

# Avancement du TER

## – Méthodes expérimentales en linguistique –

Master de Sciences du Langage  
« Linguistique Théorique & Expérimentale »  
P. Amsili

17 novembre 2015

# Plan

## 1 Linguistique et empirie

## 2 Méthodes expérimentales

## 3 Exemples détaillés

- Crowdsourcing
- Magnitude estimation
- Gries

## Linguistique et empirie

La linguistique a un rapport ancien avec l'empirie\*

- collecte de *ressources* : dictionnaires, thésaurus, concordances...
- collecte d'*attestations*
- transcriptions/enregistrements de données de terrain

\* *empirie* : ensemble des données de l'expérience (TLFi)

## Révolution chomskyenne

- Linguistique de l'attestation = linguistique de la **performance**
- Comment atteindre la **compétence** ?
  - ⇒ Nouveau type d'« évidence » linguistique : le jugement d'acceptabilité

## Concepts fondamentaux

- Grammaticalité : propriété d'un objet linguistique (compatibilité avec la grammaire d'une langue donnée)
- Acceptabilité : propriété d'un stimulus linguistique tel qu'il est perçu par un locuteur
- Jugement d'acceptabilité : réponse (comportementale) d'un locuteur à la demande d'un linguiste

Bard *et al.* (1996)

## Catégoricité vs gradabilité

- Acceptabilité relative : différence d'acceptabilité entre deux stimuli  
admise par tous les linguistes : \* \*\* ? # ?? ...
- Grammaticalité relative : différence de grammaticalité inhérente entre deux stimuli.  
controversée : l'existence d'un gradient de grammaticalité n'est pas compatible avec la plupart des théories syntaxiques actuelles

« the possibility remains that acceptability is graded because grammaticality is » (Bard *et al.* , 1996, p.33)

## Problèmes avec les jugements

- reflet indirect et infidèle (?) de la grammaticalité
- influence de facteurs non pertinents :
  - prise en compte de la fréquence d'usage (estimée)
  - conformité à une norme prescriptive
  - conformité à un registre évalué socialement
  - degré de plausibilité sémantique/pragmatique
- biais de l'expérimentateur

Labov (1972)

# Plan

1 Linguistique et empirie

2 Méthodes expérimentales

3 Exemples détaillés

- Crowdsourcing
- Magnitude estimation
- Gries



# Méthodes expérimentales

- Visant à mieux contrôler les jugements d'acceptabilité
  - multiplication des sujets (banissement de l'introspection)
  - questionnaires d'acceptabilité avec échelles
  - tâches de production induite
- Visant à réduire le caractère méta-linguistique de la tâche
  - ⇒ plus indirect mais moins biaisé
  - temps de traitement (lecture auto-segmentée, mesures oculométriques...)
  - temps de réaction (tâches de décision lexicale, de déomination)
  - imagerie (ERP, IRMf...)
- Visant à introduire de nouvelles « évidences »
  - tests de perception (phonétique par exemple)
  - enregistrements psychophysiques (laryngoscopie, glottoscopie, imagerie cérébrale...)
  - enregistrements comportementaux (eye-tracking, mouse-tracking, tâches de compréhension, map task...)
  - importation de l'arsenal expérimental des sciences cognitives

## Retour de la performance

- la linguistique de l'attesté a été rejetée dans le paradigme chomskyen, un des arguments étant le manque de données négatives
- les méthodes statistiques modernes répondent (en partie) à cet contre-argument :
  - l'étude statistique des **distributions** et des **correlations** dans
  - un échantillon **représentatif** permet de tirer des conclusions
  - même sur des phénomènes **rare**s voire **non attestés**

⇒ Linguistique quantitative sur corpus

(Gries, 2013)

# Plan

1 Linguistique et empirie

2 Méthodes expérimentales

3 Exemples détaillés

- Crowdsourcing
- Magnitude estimation
- Gries

# Crowdsourcing

Cas particulier d'une méthode expérimentale : utilisation des techniques de *crowdsourcing* (externalisation participative/ouverte) pour la constitution de ressources linguistiques ou la réalisation d'expériences.

voir les tranparents qui suivent, qui proviennent d'une présentation par l'un des auteurs du papier (Munro *et al.* , 2010).

<http://web.stanford.edu/~rmelnick/files/SchnoebelenEtAlLSA2011.pdf>

# A new generation of data: Crowdsourcing and language studies

Tyler Schnoebelen, Robert Munro, Steven  
Bethard, David Clausen, Victor Kuperman,  
Vicky Tzuyin Lai, Robin Melnick, Christopher  
Potts, and Harry Tily

# What's this presentation about?

- Chairs









# What's this presentation about?

- ~~Chairs~~
- Giving you new chairs



# What's this presentation about?

- ~~Chairs~~
- ~~Giving you new chairs~~
- Giving you fuller chairs







Victor Kuperman



Tyler Schnoebelen

Transparency of phrasal verbs

Segmentation of an audio speech stream



Harry Tily



Robin Melnick

Judgment  
gramm

Contextual predictability

stic

**EXTRA BONUS IF YOU STAY!**

Best practices!!!!

corpus trends



Robert Munro



Steven Bethard



Vicky Tzuyin Lai

Phonetic and phonological frequency analysis of  
electrophysiological responses



Christopher Potts

Filler  
informativity  
and island  
constraints



David Clausen

**BONUS!** Screening for linguistic attentiveness

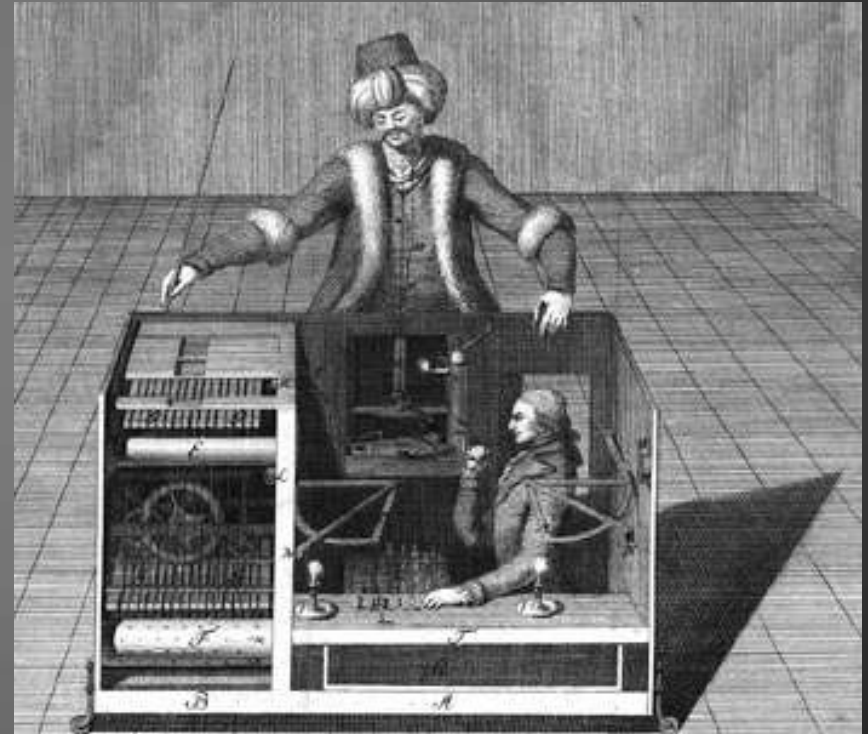
# Crowdsourcing

- First use of crowdsourcing:
  - Get everyday folks to annotate and translate sentences
  - (These are still great uses for crowdsourcing)
- Other uses:
  - Generating targeted corpora
  - Checking linguistic intuitions
  - Learning about language ideology
- Focus here today:
  - Crowdsourcing offers faster and cheaper results with broader demographics than most lab studies
  - But are the results reliable?



# Amazon Mechanical Turk

- Launched at the end of 2005
- Originally for “artificial artificial intelligence”
  - Stuff that’s easy for people to do and SHOULD be easy for computers to do...but isn’t
  - “What’s the focus of this photograph?”
- Lots of alternative services to consider:
  - CrowdFlower is especially geared towards academics
  - clickworker.com *may* be useful in Europe
  - See also <http://www.crowdsortium.org/> and <http://crowdsourcing.org>



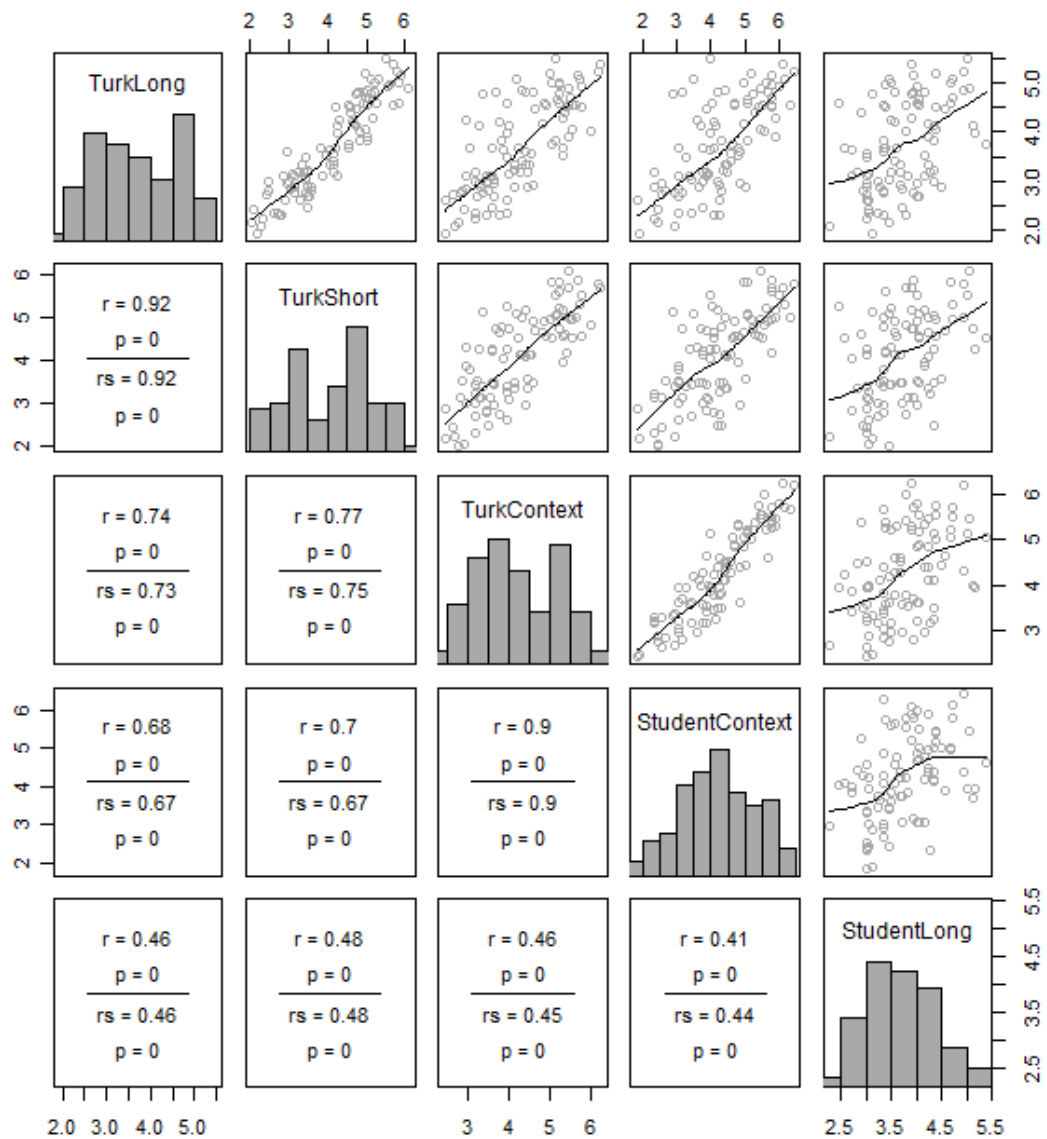
# (1) Transparency of phrasal verbs

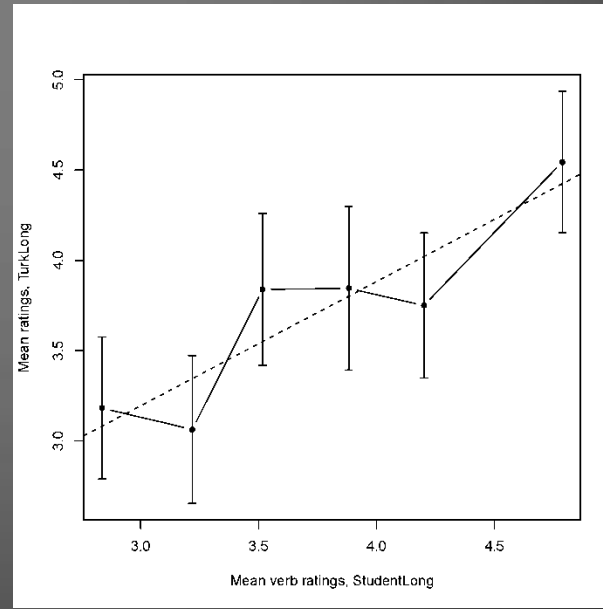
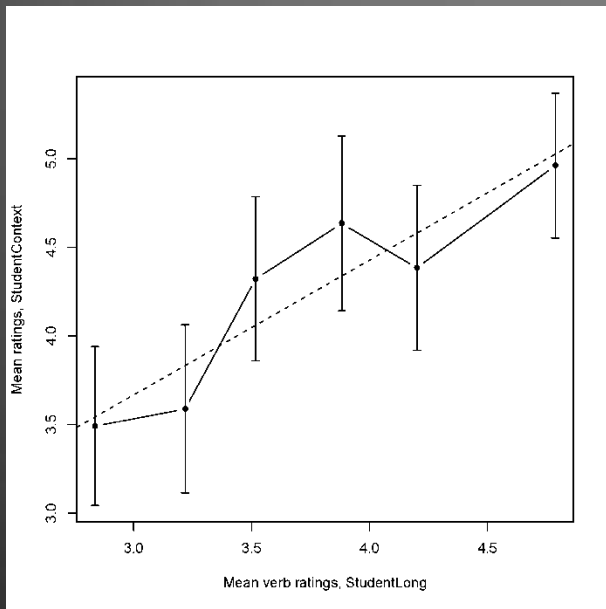
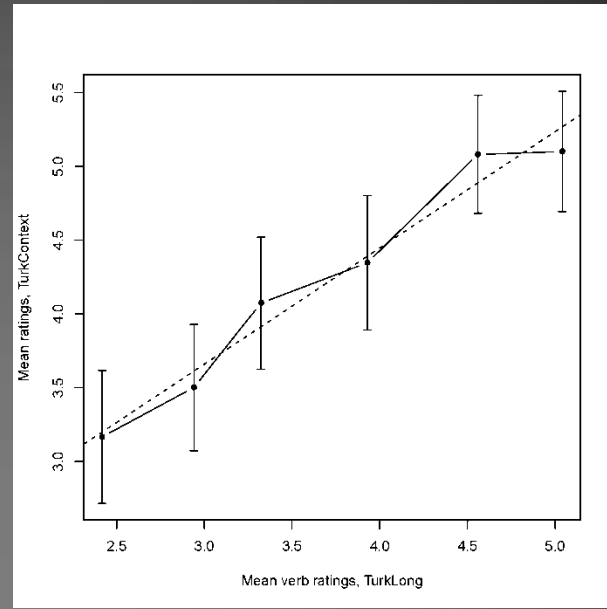
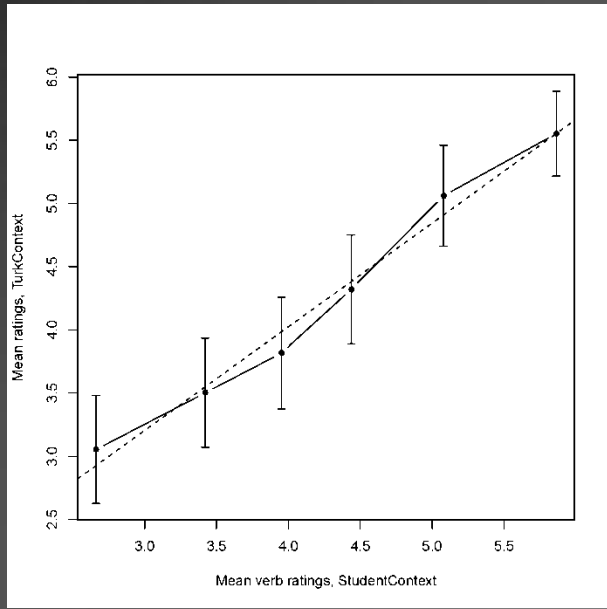
- **Transparent** phrasal verbs are easy to figure out: *lift up* entails something being lifted and going up.
- **Opaque** phrasal verbs are more confusing: when you *give up* what do you give? What goes up?
- 3 Turk experiments, 2 undergrad experiments
  - 215 participants
  - 96 phrasal verbs
  - With and without context (examples of phrasal verbs in sentences)

On a scale of 1–7, how similar are  
'cool' and 'cool down'?

# Super-cheap and fast

- The more you pay, the faster results will come in. Even if you're a cheapskate:
  - Most studies have results done in under a week, usually just a couple days
- To get 20 responses for each of 96 phrasal verbs:
  - \$11.60
- So obviously, if it's reliable that's terrific. But can we trust the results?





# Other measures show reliability

	StudentContext and TurkContext	All tests but StudentLong	StudentLong
ICC consistency	0.899	0.780	0.0934
ICC agreement	0.900	0.740	0.0854

- High intra-class correlation coefficient (ICC) values-- except for the wayward students in StudentLong
- Cohen's kappa (weighted):
  - $\kappa = 0.700$ , for all tests excluding StudentLong
  - $\kappa = 0.835$  for StudentContext and TurkContext
  - $\kappa > 0.8$  is "almost perfect agreement" (Landis and Koch 1977)

# References

- Bard, Ellen Gurman, Robertson, Dan, & Sorace, Antonella. 1996. Magnitude estimation of linguistic acceptability. *Language*, 72(1), 32–68.
- Gries, Stefan Th. 2013. *Statistics for linguistics with R : a practical introduction*. Walter de Gruyter.
- Labov, William. 1972. Some principles of linguistic methodology. *Language in society*, 1(01), 97–120.
- Munro, Robert, Bethard, Steven, Kuperman, Victor, Lai, Vicky Tzuyin, Melnick, Robin, Potts, Christopher, Schnoebelen, Tyler, & Tily, Harry. 2010. Crowdsourcing and language studies : the new generation of linguistic data. *Pages 122–130 of : Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk*. Association for Computational Linguistics.