

# L'approche Bayésienne, une approche statistique alternative en psycholinguistique\*

Cyril Perret

Semaine des DATA SHS

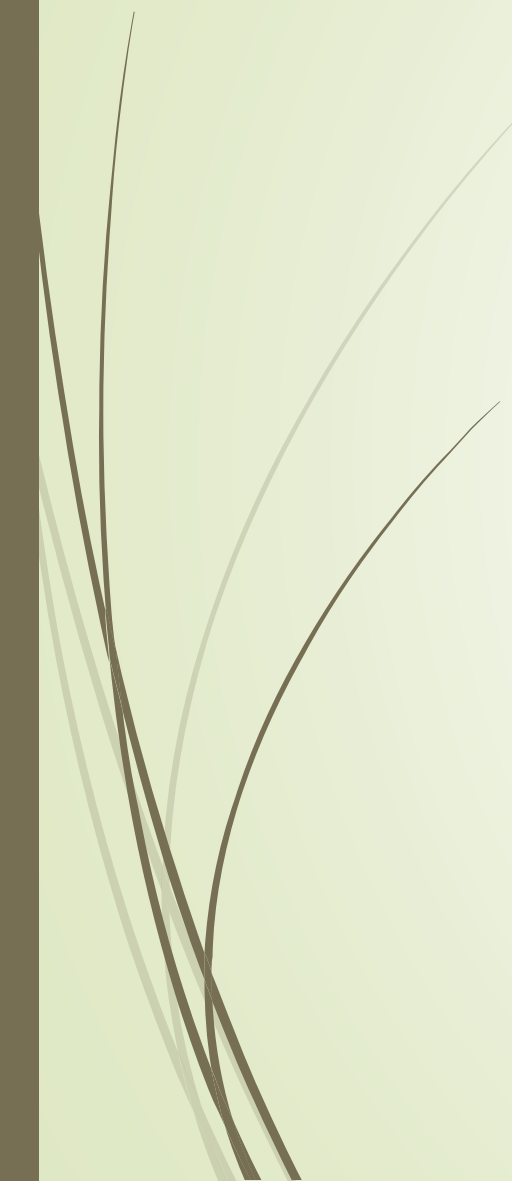
Données et méthodes quantitatives en Sciences Humaines et Sociales

7 Décembre 2022

\*Perret & Solier (2022) *Journal of Cognitive Psychology*  
Données et scripts disponibles sur OSF



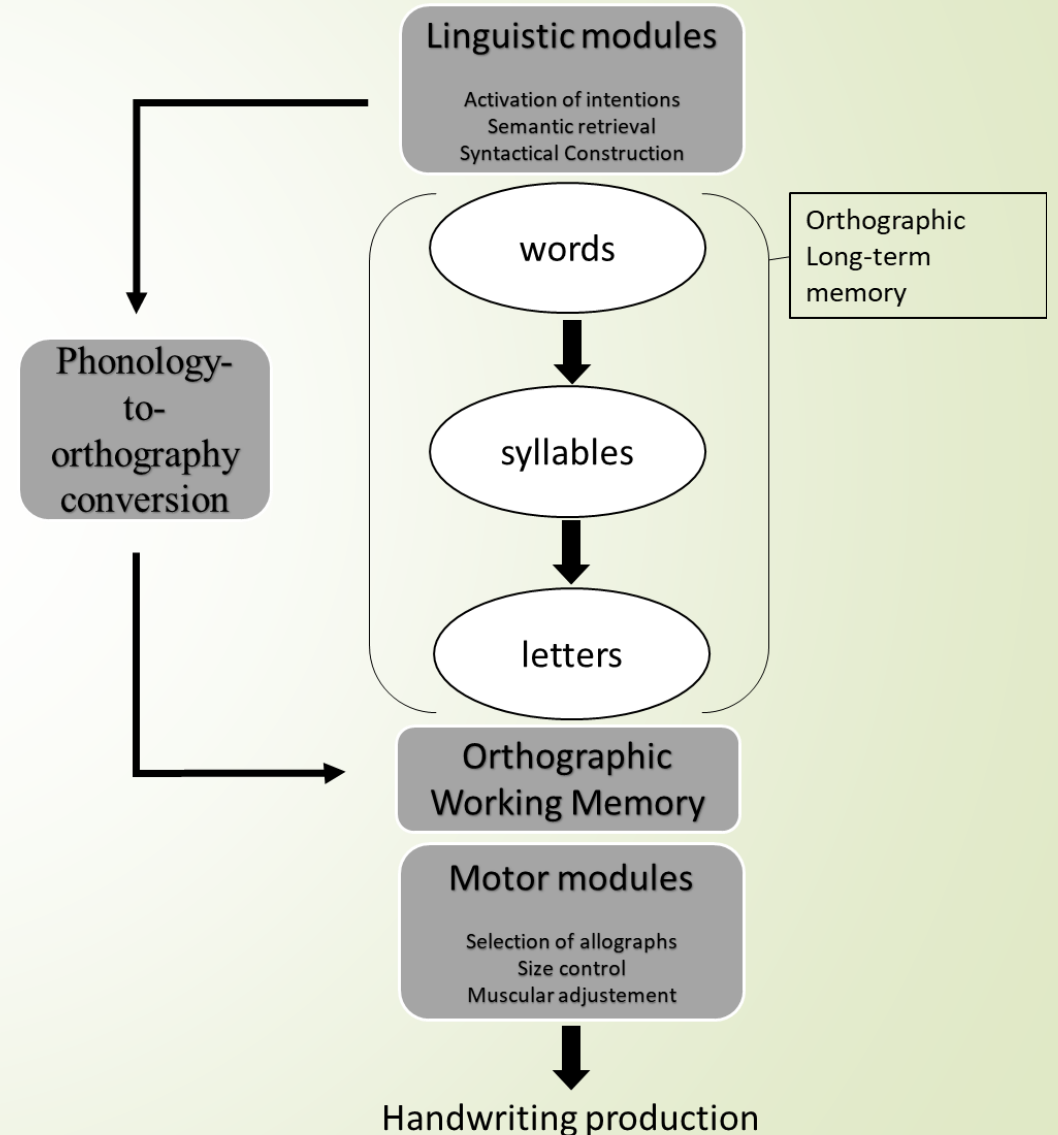
# Plan

- La syllabe : quelle est la question posée ?
  - Le problème de l'approche statistique traditionnelle
  - Le modèle Bayésien
- 

# Modèle de Kandel (2022)

- La syllabe : quelle est la question posée ?
- Le problème de l'approche statistique traditionnelle
- Le modèle Bayésien

- Un syllabaire (lexique mental de syllabes)
- Accès aux syllabes orthographiques durant l'accès à la représentation orthographique
- Modèle qui rend compte des effets rapportés dans la littérature
- **Si syllabaire alors effet de fréquence des syllabes**



# La syllabe en production manuscrite

- La syllabe : quelle est la question posée ?

- Le problème de l'approche statistique traditionnelle

- Le modèle Bayésien

- Arguments en faveur d'un rôle fonctionnel de la syllabe

- Patients avec atteinte de la mémoire de travail orthographique (Carmazza & Miceli, 1990)

- Simplification de la structure syllabique -> presso -> pesso

- Echange qui respect l'appartenance syllabique: tornava -> tronava

- Sauf que pas répliquer (Jónsdóttir et al., 1996; McCloskey et al., 1994; Ward & Romani, 2000)

- Et, aussi, alternative au syllabaire...

# Notre étude

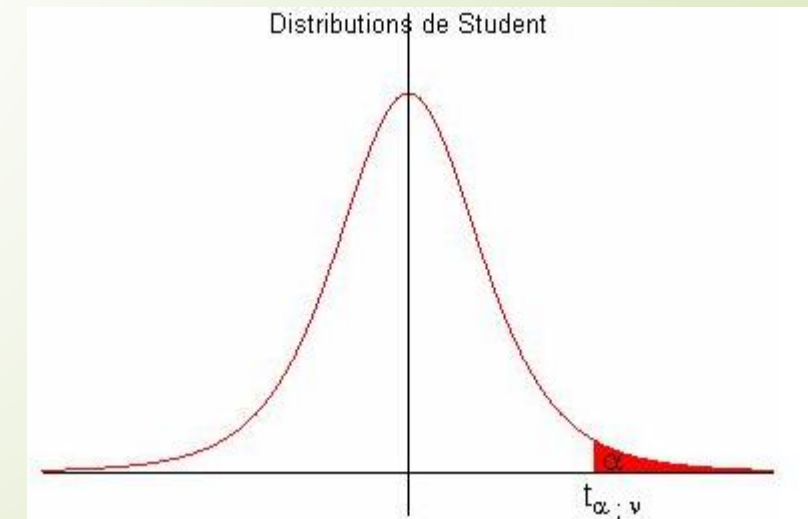
- La syllabe : quelle est la question posée ?
- Le problème de l'approche statistique traditionnelle
- Le modèle Bayésien

- 30 étudiants de L3 de l'université de Neuchâtel (SUISSE)
- 150 items
  - Huit facteurs (NA, IA, Fam, Ivar, AoA, VC, Freq, Nombre de lettres, Perret & Bonin, 2019)
  - Fréquence des syllabes
    - moyenne de fréquence (Perret et al., 2014)
    - Fréquence de la première syllabe
- Tâche de dénomination d'images
- Latences d'initialisation

# La statistique Inférentielle fréquentiste

- La syllabe : quelle est la question posée ?
- Le problème de l'approche statistique traditionnelle
- Le modèle Bayésien

- Deux conditions expérimentales  $\mu_1$  et  $\mu_2$
- Test d'hypothèse
  - $H_0 \mu_1 = \mu_2$
  - $H_1 \mu_1 \neq \mu_2$
- Calcul d'un indice (t-test  $\frac{\mu_1 - \mu_2}{\sigma_{1,2}}$ )
- Ensuite, on regarde une loi de distribution de probabilité
- Cela nous donne la p-value
- Il s'agit de  $P(H_0 | \text{données})$  soit la probabilité que  $H_0$  soit « vrai » sachant les données



# Le modèle Bayésien

- La syllabe : quelle est la question posée ?
- Le problème de l'approche statistique traditionnelle
- Le modèle Bayésien

- Exemple du bruit dans le grenier...
- Estimation de probabilités a priori
- Modification de ces probabilités => a posteriori
- Rapport de probabilité entre posteriori de  $H_0$  et posteriori de  $H_1$

$$BF_{10} = \frac{P(H_1 | \text{données})}{P(H_0 | \text{données})}$$

# Résultats : étape 1

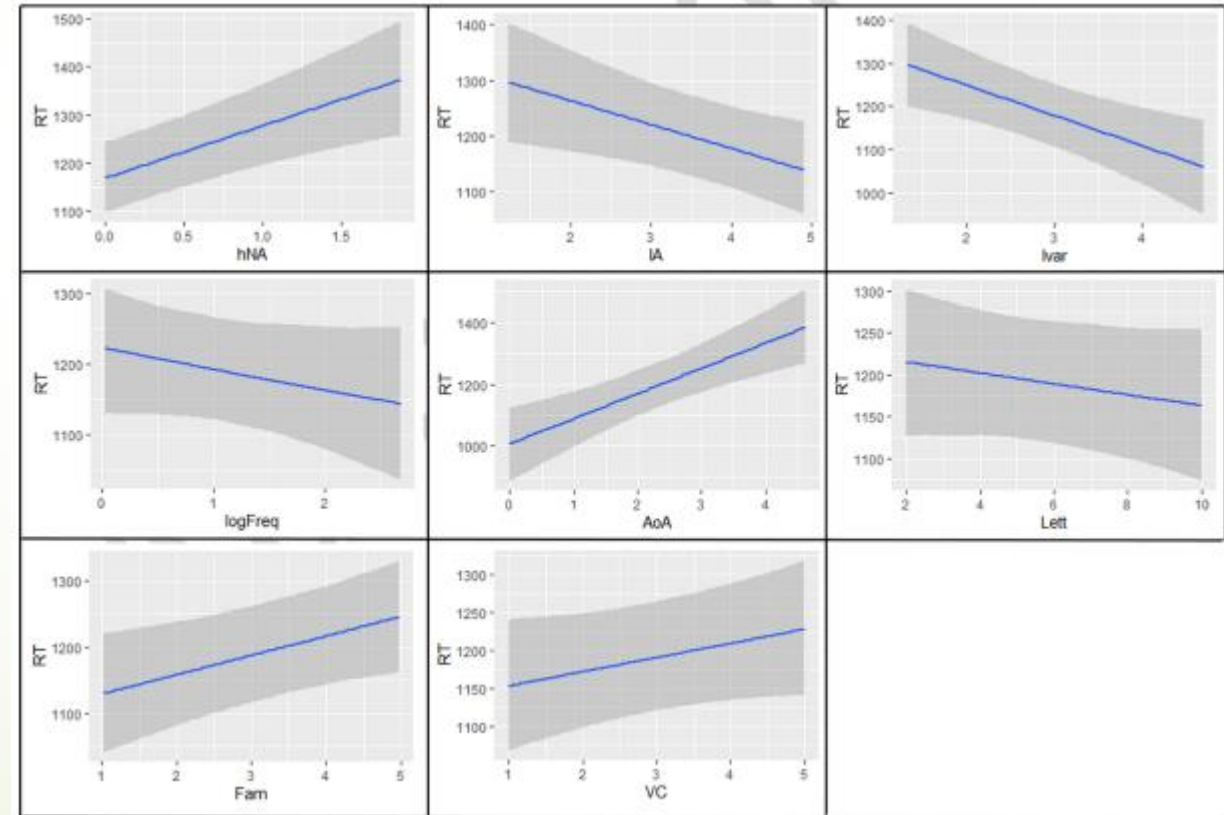
- La syllabe : quelle est la question posée ?
- Le problème de l'approche statistique traditionnelle
- Le modèle Bayésien

## Modèles de régression linéaire simultanée

$BF_{10} = 3,05 * 10^{12}$

$R^2 \text{ Bayésien}^* = .4186$

|               | Estimate      | Est.Error    | l-95% CI       | u-95% CI      | $BF_{10}$        | $BF_{hyp}$  |
|---------------|---------------|--------------|----------------|---------------|------------------|-------------|
| Intercept     | 1264,6        | 131,27       | 1000,4         | 1521,23       | $2,07 * 10^{16}$ | $+\infty$   |
| <b>NA (h)</b> | <b>108,63</b> | <b>29,39</b> | <b>51,5</b>    | <b>165,89</b> | <b>192,99</b>    | <b>7999</b> |
| <b>IA</b>     | <b>-42,74</b> | <b>16,71</b> | <b>-75,37</b>  | <b>-10,23</b> | <b>4,81</b>      | <b>199</b>  |
| <b>Ivar</b>   | <b>69,96</b>  | <b>22,36</b> | <b>-115,08</b> | <b>-26,44</b> | <b>45,63</b>     | <b>1332</b> |
| <b>AoA</b>    | <b>81,8</b>   | <b>21,03</b> | <b>41,06</b>   | <b>123,38</b> | <b>359,8</b>     | $+\infty$   |
| LogFreq       | -29,1         | 25,43        | -79,79         | 20,15         | 0,42(2,38)       |             |
| NbLett        | -6,48         | 6,75         | -19,66         | 7,02          | 0,09 (11,11)     |             |
| Fam           | 28,93         | 12,77        | 3,58           | 53,95         | 1,59             |             |
| VC            | 18,88         | 12,85        | -6,98          | 43,89         | 0,36 (2,56)      |             |





# Résultats : étape 2

- La syllabe : quelle est la question posée ?
- Le problème de l'approche statistique traditionnelle
- Le modèle Bayésien

- Effets de fréquence des syllabes
- Fréquence moyenne
  - $\beta = -16,98$  [sd = 13.98, 95%-Credible Interval {-43.38, 8.66}]
  - $BF_{10} = 0.34871$  (2,87)
  - La vraisemblance d'observer les données récoltées dans l'expérience est 2,87 fois plus importante au regard de  $H_0$  qu'au regard de  $H_1$
- Fréquence de la première syllabe
  - $\beta = -13,26$  [sd = 13.26, 95%-Credible Interval {-34,55, 7.94}]
  - $BF_{10} = 0.00018$  (5556)
  - La vraisemblance d'observer les données récoltées dans l'expérience est 5556 fois plus importante au regard de  $H_0$  qu'au regard de  $H_1$
- On peut conclure que ce facteur n'influence pas les temps de réaction.



Merci pour votre attention