



Call for papers Appel à communications



Organizers: Juliette Thuilier, Margaret Grant
Partners: ATALA, Labex EFL, INALCO

Quantitative approaches in corpus linguistics and psycholinguistics: Word order and constituent order

Paris, October 18, 2013

(Version française à suivre.) The goal of this research day is to bring together computational linguists, corpus linguists and psycholinguists who work on word order alternation phenomena in various languages.

The development and increased availability of statistical tools for sophisticated data analysis in the past decade has changed the face of quantitative linguistic analysis. In particular, alternation phenomena (in which a speaker may choose one out of several possible linguistic forms) have seen a renewed interest due to advances in statistical modeling (Gries, 2003; Benor and Levy, 2006; Bresnan et al., 2007). The recent studies have mostly modeled English corpus data, presumably in part because of the development of large tagged corpora of written and spoken English. However, there are now many research programs developing high quality, large-scale corpus resources for languages other than English.

In parallel to corpus-based studies, psycholinguistic work has also examined the factors that contribute to speakers' choices with respect to word or constituent order in production (e.g., Tanaka et al., 2011) and influence processing complexity in comprehension (e.g., Brown et al., 2012).

By integrating findings from both corpus studies and experimental research, this workshop has as its goal to make progress on understanding the underlying processes governing alternation phenomena. We also believe that developing corpus models on a diverse set of languages and alternations will give rise to interesting observations regarding the processes underlying alternation phenomena. Therefore, submissions of corpus and/or psycholinguistic work handling alternation phenomena in a variety of languages are of particular interest.

A non-exhaustive list of possible topics includes:

- Corpus studies on alternation phenomena in a variety of languages, including constituent or word order alternations and syntactic construction alternations.

- Experimental studies of alternation phenomena and the psycholinguistic factors that influence their comprehension or production.
- Theoretical issues on corpus probabilities: What do corpus probabilities represent? How do observed patterns reflect or differ from results of psycholinguistic experiments?

We invite submissions of one-page, .pdf-format abstracts in 12-point font, with an additional page for data, references and figures, for 25-minute presentations with 10 minutes for questions. The deadline for submission is August 15, 2013, and notifications of acceptance will be made by September 16. Please submit abstracts by e-mail to journee.labex.atala@gmail.com.

Approches quantitatives en corpus et psycholinguistiques: Ordre des mots et des constituants

Paris, October 18, 2013

L'objectif de cette journée scientifique est de réunir des linguistes computationnels, des linguistes de corpus et des psycholinguistes travaillant sur des phénomènes de variation d'ordre des mots ou des constituants dans différentes langues.

Le développement et la disponibilité accrue d'outils d'analyse de données sophistiqués ont modifié le visage de la linguistique quantitative durant la dernière décennie. En particulier, les phénomènes d'alternance (dans lesquels le locuteur doit choisir une forme parmi plusieurs possibles) ont connu un regain d'intérêt grâce aux avancées dans le domaine de la modélisation statistique (Gries, 2003; Benor and Levy, 2006; Bresnan et al., 2007). Les études récentes ont pour la plupart modélisé des données écrites et orales de l'anglais, mais des résultats sur d'autres langues commencent à émerger.

En parallèle aux études sur corpus, les travaux en psycholinguistique ont également examiné les facteurs qui contribuent aux choix d'un ordre de mots ou de constituants en production (par ex. Tanaka et al., 2011) et influencent les temps de traitement (processing) en compréhension (par ex. Brown et al., 2012).

En intégrant les résultats obtenus en corpus et en recherche expérimentale, cette journée espère faire avancer notre compréhension des processus sous-jacents qui régissent les phénomènes d'alternance. En développant la modélisation de données de corpus sur un ensemble varié de langues et de phénomènes, des résultats pourront également émerger en ce qui concerne les processus sous-jacents. De ce point de vue, les approches quantitatives sur corpus (écrits et oraux)

et les travaux psycholinguistes qui s'intéressent aux phénomènes d'alternance dans différentes langues seront particulièrement bienvenus.

Le but de cette journée est donc d'apporter un éclairage sur ces questions générales en rassemblant des psycholinguistes et des linguistes travaillant à partir de la modélisation statistique de données de corpus sur des thématiques relatives à des phénomènes d'alternance d'ordre dans différentes langues.

Voici une liste non-exhaustive des thèmes pouvant être abordés:

- Approche quantitative de données de corpus et modélisation statistique de phénomènes d'alternance de mots et de constructions mots dans des langues diverses ;
- Études expérimentales sur des phénomènes d'alternance et les facteurs psycholinguistiques qui influencent leur compréhension ;
- Problèmes théoriques concernant les probabilités de corpus: que représentent les probabilités en corpus ? Les observations en corpus correspondent-elles ou diffèrent-elle des résultats obtenus dans des expériences psycholinguistes ?

Les propositions de communication (durée 25 min + 10 min de discussion) sont à envoyer à journee.labex.atala@gmail.com. Les personnes intéressées devront soumettre un résumé d'une page (12 pt) avec une page supplémentaire pour les exemples, tableaux, figures et références. Les propositions devront être pdf. La date limite de soumission est fixée au 15 août 2013 et les notifications seront envoyées le 16 septembre.

References

- Benor, Sarah Bunin, and Roger Levy. 2006. The chicken or the egg? a probabilistic analysis of English binomials. *Language* 82:233–278.
- Bresnan, Joan, Anna Cueni, Tatiana Nikitina, and R. Harald Baayen. 2007. Predicting the dative alternation. In *Cognitive foundations of interpretation*. Royal Netherlands Academy of Arts and Sciences.
- Brown, Meredith, Virginia Savova, and Edward Gibson. 2012. Syntax encodes information structure: Evidence from on-line reading comprehension. *Journal of Memory and Language* 66:194–209.
- Gries, Stefan. 2003. Towards a corpus-based identification of prototypical instances of constructions. *Annual Review of Cognitive Linguistics* 1:1–27.
- Tanaka, Mikihiro N., Holly P. Branigan, Janet F. McLean, and Martin J. Pickering. 2011. Conceptual influences on word order and voice in sentence production: Evidence from Japanese. *Journal of Memory and Language* 65:318–330.