

De l'importance de l'homogénéisation des conventions de transcription pour l'alignement automatique de corpus oraux de parole spontanée

Dominique Fohr, **Odile Mella**, Denis Juvet
LORIA-INRIA
Nancy France



Introduction

- Retour d'expérience sur l'importance de l'homogénéisation des conventions de transcription pour le bon fonctionnement de l'alignement automatique texte-parole
 - transcrire un corpus est long et coûteux
 - mutualisation des corpus et réutilisabilité
- Alignement automatique de 180 heures de parole spontanée transcrite avec le logiciel Transcriber



Contexte

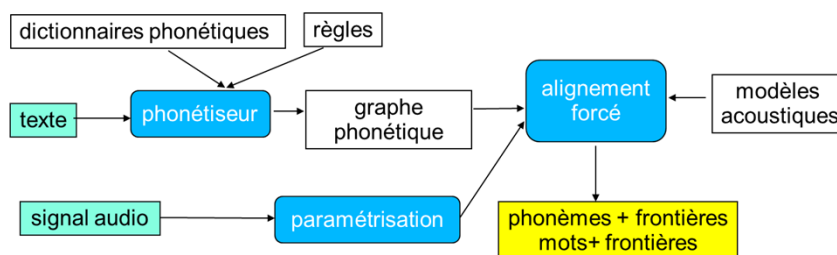
- Projet ANR ORFEO (Outils et Recherches sur le Français Ecrit et Oral)
<http://www.projet-orfeo.fr>
- coordonné par J-M Debaisieux (LATTICE)
- Offrir à la communauté scientifique un Corpus d'Étude pour le Français Contemporain écrit et oral
- Démarré début 2013 avec comme partenaires :
 - LATTICE, MODYCO, ATILF, LIF, LORIA, CLLE-ERSS, ICAR
- Rassembler sur une plate-forme (ORTOLANG)
 - des corpus oraux existants dans différents laboratoires
 - en associant à chacun un ensemble de métadonnées et de différentes couches d'annotation (phonétique, lexicale, syntaxique, sémantique,...)
- La couche d'annotation la plus proche du signal audio
 - est la segmentation en mots et en phonèmes
 - utilité :
 - annotations automatiques suivantes
 - analyse acoustique segmentale ou suprasegmentale
 - extraction et écoute avec des concordanciers
 - reconnaissance automatique de la parole : modèles acoustiques
 - a été réalisée automatiquement par deux équipes du LORIA: SyNaLP (Jtrans) et Multispeech (Astali) (selon les corpus)

03/09/2015

JLC2015-Orléans D. Fohr, O. Mella, D.Jouvet

3

Fonctionnement d'un système automatique d'alignement texte-parole



03/09/2015

JLC2015-Orléans D. Fohr, O. Mella, D.Jouvet

4

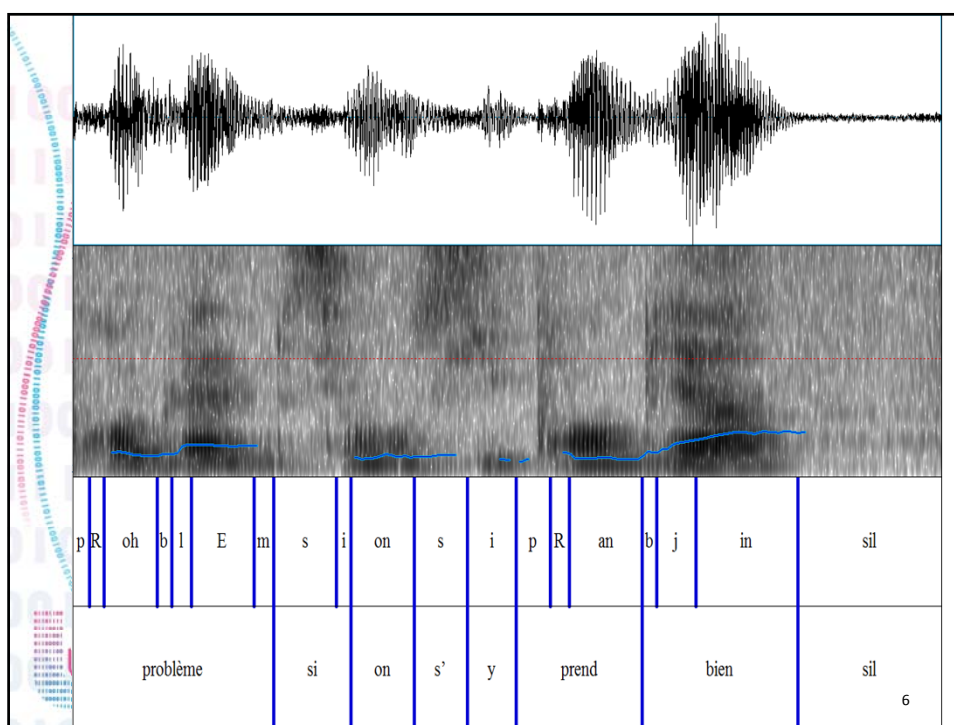
Fonctionnement d'un système automatique d'alignement texte-parole

- Fichier wav 
- Texte : « *et au décollage ça pose aucun problème si on s'y prend bien* »
- Résultat :

03/09/2015

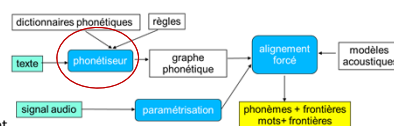
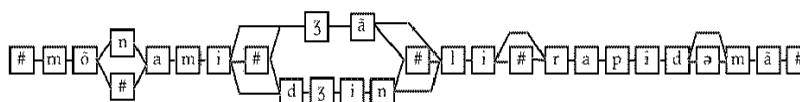
JLC2015-Orléans D. Fohr, O. Mella, D.Jouvet

5



Génération du graphe phonétique

- Générer un graphe phonétique correspondant aux différentes prononciations possibles de la phrase en utilisant éventuellement :
 - Dictionnaires phonétiques
 - Phonétiseur automatique
 - Règles
- Exemple pour la phrase « *mon ami jean lit rapidement* »



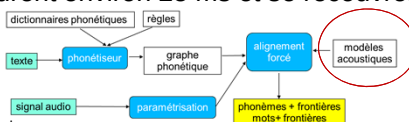
03/09/2015

JLC2015-Orléans D. Fohr, O. Mella, D.Jouvet

7

Modèles acoustiques

- Chaque élément du graphe phonétique est remplacé par son modèle acoustique
- Modèles statistiques de sons
 - phonèmes
 - non-parole (silence, respiration, bruits,...)
- Ces modèles sont appris automatiquement sur de grands corpus de parole transcrits orthographiquement
- Pour segmenter, l'algorithme affecte un modèle acoustique à chaque segment unitaire du fichier audio
 - les segments unitaires durent environ 25 ms et se recouvrent



03/09/2015

JLC2015-Orléans D. Fohr, O. Mella, D.Jouvet

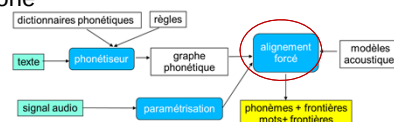
8

Alignement forcé

- L'algorithme cherche la meilleure séquence de modèles acoustiques qui correspond au signal audio :
 - « il essaie » tous les chemins possibles dans le graphe phonétique avec toutes les longueurs possibles pour chaque modèle acoustique
 - pour chacun de ces chemins il calcule un score
 - l'alignement résultat (étiquettes et frontières temporelles) est celui qui obtient le meilleur score

⇒ Inconvénient :

- il fournit toujours un alignement même si le texte ne correspond pas au signal audio
 - Si la transcription orthographique contient des erreurs → l'alignement sera erroné



03/09/2015

JLC2015-Orléans D. Fohr, O. Mella, D.Jouvet

9

ASTALI (Automatic Speech-Text Alignment Software)

- Outil d'alignement automatique développé au LORIA (Multispeech)
- Fonctionnalités dans le cadre du projet ORFEO
 - Alignement des fichiers transcrits avec Transcriber (trs)
 - Format de sortie Praat
 - Prise en compte des sigles, des noms propres, des unités lexicales contenant des chiffres
 - Phonétisation automatique des mots inconnus
 - Possibilité pour l'utilisateur de fournir son propre dictionnaire phonétique
 - Conservation de la casse et la ponctuation du texte original
 - Conservation des balises «Event» et «Comment» dans la sortie
 - Indication des locuteurs (deux « tiers » par locuteur)
 - Prise en compte de la parole superposée
 - Prise en compte des fichiers anonymisés
 - Pour obtenir des bonnes performances en cas de parole spontanée et/ou bruitée, les fichiers de transcription doivent comporter des marques temporelles environ toutes les minutes

03/09/2015

JLC2015-Orléans D. Fohr, O. Mella, D.Jouvet

10

Problèmes posés par l'hétérogénéité des conventions de transcription

- Alignement automatique avec ASTALI de 180 heures de corpus de parole spontanée
 - transcrits avec le logiciel Transcriber
 - transcriptions déjà en partie homogénéisées par l'ATILF (C. Benzitoun, ...) dans le cadre du projet ORFEO
- Malgré cette homogénéisation, certains phénomènes étaient transcrits différemment selon les corpus, voire au sein du même corpus
 - soit directement dans le texte
 - soit sous forme de balises Transcriber mais
 - de manière non homogène
 - et souvent en langage naturel

⇒ influence notable sur le résultat de l'alignement automatique



- modification de l'outil pour chaque corpus
- filtrage des informations -> impossibilité de prendre en compte certaines informations codées dans la transcription
- mauvais alignement

03/09/2015

JLC2015-Orléans

11

Exemples de conventions de transcription hétérogènes -> problèmes

- Disfluences :
 - Hésitations
 - mm mhm hum → Pb de phonétisation
 - Reprises (début de mot)
 - assem- assem*
 - assem- <Comment desc="amorcer"/> → filtrage avant phonétisation
- Bruits (rire, toux, bruits environnementaux)
 - <Event desc="pause" type="noise" extent="instantaneous" />
 - <Event desc=" bruit de chaise » type="noise" extent="instantaneous" />
 - <Event desc="conv" type="noise" extent="instantaneous" />
 - <Event desc="bea et marc enlèvent leurs manteaux et écharpes" type="noise" extent="instantaneous" />
 - <Event desc="un locuteur tousse" type="noise" extent="instantaneous"/>
 - <Event desc="toux" type="noise" extent="instantaneous"/>
 - <Event desc="tx" type="noise" extent="instantaneous"/>
 - <Event desc="L1 tousse" type="noise" extent="previous"/>
- Balise Event contenant un type « noise » et une description libre
- or la description est importante pour l'alignement → filtrage , modèles acoustiques ? 12

03/09/2015

Exemples de conventions de transcription hétérogènes -> problèmes

- Prononciations particulières (réduction, déviantes, mauvaises liaisons)
 - <Event desc="jave" type="pronounce" extent="previous"/>
 - <Event desc="[adeRi]" type="pronounce" extent="previous"/> → phonétisation
 - <Event desc="très long" type="pronounce" extent="previous"/>
- Parole incompréhensible
 - Il m'a dit hein <Event desc="suite de syllabes incompréhensibles" type="pronounce" /> je te jure
 - <Event desc="passage incompréhensible" type="pronounce" />
 - <Comment desc=" suite de syllabes incompréhensibles " />
 - Il y a un rapport avec le *** Ombres et Lumières → filtrage + modèles
 - * Ombres et Lumières
- Variantes d'interprétations
 - on (n') aura pas le temps → filtrage + phonétisation
 - il(s) /voudra, voudront/ jamais
- Indications de la langue dans laquelle a été prononcée une suite de mots
 - porque {lang = espagnol} → phonétisation
 - <Comment desc="prononcés à l'anglaise"/>
 - <Event desc="english" type="language" extent="instantaneous"/>
 - <Event desc="en" type="language" extent="instantaneous"/>
 - <Event desc="en anglais" type="language" extent="instantaneous"/>

03/09/2015

13

Exemples de conventions de transcription hétérogènes

- Anonymisation
 - Au niveau du texte
 - Deux types de fichier anonymisés
 - nom remplacé par un symbole : *P1*
 - nom remplacé par un nom générique :
 - » Prénoms -> Jean
 - » Noms -> Dupont
 - problème lors de la phase de phonétisation
 - Au niveau du signal audio
 - Variété des sons utilisés
 - Silence
 - Bip -> quel modèle acoustique utiliser ?
 - Bip trop intense -> problème d'écoute

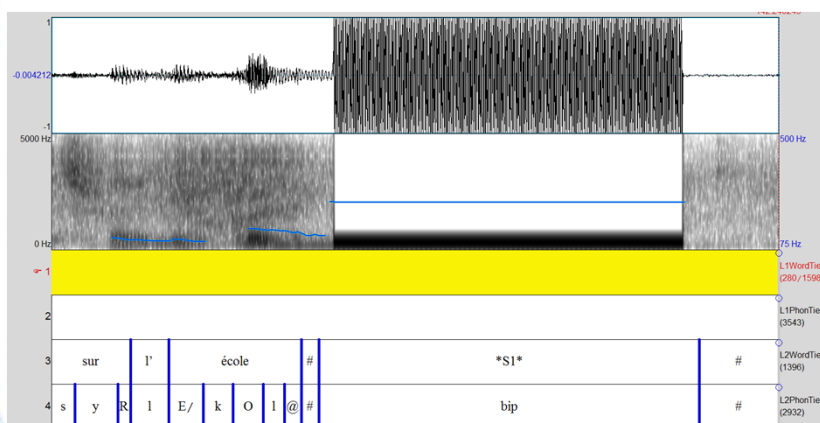
03/09/2015

JLC2015-Orléans

14

Exemple d'anonymisation

- transcription : OK
- signal audio : trop intense



03/09/2015

JLC2015-Orléans D. Fohr, O. Mella, D.Jouvet

15

Exemples de conventions de transcription hétérogènes -> problèmes

- Transcription de la parole superposée
 - Utilisation de marqueurs directement dans le texte :


```
<Turn speaker="spk1" startTime="212.664" endTime="214.476">
C'est marrant < l'essentiel
<Turn speaker="spk2" startTime="214.476" endTime="216.321">
mais euh > il y a quelqu'un qui vient
```
 - Utilisation de balises Who :


```
<Turn speaker="spk2 spk1" startTime="212.664" endTime="214.476">
<Sync time="212.664"/>
<Who nb="1"/>
de personnel de direction
<Who nb="2"/>
mais tout comme
</Turn>
```

03/09/2015

JLC2015-Orléans D. Fohr, O. Mella, D.Jouvet

16

Solutions possibles pour homogénéiser

- Avoir un format unique de transcription de l'oral
 - depuis quelques années des réflexions et des travaux sont en cours sur la définition d'un Format TEI (Text Encoding Initiative) pour l'oral :
 - au niveau international : Groupe de Travail EIT/MMI au sein du conseil technique du TEI (TEI Technical Council)
 - au niveau français : Groupe de Travail Interopérabilité du consortium IRCOM - Corpus Oraux et Multimodaux
 - <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/TS.html>
- mais aussi disposer d'outils qui permettent de transcrire les corpus oraux en respectant le format TEI
 - Transcriber, CLAN, ELAN, EXMARaLDA, FOLKER
- ... et enfin que les transpositeurs respectent le format et que le format soit suffisamment directif et que les outils vérifient

03/09/2015

JLC2015-Orléans D. Fohr, O. Mella, D.Jouvet

17

Quelques préconisations de conventions de transcription en vue d'un alignement automatique

- Primordiales :
 - ne rien coder directement dans la partie réservée au texte
 - Texte prononcé doit être à un emplacement spécifique et ne pas être mélangé avec d'autres informations
 - tout ce qui est prononcé doit être transcrit
 - Convention unique pour représenter les variantes orthographiques d'une même prononciation
 - Exemple : « il chante » ou « ils chantent »
 - Convention unique pour les bruits les plus fréquents (rires, respiration, toux, certains bruits d'environnement,...) avec un champ libre pour les autres bruits plus rares
 - dans le format TEI prévoir un type (ou un sous-type) avec une liste fermée pour les principaux bruits (avec éventuellement un item supplémentaire pour bruits « autres »)
 - En effet, le champ « desc » seul ne suffit pas car à contenu libre donc difficilement analysable de manière automatique
 - pour aligner le modèle acoustique qui correspond au bruit

03/09/2015

18

Préconisations sur les conventions de transcription en vue d'un alignement automatique

- Primordiales (suite) :
 - Convention unique pour la parole incompréhensible
 - Convention unique pour indiquer une prononciation spéciale pour un mot ou un groupe de mots avec
 - une convention de codage phonétique (unique)
 - puis [p i] cent euros [s a ~ z 2 R o] infarctus [e ~ f R a k t y s]
 - et des balises de début et de fin
 - Si un nom a été anonymisé dans l'audio il doit être indiqué dans le texte de manière explicite et non ambiguë avec éventuellement une balise en plus
 - Exemple : *L1*
- Optionnelles
 - Silences longs : spécifier début et fin (Sync time)
 - Bruits longs : spécifier début et fin (Sync time)
 - Convention unique pour les signaux d'anonymisation

03/09/2015

JLC2015-Orléans D. Fohr, O. Mella, D.Jouvet

19

Merci de votre attention !

- <http://astali.loria.fr>
 - version web du logiciel astali grâce au projet Ortolang
- <https://github.com/synalp/jtrans>
- <http://www.projet-orfeo.fr>
- <https://www.ortolang.fr/>

03/09/2015

JLC2015-Orléans D. Fohr, O. Mella, D.Jouvet

20