

# La linguistique comme laboratoire



# Partie I



Un moment dans le développement du  
paradigme génératif

# Le virage expérimental et quantitatif dans le paradigme génératif



L'adoption des méthodes quantitatives et expérimentales dans le paradigme génératif peut être décrit comme:

Un virage/ une inflexion et non pas  
une révolution

(A shift rather than a turn)

# Virage



## 1) Longue histoire

**Hill 1961**, Bard et al. 1996, Schütze 1996, Cowart 1997, Keller 2000, Edelman and Christiansen 2003, Ferreira 2005, Wasow and Arnold 2005, Featherston 2007, Gibson and Fedorenko 2010a, 2010b, .....

Cf. B. Hemforth. Experimental linguistics (to app.)

## 2) Pas de nouveau cadre grammatical ou de modification radicale d'un cadre existant. Pour le moment ...

# Pourquoi maintenant?

Réponse d'un point de vue interne



Point de vue interne à la construction des grammaires.

## Impasse

pour fixer de façon consensuelle les données dans l'analyse des questions réputées cruciales dans les programmes contemporains (bien sûr, ce qui est réputé crucial va changer selon le point de vue théorique !!).

# Pourquoi maintenant?

Réponse d'un point de vue externe



Point de vue externe à la construction des grammaires.

## Crise

∞ Symptôme: « les programmes de recherche dans les disciplines connexes ne prennent plus en compte les développements du champ grammatical »

En particulier:

- TAL

- Etude des processus langagiers (production / compréhension)

# Pourquoi maintenant ?



Trois facteurs contextuels expliquent que le virage expérimental soit rendu possible maintenant:

1. Débat argumenté et de détail sur le recours au jugement grammatical comme accès aux phénomènes empiriques,
2. Développement (relativement) autonome d'outils d'observation
  - Méthodes quantitatives dans le TAL
  - Etude du comportement langagier en psycholinguistique
3. Emergence de questions à l'interface avec des sous-disciplines ayant effectué le virage quantitatif/expérimental.

# Jugement



Donnée essentielle dans le paradigme génératif:

1) Grammaticalité:

- a. \* Chat un voit l'enfant
- b. L'enfant voit un chat

2) Interprétation:

Chaque professeur corrige un devoir

- a. Un devoir( chaque professeur( corrige
- b. Chaque professeur (un devoir (corrige

3) Appropriation:

L1: Qui ont-elles vu au cinéma?

a. L2: # Bernadette] a vu Jean-Paul au cinéma

b. L2: Bernadette a vu Jean-Paul] au cinéma

# Jugement (2)



## 1. Donnée contestée/table sur la base de l'observation:

- (1) a. Suzan whipered/yelled/mumbled/barked/muttered the news to Rachel
- b. \*Suzan whipered/yelled/mumbled/barked/muttered Rachel the news (Pinker 1989, Levin 1993, Krifka 2001)

vs ex. attestés: she muttered him a rapid apology / you just mumble him an answer ..  
(Gibson & Federenko)

## 2. Donnée peu contestable sur la base de l'observation de corpus: (Superiority condition)

- (2) a. \* Ted revealed **what who** invented (vs: **who** invented **what**)
- b.? Ted revealed **which device who** invented
- c. ? Ted revealed **what device who** invented
- d. Ted revealed **which device which participant** invented (Hofmeister et al., sub. )

Cf. complex wh-fillers lead to higher acceptability ratings than bare wh-words.

# Critiques du jugement



Le jugement grammatical n'est pas un jugement pur: il est situé et les conditions de son accomplissement déterminent son contenu.

- **Biais cognitif (theory-driven bias)**: biais pour la confirmation (confirmation bias) de la part du linguiste et des informateurs; confirmation bias → overconfidence

- **Biais social**: relation d'autorité dans la communauté des linguistes

ex. (Lambrecht)

L1: Qui a fait la vaisselle?

L2 a. C'est Bernadette

b. # Bernadette

# Contre-argumentation



Phillips 2009, Sprouse & Almeida, 2010, Sprouse, Schütze & Almeida (sub.).

**Les résultats expérimentaux corroborent les jugements récoltés de façon informels.**

Donc:

- La théorie linguistique ne peut pas être invalidée car elle serait construite sur du sable empirique ...
- Il n'est pas nécessaire de recourir à une démarche quantitative/ expérimentale. Les linguistes (bien entraînés) sont le bon instrument pour décrire les langues. De plus, c'est un instrument qui ne coûte pas cher ..

# Un résultat



- ∞ Principal résultat (Sprouse, Schütze & Almeida)
- ∞ Linguistic inquiry (2001-2010)
  - 139 contrastes
    - (1) a. \* My belief Kim is clever is sincere.
      - ∞ b. My belief that Kim is clever is sincere.
    - (2) a. \* I seem eating sushi.
      - b. I like eating sushi.

Taux de répliation : 95 %

# Mais,



- ❧ « We suggest that syntacticians abandon the idea that there is a single method for every research question (or research environment). Science is not a recipe... » (Sprouse et al.)
- ❧ Une méthodologie expérimentale et quantitative est requise pour tout jugement gradient:
  - ❧ - gradience inhérente (Sorace & Keller 2005)
  - ❧ - gradience multifactorielle

# Exemple: inversion du sujet en français



- a. On veut que soit abrogée la disposition 40.
- b. Le nouveau livre qu'a écrit Jean-Paul est sur la table.
- c. Alors sont entrés deux ou trois courtisans.

Question: Est-ce que la structure informationnelle détermine/influence/est reliée à l'inversion?

## Inversion du sujet et structure informationnelle (2)



Trois hypothèses en compétition:

H1: Le statut de focus informationnel favorise/déclenche l'inversion du sujet, (Zubizarreta, i.a.)

H2: Le statut donné du verbe ou du VP rend possible l'inversion du sujet, (Marandin 2003, 2010)

H3: La solidarité informationnelle du verbe et du sujet rend possible l'inversion du sujet (Saeboe 2004, Marandin 2011)

# Inversion du sujet et SI (3)



Moyen de tester: y a-t-il un contraste entre les paires (1) et (2)

- (1) L1: Qui voudrais-tu qui vienne?  
L2: a. Je voudrais que Paul vienne  
b. Je voudrais que vienne Paul
- (2) L1: Qu'est-ce que tu as?  
L2: a. Je voudrais que Paul vienne  
b. Je voudrais que vienne Paul

# Exemple II.

## Prosodie et association



### ∞ Association en français

(1) Il a seulement photographié les animaux dans le parc.

Hypothèse Beyssade et al. 2008: Nuclear Pitch Accent (NPA) est une barrière pour toute mise en relation (sémantique) entre un opérateur et ses arguments, en particulier pour l'association sémantique. → prosodie marque le domaine d'association

# Association (suite)



Test: contraste entre

- (2) a. Il a seulement photographié les animaux  $I_{NPA}$  dans le parc
- b. Il a seulement photographié les animaux  $H^*$  dans le parc  $I_{NPA}$

Prédiction: (3) n'est possible qu'avec (2b); impossible avec (2a)

(3) 'Il a photographié les animaux dans le parc et nulle part ailleurs'

# Association (suite, 2)



**Task:** you will listen to a series of utterances such as the following:

(8) Il a seulement photographié les animaux dans le parc.

Imagine that you do not agree with such a statement. To express your opposition, you can choose either one of the two following sentences

(29) 1. Non, il a aussi photographié les arbres dans le parc.

2. Non, il a aussi photographié les animaux dans la forêt.

# Consensus



Les phénomènes du deuxième type:

- Phénomènes multifactoriels – qui donnent lieu à des jugements gradients – : ordre des mots, choix de constructions, .. (relation complexe avec le contexte ou la sémantique lexicale), ..
- Phénomènes qui n'ont pas fait l'objet de descriptions explicites dans les savoirs (savoir savant ou non-savant) pré-linguistique: prosodie, ..
- Phénomènes rares ou limites (inaccessibles à une étude sur corpus): relation à distance, ..

# Situations effectives de transfert



Le transfert de techniques s'opère de sous-disciplines à sous-disciplines à propos de questions qui doivent être étudiées de plusieurs points de vue ...

## Historiquement:

– **Récurtivité syntaxique** : l'inacceptabilité est-elle dûe à une contrainte structurale ou bien à des difficultés de processing (source ambiguity problem)?

- Psycho-linguistique → syntaxe

- (1) a. The boy the girl the host knew brought left ← grammatical, difficile à interpréter (Miller & Chomsky 1963)
- (2) b. What did who finish ? (vs Who finished what?) ← agrammatical (Chomsky 1973)

# Situations effectives de transfert



## – **Complexité morphologique** (H. Baayen, i.a.)

Psycho-linguistique, TAL → morphologie

– [Pour la communauté de Paris 7 à l'origine de l'axe 2 du Labex « EFL »]. **Prosodie** (interface prosodie/sémantique-pragmatique)

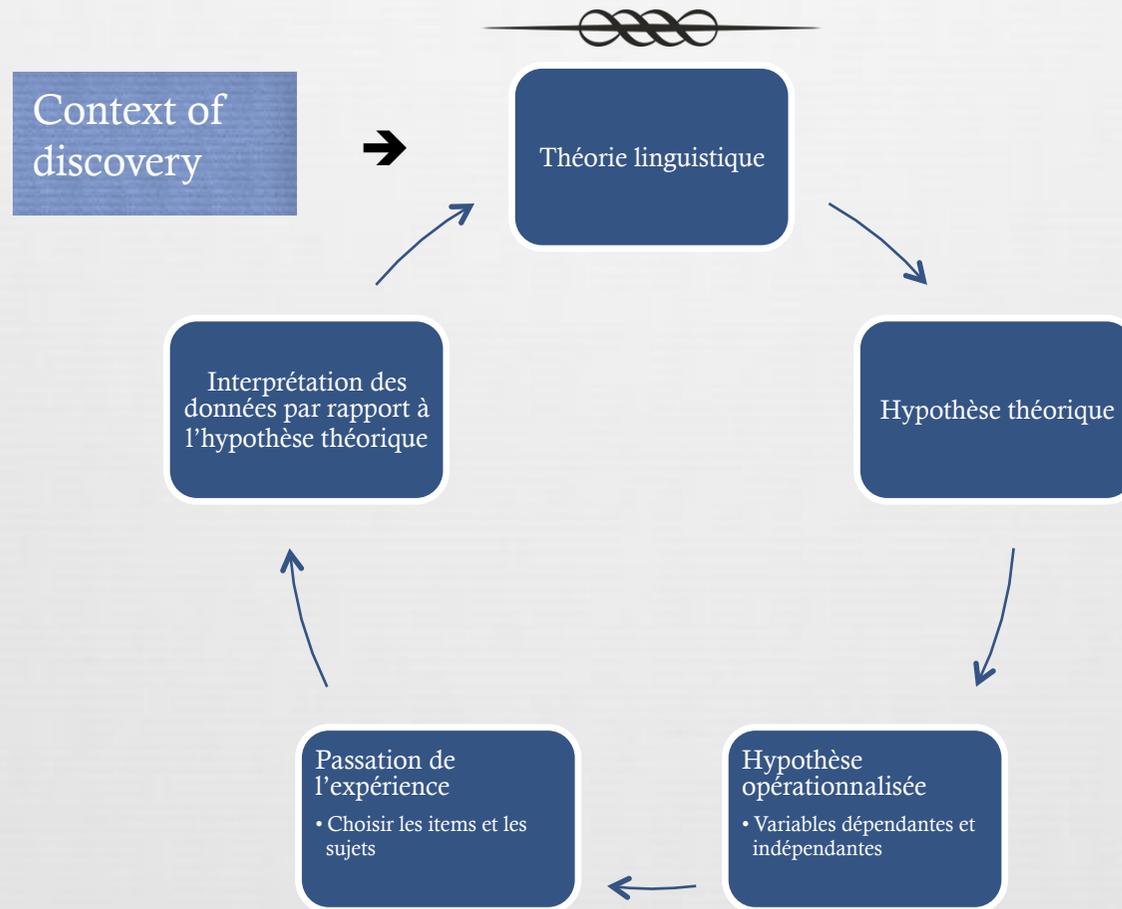
Phonologie de laboratoire → sémantique / pragmatique

# Techniques d'observation (2) : Développement des études du comportement langagier



- ❧ Qu'est-ce qu'une expérience ?
  - ❧ L'objectif d'une expérience est de trouver des relations causales entre des variables manipulées dans l'expérience et des variables observées (jugements, temps de réactions, fixations des yeux, ...).
    - ❧ Variables manipulées: variables indépendantes
    - ❧ Variables observées: variables dépendantes
  - ❧ Différence entre expérience et études de corrélation (analyse de corpus): Les corrélations dans un corpus ne peuvent jamais être interprétées comme relations causales car il est impossible de contrôler les effet des variables externes.

# Règle 1: Formuler la question de recherche en détail et en utilisant les termes de la méthode choisie



# Qu'est-ce qu'une expérience: Les contraintes indispensables



- ❧ Règle 1: Formuler la question de recherche en détail et en utilisant les termes de la démarche expérimentale choisie.
- ❧ Règle 2: Réaliser des contrastes minimaux (ne faire varier que ce que l'on veut voir varier).
- ❧ Règle 3: Réduire le « bruit », contrôler la variabilité extérieure à l'expérience.
- ❧ Règle 4: Adapter le nombre d'items / de sujets aux effets attendus

# Règle 1: Formuler la question de recherche en détail et en utilisant les termes de la méthode choisie



- ∞ L'approche expérimentale demande la variation systématique du ou des paramètres étudiés. Les variables indépendantes réalisent la question de recherche.
  - ∞ par ex.: Les dépendances de distance longue sont difficile à traiter.  
VI: nombre de mots, de syntagmes, ... ?
- ∞ Opérationnalisation des variables (voir types d'expériences)
  - ∞ Les hypothèses doivent être formulées précisément selon les variables dépendantes et indépendantes. Les variables dépendantes changent selon le paradigme expérimental.
    - ∞ Par ex: Les distance plus longue impliquent des temps de lecture élevés, sont moins acceptables, ...

## Règle 2: Réaliser des contrastes minimaux (ne faire varier que ce qu'on veut voir varier)



- ❧ Essayer de contrôler les variables externes (fréquence, longueur, familiarité, etc.).
- ❧ Si on ne peut pas les contrôler, s'assurer qu'elles ne soient pas confondues avec les variables indépendantes (VI) (qu'elles ne varient pas avec les VI).
- ❧ Exemple: Interprétation d'un pronom sujet
  - (i) Jean discute avec Marie pendant qu'elle se maquille.
  - (ii) Pierre-Edouard discute avec Anne-Sophie pendant qu'il met son costume de clown.

# Si les variables externes varient ...



## Randomiser

- ⌘ Randomiser l'attribution des sujets selon les conditions expérimentales.
- ⌘ Questionnaire: ordre de présentation.
  - ⌘ Une seule randomisation peut poser des problèmes.
  - ⌘ Une randomisation différente pour chaque sujet?
  - ⌘ Pseudo-randomisations

## ⌘ Designs contrebalancés

- ⌘ Toutes les conditions sont présentées à chaque sujet,
- ⌘ Chaque condition est présentée à des groupes parallèles de sujets.

# Contrebalancer le matériel



- ↻ Contrebalancement
  - ↻ Nombre équivalent d'items dans chaque condition,
  - ↻ Chaque item doit être réalisé dans chaque condition.
  
- ↻ Pourquoi?
  - ↻ Les items et les sujets peuvent varier d'une façon qui affectent les VD. Parfois, une variable (ou un sujet) qui se comporte très différemment des autres peut produire des effet artificiels.

# En linguistique expérimentale, on contrebalance ...



- ❧ Généralement en variant les conditions dans un design intra-sujet (chaque sujet est présenté à un nombre équilibré d'items pour chaque valeur des variables indépendantes)
  - ❧ Questionnaire, 2 conditions, N items
  - ❧ Créer 2 versions, chacune avec N items, N/2 dans la condition 1, le reste dans condition 2. Inverser pour la version 2 du questionnaire.
  - ❧ Plus généralement:
    - ❧ Pour M conditions il faut créer M versions du questionnaire, chacune avec un nombre égal d'items par condition.
    - ❧ Le nombre d'items nécessaires pour une expérience est toujours un multiple des conditions expérimentales.

# Carré Latin (Euler, 1773): 3 conditions, 9 items



Items	Liste 1	Liste 2	Liste 3
Phrase 1	Condition 1	Condition 3	Condition 2
Phrase 2	Condition 1	Condition 3	Condition 2
Phrase 3	Condition 1	Condition 3	Condition 2
Phrase 4	Condition 2	Condition 1	Condition 3
Phrase 5	Condition 2	Condition 1	Condition 3
Phrase 6	Condition 2	Condition 1	Condition 3
Phrase 7	Condition 3	Condition 2	Condition 1
Phrase 8	Condition 3	Condition 2	Condition 1
Phrase 9	Condition 3	Condition 2	Condition 1

# Variance de la variable dépendante



- ∞ Variance systématique: variabilité liée à la manipulation des VIs et d'autres variables identifiées.
- ∞ Variance aléatoire: Variabilité dont on ignore la source.
- ∞ Statistique inférentielle: Est-ce qu'il y a de la variance systématique au delà de la variabilité aléatoire ?
  - ∞ Est-ce qu'il y a un signal dans le bruit ?

# Règle 4: Adapter le nombre d'items / de sujets aux effets attendus



- ⌘ Ajouter des données pour augmenter la puissance statistique
  - ⌘ Minimiser le bruit relatif à la variabilité systématique (liée aux variables indépendantes)
  
- ⌘ Loi des grands nombres:
  - ⌘ Plus l'échantillon est grand, plus on s'approche de la population (de phrases ou de sujets).
  - ⌘ Les effets les plus fins demandent des échantillons de plus grande taille.
  - ⌘ Quand on teste (presque) toute la population, il n'y a plus de différences non-significatives. Chaque différence trouvée empiriquement est réelle.

# Types d'erreurs en linguistique expérimentale



- ❧ Erreurs de type 1 et 2:
  - ❧ Type 1: Pour chaque effet significatif, il peut être dû à des propriétés de l'échantillon choisi qui ne sont pas représentatives de la population des sujets et/ou des items testés.
  - ❧ Type 2: Pour chaque effet non-significatif, il y a une certaine probabilité que cet effet n'apparaisse pas parce que la variabilité externe, non-contrôlée, est trop importante.
- ❧ Pour réduire ces deux types d'erreurs il faut suivre les règles précédentes dans le montage des expériences.
- ❧ Même pour les meilleurs études, ces erreurs sont toujours présents.

# Techniques d'observation: Développement des études du comportement langagier



- ∞ Les résultats des expériences sont toujours à la fois liés à la représentation linguistique et au traitement.
- ∞ Problème de source (Hofmeister et al., 2011): Est-ce qu'un phénomène spécifique nous informe sur la représentation ou sur le traitement?
  - ∞ Technique on-line vs. off-line ?
  - ∞ Différences inter-individuelles ?
  - ∞ Expériences inter-langues ?

# Techniques psycholinguistiques (opérationnalisation de la variable dépendante)



- *Off-line*: Mesurer la compréhension/représentation finale, pas le processus
  - Questionnaires
    - Cloze/Completion Task
    - Jugement
      - Grammaticalité
      - Acceptabilité
      - Plausibilité
    - Magnitude Estimation

# Exemple: Cloze task



*Pierre a giflé Jean quand il était jeune.*

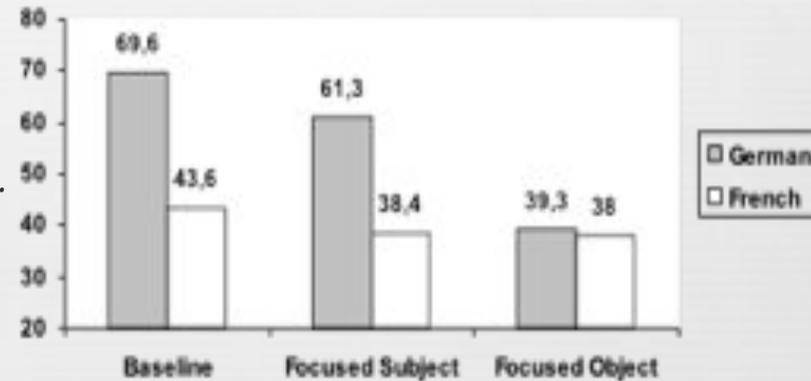
.....*était jeune.*

*C'est Pierre qui a giflé Jean quand il était jeune.*

.....*était jeune.*

*C'est Pierre que Jean a giflé quand il était jeune.*

.....*était jeune.*



# Jugements



- ❧ L'acceptabilité est influencée par
  - ❧ Familiarité
  - ❧ Facilité de réparer une erreur
  - ❧ Erreurs fréquentes (cf. speech errors, syntactic blends, etc.)
  
- ❧ Des phrases malformées peuvent être jugées acceptable, parfois même plus acceptables que des phrases bien formées.
  
- ❧ Exemples:
  - (i) This information could have been released by Gorbachev, but he chose not to. (Daniel Schorr, NPR, reported by D. Hardt)
  - (ii) Le plat mexicain que le critique culinaire que le magazine a engagé avait une odeur étrange.
  - (iii) Le plat mexicain que le critique culinaire que le magazine a engagé a goûté dans le nouveau restaurant avait une odeur étrange. (Gimenes *et al.* 2008)

# Techniques en temps réel



- *On-line*: Tester les processus, pas que l'interprétation, techniques expérimentales en temps réel, chronométrie mentale
  - Self-paced reading (auto-présentation ségmentée)
  - Eye-tracking (mouvements oculaires)
  - Event related potentials, ERPs (potentiels évoqués)

# Self-paced reading



- ❧ Rationale: La charge cognitive est reflété dans les temps de lecture.
- ❧ Les sujets lisent des phrases ou des textes sur un écran d'ordinateur.
- ❧ Présentation
  - ❧ Mot par mot/ Segment par segment segment / syntagme par syntagme/ Phrase par phrase
- ❧ Les sujets contrôlent la vitesse en appuyant sur une touche.
- ❧ Le temps entre la présentation du stimulus (mot, syntagme, ...) et la réaction du sujet est considéré comme le temps de lecture sur ce segment.

# Self-paced reading, fenêtre stationnaire



☞ Quand le public applaudissait le chanteur quitta la scène.

# Self-paced reading, fenêtre stationnaire



**Quand**

# Self-paced reading, fenêtre stationnaire



**1e**

# Self-paced reading, fenêtre stationnaire



**public**

# Self-paced reading, fenêtre stationnaire



**applaudissait**

# Self-paced reading, fenêtre stationnaire



**1e**

# Self-paced reading, fenêtre stationnaire



**chanteur**

# Self-paced reading, fenêtre stationnaire



quitta

# Self-paced reading, fenêtre stationnaire



**1a**

# Self-paced reading, fenêtre stationnaire



**scène .**

# Self-paced reading, fenêtre flexible



∞ Présentation cumulative



# Self-paced reading, fenêtre flexible



Quand -- ----- .

# Self-paced reading, fenêtre flexible



Quand le -----.

# Self-paced reading, fenêtre flexible



Quand le public ----- .

# Self-paced reading, fenêtre flexible



Quand le public applaudissait -- ----- -- -----.

# Self-paced reading, fenêtre flexible



Quand le public applaudissait le ----- - - - - - .

# Self-paced reading, fenêtre flexible



Quand le public applaudissait le chanteur ----- -- -----.

# Self-paced reading, fenêtre flexible



Quand le public applaudissait le chanteur quitta -- -----.

# Self-paced reading, fenêtre flexible



Quand le public applaudissait le chanteur quitta la -----.

# Self-paced reading, fenêtre flexible



Quand le public applaudissait le chanteur quitta la scène.

# Self-paced reading, fenêtre flexible



∞ Présentation non-cumulative



# Self-paced reading, fenêtre flexible



Quand - - - - - .

# Self-paced reading, fenêtre flexible



----- 1e ----- .

# Self-paced reading, fenêtre flexible



----- -- public ----- .

# Self-paced reading, fenêtre flexible



----- -- ----- applaudissait -- ----- ----- .

# Self-paced reading, fenêtre flexible



----- le -----.

# Self-paced reading, fenêtre flexible



----- -- chanteur -----.

# Self-paced reading, fenêtre flexible



----- quitta -----.

# Self-paced reading, fenêtre flexible



----- la -----.

# Self-paced reading, fenêtre flexible



----- scène .

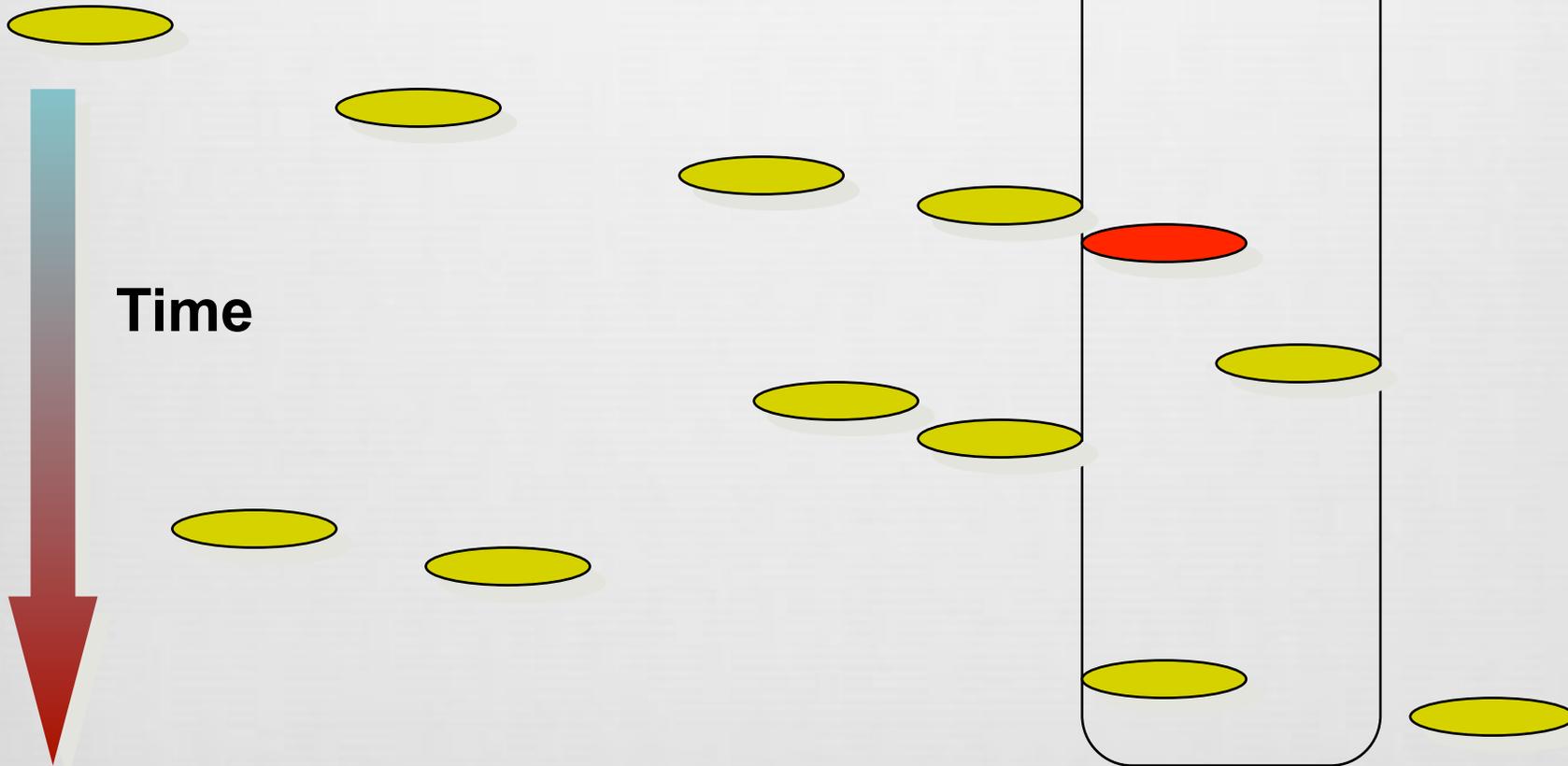
# Mouvements oculaires



- ↻ La location et le temps d'une fixation reflètent le traitement du stimulus.
  
- ↻ Mesures:
  - ↻ First fixation duration
  - ↻ First pass reading time (gaze)
  - ↻ Regression path duration (Go past, total pass)
  - ↻ Total reading time

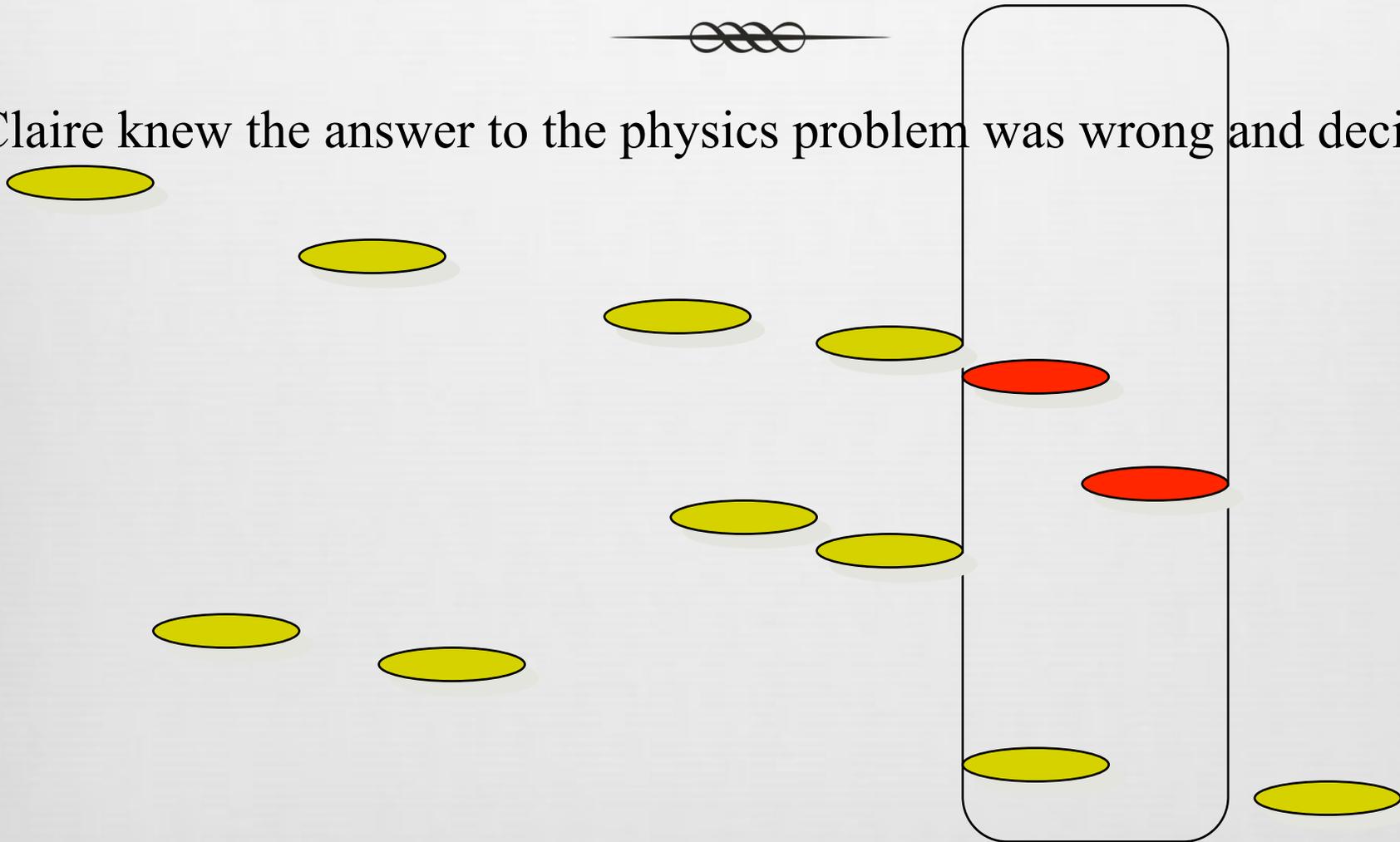
# First Fixation Duration

Claire knew the answer to the physics problem was wrong and decided ...



# First pass reading time

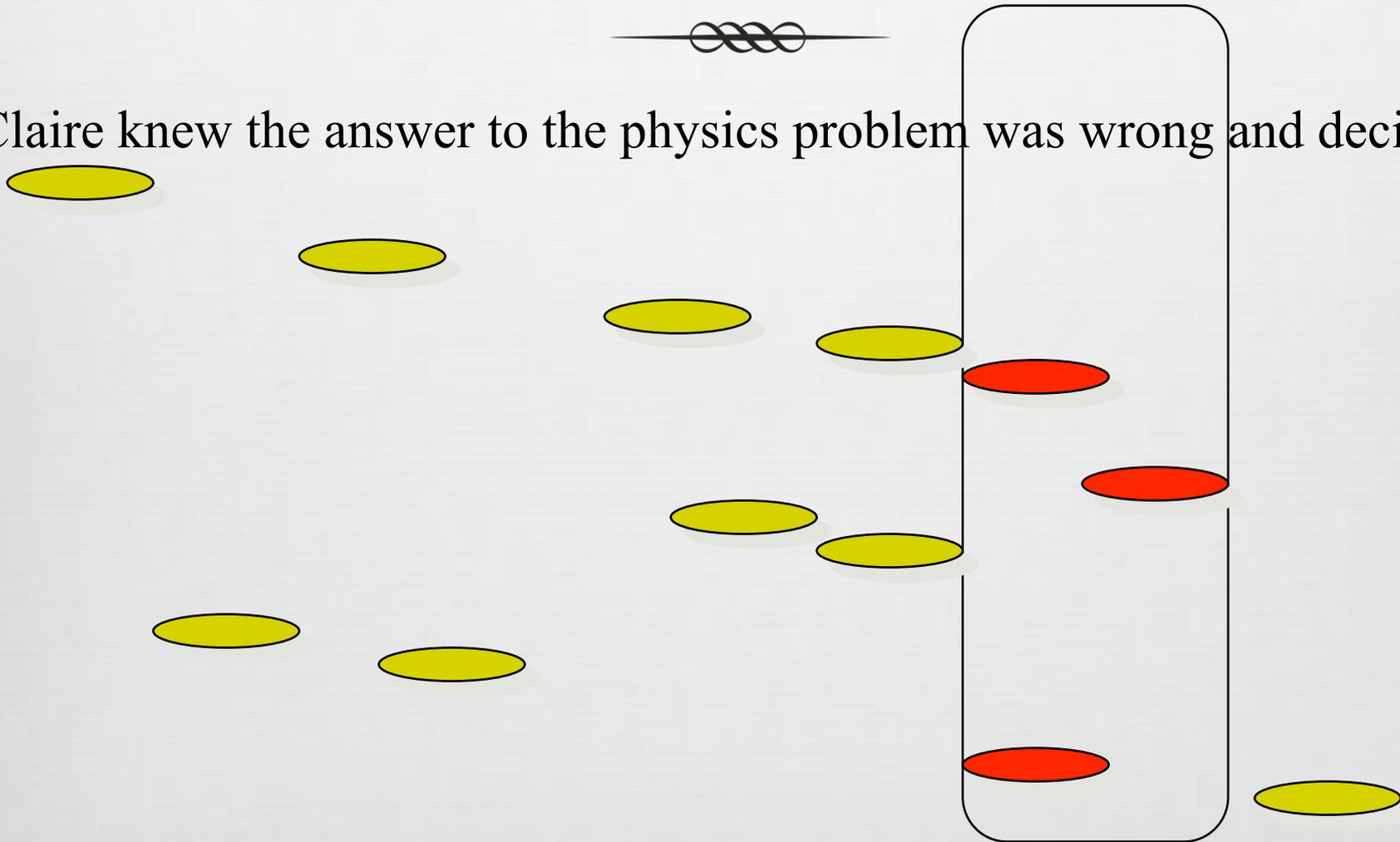
Claire knew the answer to the physics problem was wrong and decided ...



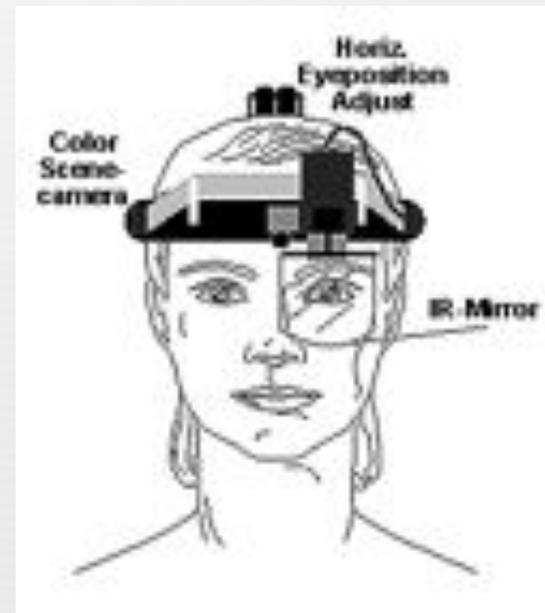
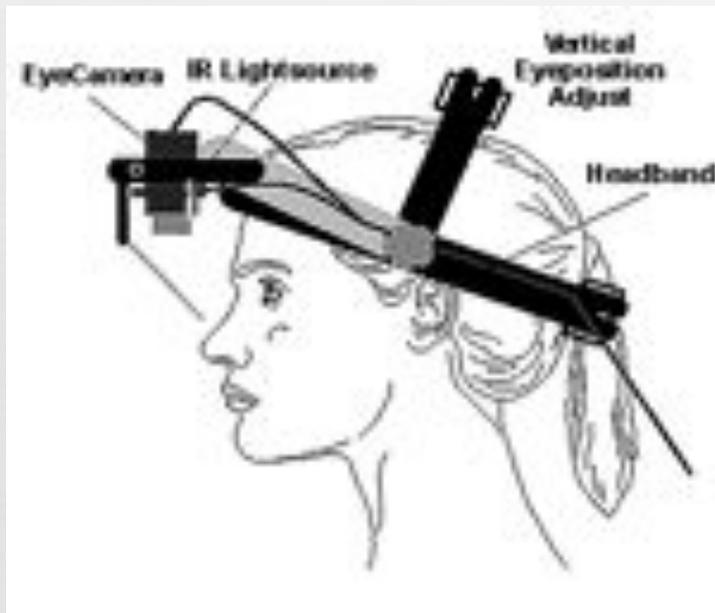


# Total reading time

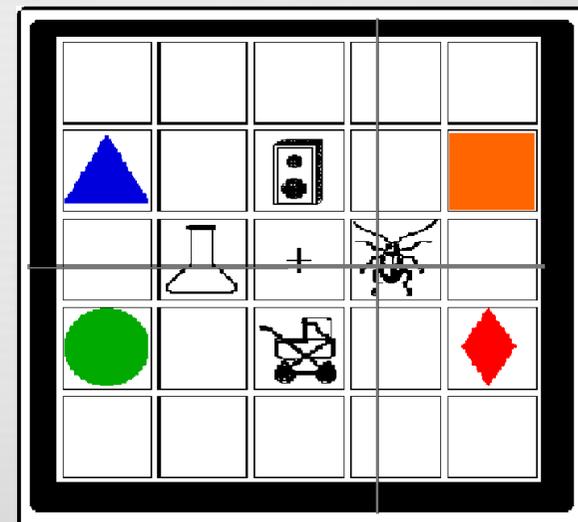
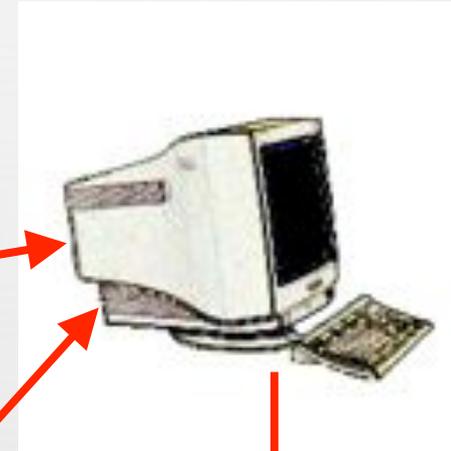
Claire knew the answer to the physics problem was wrong and decided ...



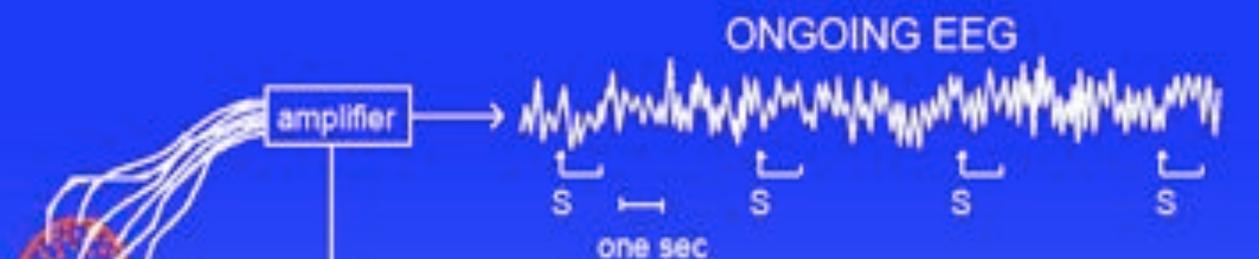
# Head Mounted Eye Tracker



# Head Mounted Eyetracker à Rochester (labo de Mike Tanenhaus)



# ERP Method

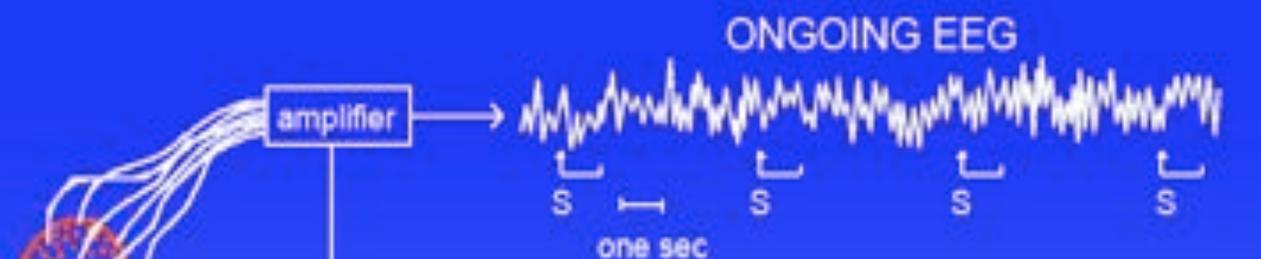


# Techniques 4: Potentiels évoqués (PEV)



- Potential shifts reflect processing characteristics
  - Semantic/lexical Integration: N400
  - Early syntactic processes: ~N200 (ELAN, LAN)
  - Reanalysis / Repair: P600
- Topography: Potential shifts come up in different electrodes
  - N200: left anterior (LAN)
  - N400: central, both hemispheres
  - P600: parietal, both hemispheres
  - Sustained frontal negativity: Working Memory

# ERP Method



# Contrôler les facteurs cognitifs



- ∞ Des facteurs cognitifs plus généraux influencent les processus linguistiques
  - ∞ Mémoire de travail
  - ∞ Expérience linguistique
  - ∞ Fonctions exécutives (attention, inhibition, ...)
  - ∞ ...

# Technique psycholinguistique et approche expérimentale



- ❧ L'utilisation de techniques psycholinguistiques n'est pas équivalent à une approche expérimentale
- ❧ Exemple:
  - ❧ Paris Corpus (de Jarnett, Ellsiepen, Hemforth, & Planq, in prep.):
    - ❧ Mouvements oculaires pendant la lecture d'une partie des articles annotés dans le French Treebank (25K mots).
      - ❧ Technique assez sophistiquée et complexe et
      - ❧ variables dépendantes assez fines qui reflètent les processus cognitifs pendant la lecture mais
      - ❧ pas de variation systématique des variables indépendantes

# Partie III



Approche expérimentale

Ex. anaphore

# Pronominal anaphora



The **postman** met the **street-sweeper** before **he** went home.

- ∞ First mentioned (Gernsbacher, 1998)
- ∞ *Sujet* (Konieczny & Hemforth 2001; Jaervikivi, van Gompel, Hyöna, & Bertram 2005)
- ∞ *Topicalité* (Givon, 1983, Schimke et al. 2009)
- ∞ *Causalité implicite, relations de cohérence* (Kehler, 2002; Sanders & Noordman, 2000)

→ Différences interlangues ?

# Approche



- ⌘ Time scale: traitement off-line ou en on-line
  - ⌘ Approche multi-méthodes et inter-langue
  - ⌘ Analyse de corpus
  - ⌘ Self-pace reading (auto-présentation segmentée)
  - ⌘ Eye-tracking (mouvements oculaires)

# Overview



- ❧ Comparaison interlangue (Anglais, Français, Allemand)
  - ❧ Analyse de corpus
  - ❧ Expérience 1: Décisions off-line (cloze task)
  - ❧ Expérience 2: Visual World eye-tracking study
  - ❧ Expérience 3: APS en Portuguais  
(Baumann, Konieczny and Hemforth, 2010)

# Analyse de corpus



- ∞ 100 phrases par langue:
  - ∞ 77% antécédents sujet en allemand (Frankfurter Rundschau)
  - ∞ 64 % antécédents sujet en anglais (Wall Street Journal)
  - ∞ 0 % (Le Monde) ou 15% (Google News groups) antécédents sujet en français.

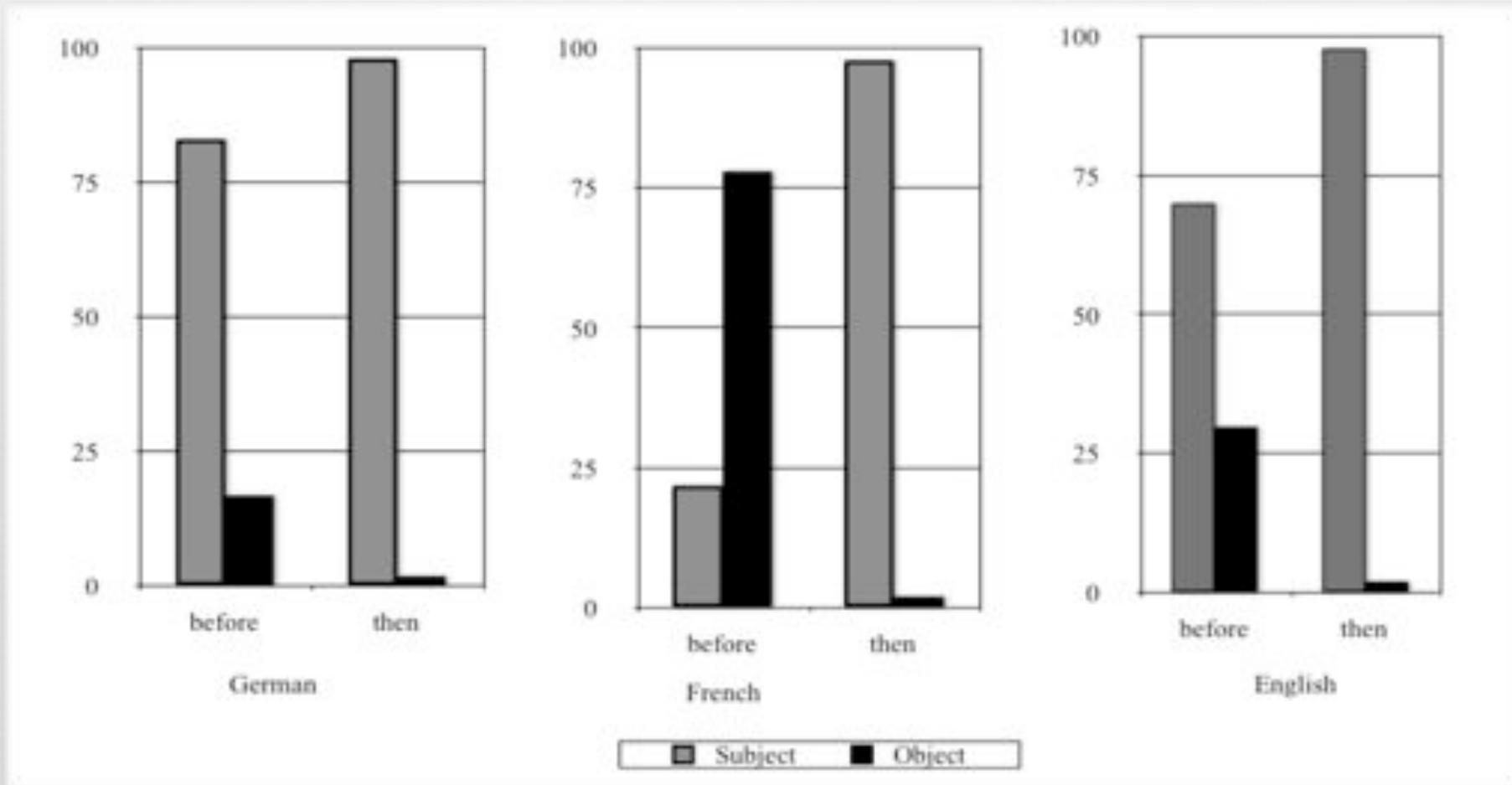
# Comparaison de préférences interlangue



1. a. The postman met the street-sweeper before **he** went home.  
The \_\_\_\_\_ went home.
- b. Der Briefträger traf den Straßenfeger, bevor **er** nach Hause ging.  
Der \_\_\_\_\_ ging nach Hause.
- c. Le facteur a rencontré le balayeur avant qu'**il** rentre chez lui.  
Le \_\_\_\_\_ rentre chez lui.
2. a. The postman met the street-sweeper. Then **he** went home..
- b. Der Briefträger traf den Straßenfeger. Dann ist **er** nach Hause gegangen.
- c. Le facteur a rencontré le balayeur. Puis **il** est rentré chez lui.

# Résultats questionnaires

(Hemforth, Konieczny, Scheepers, Collona, Schimke, Pynte, 2010)



☞ Decisions (%) for the subject vs. object of the main/first clause

# Expérience 2: Visual World

(Hemforth, Konieczny, Scheepers, Collona, Schimke, Pynte, 2010)



- ❧ Expectations
  - ❧ Préférences précoces ou tardives
  
- ❧ Matériels sémantiquement désambiguïsés
  - ❧ Logique: L'interprétation online se reflète dans le pattern de fixations

# Matériels

1. a. The postman met the street-sweeper before **he** picked up the **letters**.  
b. Der Briefträger traf den Straßenfeger, bevor **er** die **Briefe** einsammelte.  
c. Le facteur a rencontré le balayeur avant qu'**il** ramasse les **lettres**.
  2. a. The postman met the street-sweeper before **he** picked up the **trash**.  
b. Der Briefträger traf den Straßenfeger, bevor **er** den **Abfall** einsammelte.  
c. Le facteur a rencontré le balayeur avant qu'**il** ramasse la **poubelle**.
- ∞ For half of the participants, the 1st and 2nd NP were presented in the reverse order, to avoid semantic/pragmatic confounds
- 2a' The street-sweeper met the postman before **he** picked up the **letters**.
- ...

# Expérience 2: Visual World

- ↻ Participants
  - ↻ 32 locuteurs natifs français
  - ↻ 32 locuteurs natifs anglais
  - ↻ 24 locuteurs natifs allemands
- ↻ Tâche: Est-ce que la phrase que vous écoutez correspond à l'image sur l'écran (congruent) ou pas (incongruent).
- ↻ Les 16 items cibles étaient tous congruents.
- ↻ La moitié des items de la phase d'entraînement ainsi que la moitié des distracteurs étaient incongruent.

# Expérience contrôle pour le français

## ∞ Préférence visuelle

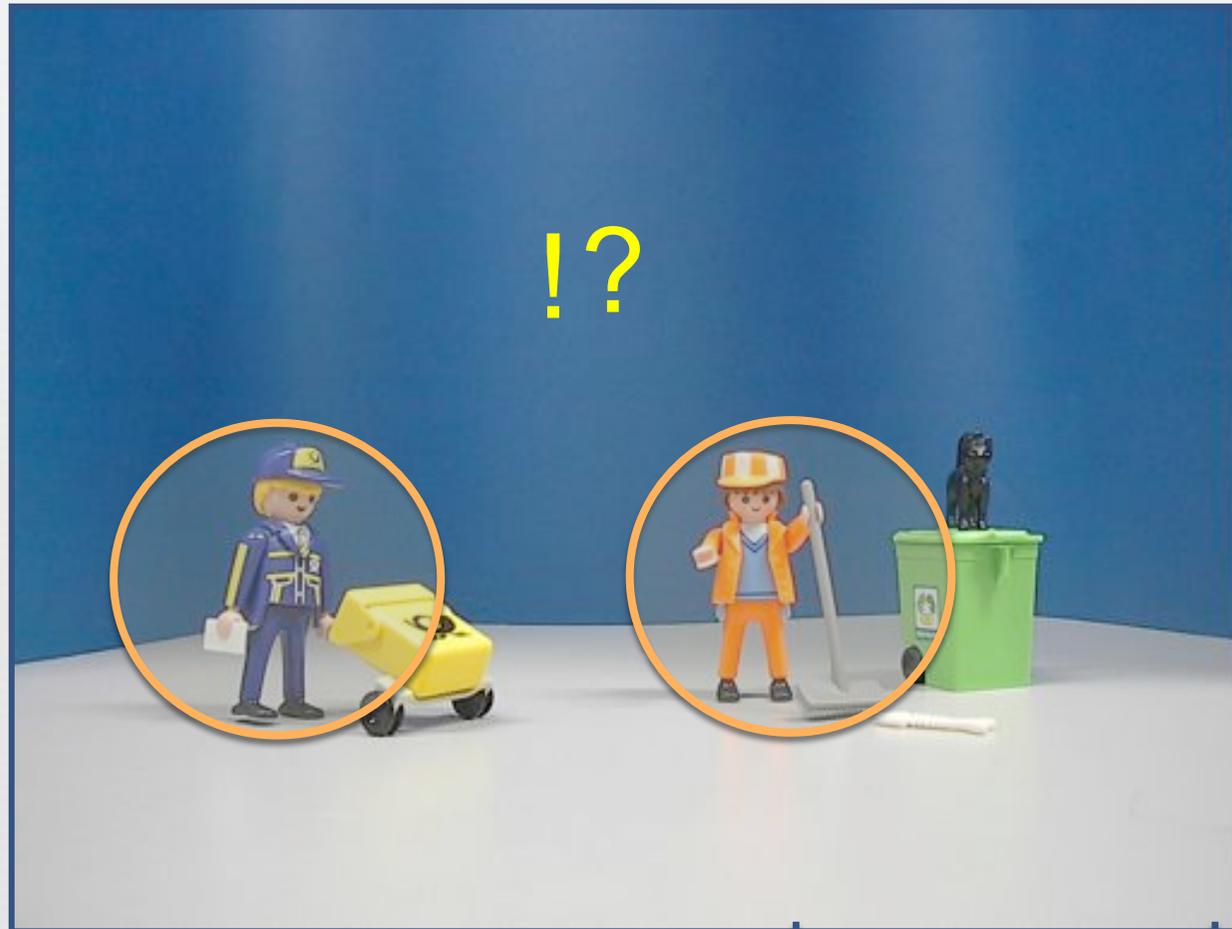
- ∞ Est-ce que les sujets français préfèrent regarder la description de l'objet pour des raisons indépendantes de la résolution des pronoms?

*Le facteur a rencontré le balayeur. Puis il a ramassé les lettres.*

The postman met the street-sweeper. Then he picked up **the letters**.

**Prédiction: Préférence pour le sujet.**

# Visual stimuli



Spoken stimuli

The postman met the street-sweeper

R1

before

R2

he picked up the

R3

letters / trash.

# Analyse de fixations

- ↻ 3 périodes d'intérêt (POI)
  - ↻ R1: 500 msec avant l'onset du pronom
  - ↻ R2: de l'onset du pronom à 480 ms après l'onset de la région désambiguïsant
  - ↻ R3: le reste de la phrase jusqu'à 2000 msec après le pronom
- ↻ Proportion de fixations sur les référents

$$\log odds = \log\left(\frac{p(i)}{p(\bar{i})}\right)$$

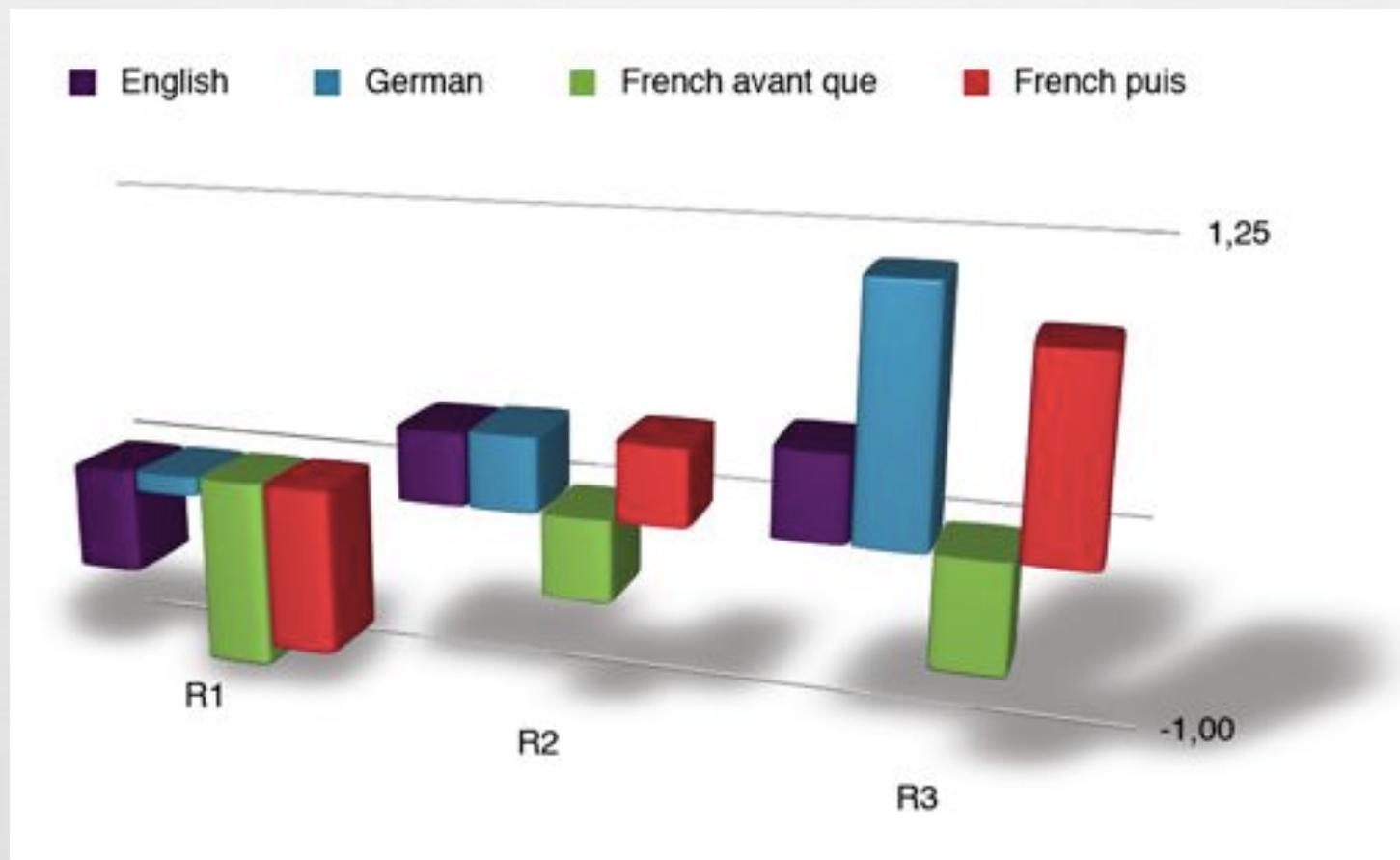
➔ Valeurs > 0 indique préférence pour le sujet

# Comparaison interlangue

(Hemforth, Konieczny, Scheepers, Collona, Schimke, Pynte, 2010)

Valeurs positives = Plus de fixations sur le SUJET

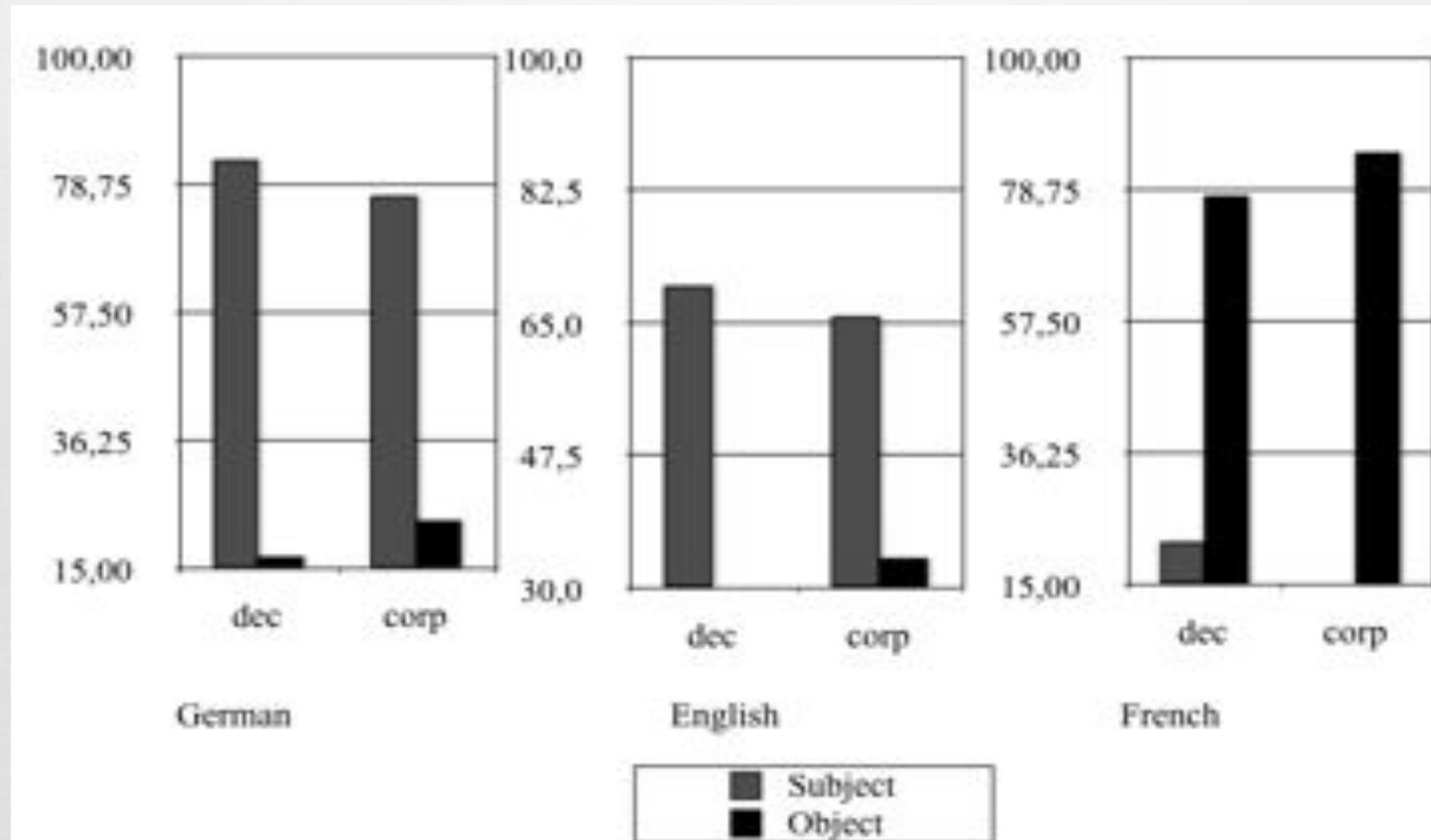
Valeurs négatives = Plus de fixations sur l'OBJET



errekt  
e in  
der  
Spr

# Decisions vs. Fréquences

(Hemforth, Konieczny, Scheepers, Collona, Schimke, Pynte, 2010)



# Anaphores en Portuguais

(Baumann, Konieczny, & Hemforth 2010)



☞ Comme en français: constructions alternatives pour *avant*

- Subordinate clause  
O João caiu *antes que* chorasse.  
John fell before he cried.
- Infinitive construction  
O João caiu *antes de* chorar.  
John fell before crying.

# Self-paced reading - Materials



1./3. O polícia encorajou a atriz, **antes que** ele/ela voltasse para casa.  
*The policeman encouraged the actress before he/she went home.*

2./4. O polícia encorajou a atriz, **quando** ele/ela voltou para casa.  
*The policeman encouraged the actress when he/she went home.*

- ☞ Crucial difference between **antes que** and **quando**:
  - ☞ alternative (infinitival) construction for **antes que**
  - ☞ No alternative construction for **quando**

Mais: Différent du français: Pro-drop

# Self-paced reading - Details

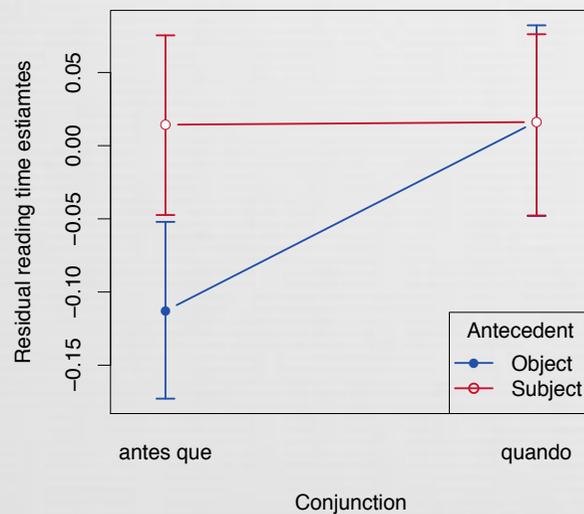


- ∞ Design 2 x 2
  - Conjonction: *antes que* (avant que) vs. *quando* (quand)
  - Antécédent: sujet vs. objet
- ∞ Auto-présentation segmentée
- ∞ 24 phrases cibles par participant, 48 fillers
- ∞ 24 locuteurs natifs de Portugal

# APS - Résultats



- Interaction significative ( $p < .05$ ) entre type de conjonction et antécédent
  - Temps de lecture plus courts pour les antécédents objets après *antes que*



# Discussion

- ⌘ Les préférences se retrouvent avec tous les paradigmes expérimentales.
  - ⌘ Les avantages et désavantages de chaque paradigme n'éliminent pas les effets.
  - ⌘ On peut alors conclure qu'il s'agit des effets très robustes.
- ⌘ **L'approche comparative permet des conclusions sur la représentation linguistique.**
  - ⌘ Pourvu, que nous testions des sujets de populations comparables et que les items soient des traductions proches, les différences inter-langues sont forcément des différences de représentation.

# Prédire l'ordre des mots

## Inférence à partir de corpus

Benoit Crabbé

- Méthodes inférentielles
  - Pour des expériences
  - Pour travail sur corpus
- **Question de fond** : Généralisation à partir de données
- $\Rightarrow$  Recours aux statistiques inférentielles

## 1 Notions intuitive d'inférence statistique

## 2 Predicting the dative alternation

- Modèle A (les variables corrélées)
- Modèle A' (variabilité des sujets)
- Modèle B (variabilité lexicale)
- Modèle C (Variabilité de corpus)

- Le paradigme expérimental utilise :
  - De l'analyse de corpus : permet d'observer globalement les données
  - Des expériences en production et compréhension.

Les deux sont complémentaires et reposent sur des méthodes de statistiques inférentielles. On fait le pari de présenter les idées clés de l'inférence statistique sans détailler les techniques mathématiques.

## Inférence statistique

L'inférence statistique est le procédé qui consiste à tirer des conclusions à partir de données sujettes à de la variation aléatoire, comme typiquement des variations d'échantillonnage

## Echantillonnage

Un **échantillon** est constitué d'un ensemble d'**individus** extraits d'une **population** (par exemple un ensemble de français issus de la population française)

Le procédé d'échantillonnage consiste à sélectionner les individus que l'on extrait de la population.

# Inférence statistique, démographie, élections présidentielles

- L'inférence statistique est utilisée systématiquement par les sondeurs (élections présidentielles).
- Les sondeurs souhaitent conclure sur le résultat de l'élection à partir d'un échantillon représentatif de la population.

## Méthodes d'échantillonnage des sondés

- Le sondeur demande à ses amis, collègues. . .
- Le sondeur tire au sort dans la liste des français un ensemble de français à sonder.
- Le sondeur constitue des échantillons soigneusement en répartissant équitablement par zone géographique, classe sociale, emploi exercé, sexe. . . de chaque individu.

# Quelques caractéristiques clés des échantillons

- **La taille de l'échantillon** : Plus l'échantillon est petit plus on a de chances que ce qu'on y compte varie d'un échantillon à un autre.

## Loi des grands nombres (Intuition)

Plus la taille de l'échantillon aléatoire augmente plus ses caractéristiques sont proches des caractéristiques de la population.

- Exemple:
  - Si on a un échantillon de 1 français, on aura soit 100% des français qui votent à gauche, soit 100% des français qui votent à droite selon le français choisi.
  - Si on a un échantillon de 30 000 000 de français, on a un échantillon qui se rapproche plus de la population française (*par contre le travail de dépouillement du sondage devient très conséquent*)
- Il faut donc choisir un échantillon pas trop petit mais pas trop grand non plus.
- Dans tous les cas, il y a toujours un risque que l'échantillon tiré soit ad hoc: risque de se tromper en analysant l'échantillon.

# Échantillonnage aléatoire et Intervalles de confiance

- **Motivation** : il y a beaucoup de chances que notre échantillon de travail soit partiellement ad hoc.
- On ne peut pas supposer que la proportion de votants à gauche dans l'échantillon est parfaitement exacte.
- L'intervalle de confiance mesure l'intervalle de valeurs dans laquelle doit se situer la vraie proportion dans 95% des cas.
- Conceptuellement, en tirant un grand nombre d'échantillons (par exemple 100) l'intervalle de confiance est l'intervalle dans laquelle 95% des estimations sont situées.

## Un peu de maths...

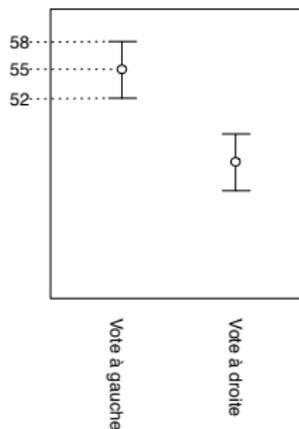
Déterminer un intervalle de confiance se fait par des méthodes mathématiques à partir de la moyenne observée pour un seul échantillon, en pratique on ne tire donc pas au sort ce grand nombre d'échantillons

# Estimation d'un paramètre

- Estimer un paramètre consiste à calculer sa moyenne sur un échantillon et à calculer son intervalle de confiance.
- Par exemple, proportion de votants à gauche :

moyenne	95bas	95haut
55	52	58

- On utilise également des représentations graphiques pour représenter l'estimation d'un paramètre:



# Tests d'hypothèses (I)

- Les statistiques inférentielles cherchent à tirer des conclusions sur une population à partir d'échantillons
- Par exemple, on veut savoir si la proportion de français qui vote à gauche est plus importante que la proportion de français qui vote à droite.
  - **Réponse facile:** compter **dans l'échantillon** le nombre de français qui votent à gauche, si il y en a plus, c'est que la population va voter à gauche, sinon la population va voter à droite.
  - **Problème:** l'échantillon est potentiellement ad hoc ! on risque de se tromper si le tirage au sort lors de la constitution de l'échantillon a été malheureux.

## Tests d'hypothèses (II)

- Un **test d'hypothèse** tient compte de la nature aléatoire de l'échantillon.
- Un test d'hypothèse consiste à poser une hypothèse nulle, et à tenter de la réfuter.
- Exemple continué:
  - **Hypothèse nulle:** la proportion de français qui vote à gauche est de 0.5.
  - **Hypothèse alternative:** la proportion de français qui vote à gauche est différente de 0.5 (et cela n'est pas lié à la manière dont nous avons tiré l'échantillon au sort)
- **Calcul:** Le test d'hypothèse se réalise par calcul. Celui-ci a pour but de vérifier que la moyenne observée dans l'échantillon est bien différente de l'hypothèse nulle en tenant compte de l'aléatoire induit par le tirage au sort.
- **Décision:** Si l'hypothèse nulle est rejetée on conclut en faveur de l'hypothèse alternative. Sinon on doit maintenir l'hypothèse nulle sans pour autant être sûr qu'elle est vraie.

# Test d'hypothèse illustré

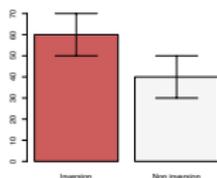
- On peut présenter le résultat d'un test d'hypothèse de la manière suivante:

moyenne	95bas	95haut	$P(t > 0.5)$	Signif
55	52	58	0.0002	***

- Ce qui complète notre présentation de l'intervalle de confiance; la probabilité représente le résultat du calcul (probabilité que l'hypothèse nulle ne puisse être rejetée), et la significativité indique la décision (on a une vraie différence par rapport à la moyenne nulle qui vaut ici 0.5)

# Test d'hypothèse dans un contexte expérimental

- Analyse des résultats expérimentaux (cas simple):
  - Habituellement on compare les moyennes mesurées pour chacune des conditions de la variable prédictive.
  - Si la différence des moyennes est significativement différente de 0, on a un effet: on en conclut que la variable prédictive influence la variable prédite.
- Illustration (variable prédictive : Inversion {oui,non}):



- Un test d'hypothèse nous permet de déterminer si la différence observée dans les moyennes de temps de lecture constitue un effet significatif.

# Inférence statistique sur corpus

- Les méthodes inférentielles utilisées sur corpus, sont conceptuellement analogues aux méthodes utilisées pour les expériences, à quelques différences près:
  - On ne mesure pas les réponses de sujets expérimentaux mais on se fonde uniquement sur des données fixées (le corpus)
  - On ne manipule pas les données
  - Les outils mathématiques sous-jacents sont différents (régression logistique)

L'usage de méthodes inférentielles sur corpus est utilisé depuis très longtemps en TAL (HMM par exemple), par contre l'usage de modèles inférentiels **interprétables** est beaucoup plus récent (Bresnan 2007)<sup>1</sup>

## Propos de l'article de Bresnan 2007

Cherche à montrer que l'on peut espérer tirer des généralisations à partir de données corpus en utilisant les méthodes de statistiques inférentielles modernes.

<sup>1</sup>Ceci dit, on trouve quelques travaux précurseurs en linguistique de corpus

## 1 Notions intuitive d'inférence statistique

## 2 Predicting the dative alternation

- Modèle A (les variables corrélées)
- Modèle A' (variabilité des sujets)
- Modèle B (variabilité lexicale)
- Modèle C (Variabilité de corpus)

# Prédire l'alternance dative: but de l'entreprise

- But : *Inférer des observations à caractère général sur le langage à partir de corpus richement annotés*
- La méthode introspective n'est pas la seule possible
- Repose sur une inférence statistique (modélisation)
- Traite un problème de préférence d'ordre des mots qui échappe à la démarche générative (écho à la strong competence hypothesis de Bresnan 82)
  - Cas gradient (on a un choix à faire qu'on pose comme non totalement arbitraire)
  - Cas où plusieurs facteurs entrent en jeu (multifactoriel)

## Bibliographie

Bresnan Joan, Anna Cueni, Tatiana Nikitina, Harald Baayen, Predicting the dative alternation, In *Cognitive Foundations of Interpretation*, G. Boume et al. eds, 2007.

Bresnan Joan, Is syntactic knowledge probabilistic ? Experiments with the English dative alternation, In S. Featherston & W. Sternefeld, Eds., *Roots : Linguistics in Search of Its Evidential Base*, 2007.

- Alternance dative en anglais :
  - 1 John gave a book to Mary (V-NP-PP)
  - 2 John gave Mary a book (dative shift, V-NP-NP)
- **Thème** ; **Bénéficiaire**

Problème trop difficile ?

**Choix** : quels facteurs interviennent pour préférer tel ou tel ordonnancement ?

- 1 Les observations sont corrélées (théories réductrices): par exemple on pourrait expliquer la préférence d'ordre des mots en fonction d'un seul facteur : la complexité syntaxique (Hawkins 94, corrélé avec l'accessibilité des référents (given/new)).  $\Rightarrow$  on veut pouvoir décorrélérer différents facteurs.
- 2 Les données groupées induisent des biais statistiques (regroupements par speaker ici)
- 3 La théorie syntaxique sera biaisée par des observations liées au seul lexique à disposition dans le corpus.
- 4 La non représentativité et les différences très fortes entre corpus bloquent toute tentative de généralisation.

# L'alternance dative : une affaire de sémantique ?

- On peut expliquer l'alternance dative en termes de sens exprimé :
  - ① Susan gives toys to the children
  - ② Susan gives children toys
- Où (1) induit un sens *changement de place* (des jouets) alors que (2) induit un sens *changement de propriétaire* (des jouets)
- Dans le cas de verbes comme *to give* (emploi idiomatique) la théorie prédit qu'on ne peut avoir de sens *changement de place*, donc pas de structure V-NP-PP
  - ① The lightning here gives me the creep
  - ② \* The lightning here gives the creep to me

- Exploration Google: beaucoup de contre-exemples à la théorie
- Montre que ces contre-exemples ne sont pas du pur bruit de Google
- Différents facteurs d'origines diverses semblent agir contre la théorie : accessibilité, définitude, complexité, pronominalisation des dépendants. . .

- 1 Notions intuitive d'inférence statistique
- 2 Predicting the dative alternation
  - Modèle A (les variables corrélées)
  - Modèle A' (variabilité des sujets)
  - Modèle B (variabilité lexicale)
  - Modèle C (Variabilité de corpus)

- **Contre les théories réductrices**
- Cite (Hawkins 94)
- **Principe général** : le plus court avant le plus long;
- Comme l'accessibilité (et l'animacité) corrèlent avec la définitude, on garde la définitude comme facteur explicatif car plus général.
- Propose une démarche de modélisation dans laquelle les données corrélées sont bien identifiées : régression logistique à effets aléatoires.

## Notion intuitive de corrélation

- Deux variables corrélées varient de manière analogue
- Par exemple longueur d'un dépendant et pronominalité sont corrélées : car quand on a un pronom (Pronominalité = vrai) le dépendant sera court (typiquement un mot) et quand on a un non pronom (Pronominalité = faux) le dépendant est potentiellement plus long.
- Si on a une corrélation parfaite, l'une des variables est redondante.
- Généralement les corrélations ne sont pas parfaites, p. ex. les noms animés sont souvent définis et *given* mais ce n'est pas toujours vrai. Il est alors difficile de démêler les deux variables ( $\Rightarrow$  donne lieu à des théories réductrices)

- **Accessibilité dans le discours** (*given, new, accessible*) (pour le thème et le bénéficiaire)
- **Définitude** (pour le thème et le bénéficiaire)
- **Pronominalité des dépendants** (pour le thème et le bénéficiaire)
- **Animacité des dépendants** (pour le thème et le bénéficiaire)
- **Classe sémantique du verbe** *abstrait, transfert de possession, futur transfert de possession, prévention de possession, communication*
- **Interaction de complexité** entre le thème et le bénéficiaire : différence de longueur
- **Personne des dépendants** (pour les pronominaux)
- **Amorçage dans la conversation**: priming.

- A extrait de Switchboard (corpus de conversations téléphoniques)  
2300+ exemples de cas de double complémentation (alternances datives)

## Extrait de Switchboard

```
( (S
  (INTJ (UH Uh) )
  ( , , )
  (NP-SBJ-1 (PRP it) )
  (VP (VBD had)
    (S
      (NP-SBJ-2 (-NONE- *-1) )
      (VP (TO to)
        (VP (VB be)
          (VP (VBN done)
            (NP (-NONE- *-2) )
            (PP (IN in)
              (NP (DT a) (NN hurry) ))))))))
  (. .) (-DFL- E_S) ))
```

# Table de données

- A créé une table de données semi-automatiquement à partir de Switchboard: chaque ligne une observation, chaque colonne, valeur de la variable pour cette observation.

## Extrait de la table de données (Bresnan 07)

	Speaker	Modality	Verb	SemanticClass	LengthOfRecipient	[...]	RealizationOfRecipient
...							
1956	S1260	spoken	pay	a	2	[...]	NP
1957	S1139	spoken	cost	p	1	[...]	PP
1958	S1212	spoken	give	t	1	[...]	NP
1959	S1236	spoken	sell	t	1	[...]	NP
1960	S1132	spoken	show	c	1	[...]	NP
1961	S1087	spoken	give	a	2	[...]	PP
...							

- Un modèle sur corpus comporte une variable dépendante à deux valeurs, par exemple pour l'ordre de l'alternance dative on a une variable

$$\textit{Realization of Recipient} = \{NP-PP, NP-NP\}$$

- Un modèle sur corpus comporte (typiquement) un grand nombre de variables prédictives.
- Pour chaque variable on estime un paramètre qui indique :
  - Si la variable vote pour l'ordre V-NP-PP. (signe positif sinon signe négatif)
  - Si la variable vote de manière significative
- Le modèle est calculé sur la table de données et peut être vu comme une élaboration de ce que l'on a présenté pour estimer une moyenne.
- Le modèle peut être utilisé sur de nouveaux exemples pour assigner une probabilité à chaque valeur de la variable prédite. (modèle prédictif)

# Fit du modèle A

---

Probability{Response = 1} =  $\frac{1}{1+e^{-X\hat{\beta}}}$ ,  
where

$$\begin{aligned} X\hat{\beta} = & 0.95 \\ & - 1.34\{c\} + 0.53\{f\} - 3.90\{p\} + 0.96\{t\} \\ (a) & + 0.99\{\text{accessibility of recipient} = \text{nongiven}\} \\ (a) & - 1.1\{\text{accessibility of theme} = \text{nongiven}\} \\ (b) & + 1.2\{\text{pronominality of recipient} = \text{nonpronoun}\} \\ (b) & - 1.2\{\text{pronominality of theme} = \text{nonpronoun}\} \\ (c) & + 0.85\{\text{definiteness of recipient} = \text{indefinite}\} \\ (c) & - 1.4\{\text{definiteness of theme} = \text{indefinite}\} \\ (d) & + 2.5\{\text{animacy of recipient} = \text{inanimate}\} \\ & + 0.48\{\text{person of recipient} = \text{nonlocal}\} \\ & - 0.03\{\text{number of recipient} = \text{plural}\} \\ & + 0.5\{\text{number of theme} = \text{plural}\} \\ & - 0.46\{\text{concreteness of theme} = \text{nonconcrete}\} \\ (e) & - 1.1\{\text{parallelism} = 1\} - 1.2 \cdot \text{length difference (log scale)} \end{aligned}$$

and  $\{c\} = 1$  if subject is in group c, 0 otherwise (and likewise for other categories).

## Conclusion (modèle A)

- Conclut que les différents facteurs ne sont pas réductibles à la seule complexité syntaxique (corrélations faibles)

### Note sur les corrélations

On voit que le modèle permet de se donner une idée du comportement de chacune des variables ainsi on a une indication pour la variable de différence de longueur indépendamment de l'indication pour la pronominalité.

Bresnan semble affirmer que les méthodes de modélisation sur corpus permettent de contourner les problèmes de corrélations, ce qui est en fait optimiste. L'interprétation des différents coefficients est plus complexe que ce que ne laisse penser l'article.

Par contre, ce qui est certain, c'est que le modèle met effectivement en lumière le rôle de facteurs multiples comme contributeurs au choix de la construction utilisée.

- 1 Notions intuitive d'inférence statistique
- 2 Predicting the dative alternation
  - Modèle A (les variables corrélées)
  - **Modèle A' (variabilité des sujets)**
  - Modèle B (variabilité lexicale)
  - Modèle C (Variabilité de corpus)

- **Seconde question posée** : est-ce que la variabilité inter-locuteur affecte le modèle obtenu ?
- Autrement dit: le modèle est-il dépendant des locuteurs impliqués dans l'échantillon
- Rappel: dans la table de données, chacune des lignes de données indique l'identifiant du locuteur

## Méthodologie présentée dans l'article

L'idée est de constituer des corpus alternatifs potentiellement biaisés (ou certains locuteurs seront potentiellement surreprésentés) de manière à vérifier si les modèles se trouvent biaisés de manière récurrente et significativement modifiés (**bootstrapping avec remplacement** sur les locuteurs)

- Bresnan calcule des intervalles de confiance sur les poids associés à chaque variable calculés sur un grand nombre de corpus resamplés

Model A: Relative magnitudes of significant effects with corrected error estimates

	Coefficient	Odds ratio PP	95% C.I.
Inanimacy of recipient	2.54	12.67	5.56–28.87
Nonpronominality of recipient	1.17	3.22	1.70–6.09
Nongiveness of recipient	0.99	2.69	1.37–5.3
Transfer semantic class	0.96	2.61	1.44–4.69
Indefiniteness of recipient	0.85	2.35	1.25–4.43
Plural number of theme	0.50	1.65	1.05–2.59
Person of recipient	0.48	1.62	1.06–2.46
Nongiveness of theme	-1.05	0.35	0.19–0.63
Structural parallelism in dialogue	-1.13	0.32	0.22–0.47
Nonpronominality of theme	-1.18	0.31	0.19–0.50
Length difference (log scale)	-1.21	0.3	0.22–0.4
Communication semantic class	-1.34	0.26	0.13–0.55
Indefiniteness of theme	-1.37	0.25	0.15–0.44

- Elle conclut que la variation inter-locuteur ne change pas fondamentalement les propriétés du modèle, car l'incertitude sur les poids ne passe jamais la barre de 1. (pas de changement de signe)

- 1 Notions intuitive d'inférence statistique
- 2 Predicting the dative alternation
  - Modèle A (les variables corrélées)
  - Modèle A' (variabilité des sujets)
  - **Modèle B (variabilité lexicale)**
  - Modèle C (Variabilité de corpus)

- Question : le modèle ainsi construit n'est-il pas biaisé par les spécificités des items lexicaux utilisés. (Une très large proportion des exemples font intervenir le verbe *give*).
- Méthode : régression logistique à effets aléatoires:
  - Utilise un modèle à effets aléatoires (intercept aléatoire)

- Crée une nouvelle variable "sens du verbe" faite de la concaténation des variables Verbe et Classe sémantique
- Par exemple, le verbe *give* de classe sémantique *t* aura le sens *give+t*.

# Table de Bresnan

Model B: Relative magnitudes of significant effects.

	Coefficient	Odds ratio PP	95% C.I.
Nonpronominality of recipient	1.73	5.67	3.25–9.89
Inanimacy of recipient	1.53	5.62	2.08–10.29
Nongiveness of recipient	1.45	4.28	2.42–7.59
Indefiniteness of recipient	0.72	2.05	1.20–3.5
Plural number of theme	0.72	2.06	1.37–3.11
Structural parallelism in dialogue	-1.13	0.32	0.23–0.46
Nongiveness of theme	-1.17	0.31	0.18–0.54
Length difference (log scale)	-1.16	0.31	0.25–0.4
Indefiniteness of theme	-1.74	0.18	0.11–0.28
Nonpronominality of theme	-2.17	0.11	0.07–0.19

- Bresnan n'observe pas de changement de directionnalité ni de significativité dans les coefficients, quand on conditionne sur des sens spécifiques des verbes. Donc on a effectivement un modèle largement indépendant des idiosyncrasies lexicales.

- 1 Notions intuitive d'inférence statistique
- 2 Predicting the dative alternation
  - Modèle A (les variables corrélées)
  - Modèle A' (variabilité des sujets)
  - Modèle B (variabilité lexicale)
  - Modèle C (Variabilité de corpus)

- Question : le modèle se transfère-t-il à un corpus de genre différent (journalistique) ?
- Problème de l'oral : le modèle ne reflète-t-il pas de propriétés extra-linguistiques ? comme le contexte, les limitations de mémoire. . .
- Méthode : teste le modèle sur un corpus écrit (le Wall street journal)

# Différences entre les deux corpus

- Remarque que les proportions de pronoms et de dative shift à l'oral et à l'écrit sont différentes.
- Teste si le même modèle prédit aussi bien sur un genre que sur l'autre malgré les différences de proportions.
- Montre qu'inclure une variable de genre (de corpus) dans le modèle n'a pratiquement aucun impact sur sa capacité à prédire correctement l'alternance dative.
- Conclusion : modèle qui généralise à travers différents types de corpus.

- Statistiques inférentielles permettent de tirer des conclusions généralisantes en syntaxe à partir d'échantillons finis de langue :
  - Permet d'adresser le problème des théories réductrices en mettant en lumière la contribution de facteurs multiples
  - Modèle qui généralise au-delà des spécificités des locuteurs (modèle A, bootstrapping)
  - Modèle qui n'est pas "collé" dans les idiosyncrasies lexicales (modèle B)
  - Modèle capable de généraliser au delà des variations de proportions dans les corpus (modèle C)

- **Question:** Le modèle construit sur corpus est-il également valable en production ? Reflète-t-il les préférences de locuteurs de l'anglais ?
- Tester que le modèle appris sur le corpus switchboard a des propriétés de prédiction analogues à celle d'un humain qui réalise la même tâche en production.

## Rappel: prédiction du modèle

Le modèle inférentiel sur corpus est capable d'assigner une probabilité à chacun des deux ordres (V-NP-PP, V-NP-NP) On peut faire prédire le modèle inférentiel sur des exemples qu'il n'a jamais vus auparavant.

- **Hypothèse:** En production, les sujets vont marquer des préférences qui vont refléter les probabilités assignées par le modèle sur ce même exemple
- **Méthode:** On demande aux sujets de distribuer 100 points sur les deux alternatives d'items du type:

-----  
Speaker:

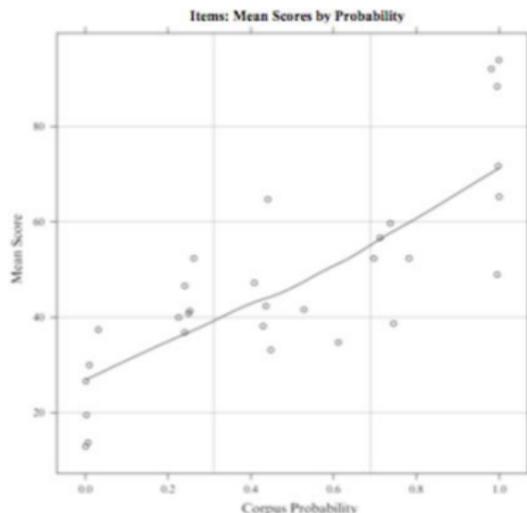
About twenty-five, twenty-six years ago, my brother-in-law showed up in my front yard pulling a trailer. And in this trailer he had a pony, which I didn't know he was bringing. And so over the weekend I had to go out and find some wood and put up some kind of a structure to house that pony,

(1) because he brought the pony to my children.

(2) because he brought my children the pony.  
-----

# Résultat (I)

On a une corrélation non nulle.



*On cherche à identifier une corrélation : quand le modèle vote bas pour l'ordre V-NP-PP on espère que les sujets votent bas pour cet ordre et vice-versa. La corrélation parfaite est une droite inclinée à 45 degrés.*

# Résultat (II)

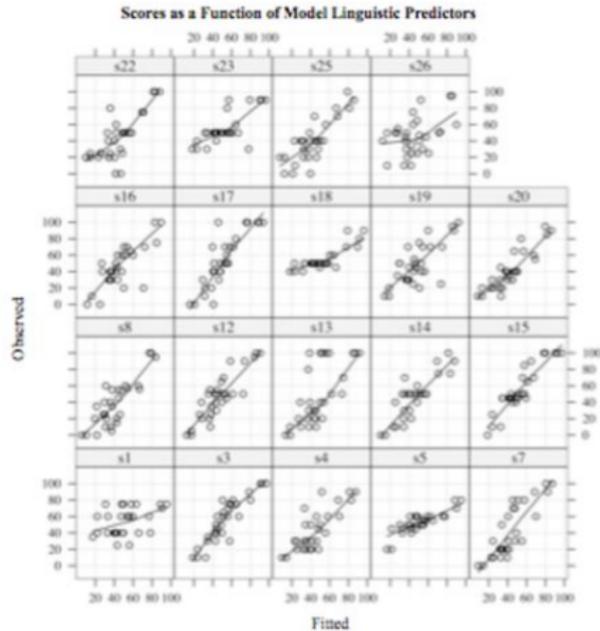


Figure 6. Fit of linear mixed effects model to Experiment 1 scores

*On a également une corrélation pour chaque sujet de l'expérience, ce qui renforce l'idée de la strong competence hypothesis*

- Corpus permet d'observer les données (toutes les variables varient librement) mais pose problème pour les corrélations. Les modèles construits sont prédictifs: il est possible de tester (valider, invalider) un modèle sur des données non utilisées pour le construire et de manière générale estimer sa correction.
- Expériences permettent de tester un seul facteur en fixant les autres. Permet de tester explicitement une hypothèse en procédant à une manipulation des données (test à caractère causal).
- Par exemple, (Brown et al. 2012) proposent une expérience de self paced reading destinée à tester spécifiquement si la structure informationnelle (réduite à given/new) constitue une contrainte sur le choix de la structure dative en étant certains que les corrélations sont explicitement contrôlées (conception de l'expérience qui contrôle définitude et animacy).

# Conclusion



# Les chausse-trapes et les bénéfices...



Déjà une histoire et donc du recul dans l'utilisation des différentes techniques ... et une intense réflexion sur les biais menaçant la production et l'interprétation des résultats.

**Il n'y a pas de résultats définitifs ou de résultats vrais. Il n'y a pas d'expérience cruciale. Il n'y a que des résultats dont l'obtention ou l'interprétation sont contestables.**

# Les chausse-trapes



## 1. L'échantillonnage

- 1.1: des sujets,
- 1.2: des « stimuli »

## 2. L'écologie de la tâche:

- 2.1: compréhension de la tâche par les sujets,
- 2.2: naturalité de la tâche,
- 2.3: biais de présentation

## 3. Jugements vs grammaire en acte (compréhension/ production)

# Bénéfices



- ∞ - Stabilisation des observations
- ∞ - Découverte de nouveaux facteurs (relevant soit des processus langagiers soit du phénomène étudié) et donc de nouvelles hypothèses explicatives.
- ∞ - Explicitation maximale des analyses

# Vœu (pieux ?)



L'expérimentation (incluant la manipulation de corpus) devrait jouer dans le futur proche le rôle que la formalisation a joué à l'origine du programme dans les années 50-80.