

UN EXEMPLE D'ANNOTATION SUR CORPUS ORAL TRANSCRIT: PFC

Le projet PFC

- Projet « socio-phonologique » visant à décrire la prononciation du français et la réalité de ses usages attestés, dans un cadre variationniste
 - variation diatopique avant tout
 - variations sociales ou variations diaphasiques dans une moindre mesure
- Constitution d'une base de données orales
- Exploitation de ces données
- <http://www.projet-pfc.net/>

Recueil des données

- Différents points d'enquête
- Une douzaine de personnes (h/f, âges variés, etc.)
- Deux séquences de lecture
 - liste de mots
 - texte
- Deux séquences de conversation
 - guidée
 - libre
- Homogénéité et comparabilité des enquêtes

Exploitation des données

- Analyses phonétiques et phonologiques
- Analyses syntaxiques, pragmatiques, sociolinguistiques, conversationnalistes...
- Didactique du français
- Etc.

- Exploitation initialement visée
 - schwa
 - liaison

Préparation des données

□ Transcription

- transcription orthographique standard
- séquences de lecture + extraits des séquences de conversation
- logiciel Praat

□ Annotation

- tires distinctes pour les deux phénomènes à annoter
- système de codage chiffré

Transcription



Son (OLE2)

File Edit Query View Select Interval Boundary Tier Spectrum Pitch Intensity Formant Pulses Help

ZB : C'est-à-dire tellement le, le volume est trop élevé donc on, pour euh, (rires)

3.964

1

ZB : Exactement, tellement le volume de travail est <X : Il y a toujours>.

X : du monde, il y a toujours euh, on a pas le temps.

acazb1qanscr (10/97)

159.311828 159.311828 Visible part 2.596132 seconds 161.907960 736.812040

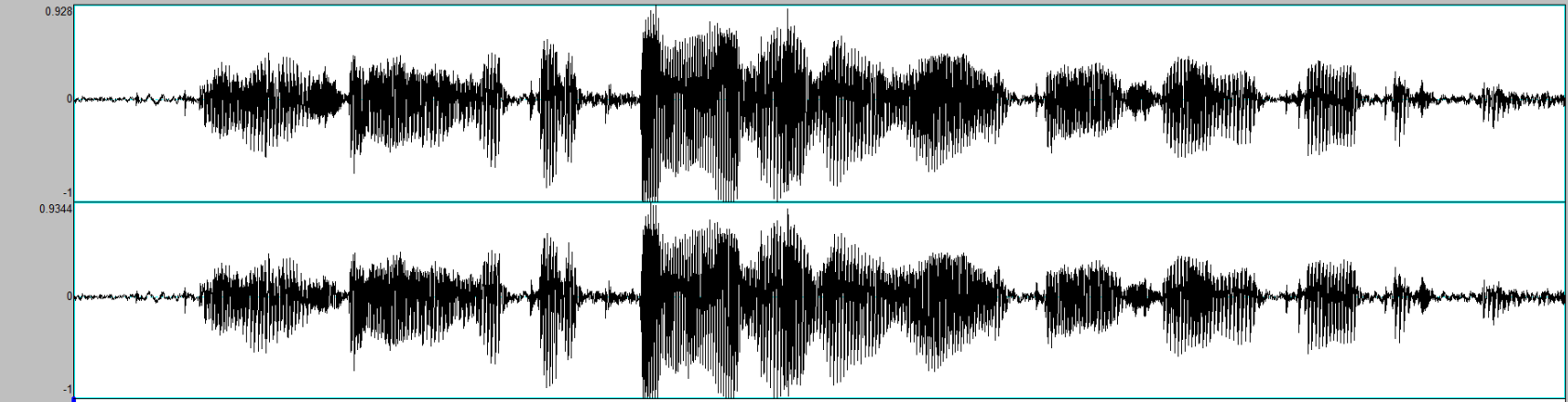
Total duration 898.720000 seconds

all in out sel Group

Annotation

File Edit Query View Select Interval Boundary Tier Spectrum Pitch Intensity Formant Pulses Help

MA : je0132 connais tout le0112 monde0414,



1 MA : Pendant une semaine, j'étais le premier de la région à commencer avec eux, parce que,

2 MA : Pendant une0412 se1212maine0414, j'étais le2112 pre1212mier de0112 la région à commencer avec0411 eux, parce0412 que1124,

3 MA : Pendant20 une semaine, j'étais le premier de la région à commencer20 avec eux, parce que,

364.217539 364.217539 Visible part 3.576995 seconds 367.794534 645.880133

Total duration 1013.674667 seconds

all in out sel Group

Annotations

- Système chiffré dans lequel chaque chiffre mesure un trait
- Schwa = 4 paramètres
je veux [ʒvœ] => *je0132 veux*
- Liaison = entre 2 et 6 paramètres
pâtes italiennes [patitaljɛn] => *pâtes10 italiennes*
mon ami [mɔnami] => *mon11nVO ami*
il pourrait préférer [ilpʁɪkɛtPɪpʁɛfɛʁɛ] => *il*
pourrait21thC préférer

Principes généraux de l'annotation

- Manuelle
- Descriptive
- Limitée à quelques traits essentiels
- Fondée sur les perspectives théoriques les plus largement partagées
- À grande échelle – approche globale des données
- En « détournant » un outil non dédié

- Réalisation par des non-spécialistes (éventuellement)
- Homogénéité et comparabilité des annotations

Lieux d'annotation (1)

- Définition du phénomène à annoter
 - liaison => consonne latente prononcée lorsque le mot suivant a une initiale vocalique
 - schwa => élément vocalique à prononciation facultative correspondant au e de l'orthographe non suivi de consonne appartenant à la même syllabe

Lieux d'annotation (2)

- Réalisation du phénomène à annoter
 - liaison réalisée => *les_amis* [lezami], *ils sont_amis* [ilsõtami]
 - schwa réalisé => *je viens* [ʒəvjẽ]
- Non réalisation du phénomène
 - liaison non réalisée => *très apprécié* [tʁeapʁesje]
 - schwa non réalisé => *j'viens* [ʒvjẽ]

Lieux d'annotation (3)

- Réalisation imprévue du phénomène
 - liaison épenthétique ou devant consonne => *il va à Paris* [ilvatapaxi] , *il allait vers Paris* [ilalɛtvɛpaxi]
 - schwa non marqué par un e graphique
 - à l'initiale pure d'un mot (métathèse) => *je viens* [əʒvjɛ̃]
 - à la finale d'un groupe rythmique (e prépausal) => *un concours* [œ̃kõkuʁə], *salut* [salyə]
- Annoter la non réalisation imprévue ?
 - liaison épenthétique ou devant consonne => annotée si réalisée
 - schwa non marqué par un e graphique =>
 - métathèse => annotée si réalisée
 - e prépausal => annotation de toutes les consonnes finales prononcées

Lieux d'annotation (4)

- Annotation idéale: tous les lieux possibles de schwas et de liaisons
- Annotation réelle: les réalisations imprévues ne sont annotées que lorsqu'elles sont réalisées.

- Lieux possibles:
 - connus et indiqués pour le texte (apprentissage)
 - devant être repérés par l'annotateur pour les extraits de conversations (risque d'erreurs et d'oublis)

Traits à annoter: schwa (1)

- Visée générale: présence / absence du schwa => 1^{er} trait à annoter
- Repère: e graphique, quelle que soit la réalisation phonétique (+ consonnes finales prononcées)
- Chiffre 1 => présence / absence (ou incertitude)

Traits à annoter: schwa (2)

- Visées plus précises : mettre la présence / absence en relation avec le contexte phonologique
- Position:
 - schwa en syllabe initiale (ou unique) => réalisations variées selon contexte et rythme
 - schwa en syllabe interne ou finale => non réalisation
- Chiffre 2 => syllabe initiale (ou unique) / interne / finale

Traits à annoter: schwa (3)

- Contextes droit et gauche de la syllabe concernée:
 - règle du contexte droit: si le schwa est suivi d'une voyelle ou se trouve en fin de mot, il n'est pas prononcé
 - règle des trois consonnes: si le schwa est précédé d'au moins deux consonnes et suivi d'au moins une consonne, il est réalisé; sinon, il n'est pas réalisé
- Chiffre 3 => contexte gauche (voyelle, consonne, début de groupe intonatif, cas particuliers)
- Chiffre 4 => contexte droit (voyelle, consonne, frontière intonative)



Traits à annoter: schwa (exemple)



- ZB : *On parle français oui, on parle* <X : *On parle français, entre nous.*><E2 : *Même les conversations intimes ?*>.
- ZB : *On parle français oui, on parle* <X : *On parle français, entre nous.*><E2 : *Même les conversations intimes ?*>.
 - ZB : *On parle1 422 français oui, on parle0423*
 - X : *On parle0422 français, entre1 41 2 nous.*
 - E2 : *Même041 2 les conversations intime041 3s ?.*

Analyses à partir de l'annotation

□ Variation diatopique

- français méridional => schwa normalement prononcé en fin de polysyllabe (suivi d'une consonne) 
Son Wave
- français canadien => tendance à l'effacement du schwa et de là à la réduction consonantique 
Son Wave

□ Variation diaphasique

- plus le registre de langue est soutenu, plus le taux de réalisation du schwa est élevée
- plus le registre est informel/familier, plus le taux de réalisation du schwa tend à être bas

Traits à annoter: liaison (1)

□ Deux traits nécessaires:

- Chiffre 1 => caractérisation syllabique du mot liaisonnant (mono ou polysyllabe)
- Chiffre 2 => absence / présence de la liaison + caractérisations supplémentaires

Soyez efficace [swajeefikas] => *soyez20 efficace*

Soyez efficace [swajePHefikas] => *soyez20h efficace*

Soyez efficace [swaje?efikas] => *soyez23 efficace*

Soyez efficace [swajezeфикas] => *soyez21z efficace*

Soyez efficace [swajePHzefikas] => *soyez21zh efficace*

Soyez efficace [swajezPHefikas] => *soyez22z efficace*

Traits à annoter: liaison (2)

- Caractérisations supplémentaires:
 - Enchaînement ou non-enchaînement => chiffre
 - Si présence de la liaison => consonne de liaison
 - Si pause-hésitation => *h*
 - Si initiale consonantique au mot suivant => C
 - Si incertitude, liaison épenthétique => chiffre
 - Si liaison en *n* => indication du caractère nasal ou pas
- Annotation moins régulière, plus précise, pour un phénomène moins fréquent

Traits à annoter: liaison (exemples)

- *X : mais par contre nous, on a souffert avec nos enfants,*
- *X : mais par contre nous, **on a** souffert avec nos **enfants,***
 - *X : mais par contre nous, on1 1NVN a souffert avec nos1 1Z enfants,*



Son Wave

Traits à annoter: liaison (exemples)

- ZB : C'est-à-dire tellement le, le volume est trop élevé donc on,
- ZB : C'est-à-dire tellement le, le volume est trop élevé donc on,
 - ZB : C'est-à-dire tellement le, le volume est trop élevé donc on,



Son Wave

Analyses à partir de l'annotation?

- Nombreux facteurs de variation
- Contextes linguistiques
 - Lexique (certains mots liaisonnants)
 - Syntaxe (liaison à l'intérieur de syntagmes)
 - Morphologie (marquage du pluriel)
 - Prosodie (liaison à l'intérieur de groupes rythmiques)
 - Phonétique (certaines consonnes finales bloquent la liaison)
 - Longueur du mot (mots liaisonnants courts)
 - Fréquence des contacts (locutions par exemple)
 - Nature de la consonne ([n], [z], [t])
- Paramètres sociolinguistiques
 - Diaphasie (registre soutenu liaisonne plus)
 - Diachronie (âge du locuteur)
 - Diatopie (peu pertinente)

Problèmes rencontrés

- Qualité sonore
- Importance de l'annotateur
- Fiabilité des annotations ?

- Imprécisions dans l'annotation (*euh*)
- Manques dans l'annotation (contexte syntaxique liaison)

- Nécessité de raffiner l'annotation pour des analyses précises
- Utilité d'un « balayage annotatif »?