

Outils de description prosodique et phonostylistique de grands corpus de français parlé

Jean-Philippe Goldman
Antoine Auchlin
Dpt de Linguistique



UNIVERSITÉ
DE GENÈVE

Judi 17 Janvier 2008
Journée d'étude à l'ENS Lyon
« Analyse Linguistique des français parlés »

objectifs > approches > outils > easyalign > prosoprom > prosoreport > remarques

Objectifs

- Observer, analyser, caractériser, modéliser:
 1. l'interface discours-syntaxe ⇔ prosodie
 2. l'expressivité: émotions, attitudes, phonostyles, accents régionaux
 - L'expressivité se caractérise-t-elle par un ensemble de traits tout au long de l'énoncé ou par des manifestations ponctuelles et ciblées ?
 - Quelle perception de l'expressivité ?
- Deux principaux objets d'étude prosodiques:
 - Proéminences syllabiques et Unités Intonatives Majeures
 - Objectifs : repérage et catégorisation automatique

objectifs > approches > outils > easyalign > prosoprom > prosoreport > remarques

Approches, méthodes

- Outillage
 - gain de temps et fiabilité, surtout pour de gros corpus, reproductibilité
 - validé par le manuel
 - aller-retour avec les modèles sous-jacents
 - modularité des outils (en fonction des cadres théoriques)
 - ergonomie et mutualisation des données et outils
- Sans modèle ou théorie prosodique précise.
Approche inductive, essentiellement acoustique

objectifs > approches > outils > easyalign > prosoprom > prosoreport > remarques

Outils (quasi-)automatiques

1. *EasyAlign (Aligner)*
 - segmentation phonétique, syllabique, lexicale
2. *ProsoProm*
 - détection de syllabes proéminentes
 - Choix des paramètres acoustiques, de l'empan de contraste, de la stratégie de décision (PFC06, Interspeech07)
 - exemples
 - étude « CPF » (CMLF - juil08)
3. *ProsoReport (Phonostylographe)*
 - description phonostylistique globale
 - étude « Radio vs. Lecture » (IDP - sept07, SP - mai08)
 - étude « Régionalisme Lyon vs. Nyon » (PFC - déc07)

easyalign

Outil automatique de segmentation phonétique, syllabique et lexicale, sous Praat, à partir d'une transcription orthographique (ou phonétique)



Série de scripts organisés dans un sous-menu «EasyAlign»

1. Segmentation en « pseudo-phrases »
Texte + Son → tire *ortho*
2. Phonétisation des « pseudo-phrases »
tire *ortho* → tire *phono*
3. Segmentation en phonèmes (alignement)
tire *phono* → tire *phones + syll + words*

DEMO

Utilisation, détails internes

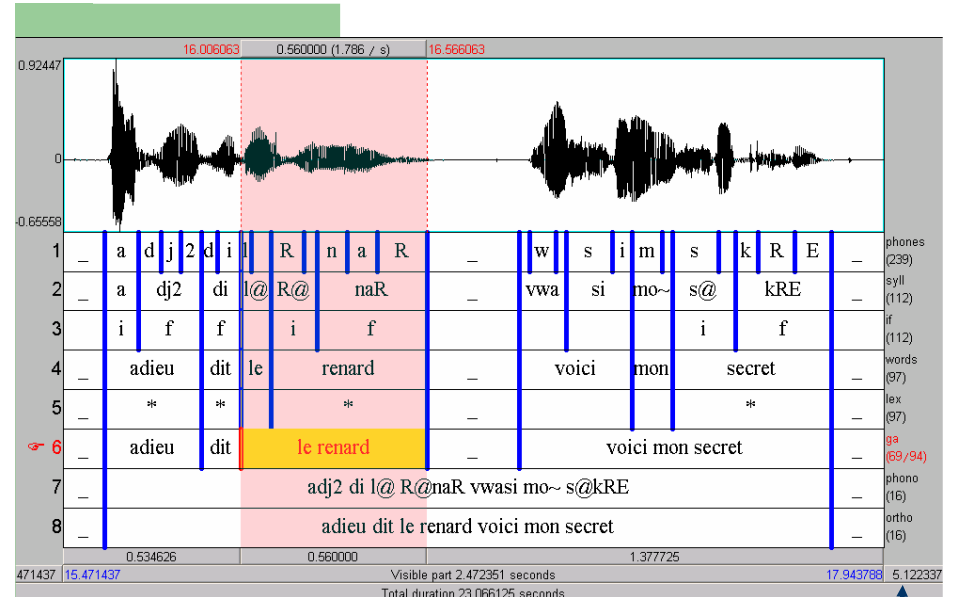
1. Segmentation en pseudo-phrases

- Pose des frontières entre chaque pseudo-phrase
Après reformatage par l'expert du texte en une pseudo-phrase par ligne, la durée de chaque pseudo-phrase est estimée par sa longueur orthographique et la durée des pauses
- Nécessite une vérification manuelle des frontières de phrases

2. Phonétisation

Performances et développements

- Evaluation : différence moyenne des positions de frontières segmentales (5mn)
 - 2 aligneurs humains: 23 ms
 - easyalign vs. humain: 28 ms
- Efficacité
 - de 2x à 5x le temps réel (vs. 80x)
- Développement en cours
 - Autres langues prises en compte (nl,hb,en,pt-br)
 - Automatisation de la détection des variantes phonologiques (élision, liaison)



tires : ortho, phono, ga, lex, words, if, syll, phones

ProsoProm

- détection de syll. proéminentes d'un enregistrement segmenté en syllabe
- décrire un enregistrement, aider l'annotateur, confirmer/infirmier un modèle, aider pour la segmentation en macro-unité

Quels paramètres prosodiques ? Pour quels seuils ?

- **Cible** mélodique après stylisation (Prosogram, Mertens), **Mouvement** mélodique
- **Durée** syllabique, **Pause** subséquente
- (...et intensité, qualité de la voix, degré d'articulation)

Dans quel contexte ? (empan de contraste)

- relativisation des paramètres
- universalité de ces paramètres (i.e. indépendance du locuteur, du style; et indépendance de la langue...)

Quelle stratégie de décision ?

- explicite : de 1 à n critères suffisants (i.e dépassant un seuil)
- statistique : analyse discriminante

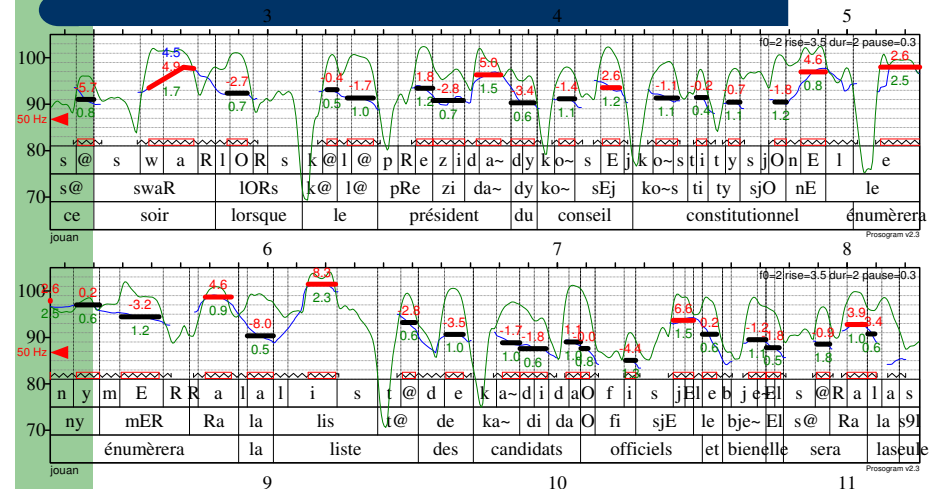
Quelle évaluation ? comparaison syllabe-à-syllabe avec un corpus annoté

Corpus

- CPF: Corpus Phonostyles du Français (Avanzi, Goldman, Simon CMLF-08)
 - Parole publique : discours officiels, média
 - Parole privée: prescription d'action, récit de vie, conversation familiale
 - Actuellement 50 mn, 21 locuteurs, F / B / CH
 - Annotation:
 - Segmenté en phonèmes, syllabes, mots (auto + vérif.)
 - Courbe mélodique vérifiée manuellement
 - Méthodologie stricte d'annotation des proéminences (o p P)
 - Tire « delivery » (hésitation, schwa post-tonique, appendice, interruptions, inaudible, prise de souffle)

ProsoProm : exemples : radio

Hélène Jouan : « Dominique Voynet qui a gagné, Nicolas, oui, **ce soir, lorsque le président du conseil constitutionnel énumèrera la liste des candidats officiels, et bien, elle sera la seule survivante du jeu écologiste.** »



ProsoProm : exemples : CPF

- Etude CMLF08: 50mn (CPF) soit 11300 syllabes

		Codage manuel		Total
		Non-proém.	Proéminent	
détection auto.	Non-proéminent	65 %	7.8 %	73.3 %
	Proéminent	6.6 %	20.1 %	26.7 %
Total		72.1 %	27.9 %	

Score: 85.1%
Précision: 75.2%
Rappel: 72.1%
F-score: 73.6%

ProsoReport (Phonostylographe)

- outil de description phonostylistique globale
- But: objectiver les propriétés prosodiques et rendre comparable la description
- Limité ici à une "photographie" globale sans l'évolution du style dans le temps de la parole («détemporalisation »)
- Basés sur des outils associés

ProsoReport : outils associés

- ▶ **alignement (syllabes)** [EasyAlign](#)
 - objectifs: mesures de durée; stylisation
- ▶ **Stylisation de la f0** [Prosogramme](#)
 - f0 segmentée = nuclei (partie voisée et intense de la syllabe)
 - f0 représentée par des segments de droite
 - objectifs: neutraliser les erreurs de détection
- ▶ **détection des proéminences** [ProsoProm](#)
 - objectif: décrire phénomènes accentuels / rythmiques / intonatifs
 - algorithme de détection automatique satisfaisant
- ▶ **annotation morphosyntaxique** [Lex](#)
 - codage : mots clitiques / non clitiques
 - objectifs : catégoriser les proéminences (finales vs non finales)

Mesures acoustiques alignement + stylisation

Alignement syllabique et stylisation mélodique

syllabes	o nombre de syllabes	
durée	o durée de parole	s. (avec pauses)
	o durée d'articulation	s. (sans pauses)
	o taux d'articulation	% (articul./ parole)
	o débit de parole	syll./s. (pour parole)
	o débit d'articulation	syll./s. (pour articul.)
	o durée moyenne syllabes + déviation std	s.
	o durée moyenne nuclei ⁴ syl. + dév. std	s.
f0 globale	o proportion des nuclei pdt l'articulation	%
	o moyenne + déviation standard	ST (relatif à 1 Hz)
	o étendue (inter quantile range)	ST
f0 tons	o proportion des tons statiques,	%
	o prop. des tons dynamiques montants,	%
	o prop. des tons dynamiques descendants	%
	o prop. des tons dynamiques complexes	%
f0 mvt	o mouvement intra-nuclei des tons	
	o dynamiques (excluant les statiques)	ST/syll. et ST/sec.
	o mouvement absolu intra-nuclei (pour tous les tons)	ST/syll. et ST/sec.
	o mouvement absolu inter-nuclei (i.e mvt moyen entre les tons)	ST/syll. et ST/sec.
	o agitation mélodique	ST/syll. et ST/sec.
intensité	o intensité moyenne dans les nuclei	dB
	o intensité moyenne hors des nuclei	dB
	o différence d'intensité entre nuclei et non-nuclei	dB

Mesures acoustiques alignement + stylisation + proéminences

Détection de proéminences

syllabes	proportion de syllabes proéminentes	%
	proportion de syllabes non proéminentes	%

durée, f0, intensité, ... des syllabes proéminentes vs non proéminentes

alignement + stylisation + lex

Annotation morphosyntaxique

syllabes	proportion de syllabes en position initiale de mot lexical	%
	proportion de syllabes en position finale de mot lexical	%

durée, f0, intensité, ... des syllabes initiales vs finales

alignement + stylisation + lex + proéminences

Annotation morphosyntaxique et détection de proéminences

syllabes	proportion de syllabes proéminentes en position initiale de mot lexical	%
	proportion de syllabes proéminentes en position finale de mot lexical	%

ProsoReport : étude IDP07

- hypothèse générale
 - Callamand 1973; Léon; Burger & Auchlin 2007
 - phonostyles spécifiques: au genre radiophonique; au genre "chronique"; à certaines chaînes de radio (FI)
- hypothèses particulières à notre étude
 - par rapport à la "lecture neutre", le style radio lu se caractérise par (Goldman & Auchlin 2006)
 - H1: surarticulation
 - H2: plus grande agitation mélodique
 - H3: quantité plus élevée d'accents initiaux
- matériau
 - 3 chroniques radiophoniques de 2mn (2 femmes, 1 homme; France Inter, France Info)
 - relues de manière neutre par une locutrice unique

H1 taux global d'articulation

Tableau 2 : Durées de parole, durées d'articulation (en sec.) et taux d'articulation (en %)

	L-j	L-a	L-g	Lmoy	R-moy	FR-j	FI-a	FI-g
Durée de parole	161,2	138,9	122,2	140,8	128,1	149,6	123,2	111,5
Durée d'art.	127,8	110,9	99,9	113	112,4	132,8	110,6	93,7
Taux d'art.	79,2	79,8	81,8	80,2	87,5	88,8	89,8	84

→ taux d'occupation du canal de parole (ratio articulation / silences)

H1 durée des noyaux vs. non noyaux

Tableau 10 : Durée des noyaux (en secondes) et proportion des noyaux sur la durée totale d'articulation (sans les pauses) en %

	L-j	L-a	L-g	L-moy	R-moy	FR-j	FI-a	FI-g
Durée noy.	60,09	51,767	41,747	153,6	138,5	56,588	47,342	34,574
% noy/artic.	47	46,7	41,8	45,17	40,77	42,6	42,8	36,9

→ proportion des "consonnes" par rapport aux noyaux voisés des syll. (sur le temps d'articulation)

H2: étendue du registre

Étendue du registre tonal : étendue totale et étendue calculée en exceptant les 5% de syllabes les plus hautes et les 5% les plus basses (en demi-tons relatifs à 1 Hz)

	L-j	L-a	L-g	L-moy	R-moy	FR-j	FI-a	FI-g
Étendue f0 max-min	15,8	10,9	23	16,57	18,87	23	15,4	18,2
95%-5%	8,6	8,4	8,7	8,57	12,90	13,8	12,2	12,7

H2: mvts mélod. intra- et inter-syll.

: Mouvements mélodiques (en ST/sec.)

	L-j	L-a	L-g	L-moy	R-moy	FR-j	FI-a	FI-g
Mouvements intra-nuclei des tonsdyn.	18,2	20	20	19,40	25,67	26,3	26,1	24,6
Intra-mvt	2	3	2,9	2,63	4,40	4,5	5,3	3,4
Inter-mvt	13,3	14,5	14,9	14,23	17,23	18,4	15	18,3
Agitation	15,3	17,5	17,8	16,87	21,63	22,9	20,3	21,7

→ agitation mélodique: parcours tonal cumulé, en ST par seconde

H3 : proportion de syllabes proéminentes en position finale vs non finale

Syllabes proéminentes : pourcentage de syllabes proéminentes par rapport aux syllabes non proéminentes ; pourcentage de syllabes proéminentes en position d'accent initial (syllabe initiale de mot lexical polysyllabique) ; pourcentage de syllabes proéminentes en position d'accent final (dernière syllabe de mot lexical)

	L-j	L-a	L-g	L-moy	R-moy	FR-j	FI-a	FI-g
Prom.	29,5	33,2	34,5	32,40	37,03	35,6	35,9	39,6
	(228/772)	(214/644)	(181/525)			(269/755)	(229/638)	(201/508)
Prom/i	19,7	18,1	26,3	21,37	31,43	30,1	30,3	33,9
Prom/f	58,6	64,5	58,8	60,63	59,6	59,3	59,4	60,1

→ davantage de proéminences en position initiale de mots lexicaux en parole radio

Phonostylographe : résultats

- H1 (surarticulation)
 - 😊 taux d'articulation (%)
 - 😊 durée des consonnes / durée d'articulation
 - 😞 intensité moyenne noyaux / intensité moy. parole
- H2 (agitation mélodique)
 - 😊 étendue du registre de f0
 - 😊 étendue du registre tonal
 - 😊 "chemin mélodique" (parcours tonal cumulé)
 - 😊 + de tons dynamiques en radio
- H3 (quantité d'accents initiaux)
 - 😊 proportion des syllabes proéminentes
 - 😊 proportion des syllabes proéminentes en position AI

ProsoReport

- méthode d'analyse contraire à l'idée qu'un phonostyle peut être marqué par un petit nombre de traits disséminés (émaillage, Ch. Bally)
- méthode d'analyse qui "aplatit" en une dimension la variation ou l'évolution temporelle au cours d'un enregistrement des phénomènes étudiés
- permet la comparabilité de corpus (tableaux)
- permet de faire émerger des différences non prévues (bottom up)
- permet d'utiliser ou non de l'information linguistique (description "substance" → description "phonologique")

Conclusion/Remarques

- Approche objective d'aspects locaux et globaux de la description prosodique et phonostylistique
- Injection maîtrisée de connaissances linguistiques
- (in)dépendance des outils et de l'annotation manuelle vis-à-vis du langage
- Outils modulables mutualisés
- **Développements supplémentaires**
 - meilleure précision temporelle dans la syllabe
 - modèle dynamique intra-corpus
 - ancrage linguistique et discursif plus sophistiqué que « lex »