

Annotation prosodiques de corpus oraux

A. Lacheret

Le projet Rhapsodie

<http://rhapsodie.risc.cnrs.fr/fr/>

Journées PFC, Paris, décembre 2011

le projet Rhapsodie

Un travail collaboratif

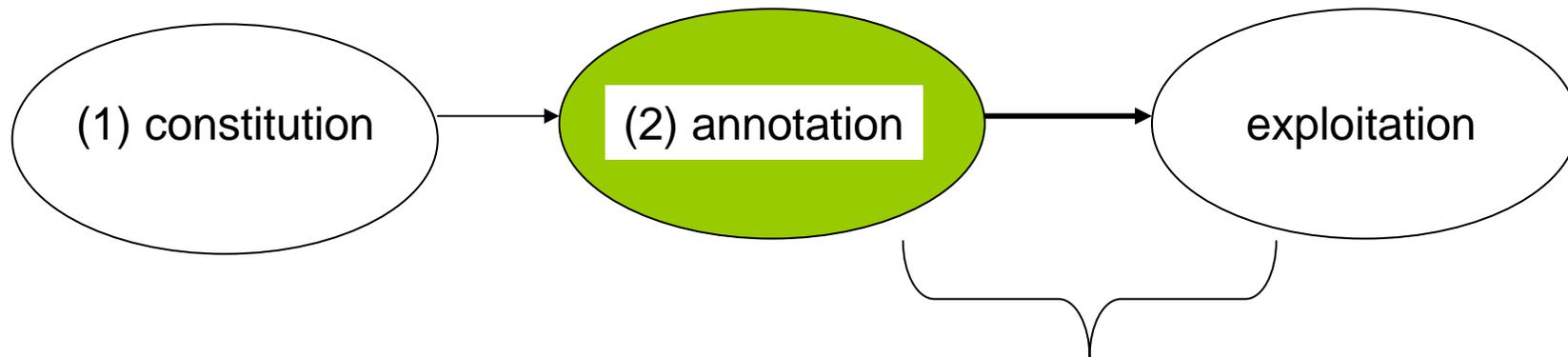
- *Syntaxe : Ch. Benzitoun, C. Bolly, J.M. Debaisieux, E. de la Clergerie, J. Deulofeu, A. Dister, K. Gerdes, S. Kahane, Fl. Lefeuvre, R. Marlet, P. Pietrandrea, N. Rossi, N. Tanguy, F. Sabio, B. Victorri*
- *Prosodie : C. Astesano, M. Avanzi, R. Bertrand, X. M.P. D'imperio, G. Ferré, J.Ph. Goldman, A. Lacheret, Ph. Martin, P. Mertens, N. Obin, V. Pasdeloup, Ch. Portes, A.C. Simon, B. Victorri*
- *Ingenierie : J. Beliao , B. Marcatté, X. Rodet, A. Tchobanov, A. Truong, Ch. Veaux*

Contexte, enjeux, objectifs et méthodes

- Le projet Rhapsodie : un projet pilote
 - Contexte : donné à manipuler en entrée = continuum sonore (enregistrement audio)
- Enjeux & objectifs
 - Pourquoi faire ?
 - Enjeux théoriques (linguistique de corpus, linguistique fonctionnelle) technologiques
 - Enjeux technologiques
 - Comment faire (tâches) ?
- Méthodes et outils pour le faire (manuel vs automatique) ?
 - Transcription orthographique
 - prosodie
 - (Syntaxe)
 - assemblage intonosyntaxique

Le projet Rhapsodie : un projet pilote (1/2)

Chaîne de traitement pour une exploitation mutualisée



Principe de réalité : (1) création des ressources (préparation du matériel, transcription alignée au son, etc), (2) annotations (outils d'aide à l'annotation)

→ temps + argent

Principe d'interopérabilité et d'échange (interrogation et traitement des données annotées par d'autres ; reproductibilité, y compris par des machines)

→ Partage des tâches (1), (2), (3) : focus sur (2)

Le projet Rhapsodie : un projet pilote (2/2)

- Points saillants de l'approche (hypothèses/méthodes)
 - discours oral se constitue par morceaux successifs (séquençage)
 - morceaux qui s'appellent les uns les autres (clauses, syntagmes, bouts de syntagmes, etc) : processus de liage, rupture non explicitement marqués → quels indices ?
 - » *Le vélo de Pierre est tordu vs. Pierre son vélo le guidon il est tordu*
 - Topologie (position des éléments/pièces dans la structure d'ensemble > nature disloquée, parenthèse, etc) + gram de construction
 - » Forme intonosyntaxique-Fonction sémantico-pragmatique
 - Multilinéarité (bidimensionnalité) : double projection structurelle : combinaison syntagmatique vs substitution paradigmaticque
 - » *Chaque région a son propre comportement vis-à-vis de l'église vis-à-vis de la fois vis à vie de la vie chrétienne (Eslo)*
 - Segmenter en unités, Typer ces unités, typer des relations entre unités
 - Autonomie des traitements prosodiques et syntaxiques

Pour quoi faire ?

Pour quoi faire : enjeux théoriques (1/3)

- **Linguistique de corpus**

- constitution, **annotation**
 - Archivage
 - **Requêtes sur les données**

→ **Méthode d'annotation mutualisée**

- Mise au point de formats pour l'annotation prosodique et syntaxique de données orales dans une perspective d'interopérabilité et d'échange ;
- Interface web de consultation et d'annotations de données orales structurées (requêtes + affichages) ;
- Distribution en ligne d'outils pour traiter et analyser ces ressources avec manuels d'annotation et d'utilisation (requêtes en langage naturel et en Rhapsodie-QL);

Pour quoi faire : enjeux théoriques (2/3)

- Linguistique de corpus **oraux** & transfert de connaissances
 - Une nouvelle lunette d'observation
 - aller-retour entre traitement manuel et automatique: adapter aux spécificités de l'oral outils conçus pour traitement de l'écrit (traitement syntaxique) ou élaborés pour lecture oralisée, ctx = phonétique/phonologie de laboratoire (annotations prosodiques).
 - questions posées par annotation des données orales tout venant : pas seulement objectifs pratiques mais aussi levier pour impulser propositions théoriques avec nouveaux modes de représentation à l'interface de syntaxe et prosodie
 - segmentation en groupes : comment couper ? Comment catégoriser les unités qui émergent ? Comment définir les relations entre les unités

Pour quoi faire : enjeux théoriques (3/3)

- Linguistique **fonctionnelle** (interprétative, vs formelle): activité de langage fondamentalement déterminée par les fonctions qu'elle sert: **Discours** :
 - fonctionnement des formes intonosyntaxiques rencontrées dans le discours
 - **corrélations** entre unités syntaxiques et marqueurs prosodiques dans la structuration des discours
 - » Identification d'unités de discours de rang et de portée variable (ex l'équivalent de la *phrase* à l'écrit)
 - » Constructions grammaticales prototypiques sous l'angle intonosyntaxique (ex. parataxe, greffes, positions détachées, parenthèses, etc)
 - hypothèses sur les raisons énonciatives qui motivent leur choix (corrélations formes/marques-fonctions : description (constructions attestées) & généralisations : prouver qu'une construction donnée appartient à une famille de construction)
 - » Point de vue, expressivité
 - » Discours autre (l'hypothèse du faible marquage de la clôture d'un discours rapporté est-elle vérifiée?)
 - » Empaquetage informationnel
 - » etc

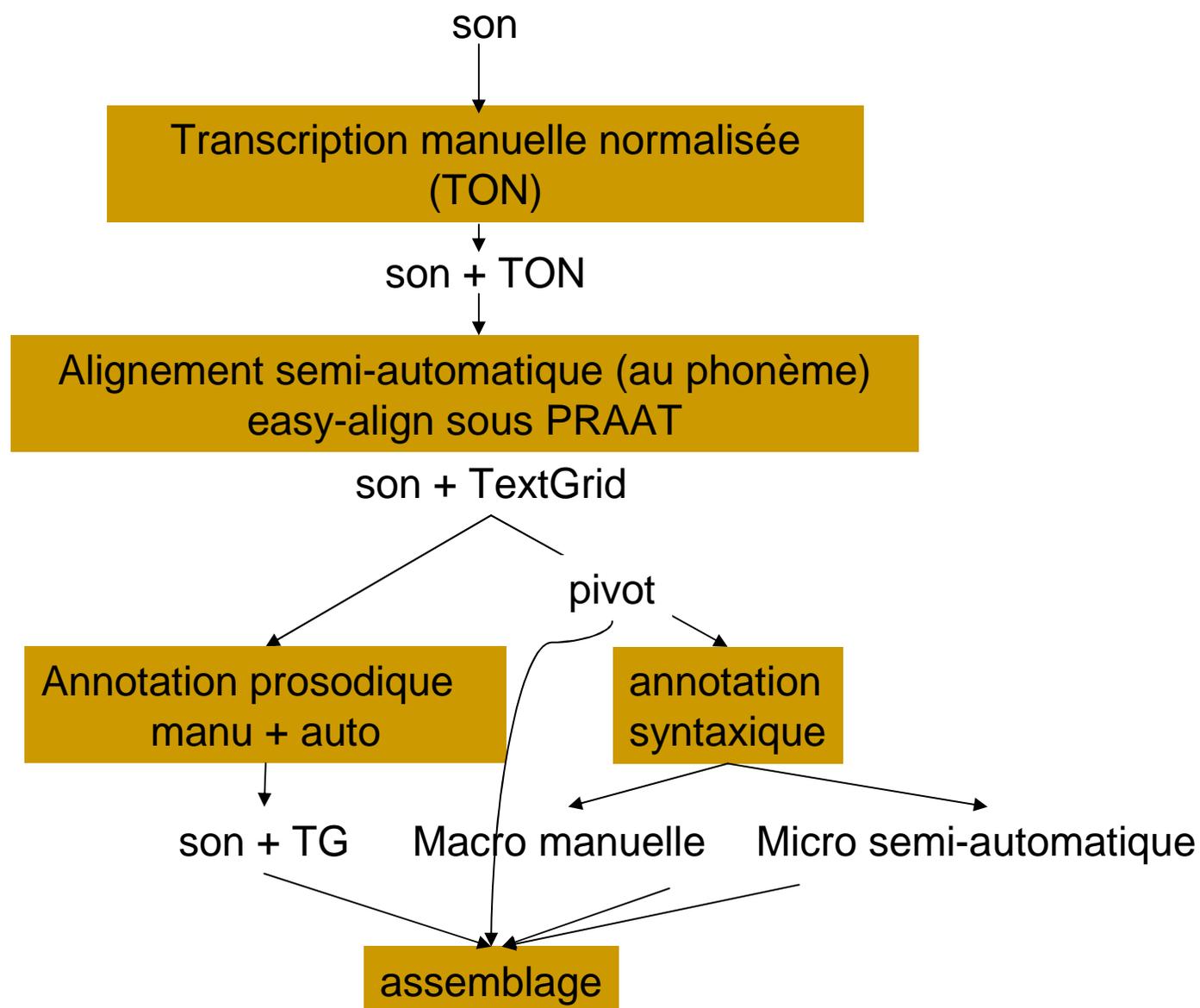
Pour quoi faire : enjeux technologiques

- Grands corpus oraux annotés : **passage à l'échelle** → besoin de corpus oraux annotés manuellement (syntaxe+prosodie)
- TAP
 - Apprentissage automatique (entraîner des systèmes en vue d'annotations automatiques ultérieures) → corpus annotés manuellement

Comment faire ?

- Encore trop tôt pour faire trop grand : privilégier l'approche **qualitative** (vs approche quantitative (\neq grands corpus oraux type PFC))
 - Faire peu pour faire bien, propre et cohérent ; peu = déjà beaucoup !
 - **annotation exhaustive** : corpus entier (3h de corpus) annoté et **vérifié**
 - Privilégier la variété des sources (ESLO, PFC, CFPP, corpus de thèses)
 - monologues et dialogues
 - parole privée, professionnelle, publique
 - Entretiens semi-dirigés, conversations libres
 - Descriptions, prescriptions, narrations, argumentation
 - **Hétérogénéité vs représentativité**
 - chaque type de discours engendre des marqueurs et des constructions caractéristiques du type en question
 - richesse des annotations fournies en sortie relève de la diversité des situations discursives, *i.e.* de la pluralité des constructions à gérer (constructions « standards » mais aussi « non standards »)
 - !! **Petit corpus \neq de couverture faible : nbre des échantillons (58)**
- En sortie : des annotations utilisables pour des systèmes **d'apprentissage automatique** de structures sur des genres variés en syntaxe comme en prosodie

Architecture (J.Ph. Goldman)



Méthodes et outils pour le faire

Transcription orthographique

1^{re} étape : transcription orthographique

Transcrire : un jeu d'enfant ?

“Travail de copiste, de secrétaire”. “Il suffit d’écouter.”

Cf. La petite phrase de Martine Aubry à propos de Ségolène Royal a bien montré qu’il ne suffit pas d’écouter...

<http://www.youtube.com/watch?v=Hi6Kt9STQuM>

Travail sous-estimé, dévalorisé.

Pourtant “transcription as theory” (Ochs 1979)

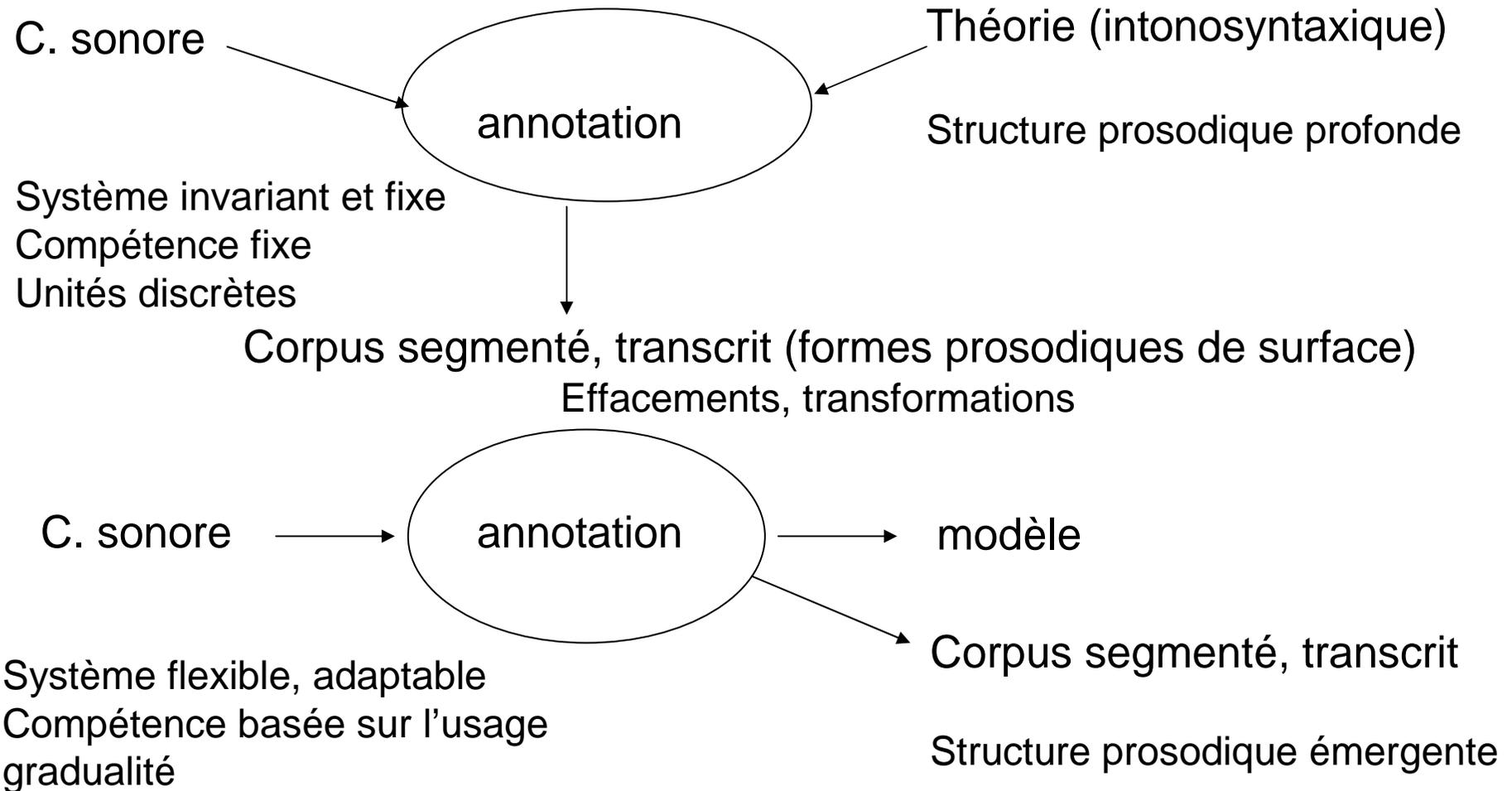
→ Guide pour une transcription orthographique normalisée

Ecole corpus Biarritz, septembre 2011 <http://annotationlinguistique.fr/>

Principes et objectifs d'annotation

- Base perceptive (ce que j'entends, (ce que je vois) sur le signal)
- Annotations mutualisées : novices (étudiants en SDL), experts
- Annotation automatisable
 - annotation explicite
 - Guide d'annotation
 - Vers une annotation de référence

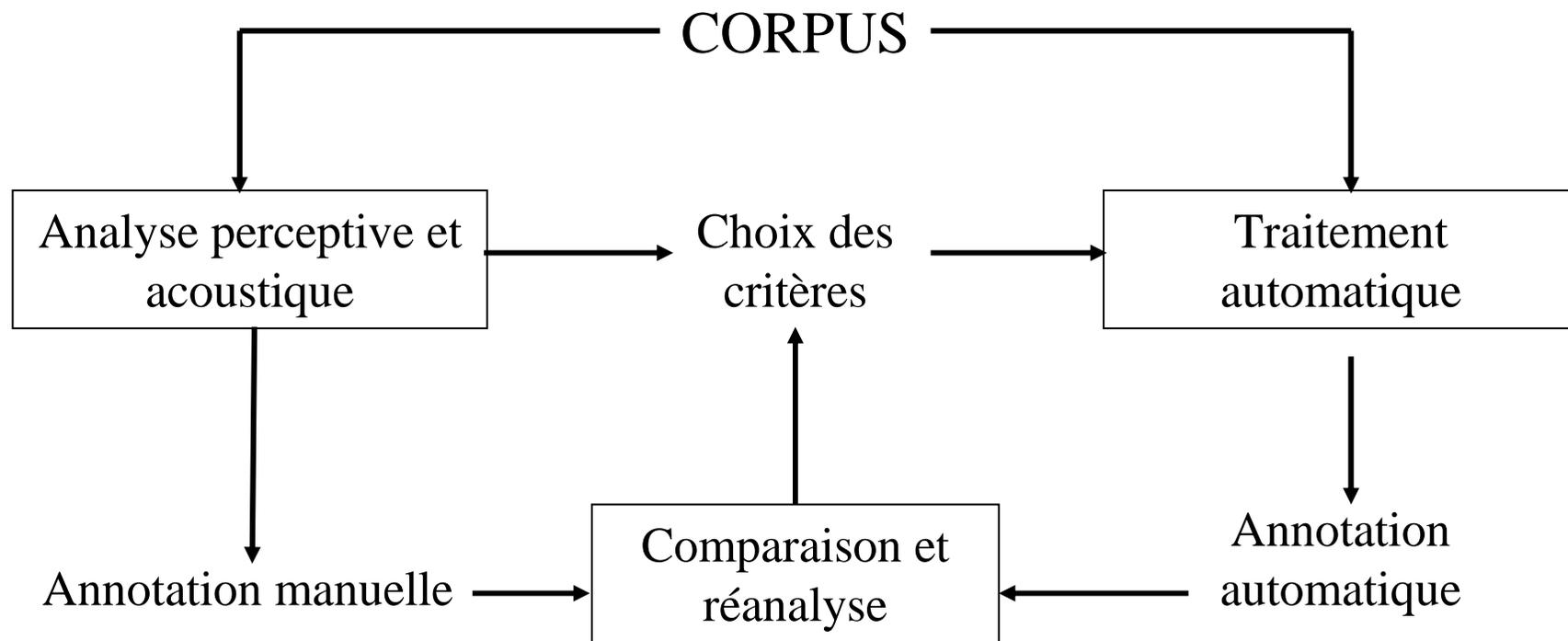
Annotation : déduction ou induction ?



!! Pas de théorie ne veut pas dire ni hypothèses, ni intuition (modèle mental du locuteur-transcripteur)

La place de la modélisation informatique

1. simulation
2. Génération



Construction pas à pas du modèle

Méthode

Méthode bottom-up (du perceptif au fonctionnel), Intégration progressive de la complexité (phase d'annotation précède la phase de caractérisation et d'interprétation linguistique)

1. Repérer des objets perceptifs : proéminences syllabiques à partir desquelles la structure va émerger (unités et relations entre unités)
2. Caractériser acoustiquement ces objets
 1. Configuration tonale d'une proéminence sur un empan donné (ex. proéminence syll) → annotation tonale
 2. Contours prototypiques
3. Interpréter fonctionnellement ces objets
 1. Proéminence = accent
 2. Proéminence = frontière
 3. Interprétation discursive : domaine de projection, portée, etc

Structure émergente

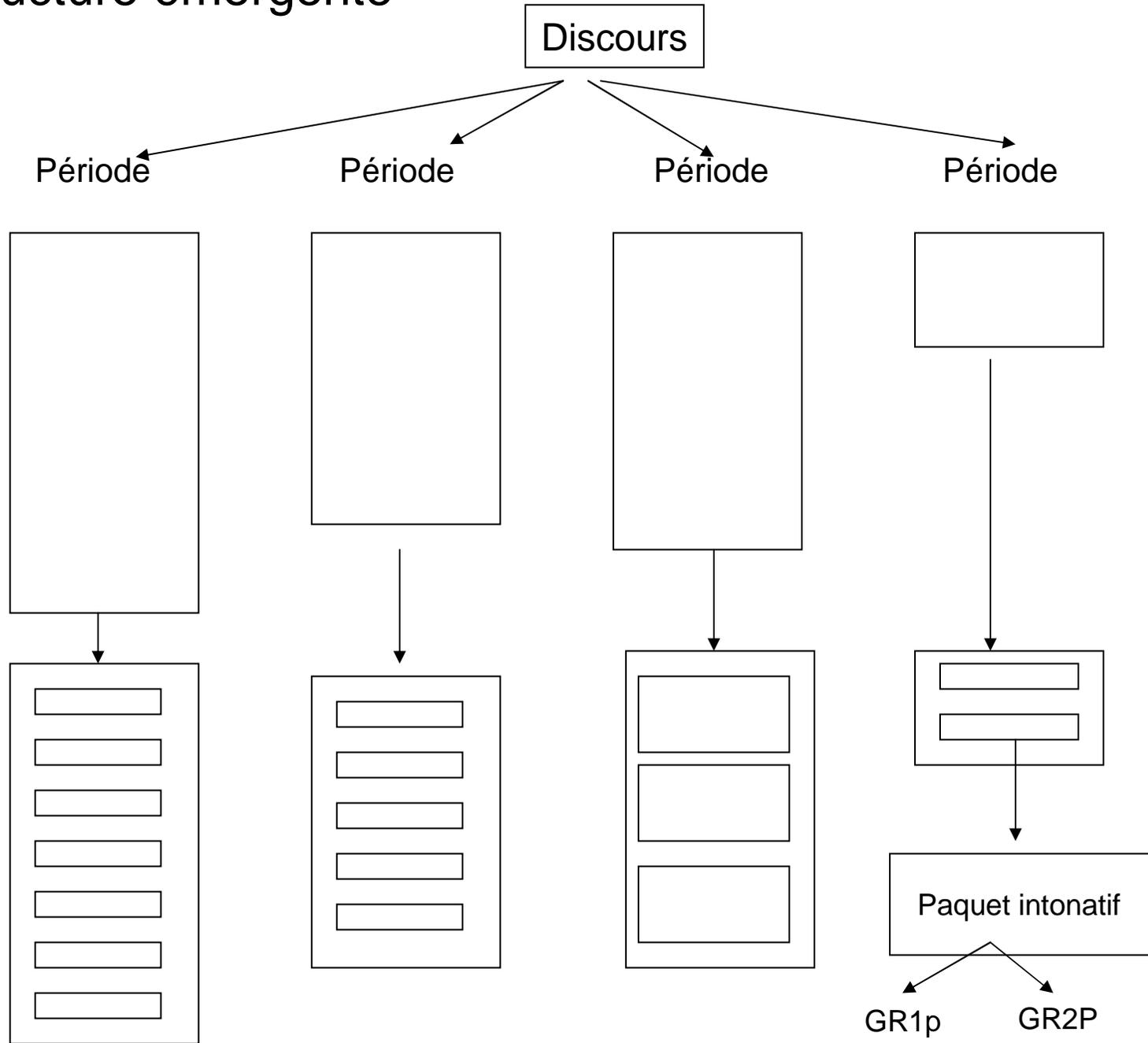
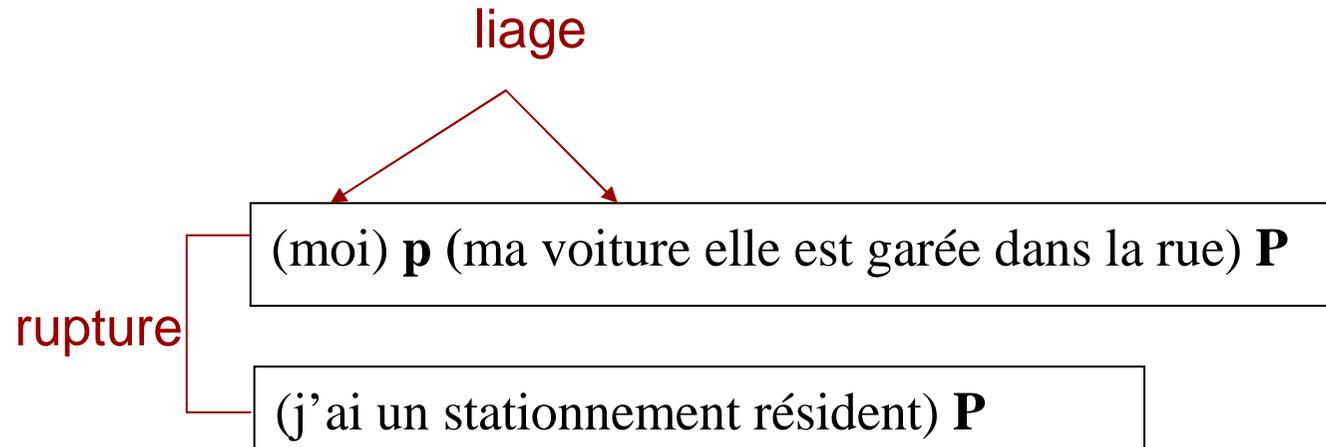
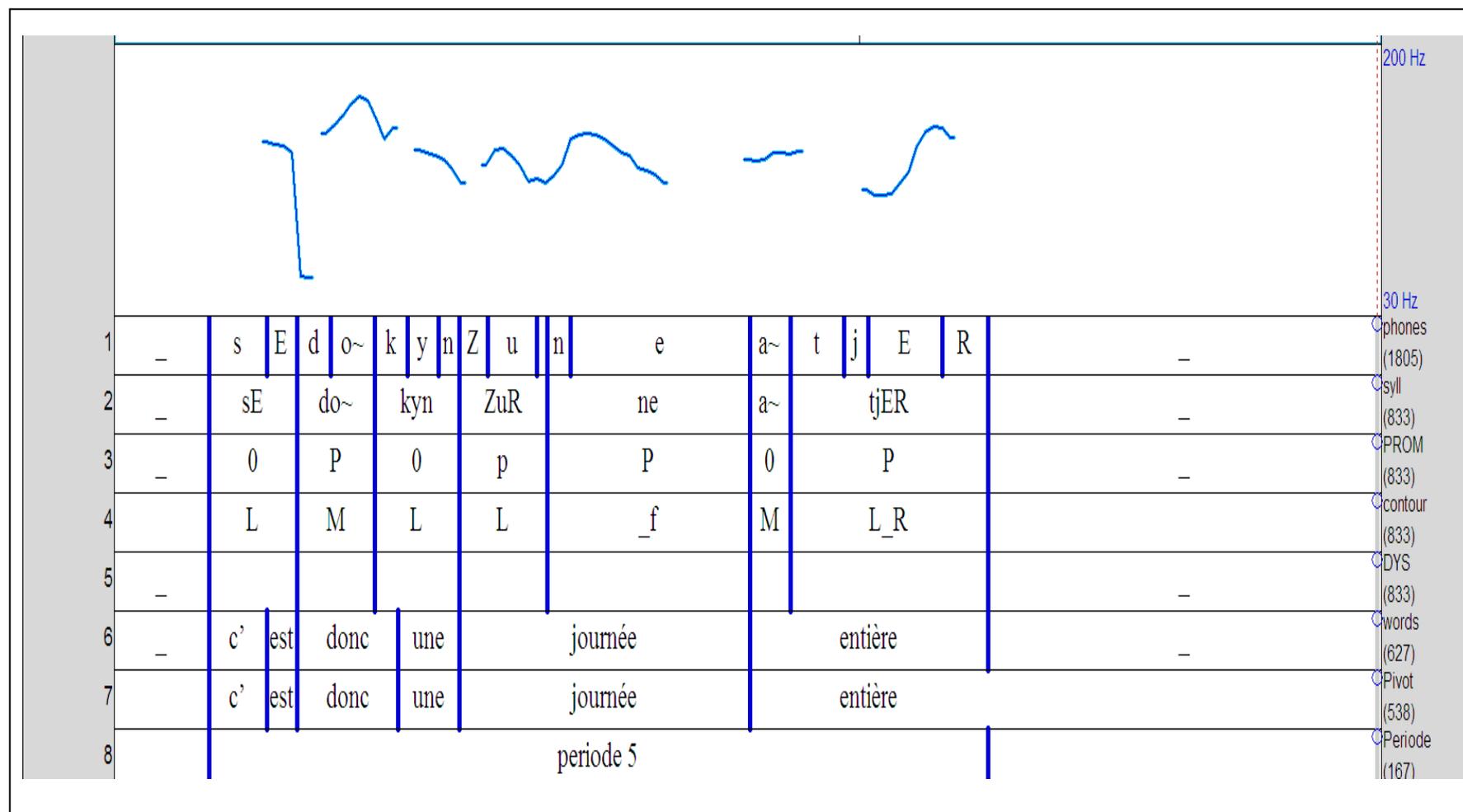


Illustration (CFPP 2000)



Sortie textgrid



Autonomie des couches d'annotation → degrés de granularité variable (standard vs enrichie)

En pratique : proéminences (1/2)

- Proéminences **et principe de contraste (perception syllabique)** : une syllabe peut se détacher dans le flux verbal par rapport à ce qui précède et ainsi contraster avec son environnement. elle se détache perceptivement comme une figure sur un fond sonore.
- Perception des proéminences continue, et non catégorielle.
 - Différents degrés de proéminence (du non proéminent au très proéminent).
 - Pour rendre compte de manière rudimentaire de ce processus continu, nous proposons d'utiliser deux niveaux : proéminence normale (p ou W), proéminence forte (P ou S)

En pratique : proéminences (2/2)

- Traitement manuel
 - Novices formés > vérification experts
- Annotation de référence
- Traitement automatique : Analor
- Pb méthodologiques (site Rhapsodie <http://rhapsodie.risc.cnrs.fr/fr/archives.html>, jnées Toulouse)
 - Comment établir une annotation de référence
 - Contrôle des annotations novices et stats sur les annotations (sous-annotations, sur-annotations, mouton noir) : kappa
 - Annotation manuelle vs automatique
 - Repérer les distributions peu consensuelles
 - Proéminence en contexte de schwa post-tonique
 - Hésitation et ou proéminence ?
 - Degré de proéminence

En pratique : disfluences

- Disfluences (« éléments qui brisent le déroulement syntagmatique de l'énoncé », CBB)
 - Segments abandonnés, amorces de morphèmes : c'est pas loin **tu** j'y vais
 - Segments repris/répétés : c'est pas loin **tu tu** y vas
 - « Euh »
 - Allongement syllabique excessif (ne correspondant a priori pas à une démarcation structurale (il y **a::::::**))
- Questions : corrélats acoustiques toujours présents ? Nature ?
- Objectif : matériel pour apprentissage automatique sur des corpus non annotés

En pratique : périodes (1/2)

- Principes théoriques
 - Constructions prosodiques globales discursivement contraintes : Indices temporels et fréquentiels associés à la perception de ruptures fortes dans le discours
 - Unité de segmentation majeure dans le discours ?
 - Relations entre périodes : notion de *macro-période* : séquençage discursif
 - Explication-but-résultat
 - Saillance discursives : Arrières plans vs premiers plans; encadrements du discours (portée, cf. Charolles)
 - Enchaînements thématiques
 - etc
- En pratique
 - Segmentation automatique avec le logiciel Analor revue manuellement
 - Pb en contexte de disfluence (fonction du marqueur « euh », allongements

En pratique : périodes (2/2)

Illustration (Rhap-M0001, corpus Avanzi)

_ et ben euh _ tu prends le boulevard euh _ là qui part de Nef Chavant
là le boulevard qui passe à côté d'Habitat _

tu continues tu vas arriver sur la place euh Victor Hugo _

(à) la place Victor Hugo à la banque euh _ qui fait l'angle tu prends à
droite _

tu longes les rails du tram _ jusqu'à la place Grenette _

tu continues dans la vieille ville tu prends la grande rue _

et euh _ après tu bifurques euh _ y a une petite bifurcation euh juste
avant la place du Tribunal

tu passes à côté d'une petite fontaine

tu arrives place aux Herbes avec une _ une sorte de halle _

quoi de _ de _ de structure métallique _

tu continues la rue _

En pratique : contours syllabiques (1/2)

- Quelle étiquette : ton, contour, geste ?
- Pratique de l'annotation
 - Pas d'annotation unifiée
 - Un système de référence : TOBI
 - Non automatisable
 - Gomme la dimension temporelle
 - » Prosodie = F0, tps (durée segments pleins et pauses), qualité vocale, intensité
 - Utilise une théorie linguistique implicite → * mutualisation
 - Annotation d'emblée fonctionnelle (tons accents vs. tons de frontières)

En pratique : contours syllabiques (2/2)

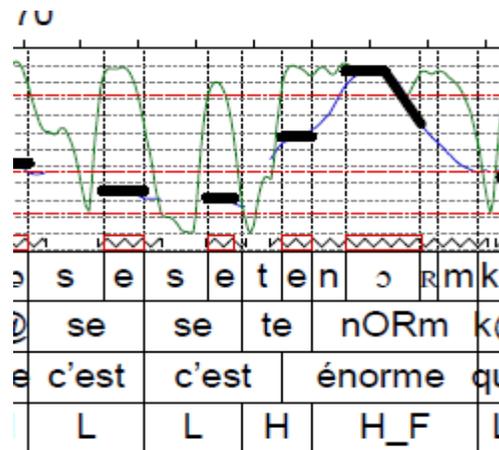
Description riche des contours → **Gap descriptif** (> contour majeur, mineur, H*, H%) → **Gap interprétatif**

- Contours syllabiques

- Suivi de contours sur modèle perceptif : simulation de la perception des variations de hauteur dans le temps, le Prosogram de P. Mertens

<http://bach.arts.kuleuven.be/pmertens/prosogram/>

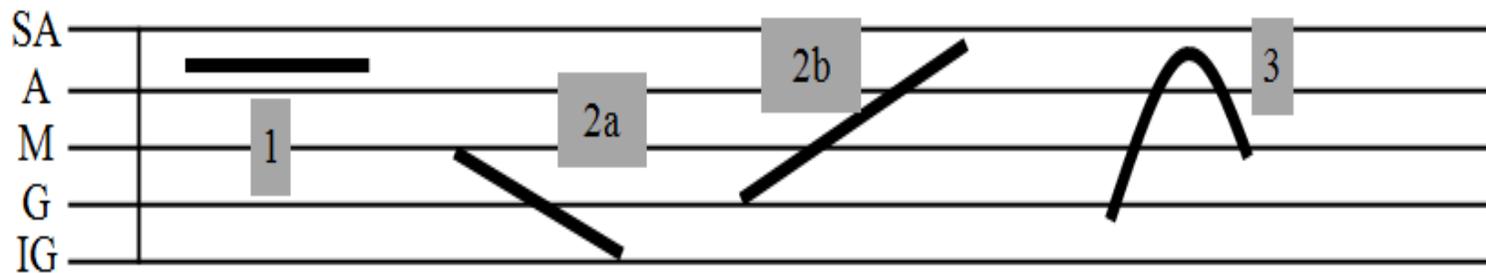
- niveau de hauteur au début du noyau syllabique,
- variation mélodique à l'intérieur de la syllabe,
- niveau infra-bas (*bottom*) atteint à la fin du noyau syllabique



- "HF_" : chute à grand intervalle à partir du niveau haut suivi d'un palier ;

En pratique : contours globaux (Ircam)

- Contours globaux sur des unités d'empan variable (visée fonctionnelle)



Avanzi (2011)

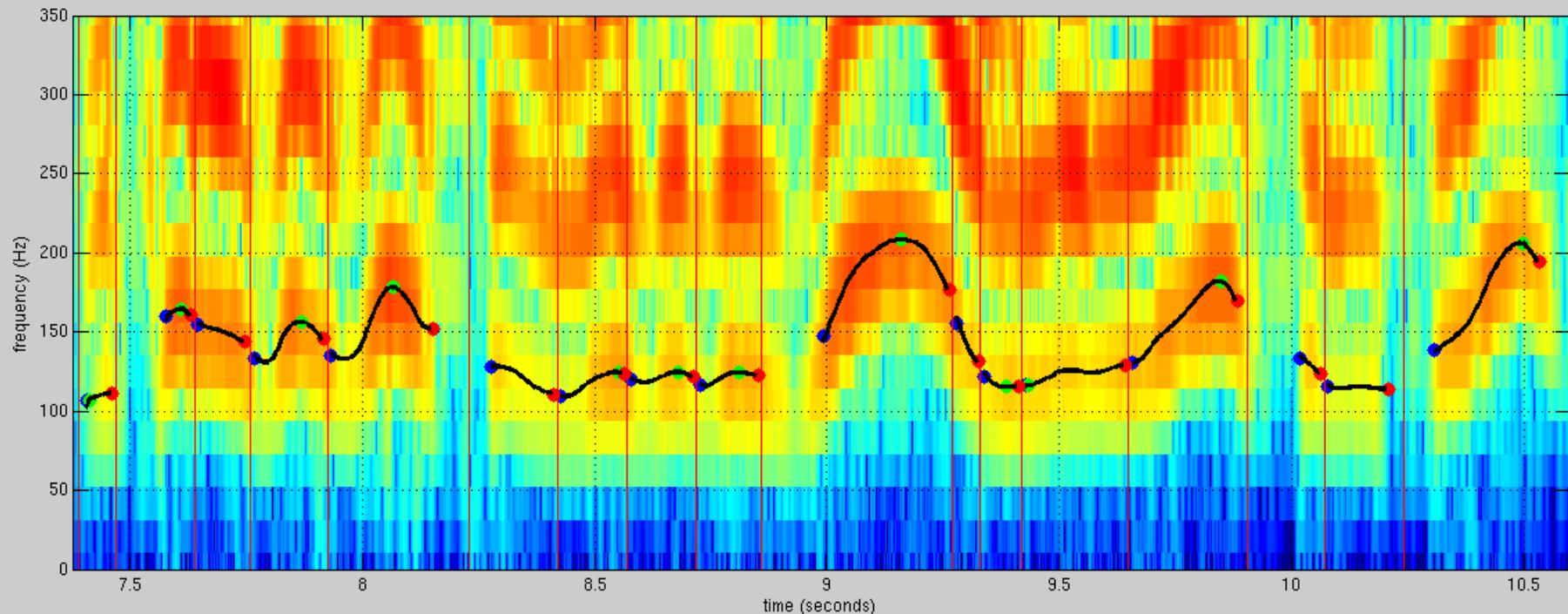
- Morphologie de contours à partir d'une courbe lissée sur les bases de :
 - Modèle mathématique (approximation des variations de F0 sur le signal) : variations "lentes" considérées comme prosodiques, variations "rapides" considérées comme du bruit ou un résidu phonétique → gommées/lissées/effacées

Après ma visite à Landovisiau et à l'île longue ce matin

- Extraction des points valeurs caractéristiques (Initial, Final, Extrême, Ancrage et Registre Local) pour chaque empan souhaité
- Conversion en alphabet symbolique

Exemple avec les empans syllabe, paquet intonatif et période :

a	pRE	ma	vi	zi	ta	la~	di	vi	zjo	e	a	lil	lo~	s@	ma	t9~
0	p	p	0	P	0	0	0	p	P	0	0	0	P	0	0	P
ll	hh	hm	mm	mhh2	ml	lm	mm	lm	mhH3	hm	ml	lm	mh	mm	ll	mH
lhh3				mhH3					hh			mH				
IH																



Signification de l'annotation tonale

Le contour de f0 **sur un empan donné** est représenté par les valeurs caractéristiques suivantes :

-Initial

-Final

-Extrême

-Ancrage du point extrême (relativement aux points initial et final)

-Registre local (moyenne du f0 exprimé en demi-tons , sur l'empan considéré)

La représentation symbolique d'un contour prend ainsi la forme suivante :

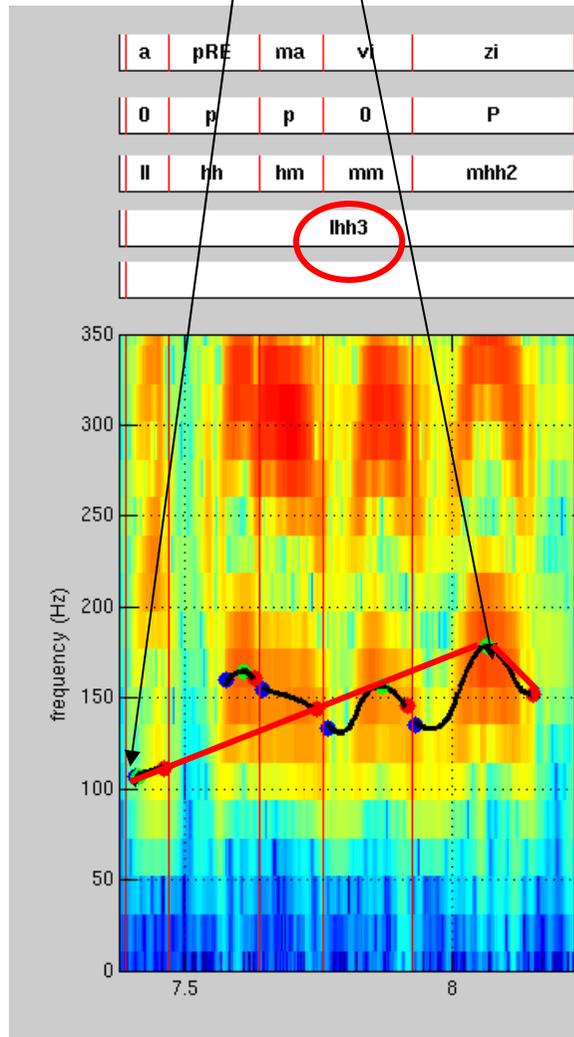
Initial Final [Extrême] [Position]

où les 2 derniers champs peuvent être absent (si point extrême non significatif)

L'intérêt de cette caractérisation par un ensemble de points est qu'elle peut s'appliquer à n'importe quel type d'empan, et pas seulement à la syllabe (en particulier, on ne suppose pas que le contour de f0 est continu sur toute la longueur de l'empan)

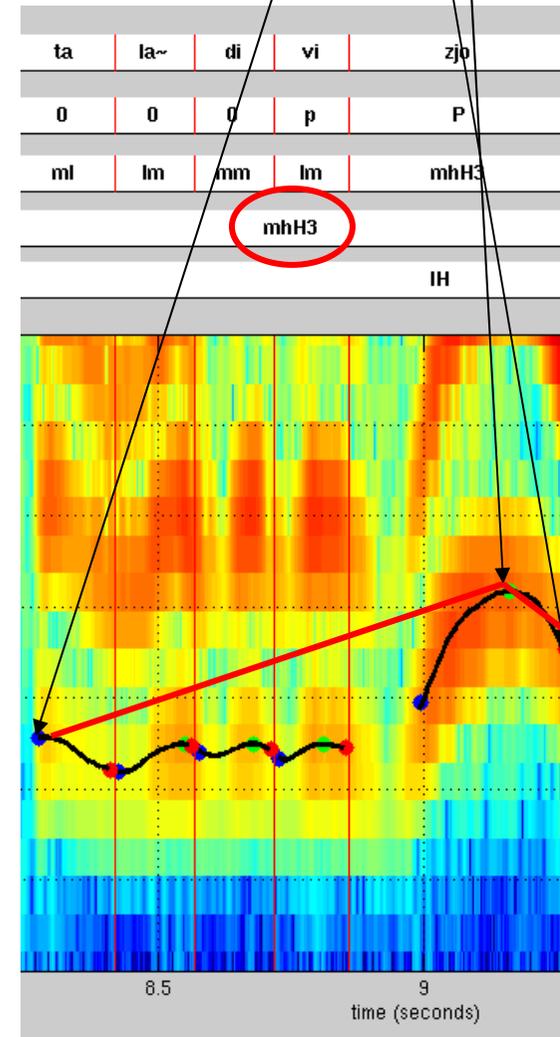
Après ma visite : l hh3

Point initial de registre bas, point final de registre haut avec un xtremum de registre haut situé dans le dernier tiers de l'empan



À Landovisiau : m h H3

Point initial de registre moyen, point final de registre haut avec un extrêmu de niveau extrême haut situé dans le dernier tiers de l'empan



Sortie textgrid

et vous continuez tout droit jusqu'au deuxième arrêt

2. TextGrid Rhap-M0017-contour

File Edit Query View Select Interval Boundary Tier Help

15.513300 0.635500 (1.574 / s)

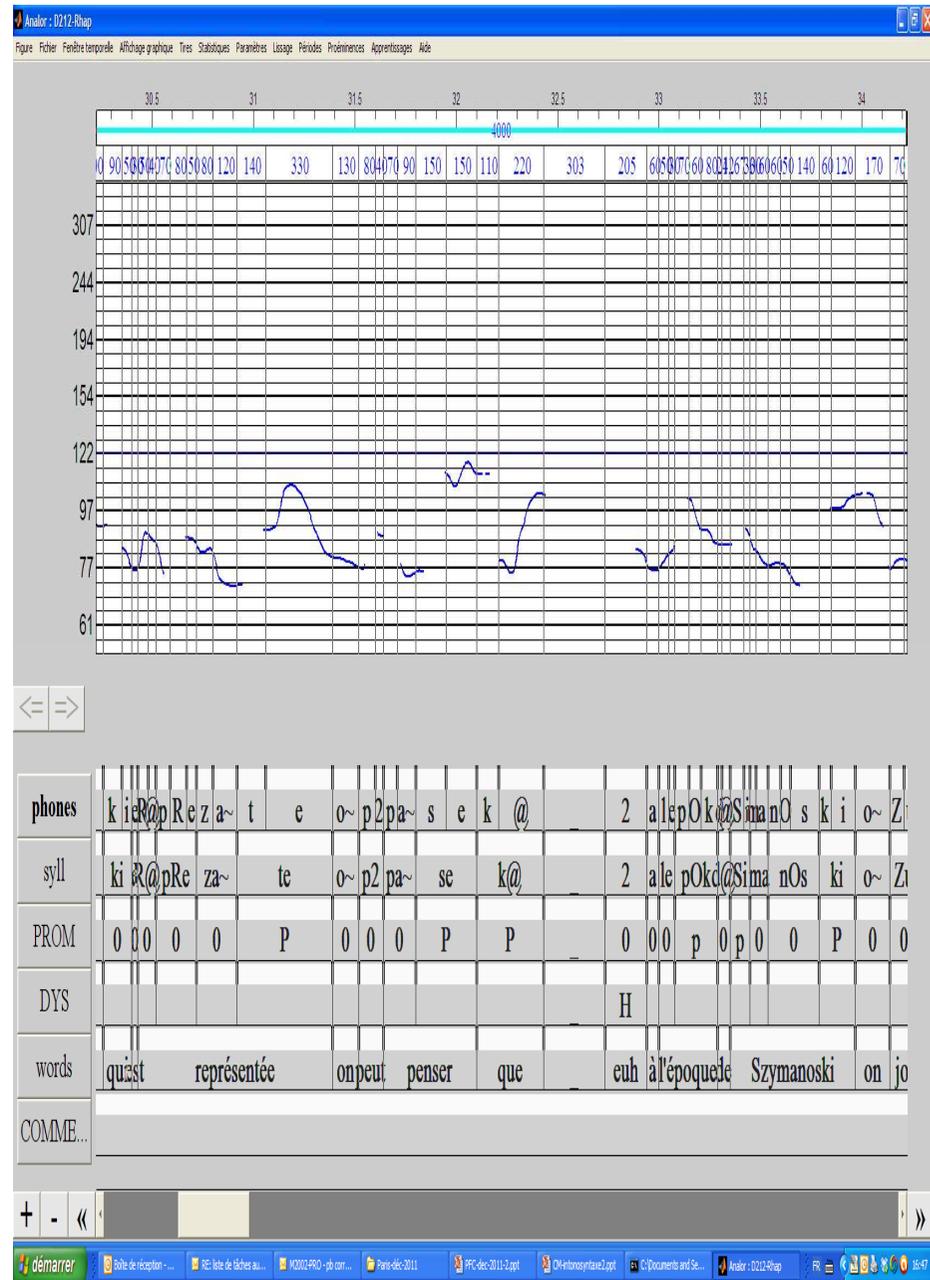
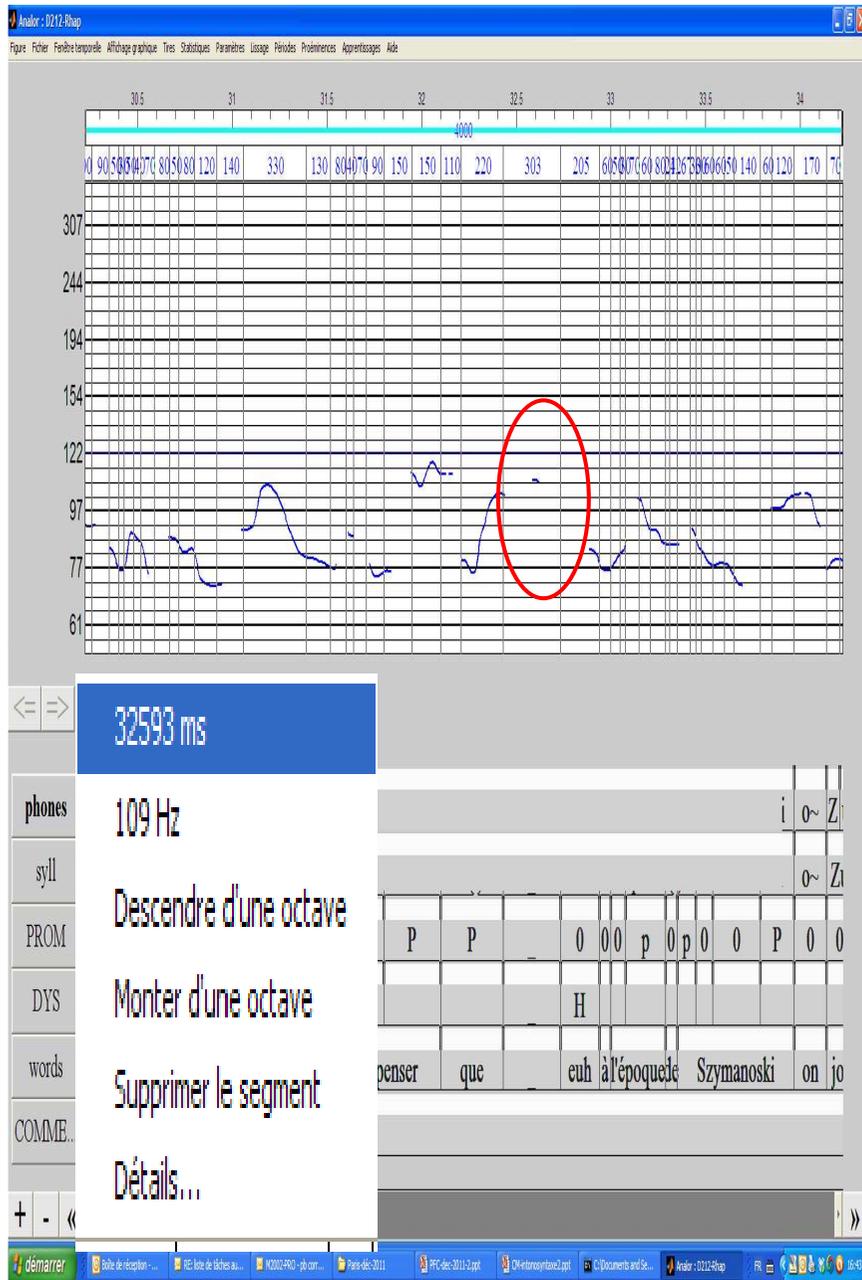
1	e	vu	ko~	ti	nHe	tu	dRwa	Zys	ko	d2	zjE	ma	RE	-	e	syll (49/162)	
2	ll	ll	mh	h h	hh	hh	hh	hm	mh	hh	hH	Hm	ml	ll	lm	contour (162)	
3	l	l	m	h	h	h	h	h	h	m	h	h	m	l	l	local,register (162)	
4	rhythmic			rhythmic			rhythmic						-	rhythmic	type (50)		
5	lh			hh			hH2						ll	lh	contour (50)		
6	m			h			h						l	m	local,register (50)		
7	intonative														-	intonative	type (34)
8	llH3														ll	llH2	contour (34)
9	m														l	m	local,register (34)

Une annotation outillée

Prétraitement

Nettoyage courbes

- Courbes nettoyées nécessaires pour les traitements automatiques (détection périodes, proéminences, annotation tonale) :
 - creaky voice, sauts d'octave, echo et bruit de fond (pb de la qualité des corpus d'entrée) 



Alignement texte son

- Méthode (requêtage) :
 - **alignement au phonème**, la syllabe ne suffit pas ! Si on veut faire des requêtes dans une structure arborescente et distribuée alignant token syntaxiques et unités prosodiques
 - Pb de la non correspondance entre mots phonologiques et mots orthographiques (épenthèse, liaison, élision, *ex il y a → ja*)

Chaîne de traitement sous easyalign

- Easyalign standard :
- **Easyalign enrichi : Gestion multilocuteurs plus chevauchements parole**
 - Locuteurs transcrits sur deux tires de transcription
 - Merging des tires
 - Marquage des segments chevauchés
 - traitement (alignement) de ces segments



Principes pour l'annotation des proéminences et des périodes sous Analor

- Un ensemble de **paramètres** acoustiques,
 - Périodes
 - Pause
 - Geste terminal
 - Saut
 - * « euh »
 - Proéminences
 - Durée syllabique
 - Hauteur F0
 - Point extrême
 - Moyenne sur empan syllabique
 - Amplitude montée
- Un ensemble de critères de segmentation (**seuils** d'activation du paramètre)
 - **Compensation** des seuils (un des paramètres en deçà du seuil)
- Tableau de paramètres (par défaut vs session en cours)
- Corpus à traiter un à un ou en série

Méthodes et outils pour le faire

Syntaxe

Annotation syntaxique

A l'heure actuelle, aucun outil d'annotation syntaxique automatique pour de l'oral transcrit

- Adapter des outils existants
- Fournir à des outils existants des entrées qu'ils peuvent traiter

Principes :

annotation manuelle et annotation automatique

2 niveaux d'annotation :

- microsyntaxe (automatique) : dépendance, fonctions, classes grammaticales
- macrosyntaxique (manuel)

Principes de l'annotation syntaxique

Rhapsodie

Annotation sémi-automatique complète d'un corpus oral en

- lemmes
- catégories
- fonctions
- dépendances
- constituants

En plus annotation manuelle de la structure macrosyntaxique des entassements (reformulations, répétitions, coordinations).

Assemblage des annotations

intonosyntaxe

Objectif : requêtes intonosyntaxiques
(tester des hypothèses, faire émerger des
constructions nouvelles (spécifiques à des
genres vs génériques, etc)

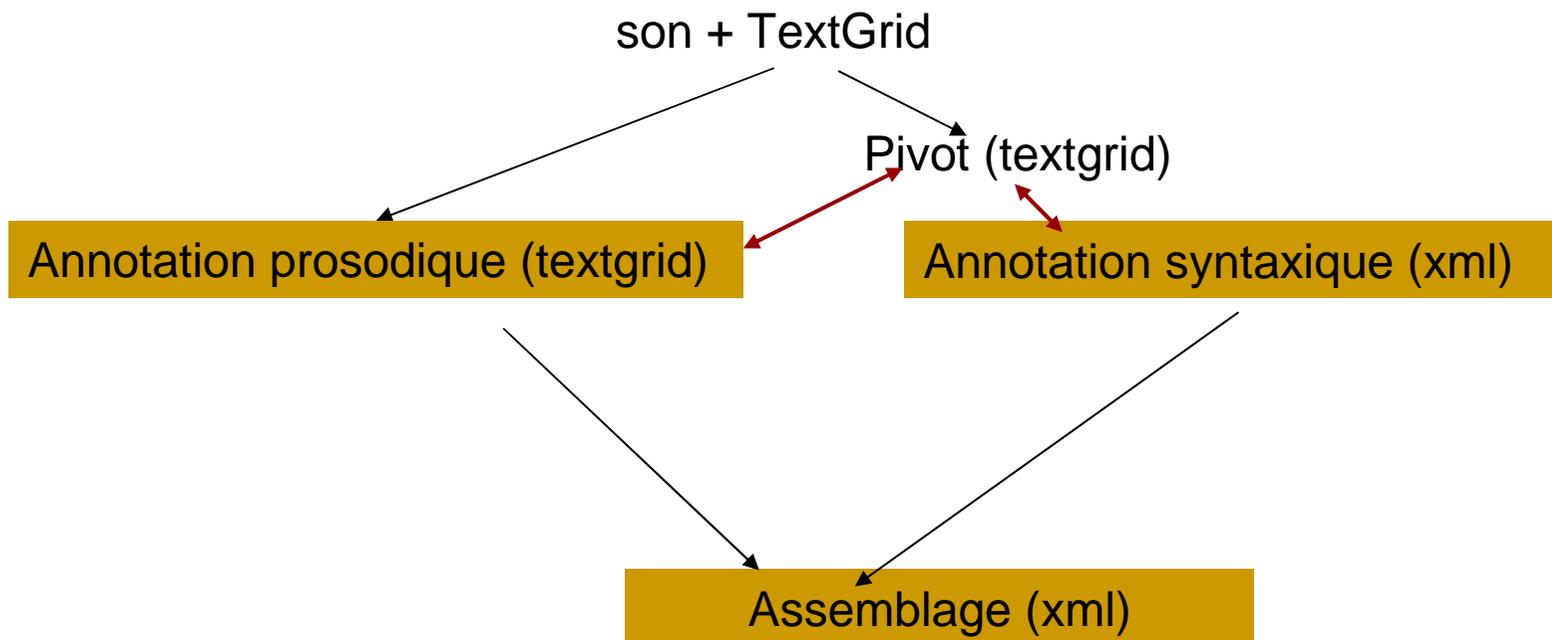
- Liste des contours variants associés à telle unité
(ex pré-noyau (cadre de discours (portée du
cadre), SN <PRO+V>, etc

Assemblage

→ vérifier tires correctement alignées temporellement et contenus cohérents d'une tire à une autre

Principe d'annotations modulaires → **autosegments**

Question non triviale : assemblage temporel des autosegments + cohérence des segmentations et transcriptions pour requêtes



Logiciel de vérification

Prosodie - Vérification - Mozilla Firefox

Fichier Édition Affichage Historique Marque-pages Outils ?

http://localhost/logiciel_corrections/

http://www.vulgaris-... Les plus visités Hotmail Maps Personnaliser les liens Windows Media Windows

Google Rechercher Partager Mes favoris Orthographe Tra

Prosodie - Vérification

Fichier à analyser

TP4-M0003-to-struct-proso.textgrid

Vérification à faire

VERIFICATION DES PROEMINENCES

OPERATIONS SPECIALES

VERIFICATION DES PROEMINENCES

VERIFICATION DES MOTS

ALIGNEMENT DES SYLLABES SUR LES PHONEMES

ALIGNEMENT DES SYLLABES SUR LES PHONEMES 1: les temps d'une syllabe sont des temps de phonème(s)

ALIGNEMENT DES SYLLABES SUR LES PHONEMES 2: tous les phonèmes ont un parent syllabe:

ALIGNEMENT DES SYLLABES AVEC...

ALIGNEMENT DES SYLLABES AVEC LES PROEMINENCES

ALIGNEMENT DES SYLLABES AVEC LES CONTOURS

ALIGNEMENT DES SYLLABES AVEC LES DISFLUENCES

ALIGNEMENT DES MOTS SUR LES SYLLABES

ALIGNEMENT DES MOTS SUR LES SYLLABES: les temps d'un mot sont des temps de syllabe(s)

ALIGNEMENT DES PIVOTS SUR LES SYLLABES

ALIGNEMENT DES PIVOTS SUR LES SYLLABES: les temps d'un pivot sont des temps de syllabe(s)

ALIGNEMENT DES PIVOTS SUR LES MOTS

ALIGNEMENT DES PIVOTS SUR LES MOTS: les temps d'un pivot sont des temps de mot(s)

ALIGNEMENT DES LOCUTEURS SUR LES PIVOTS

ALIGNEMENT DES LOCUTEURS SUR LES PIVOTS: les temps d'un locuteur sont des temps de pivot(s)

ALIGNEMENT DES PERIODES SUR LES MOTS

ALIGNEMENT DES PERIODES SUR LES MOTS: les temps d'une période sont des temps de mot(s)

[modifier un intervalle dans une tire](#)

[supprimer un intervalle dans une tire](#)

Vérification tire *proéminence*



Exemples d'erreur d'alignement token syntaxique (pivot)-syllabe

TextGrid showing phonetic segments and syllabic tokens:

d	e	z	a~	f	a~
de		za~		fa~	
0	0		S		
des			enfants		
des			enfants		

\$L1

periode n°1

Visible part 0.937500 seconds
Total duration 291.803107 seconds



TextGrid showing phonetic segments and syllabic tokens with misalignment:

d	e	z	a~	f	a~	phone (2876)
de		za~		fa~		syllabe (1359)
0	0		S			prom (1359)
des			enfants			hes (1359)
des			enfants			word (1076)
des			enfants			pivot (879)
locuteur (1/80)						
comment (53)						
periode (199)						
						2.296875
						289.506232

logiciel_corrections

C'est fini !