

**OUTILS POUR LE TRAITEMENT DES
PROÉMINENCES ACCENTUELLES
DANS LES CORPUS PFC :
AVANCEMENT DES TRAVAUX**

**Anne Lacheret-DUJOUR¹ ; Mathieu AVANZI^{1,2} ;
Bernard VICTORRI³**

¹Modyco, Paris Ouest Nanterre ; ² FNS, Neuchâtel ;

³Lattice/CNRS, Paris

Journées PFC, Paris, 10-13 décembre 2009

OBJECTIFS DE LA PRÉSENTATION

- Baliser différentes étapes d'un programme de recherche consacré à l'annotation prosodique des corpus oraux (spontanés) :
 - Étiquetage des proéminences syllabiques et des dysfluences
 - Acteurs concernés
 - Contribution scientifique de chaque étape dans avancement travaux
- Présenter un des outils d'annotation semi-automatique de la prosodie : ANALOR
 - Qui a profité de ces différentes contributions pour fournir les résultats qu'il fournit aujourd'hui
 - Peut s'avérer utile pour la communauté PFC





PARTIE 1
LES ÉTAPES

ETAPE 1 : DESCRIPTION

- Janvier 2003 : rencontre Lacheret-Lyche
 - Définir un codage prosodique mutualisé qui puisse éclairer le comportement du schwa dans les corpus PFC
 - Mise en place d'un protocole de codage (Lacheret & al. 2003, 2004) méthode + champs à coder + nature des champs
 - Codage CL,CD, réalisé sur 5 points d'enquêtes, 5 loc., trois codeuses pseudo-expertes (entre 2004 et 2006)



ETAPE 1 : RETOUR D'EXPÉRIENCE

○ Contributions scientifiques

→ Interprétation résultats avec Hypothèses en phonologie cognitive (fonctions équivalentes du schwa prononcé en surface) et de l'accent (Lacheret & Lyche 2008)

→ Implémentation outil consultation de la base par A. Tchobanov (colloque N^{velle} Orléans, Lacheret & al. à par., Langue Française)

→ Problèmes

→ Nature du codage : **flou (post)-lexical**

→ Instructions aux codeurs quant à la zone de codage



ETAPE 2 : DESCRIPTION

- Atelier Prosodie Caen 2004 et expérience Poiré (2005/2006)
 - réfléchir ensemble sur les problèmes posés par l'annotation prosodique et sur la façon de s'y prendre
 - définir un minimum épistémologique commun



ETAPE 2 : RETOUR D'EXPÉRIENCE

- Émergence de la notion de *proéminence* (événement prosodique perceptivement remarquable)
- Annotation pour tous et par tous pose la question des relations entre théories et données
- Expérience d'annotation prise en charge par F. Poiré



ETAPE 2' : DESCRIPTION

- Expérience **Poiré (2006)**

- Tâche : identification PROM en position finale et non finale sur 3 min parole spontanée, locuteur belge; 7 codeurs experts
- Analyse : 165 syllabes (notées '1' par au moins un juge ou en position accentuable)



ETAPE 2' : RETOUR D'EXPÉRIENCE

Morel & al 2006

- Consensus inter annotateur [syll term] > [- term & clit]
- Consensus inter-annotateur toute distribution confondue : mauvais
 - 1. Les experts sont-ils si experts que ça ?
 - 2. Performance de chaque juge calculable
- Etant donné 1 & 2 :
 - faire réaliser la tâche par un automate : outil de détection automatique des proéminences
 - Consigne mal circonscrite ?
 - Ecoutez et codez ce que vous percevez comme proéminent
 - Mais quand-même des pistes acoustiques intéressantes
 - Accord inter-juge proportionnel aux variations de F0, moins bon pour la durée ; cf. rôle perturbateur des hésitations : allongements importants
 - Pb des hésitations !!



ETAPE 3 : DESCRIPTION (1/3)

DÉVELOPPEMENT D'UN CORPUS D'ANNOTATION PROSODIQUE DE RÉFÉRENCE POUR ENTRAÎNER DES SYSTEMES

- Un corpus accède au statut de référence parce qu'il s'agit de ressources clairement référencées, mises à disposition et réutilisables par d'autres. Il garde ce statut jusqu'à ce qu'un autre plus performant, plus complet, plus riche, plus rigoureux dans sa construction etc. vienne le supplanter.



ETAPE 3 : DESCRIPTION (2/3)

DÉVELOPPEMENT D'UN CORPUS D'ANNOTATION PROSODIQUE DE RÉFÉRENCE

- Mise en place d'un consortium informel d'experts (Louvain-la-Neuve, octobre 2005: Goldman, Lacheret, Mertens, Simon) puis Auchlin et Avanzi → **Goldman & al. 2006**
- Redéfinition de la notion de *proéminence* (pas de corrélat fonctionnel a priori, (accentuabilité))
- Protocole codage
 - Empan de codage (3",5")
 - Nombre d'écoutes (3)
 - Symboles OpP (perception continue)
 - Symboles delivery
 - Allongements liés à une hésitation
 - Schwa post-toniques
 - Appendices Statut phonologique particulier

Ne pas perturber
les mesures de
durée relatives

Perspective de
reconnaissance automatique



ETAPE 3 : DESCRIPTION (3/3)

DÉVELOPPEMENT D'UN CORPUS D'ANNOTATION PROSODIQUE DE RÉFÉRENCE

- Phase 1 : corpus de 20' (demandes itinéraires, interview radios)
 - 2 experts avec calcul taux accord inter-annotateurs, aller-retours entre experts et mise en place d'une annotation consensuelle
 - Utilisation du corpus pour développement PROSOPROM, soft Goldman détection automatique PROM

- Phase 2 sqq. : Prolongation du travail
 - 20' > 50' > 70' : corpus **C-PROM** : interview radio, map_task > discours politiques, journaux radio, récit_vie > conférences universitaires, lectures
 - Locuteurs belges, suisses et français (métropole)
 - **Publications diverses : Avanzi, Simon, Goldman, Auchlin (Interspeech, Bristol, CMLF, Speech Prosody 2010)**



ETAPE 3 : RETOUR D'EXPÉRIENCE

Linguistique générale

- Genres de discours et phonostylistique
 - Catégorisation prosodique des genres de discours
 - Effets des genres discursifs sur la prosodie
 - Possibilité de discriminer des genres à partir de propriétés prosodiques (Prom, débit, dysfluences, pauses, etc)
 - map_task, récit_vie / radio-interview // [DP](#), [news](#)
 - *Obin & al. 2008 ; Goldman & al. 2009 ; Simon & al. à par.*

Linguistique de corpus

- Codage = travail long, charge cognitive, fatigue, etc
 - Pas réaliste d'envisager la tâche sur gros corpus
- Consensus inter-annotateur (entre 83% et 86% selon la mesure F-mesure vs. Kappa, cf. [Avanzi et al. 2010](#))
- Objectif d'automatisation réaliste



ETAPE 4 : DESCRIPTION (1/2)

COMPARAISON D'OUTILS POUR LA DÉTECTION AUTOMATIQUE

Obin, Goldman & al. (2008)

- Objectif : comparer performances sur tâche de détection automatique de proéminences.

- Matériel
 - Corpus C-PROM
 - 5 genres : Map Task, Interviews radio, discours politiques, journaux radiophoniques et récits de vie
 - 50' (5 x 10 minutes)
 - 12851 intervalles syllabiques
 - 973 exclus via delivery
 - 3244 syllabes proéminentes
 - 11878 ni prom ni delivery



ETAPE 4 : DESCRIPTION (2/2)

COMPARAISON D'OUTILS POUR LA DÉTECTION AUTOMATIQUE

- Système à base de règle vs systèmes entraînés:
 - Système de décision à base de règles
 - **Analor** (Avanzi, Lacheret & Victorri 2008a et b)
 - Systèmes entraînés
 - **ProsoProm** (Goldman, Avanzi, Lacheret & Simon 2007 ; Simon, Avanzi & Goldman 2008)
 - **IrcamProm** (Obin, Lacheret & Rodet 2008, 2009)

- Autres différences :
 - F0, durée et pause (ANALOR, PROSOPOM) vs beaucoup de paramètres (IRCAMPROM)
 - Courbe F0 brute ou stylisée
 - Empan de traitement pour relativisation (large et dynamique pour Analor, restreint et fixe pour ProsoProm, moyen et fixe pour IrcamProm)



ETAPE 4 : RETOUR D'EXPERIENCE (1/2)

- Info sur genre (distribution des erreurs)
- Variation selon les algorithmes (F-mesure)

	Map Task	Interview	Récit	Disc. Pol.	Jrnx Radio	Tot.
Analor	68,7	71,5	63,2	74,5	70,8	69,7
ProsoProm	74,4	73,4	71,0	74,9	72,5	73,2
IrcamProm	75,8	74,3	76,3	76,00	75,0	75,4

- Analor : délétion de proéminence (sous-détection)
- rappel = 63.6%, précision = 77.2%
- ProsoProm : insertion de proéminence (sur-détection)
- rappel = 78.9%, préc. = 68.2%
- IrcamProm : bon compromis (*well-balanced*)
- rappel = 76.4%, préc. = 74.5%



ETAPE 4 : RETOUR D'EXPÉRIENCE (2/2)

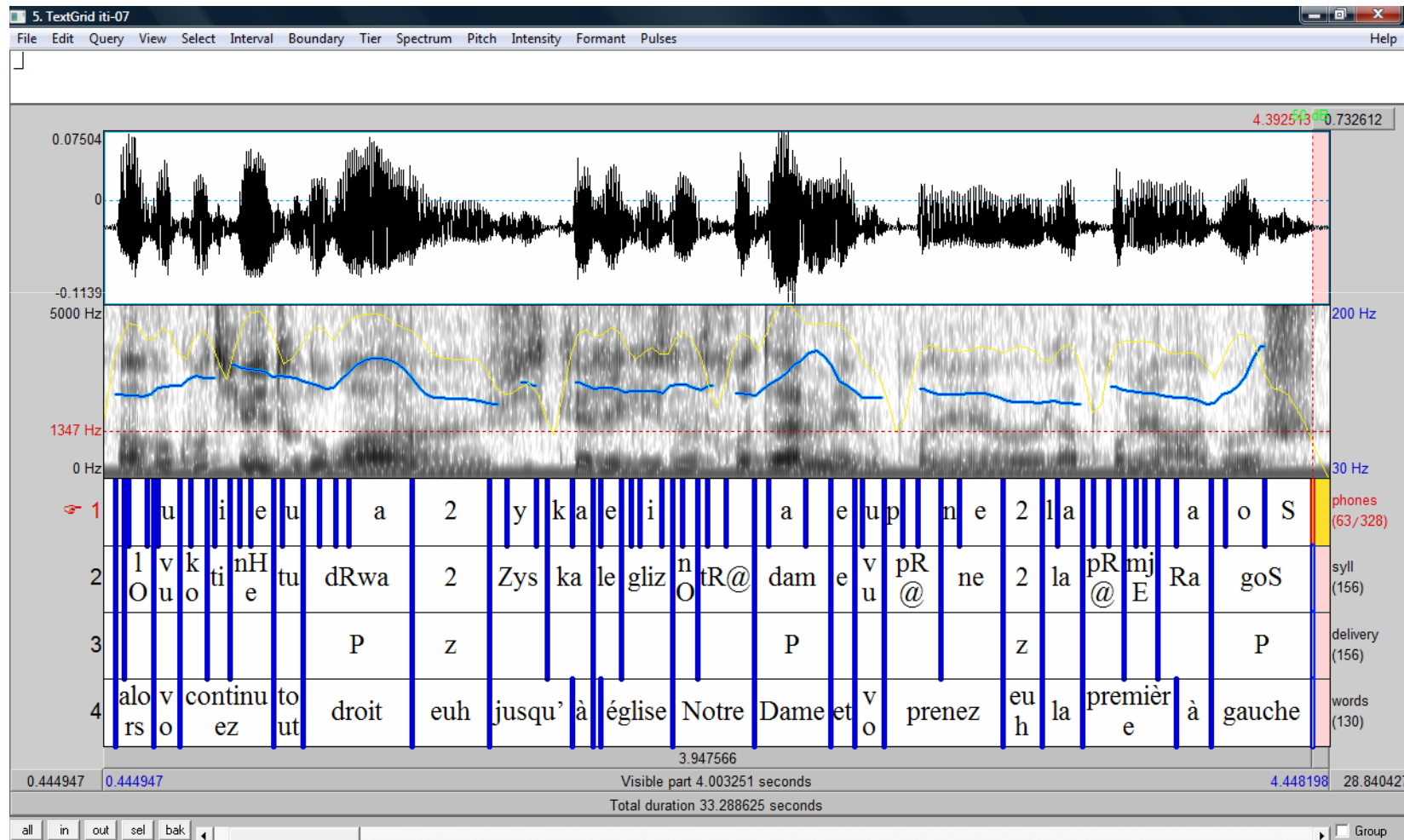
- Question de la **fenêtre contextuelle** pour les calculs relatifs des variations de durée et de hauteur
 - Trop large : sous-détection
 - Trop restreinte : sur-détection
- Limites capacité généralisation à base de règles vs système apprentissage
- Meilleur compromis entre résultats obtenus et complexité algorithmique
 - Ce compromis doit être trouvé par l'expert lui-même (manipulation et réglage paramètres)
 - Programme ANALOR : B. Victorri & M. Avanzi
 - Système entraîné
 - Corpus apprentissage enrichi
 - Tire « delivery » épurée pour être utilisable





PARTIE 2
L'OUTIL ANALOR

FICHIERS EN ENTRÉE PRAAT & EASYALIGN



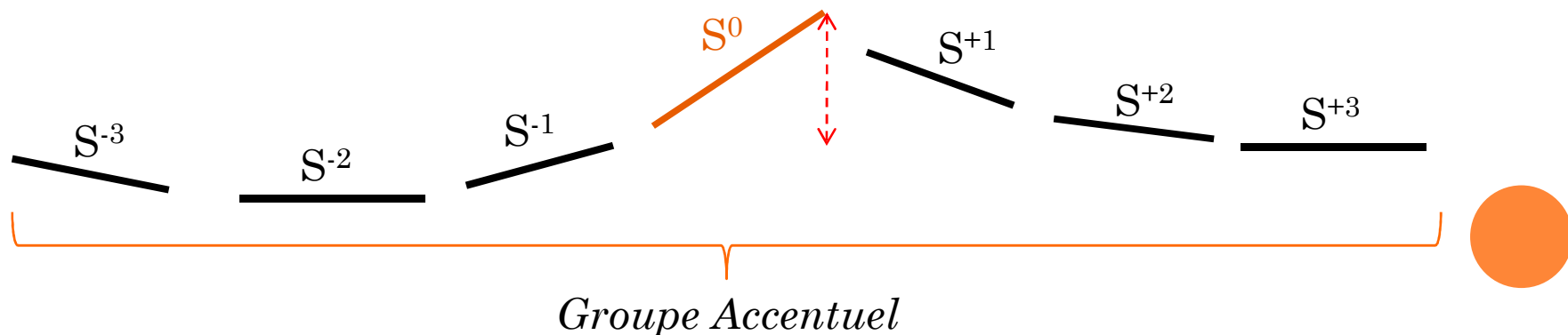
PARAMÈTRES ACOUSTIQUES MANIPULÉS

o Mesures acoustiques choisies

- **durée** syllabique (moyenne)
- **hauteur** de la syllabe (moyenne)
- **montée mélodique** intra-vocalique
- **pause** (subséquente)

Poids syllabique

o Empan de **relativisation**



MÉTHODE D'APPRENTISSAGE - RAPPELS

- Approche par règles peu robuste
 - Grande variation dans les genres de discours
- Entraînement des seuils sur un corpus annoté manuellement par des experts (C-PROM)



MÉTHODE D'APPRENTISSAGE – PRINCIPES

- Une méthode d'apprentissage stochastique
 - On parcourt aléatoirement l'espace des paramètres à la recherche d'un jeu de paramètres optimal.
 - Les performances sont évaluées à l'aide de la F-mesure (privilégiant un bon équilibre entre précision et rappel).
- Une méthode d'apprentissage utilisant pleinement les connaissances linguistiques
 - Les valeurs initiales des paramètres ne sont pas choisies aléatoirement, mais en fonction des connaissances sur les phénomènes prosodiques pris en compte
 - Il s'agit donc d'une recherche locale dans l'espace des paramètres : l'optimum recherché doit être relativement proche des valeurs initiales.



MÉTHODE D'APPRENTISSAGE – ALGORITHME

- Chaque essai est caractérisé par la donnée de :
 - un jeu de paramètres courant : (p_1, p_2, p_3, p_4) ;
 - un pas d'exploration, c'est-à-dire un ordre de grandeur des modifications à effectuer (10% par exemple).
- Chaque essai consiste à modifier aléatoirement chaque paramètre :
 - si la F-mesure est meilleure, on remplace le jeu de paramètres courant par le nouveau jeu, sinon on garde l'ancien.
 - Si l'on n'a pas trouvé de meilleur jeu au bout d'un certain nombre d'essais (250 par exemple), on recommence la procédure avec un pas d'exploration plus petit (5% au lieu de 10% par exemple)
- Au démarrage :
 - Le jeu initial choisi doit être linguistiquement pertinent
 - Le pas d'exploration initial est assez grand (40%)
- A la fin :
 - On arrête quand le pas d'exploration devient très petit (<1%)



VALEURS INITIALE DES SEUILS

- Valeurs initiales doivent être définies **AUPARAVANT** pour être efficaces
 - On cherche dans un espace données, pas dans le vide
- **Mesures acoustiques** choisies
 - **durée** syllabique (moyenne)
 - **poids** phonèmes syllabes
 - **hauteur** de la syllabe (moyenne)
 - **montée** mélodique intra-vocalique
 - **pause** (subséquente)
- **Seuils** optimaux
 - = **2** (sans unité)
 - = **0.3** (sans unité)
 - = **2** (en demi-tons)
 - = **3** (en demi-tons)
 - = **0.1** (en ms)
- Justification de ces seuils de départ *a priori*
 - Rossi *et al.* 1981 ; Mertens & D'Alessandro 1994 ; Garnier-Rizet 1994 ; Lacheret & Beaugendre 1999 ; Astésano *et al.* 2003 ; Goldman *et al.* 2007



CORPUS D'ENTRAÎNEMENT

- Corpus C-PROM (70 min.)
- Recodage par un expert des syllabes NP et delivery :
 - Syllabes 0 suivies d'une pause dues à une interruption syntaxique sont codées « Z »
 - Élargissement de la notion de **dysfluence**, cf. Rhapsodie
 - Les schwas sont resyllabés ou codés « P » (< macro-syllabe)
 - Les syllabes appartenant à un appendice sont recodées 0 ou notées P
- Tableau récap.:

Disc. genre	Duration (sec.)	Nb. Syll.	P	NP	Z	Valid syll.
Lectures	401	1830	470	1357	0	470
Politiques	635	2174	632	1539	1	633
Jnx radio	621	3165	825	2279	58	883
Conf.	687	3133	818	2202	108	926
Interview	627	2591	684	1806	90	774
Map Task	590	2222	562	1490	162	724
Récits	622	2663	685	1763	199	884
TOTAL	4183	17730	4676	12436	618	17112



ÉVALUATION

- Entraînement pour chaque genre à partir des seuils initiaux fixés *a priori*
- Performances

Genre	Initial performance			Trained performance			Gain
	Prec.	Rec.	F-ms	Prec.	Rec.	F-ms	
Lectures	79.86	71.7	75.56	76.41	77.87	77.13	1.57
Disc. Pol.	75.07	83.39	79.01	82.35	81.86	82.16	3.15
Journaux	74.57	73.58	74.07	75.7	82.3	78.86	4.79
Conf.	76.11	73.23	74.64	79.18	79.95	79.56	4.92
Interviews	71.88	82.6	76.87	79.3	80.89	80	3.13
Map-tasks	75.31	76.51	75.9	79.86	79	79.43	3.53
Récits	83.27	61.75	70.91	73.44	80.73	76.91	6
TOT.	76.58	74.68	75.28	78.03	80.37	79.15	3.87





PARTIE 3
CONCLUSION

POUR CONCLURE

- Annotation automatique
 - Seule ressource viable pour exploiter information prosodique sur grands corpus
- Différents retours d'expérience
 - Modèles d'annotation et implémentation : compromis entre annotation sans filets (Poiré) et étiquetage trop fin avec présupposés implicites (phénomènes de delivery chez Avanzi & Simon)
 - Corpus de référence pour l'apprentissage des systèmes
 - 2 codeurs < 5 codeurs < etc.
 - Codeur expert vs naïfs + contrôle experts
 - 1 h < 3h (Rhapsodie < n heures : corpus PIII)
- Interface-PFC-ANR-Rhapsodie (corpus prosodique de référence du fr. parlé)
 - Travail de thèse Guri Bordal : croisement méthodologie Rhapsodie, protocole codage PFC : problématique de l'interlangue Loc Sango (français centre africain) : interface lexical/postlexical

