

Outils Linguistiques

Jean-Yves Antoine
LI - Université Rabelais de Tours
Jean-Yves.Antoine@univ-tours.fr

Mastere L&R — U. Tours



Outils Linguistiques

**TRAITEMENT DE LA PAROLE ET CORPUS ORAUX :
INTRODUCTION**

Mastere L&R — U. Tours



Pourquoi la parole ?

- **Domaine d'étude de plus en plus reconnu en linguistique**

- **Phonétique** : domaine de recherche ancien très bien établi
- **Linguistique de corpus** : redécouverte du langage parlé après un long purgatoire dû à la vision normative des grammairiens (compétence = écrit vs. Performance = oral)



- **Type de langage de plus en plus étudié en TAL**

- **Technologies langagières** : applications vocales commercialisées (synthèse de la parole, dictée vocale, serveurs vocaux interactifs)
- **Phonétique** : connaissances théoriques requises pour la modélisation des phonèmes et de la coarticulation (diphones...)
- **Corpus oraux** : apprentissage de modèles statistiques sur des corpus audio mais également sur leurs transcription orthographique



Pourquoi la parole ?

- **Parole, le langage par essence**

- **Parole spontanée**

- production en direct, le plus souvent sous la pression d'un dialogue, d'un message langagier : « ratés » de production (répétitions, reprises, lapsus)
- indices de la cognition humaine impliquée dans les activités langagières
- révélateurs du système de la langue : ces « ratés » et leurs stratégies de rattrapage répondent à une logique de la langue.

- **Langage non écrit**

- production plus libre, plus indépendante de la norme sociolinguistique
- témoin très précieux de l'évolution diachronique de la langue : l'oral précède généralement l'écrit dans ses évolutions.

Parole et corpus

- **Signal de parole**
 - **Corpus** : signal audio
 - **Analyse** : caractéristiques spectrales et phonétiques du signal de parole
- **Langage parlé**
 - **Corpus** : transcription phonétique ou orthographique (+ corpus augmentés)
 - **Transcription** : valeur ajoutée au corpus audio
 - **Analyse** : caractéristiques linguistiques (lexique, syntaxe, pragmatique) du langage parlé

Cours « outils linguistiques »

- **Objectifs**
 - Compréhension du fait de parole au niveau linguistique
 - Outils pour la gestion des corpus oraux (et non pas pour le TALP)
- **Signal de parole**
 - *approche théorique* production du signal de parole, caractéristiques spectrales et phonétiques du signal de parole
 - *approche pratique* outils d'édition de signaux de parole : spectrogramme, détection de formants, de fréquence fondamentale (pitch)
- **Langage parlé**
 - *approche théorique* langage parlé: spécificités syntaxiques de la parole spontanée
transcription de corpus oraux: problèmes méthodologiques
 - *approche pratique* outils de transcription de corpus oraux : *Transcriber*
- **Un absent...**
 - Méthodologie de recueil de corpus oraux

