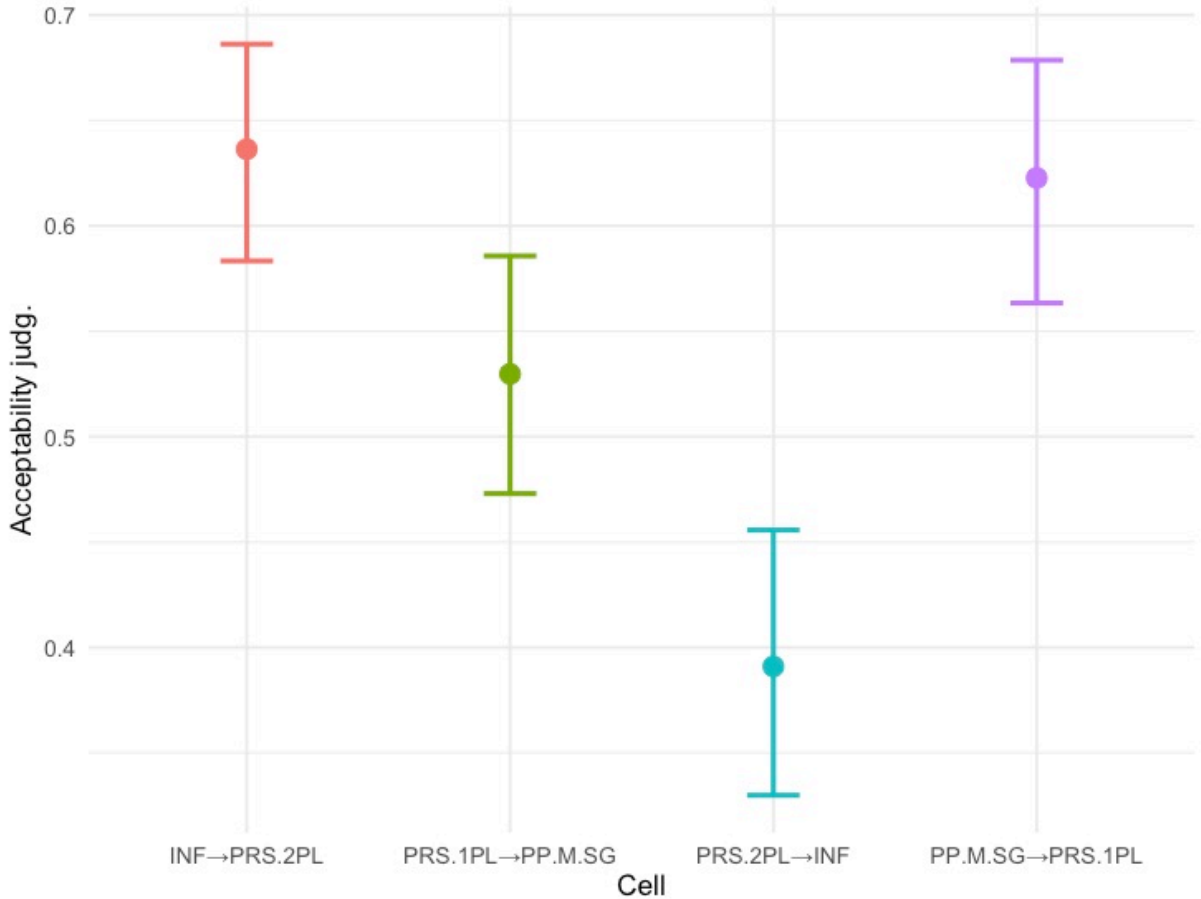
$$judgment \sim MGL\ score * cell + wellformedness + (MGL\ score * cell + wellformedness|participant) + (MGL\ score|item)$$

# Results - form predictability

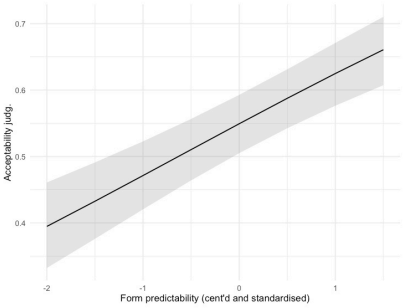


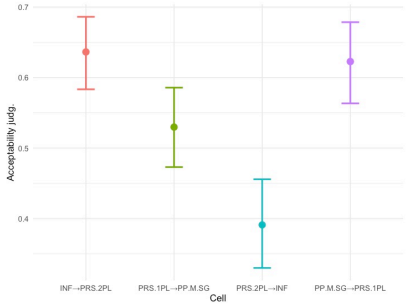
# Results - pairs of cells

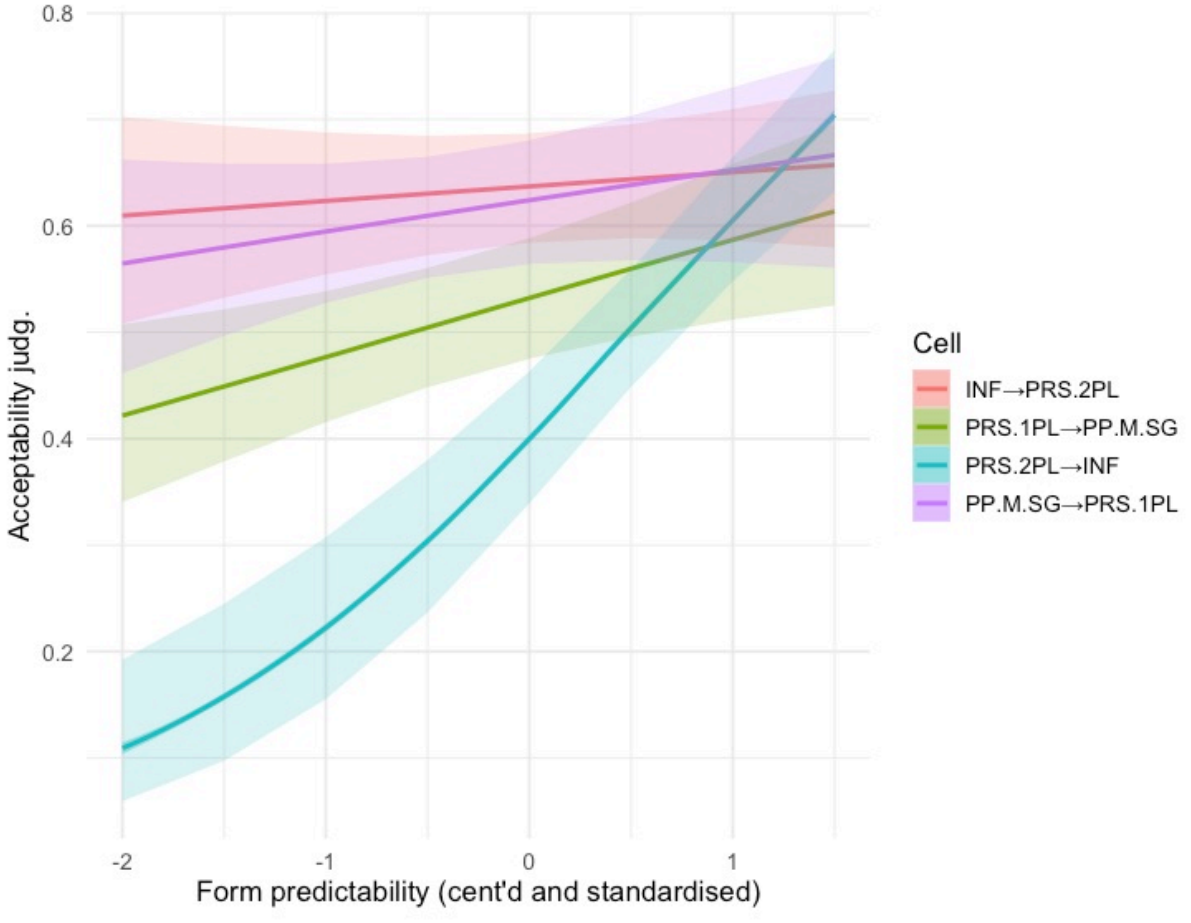


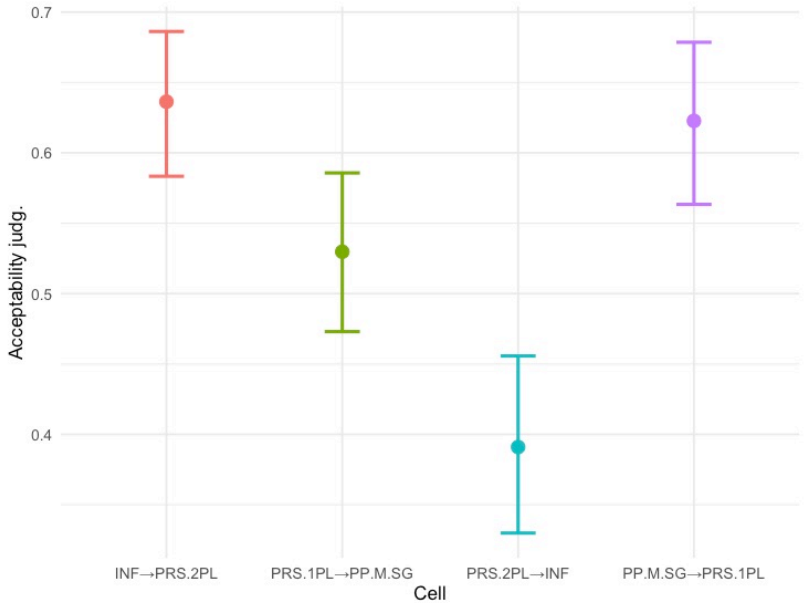
# Results - interaction

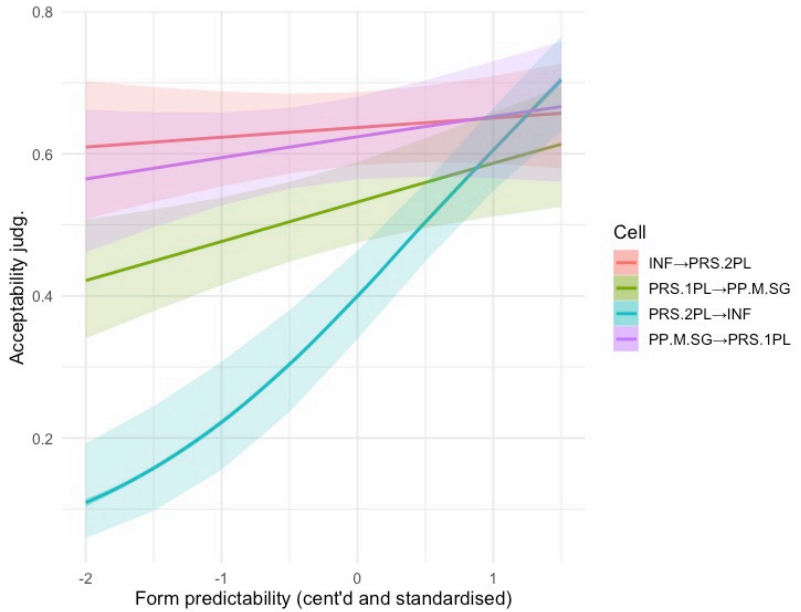


# An effect of cell frequency?

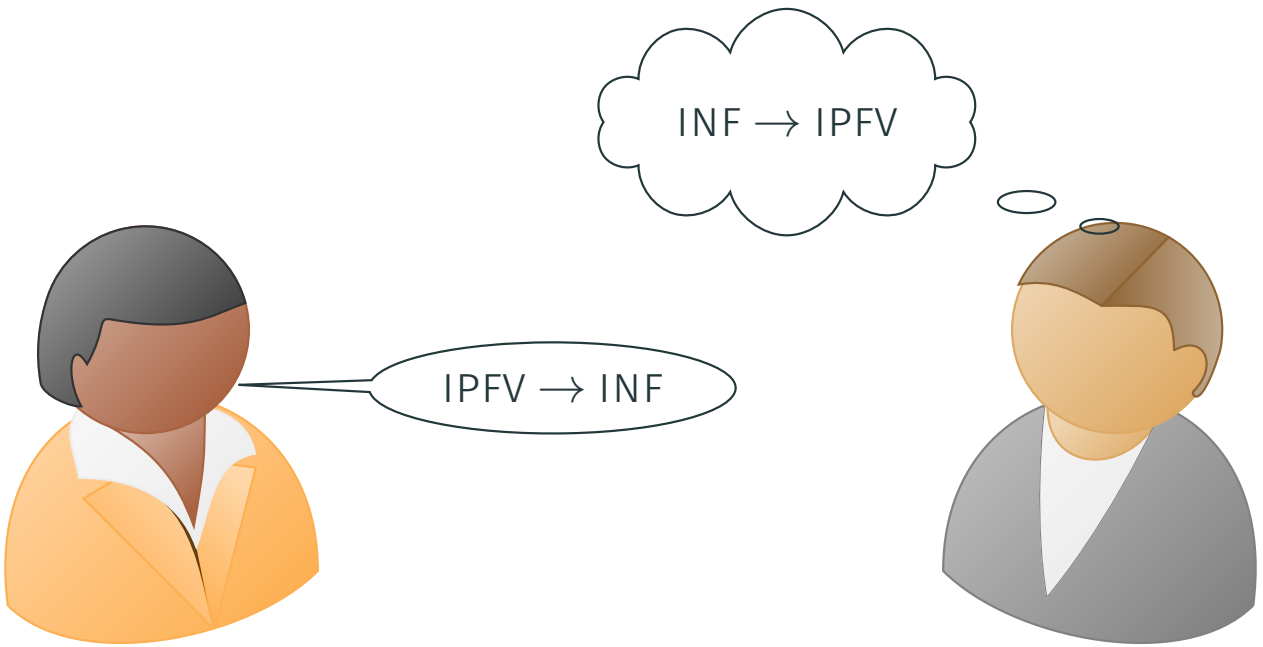



- The more frequent the predicted cell is…
  - The lower the average score
    Speakers are more tolerant in cells they are less familiar with?
  - The more extreme the effect of predictability
    Violation of stronger expectations is penalised more harshly?

# Direction of prediction

- Jun & Albright (2016) claim that speakers predict from the base, even when making predictions towards it.



# Direction of prediction

- The INFINITIVE is thought to be the base in the French verbal system.
- Trained two models only on IPFV.2PL → INF data. Same model structure, MGL scores substituted.

| MGL scores | LOO-CV error | AIC | MSE |
|---|---|---|---|
| IPFV.2PL → INF | 0.355 | -371 | 0.12 |
| INF → IPFV.2PL | 0.398 | -348 | 0.14 |

# Direction of prediction

- The INFINITIVE is thought to be the base in the French verbal system.
- Trained two models only on IPFV.2PL → INF data. Same model structure, MGL scores substituted.

| MGL scores | LOO-CV error | AIC | MSE |
|---|---|---|---|
| IPFV.2PL → INF | 0.355 | -371 | 0.12 |
| INF → IPFV.2PL | 0.398 | -348 | 0.14 |

- Speakers are using knowledge about implicative relations in the required direction
- Some cases considered by our experiment did not involve the base at all

# Conclusion

- Speakers make use of paradigmatic implicative relationships
  - In all directions of prediction
  - Without necessarily relying on the base/citation form
  - In larger paradigmatic systems
- A potential effect of cell frequency
  - Participants are more generous when scoring forms in less frequent cells
    - less familiarity with possible patterns, more tolerance
  - Form predictability matters more when predicting towards frequent cells
    - Speakers have clearer expectations, violations are penalised more harshly

# Appendix

# Results

| | Estimate | Std. Error | z value | Pr(>\|z\|) | |
|---|---|---|---|---|---|
| (Intercept) | 0.21 | 0.09 | 2.27 | 0.02 | * |
| MGL score | 0.31 | 0.05 | 6.11 | <0.001 | *** |
| PRS.1PL→PP.M.SG[2] | -0.43 | 0.12 | -3.57 | <0.001 | *** |
| PRS.2PL→INF[2] | -0.97 | 0.14 | -7.05 | <0.001 | *** |
| PP.M.SG→PRS.1PL[2] | -0.06 | 0.13 | -0.44 | 0.66 | |
| phon. wellformedness | 0.17 | 0.10 | 1.80 | 0.07 | . |
| MGL score:PRS.1PL→PP.M.SG[2] | 0.16 | 0.12 | 1.41 | 0.16 | |
| MGL score:PRS.2PL→INF[2] | 0.79 | 0.15 | 5.21 | <0.001 | *** |
| MGL score:PP.M.SG→PRS.1PL[2] | 0.06 | 0.14 | 0.48 | 0.63 | |

[2]The directed cell variable is sum-coded, with INF→PRS.2PL as baseline

# Sample items

| Condition | Attested example | | | Sample pseudoword pair | | |
|---|---|---|---|---|---|---|
| | Lexeme | Predictor | Target | Predictor | Target | Score |
| INF ↓ IPFV.2PL | VENIR ‘come’ | vəniʁ | vənje | pɥistəniʁ | pɥistənje | 0.96 |
| | AMORTIR ‘damp’ | amɔʁtiʁ | amɔʁtisje | saʁniʁ | saʁnisje | 0.90 |
| | ENSEVELIR ‘bury’ | ɑ̃səvəliʁ | ɑ̃səvəlisje | ʃeləniʁ | ʃeləsisje | 0.70 |
| | SORTIR ‘exit’ | sɔʁtiʁ | sɔʁtje | dekɑ̃fiʁ | dekɑ̃ffje | 0.09 |
| IPFV.2PL ↓ INF | TAPIR ‘hide’ | tapisje | tapiʁ | plasisje | plasiʁ | 0.96 |
| | TAPISSER ‘paper’ | tapisje | tapise | nɔlvisje | nɔlvise | 0.69 |
| | *unattested* | | | klɑ̃sisje | klɑ̃se | 0.00 |
| PST.PTCP ↓ PRS.1PL | BLÊMIR ‘go pale’ | blemi | blemisɔ̃ | vemi | vemisɔ̃ | 0.73 |
| | DORMIR ‘sleep’ | dɔʁmi | dɔʁmɔ̃ | plomi | plomɔ̃ | 0.05 |
| PRS.1PL ↓ PST.PTCP | LAVER ‘wash’ | lavɔ̃ | lave | lanɔ̃ | lane | 0.97 |
| | VENIR ‘come’ | vənɔ̃ | vəny | ɛgzisənɔ̃ | ɛgzisəny | 0.34 |
| | BATTRE ‘beat’ | batɔ̃ | baty | ɛspatɔ̃ | ɛspaty | 0.01 |
| | SUIVRE ‘follow’ | sɥivɔ̃ | sɥivi | kʁiʒɔ̃ | kʁiʒi | 0.004 |

# Procedure

60 French native speakers from Prolific.co

⇓

Given instructions for the task

⇓

24 crucial items + 24 distractors in a randomised order

Crucial items uniformly distributed between the four cell pairs and the three levels of predictability