

Comment tester expérimentalement ses hypothèses théoriques?

L'exemple du jugement d'acceptabilité

Céline Pozniak

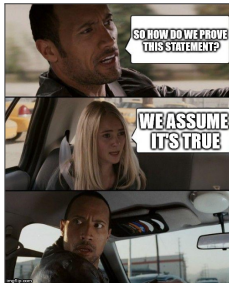
27 mai 2024





Les théories en linguistique formelle se sont souvent appuyées sur des méthodes non quantitatives :

- ▶ intuitions venant de la personne qui propose la théorie
- ▶ jugements venant de quelques collègues de la personne (~2, 3)
- ▶ portant sur très peu de stimuli linguistiques (~1)



⇒ **Biais cognitifs + Biais sur les stimuli**



Pourquoi avoir besoin de méthodes quantitatives ?

Comment faire ?

Le jugement

Take Home message



Pourquoi avoir besoin de méthodes quantitatives ?

Pourquoi des méthodes quantitatives ?

Les données et les intuitions sont-elles redondantes entre elles ?

Comment faire ?

Le jugement

Jugement : tâche pertinente ?

Quel type de jugement ?

Quelle échelle ?

Temps limité ou pas ?

Take Home message



Pourquoi avoir besoin de méthodes quantitatives ?

Pourquoi des méthodes quantitatives ?

Les données et les intuitions sont-elles redondantes entre elles ?

Comment faire ?

Le jugement

Jugement : tâche pertinente ?

Quel type de jugement ?

Quelle échelle ?

Temps limité ou pas ?

Take Home message







1. La linguistique serait différente des autres domaines (psycholinguistique) et les méthodes quantitatives ne seraient pas appropriées
2. Ce que l'on obtient en quantitatif est souvent le reflet de nos intuitions (Phillips, 2009)



Arguments contre les méthodes quantitatives

Gibson and Fedorenko (2013)



1. *Rassembler beaucoup de données est généralement difficile* 
 - ↳ Pour beaucoup de langues, non + possibilité de compenser le nombre de sujets avec le nombre d'items et inversement
2. *Dans d'autres domaines, on accepte des résultats avec peu de participant·es* 
 - ↳ Souvent compensé par le nombre d'items + pas des processus cognitifs de haut niveau
3. *Les données ne permettent pas tout le temps de répondre à une question de recherche* 
 - ↳ Il y a tout de même un contrôle des facteurs confondus contrairement à l'intuition
4. *Cela n'est pas suffisant et utiliser ces méthodes va seulement ralentir la recherche* 
 - ↳ Avoir des théories appuyées par des données permet justement d'avoir une véritable base scientifique sur laquelle avancer



5. *Les linguistes sont mieux que les participant-es pour évaluer de stimuli linguistiques* 🙄
 - ↳ Non car biais de confirmation de la part de la chercheuse mais aussi de la participante
6. *Pas besoin d'expérience, on peut juste échanger entre collègues* 🙄
 - ↳ Effet clever Hans + Conformité sociale
7. *Continuer de travailler sans des méthodes quantitatives ne fait pas de mal* 🙄
 - ↳ La méthodologie traditionnelle non quantitative est problématique et éloigne l'intérêt de la linguistique formelle par d'autres disciplines + développement de théories basé sur des hypothèses infirmées par les données



Pourquoi avoir besoin de méthodes quantitatives ?

Pourquoi des méthodes quantitatives ?

Les données et les intuitions sont-elles redondantes entre elles ?

Comment faire ?

Le jugement

Jugement : tâche pertinente ?

Quel type de jugement ?

Quelle échelle ?

Temps limité ou pas ?

Take Home message

Données = Intuitions dans tous les cas ?



- ▶ De manière générale, les théories vont faire des prédictions sur l'acceptabilité et/ou la facilité de traitement de certaines phrases
- ▶ Cependant, les phrases ne sont jamais testées et sont soutenues principalement par des intuitions...
- ▶ ... qui ne sont finalement pas vérifiées en utilisant des méthodes quantitatives

Intuitions = réalité ?

(Gibson, 1991)



(1) *Relative modifiant le sujet de la phrase racine*

The man [that the woman [that the dog bit] likes] likes fish.

(2) *Relative modifiant l'objet de la phrase racine*

I saw **the man** that [the woman [that the dog bit] likes].

- ▶ 1 est supposé être plus difficile que 2 car il y a moins de relations de dépendance non résolues (Miller, 1963; Yngve, 1960)
- ▶ Mais qu'en est-il lorsque l'on teste avec des méthodes quantitatives ?



- ▶ Données de jugements d'acceptabilité ne montrent pas de différences entre les deux constructions (Gibson and Thomas, 1996)
 - ▶ Tâche de self-paced testant la différence entre relatives modifiant le sujet de la phrase racine et relatives modifiant l'objet de la phrase racine (Gibson et al., 2005) (3-4)
- (3) The reporter who the senator attacked on Tuesday ignored the president
 - (4) The president ignored the reporter who the senator attacked on Tuesday

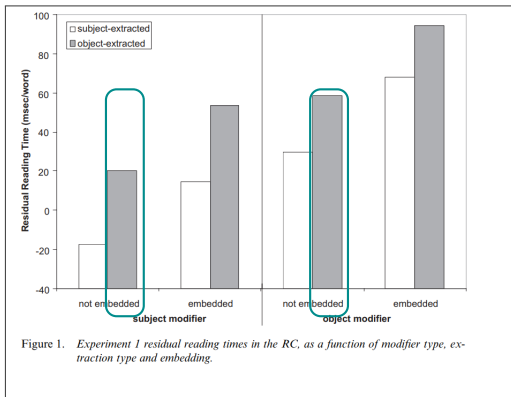


Figure 1. Experiment 1 residual reading times in the RC, as a function of modifier type, extraction type and embedding.

- ▶ Les relatives modifiant le sujet sont en effet plus rapides à traiter que les relatives modifiant l'objet, contra Miller (1963) ou encore Yngve (1960)



- ▶ Même si partir des intuitions est important, les résultats obtenus ne sont pas nécessairement généralisables (Schütze, 2016)
- ▶ Afin de faire avancer les théories en linguistique, il est nécessaire d'adopter d'autres méthodes plus quantitatives ...
- ▶ ... Comme des expériences



Pourquoi avoir besoin de méthodes quantitatives ?

Pourquoi des méthodes quantitatives ?

Les données et les intuitions sont-elles redondantes entre elles ?

Comment faire ?

Le jugement

Jugement : tâche pertinente ?

Quel type de jugement ?

Quelle échelle ?

Temps limité ou pas ?

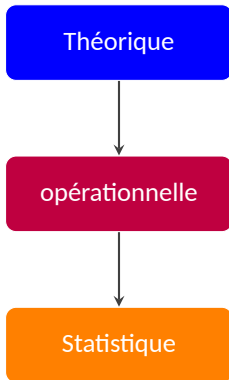
Take Home message



1. Formule clairement ta question de recherche
2. Contrôle la variabilité autant que possible
3. Aie suffisamment de puissance
4. Tiens compte de tes données, et pas juste des analyses statistiques

Formuler sa question de recherche

Déclinaison en hypothèse(s)





Variable dépendante (VD) ~ Variable(s) indépendante(s) (VI)

VD : que
voulons-nous
observer pré-
cisément ?

VI : bien préci-
ser les modali-
tés de ce que
l'on manipule



Variable dépendante (VD) \sim Variable(s) indépendante(s) (VI) + bruit(s)



► Variables parasites

↳ Variables non directement testées dans l'expérience et qui peuvent avoir une action sur le comportement observé

↳ Il va falloir contrôler ces variables pour éviter qu'elles aient un effet sur la VD ou du moins pour minimiser son effet

► Variabilité extérieure

↳ Tout doit rester constant

↳ Il va falloir réduire le bruit expérimental autant que possible (environnement, instructions, ordre passation, etc)



- ▶ Avoir pas mal de données !
 - Suffisamment de participant-es
 - Suffisamment d'items
- ▶ Loi des grands nombres
- ▶ Combien exactement ? Tests d'analyse de puissance (*power analysis*)

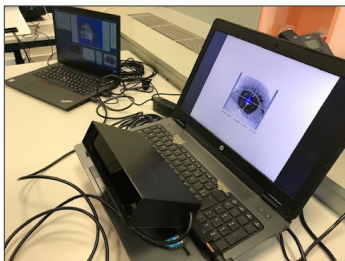


- ▶ Importance de la visualisation descriptive des données
- ▶ Les analyses inférentielles évoluent, mais pas vos données récoltées
- ▶ Stop hunting p-value :)

À partir de là, quelles tâches ?



Online



Offline



À partir de là, quelles tâches ?



progress

Les résultats des matchs de la Russie pour les qualifications de la Coupe du Monde sont diffusés.

(Mauvais)

1

2

3

4

5

6

7

8

9

10

(bon)



Pourquoi avoir besoin de méthodes quantitatives ?

Pourquoi des méthodes quantitatives ?

Les données et les intuitions sont-elles redondantes entre elles ?

Comment faire ?

Le jugement

Jugement : tâche pertinente ?

Quel type de jugement ?

Quelle échelle ?

Temps limité ou pas ?

Take Home message



Pourquoi avoir besoin de méthodes quantitatives ?

Pourquoi des méthodes quantitatives ?

Les données et les intuitions sont-elles redondantes entre elles ?

Comment faire ?

Le jugement

Jugement : tâche pertinente ?

Quel type de jugement ?

Quelle échelle ?

Temps limité ou pas ?

Take Home message

Le jugement : un peu trop méta ?

Gibson and Fedorenko (2013)



- ▶ Critique régulière pertinente : langage comme objet d'évaluation et non moyen de communication (Edelman and Christiansen, 2003)
- ▶ Il est vrai qu'il faut tenir compte de cela et utiliser cette tâche dépendra de votre question de recherche
- ▶ **Première étape** d'un projet de recherche

Avantages du jugement

Gibson and Fedorenko (2013)



- ▶ Simple à mettre en place
- ▶ Résultats *souvent* systématiques entre participant-es
- ▶ Il y a *souvent* une corrélation entre le jugement et d'autres VD



Pourquoi avoir besoin de méthodes quantitatives ?

Pourquoi des méthodes quantitatives ?

Les données et les intuitions sont-elles redondantes entre elles ?

Comment faire ?

Le jugement

Jugement : tâche pertinente ?

Quel type de jugement ?

Quelle échelle ?

Temps limité ou pas ?

Take Home message



- ▶ Demander à un·e participant·e un jugement de grammaticalité
 - ↳ Juger selon la grammaire prescriptive et non selon les intuitions
 - ↳ La·e participant·e aura du mal à séparer les facteurs syntaxiques des autres facteurs (plausibilité, fréquence)

- ▶ Demander à un·e participant·e un jugement d'acceptabilité
 - ↳ Intuitions plus spontanées
 - ↳ Meilleure prise en compte de l'interaction avec d'autres facteurs lors du jugement

Le jugement : grammaticalité ou acceptabilité ?

Huang and Ferreira (2020)



Grammaticalité

⇒ propriété propre au stimulus linguistique

Acceptabilité

⇒ perception du stimulus linguistique par des participant·es naïves et naïfs (Schütze and Sprouse, 2014)



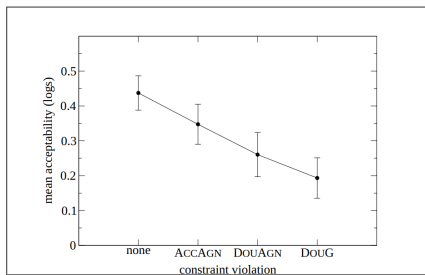
- ▶ Interactions complexes entre (in)acceptabilité et (a)grammaticalité
- ▶ Intermédiaire entre grammaticalité et agrammaticalité, avec des phrases acceptables même si contraire aux règles grammaticales
- ▶ Keller and Alexopoulou (2001) dans Sorace and Keller (2005)

a. I Maria apelise to YANI.
the Maria-NOM fired-3SG the Yanis-ACC
'Maria fi red Yanis.'

b. ? I MARIA apelise to Yani.
the Maria-NOM fired-3SG the Yanis-ACC

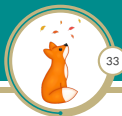
c. ?? To Yani apelise i MARIA.
the Yanis-ACC fired-3SG the Maria-NOM

d. * I Maria ton apelise to YANI.
the Maria-NOM him-CL fired-3SG the Yanis-ACC





- ▶ Les participant·es devaient juger des phrases selon si ils étaient :
 - professeur·es
 - Tuteur·ices
 - Elles-mêmes ou eux-mêmes (condition intuitive)
 - ▶ Le jugement diffère selon la position prise...
- ↳ Notion **continue** de l'acceptabilité ?



Pourquoi avoir besoin de méthodes quantitatives ?

Pourquoi des méthodes quantitatives ?

Les données et les intuitions sont-elles redondantes entre elles ?

Comment faire ?

Le jugement

Jugement : tâche pertinente ?

Quel type de jugement ?

Quelle échelle ?

Temps limité ou pas ?

Take Home message



- ▶ La notion d'acceptabilité recouvre plusieurs facteurs :
 - linguistiques (pragmatique, sémantique, syntaxique)
 - cognitifs (difficulté de traitement etc)
 - sociolinguistiques (rapport à la norme)

- ▶ Nécessité de spécifier ce que vous voulez tester :
 - pour vous (quelle hypothèse, quel(s) effet(s) attendu(s))
 - pour les participant-es (comment allez-vous définir l'acceptabilité)



- ▶ Échelle de Likert
- ▶ Magnitude estimation
- ▶ Échelle coulissante ?

- ▶ Voir [Langsford et al. \(2018\)](#) pour plus de détails sur les mesures



Please rate the acceptability of this sentence.

Which coworker did George yawn before insulting?

Bad Good

- ▶ Peut-être la plus utilisée (Schütze and Sprouse, 2014)
- ▶ 5 points ? 7 points ? 10 points ? Plus ?

Magnitude estimation

Langsford et al. (2018)



If this sentence gets an acceptability rating of one hundred...

Who said that my brother was kept tabs on by the FBI?

...what should these get?

The book is long and the essay is short.

The virtuoso practices any pieces only rarely.

- ▶ Échelle continue (contra Likert)
- ▶ Plus fin ? A priori pas de différence (Gibson and Fedorenko, 2013)



Progression

Le cerceau est attrapé par le voltigeur.

Pas du tout acceptable



Tout à fait acceptable



Pourquoi avoir besoin de méthodes quantitatives ?

Pourquoi des méthodes quantitatives ?

Les données et les intuitions sont-elles redondantes entre elles ?

Comment faire ?

Le jugement

Jugement : tâche pertinente ?

Quel type de jugement ?

Quelle échelle ?

Temps limité ou pas ?

Take Home message



- ▶ Illustration en vidéo
- ▶ Présentation de la phrase segment par segment
- ▶ Quelques secondes pour juger (échelle binaire)
- ▶ Degré de confiance évalué

- ▶ 2 VDs : jugement et temps de réaction
- ▶ Des réponses plus *online* et moins méta



Pourquoi avoir besoin de méthodes quantitatives ?

Pourquoi des méthodes quantitatives ?

Les données et les intuitions sont-elles redondantes entre elles ?

Comment faire ?

Le jugement

Jugement : tâche pertinente ?

Quel type de jugement ?

Quelle échelle ?

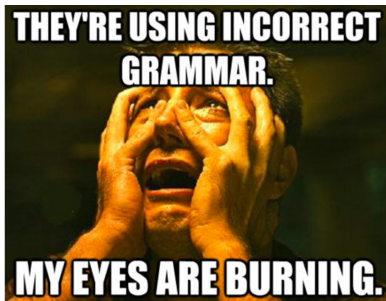
Temps limité ou pas ?

Take Home message



1. Les méthodes quantitatives sont nécessaires en linguistique formelle
2. Les expériences, et notamment les jugements d'acceptabilité sont un moyen efficace et commode pour cela
3. Attention à bien préciser vos hypothèses...
4. ... Et à réfléchir sur le type de méthode que vous voulez pour le jugement

Merci pour votre attention !





- Clifton, C. (n.d.). Experimental design for linguists. <https://people.umass.edu/cec/Experimental%20Design%20for%20Linguists.pdf>. Accessed : 2024-05-16.
- Edelman, S. and Christiansen, M. H. (2003). How seriously should we take minimalist syntax? *Trends in cognitive sciences*, 7(2):60-61.
- Gibson, E., Desmet, T., Grodner, D., Watson, D., and Ko, K. (2005). Reading relative clauses in english. *Cognitive Linguistics*, 16(2).
- Gibson, E. and Fedorenko, E. (2013). The need for quantitative methods in syntax and semantics research. *Language and Cognitive Processes*, 28(1-2):88-124.
- Gibson, E. and Thomas, J. (1996). The processing complexity of english center-embedded and self-embedded structures. In *Proceedings of the NELS 26 workshop on language processing : MIT Working Papers in Linguistics*.
- Gibson, E. A. F. (1991). *A computational theory of human linguistic processing : Memory limitations and processing breakdown*. Carnegie Mellon University.
- Huang, Y. and Ferreira, F. (2020). The application of signal detection theory to acceptability judgments. *Frontiers in Psychology*, 11:498268.
- Keller, F. and Alexopoulou, T. (2001). Phonology competes with syntax : Experimental evidence for the interaction of word order and accent placement in the realization of information structure. *Cognition*, 79(3):301-372.
- Langsford, S., Perfors, A., Hendrickson, A. T., Kennedy, L. A., and Navarro, D. J. (2018). Quantifying sentence acceptability measures : Reliability, bias, and variability. *Glossa : a journal of general linguistics*, 3(1):1-34.
- Miller, N. C.-G. (1963). Introduction to the formal analysis of natural languages' in : Handbook of mathematical psychology. *New York*, pages 269-321.
- Phillips, C. (2009). Should we impeach armchair linguists. *Japanese/Korean Linguistics*, 17:49-64.
- Robbins, P. T. (2019). The effect of prescriptive rules and instructions on the grammaticality judgment task.
- Schütze, C. (2016). *The empirical base of linguistics : Grammaticality judgments and linguistic methodology*. Language Science Press.
- Schütze, C. and Sprouse, J. (2014). Judgment data. in *research methods in linguistics*, ed. rj podesva, and d. sharma, 27-50.
- Sorace, A. and Keller, F. (2005). Gradience in linguistic data. *Lingua*, 115(11):1497-1524.
- Wayne, C. (1997). *Experimental syntax : Applying objective methods to sentence judgements*.
- Yngve, V. H. (1960). A model and an hypothesis for language structure. *Proceedings of the American philosophical society*, 104(5):444-466.



- Slide 1 : [https://imgflip.com/meme/44287307/10th-Doctor?sort=top-2022-06\(sciencewhocares\)](https://imgflip.com/meme/44287307/10th-Doctor?sort=top-2022-06(sciencewhocares))
- Slide 5 : <https://www.cs.toronto.edu/~ylzhang/csc236/memes.html#2>
- Slide de fin : <https://www.pinterest.com/pin/the-most-common-mistakes-students-make-when-writing--40039884179822448/>