

Guide méthodologique

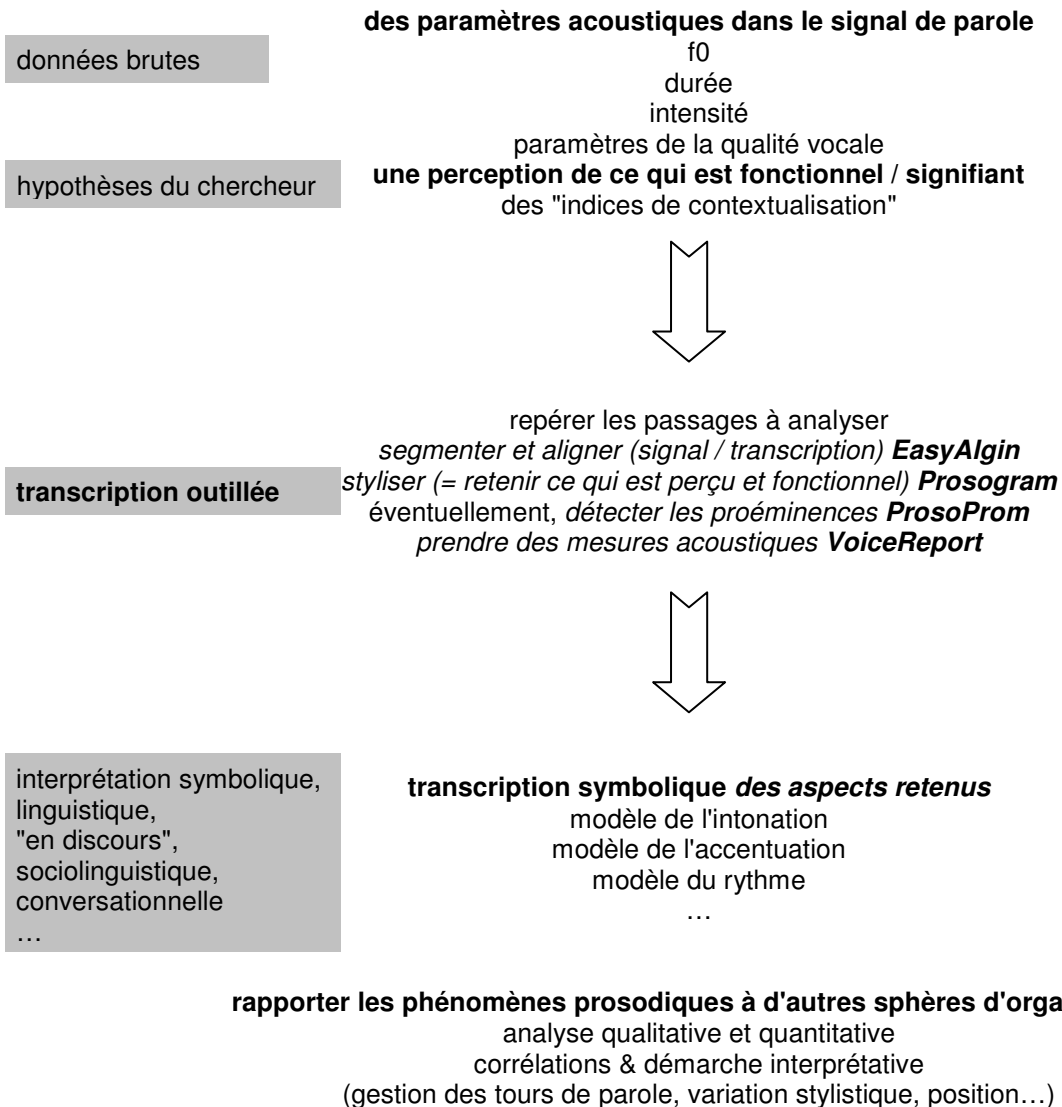
Transcription outillée – prosodies

Anne Catherine Simon
UCLouvain – juin 2007

Table des matières

1.	DANS QUELLE DÉMARCHE S'INSCRIT UNE TRANSCRIPTION <i>OUTILLÉE</i> ?	3
2.	PARAMÈTRES PROSODIQUES ET OUTILS D'ANALYSE (SEMI)AUTOMATIQUE	4
3.	ALIGNEMENT: <i>EASYALIGN</i>	5
4.	FRÉQUENCE FONDAMENTALE (STYLISATION)	7
5.	STYLISATION: <i>PROSOGRAM</i>	7
6.	DURÉE (DÉBIT, ALLONGEMENT)	8
7.	INTENSITÉ	9
8.	QUALITÉ VOCALE	9
9.	REGISTRE ET ÉTENDUE TONALE	10
10.	PROÉMINENCE: <i>PROSOPROM</i>	11
11.	ACCENTUATION	12
12.	INTONATION (TONS ET CONTOURS)	13
13.	RYTHME	14
13.1.	DENSITÉ ACCENTUELLE	14
13.2.	RYTHME: ISOCHRONIE ET ISOMÉTRIE	14
13.3.	TRANSCRIPTION	15
13.4.	GRILLE D'ANALYSE (TEXTGRID) POUR LES SCANSIONS RYTHMIQUES	16
14.	MESURES ACOUSTIQUES AUTOMATIQUES: <i>VOICE REPORT</i>	17

1. Dans quelle démarche s'inscrit une transcription *outillée*?



2. Paramètres prosodiques et outils d'analyse (semi)automatique

fréquence fondamentale	durée (vitesse)	intensité sonie	qualité vocale voir d'Alessandro 2006
<i>segmentation & alignement</i> (phones, syllabes, mots, macro-unités) EASYALIGN <i>stylisation</i> PROSOGRAMME			
registre (hauteur moyenne) étendue (plage utilisée)	temps d'articulation débit allongement	?	
<i>détection des proéminences</i> PROSO PROM <i>interprétation fonctionnelle (symbolique)</i>			
accentuation types d'accents (selon leur forme, leur localisation, leur fonction)			
intonation [représentation symbolique] tons / morphèmes / contours intonatifs	rythme densité accentuelle tempo isochronie – isométrie		
<i>mesures acoustiques globales</i> sur le signal aligné VOICE REPORT <i>interprétation séquentielle des phénomènes perçus</i> "indices de contextualisation"			

3. Alignement: *EasyAlign*

L'annotation outillée de la prosodie implique un **alignement**, c'est-à-dire une synchronisation entre le signal acoustique (paroles, bruits de fond) et une annotation, via l'indication fine du temps écoulé (time codes). Des logiciels comme Praat ou WinPitch permettent de créer de tels fichiers d'alignement. Les unités choisies pour l'alignement dépendent des paramètres qu'on souhaite analyser:

Tableau (1) Unités linguistiques d'alignement du signal sonore (des plus micro aux plus macro)

unité d'alignement	définition, commentaires
segments	sons (phones décrits d'après l'alphabet SAMPA) phonèmes (ensemble réduit aux symboles fonctionnels)
silences	détectés manuellement ou selon le "modèle de silence" du logiciel utilisé pour les détecter (traitement des prises de souffle, etc.)
syllabes	recomposées à partir de l'alignement phonétique selon un système de règles, ou segmentées manuellement
mots	mots graphiques (séparés par une espace ou une apostrophe) ou mots selon le lemme (mots composés) – les frontières de mots ne coïncident pas avec les frontières de syllabes
marques du travail de formulation	hésitation (euh ou syllabes allongées), parties non traitables automatiquement (parole superposée, rires, paralinguistique ...)
groupes accentuables	mots accentuables + clitiques dépendants
macro-unités	IPU, période, tour de parole, pseudo-énoncé

Le choix d'une **macro-unité** d'alignement se révélera crucial pour certains traitements qui nécessitent un domaine inférieur à l'enregistrement complet. Soit on identifie ces unités manuellement, soit on les découpe de manière automatique, avant ou après la transcription orthographique.

Tableau (2) Macro-unités d'alignement (de type "discours" ou "conversationnelles")

	segmentation manuelle	segmentation automatique (critères acoustiques)
unité relativement brève	pseudo-énoncé critères = frontière intonative majeure + autonomie syntaxique + absence d'hésitation	inter-pausal unit critère = pause silencieuse > 200ms
unité de durée variable	production verbale ¹ critère = changement de locuteur	période critères = pause silencieuse > 300ms + geste et saut f0 + absence d'hésitation

¹ Les trois autres types (pseudo-énoncé, inter-pausal unit et période) présupposent une segmentation en productions verbales: on identifie ces unités dans la parole de chaque locuteur.

Unités séparées par des pauses (IPU)

Les transcriptions sont effectuées à partir d'un pré-découpage automatique du signal de parole en Interpausal-Units (IPU). Les IPU sont des blocs de parole bornés par des pauses silencieuses d'au moins 200 ms (durée pouvant varier selon les langues). L'IPU est souvent utilisée sur des corpus de taille importante. Par sa nature formelle et objective, elle se distingue d'autres unités « prosodiques » telles que les unités intonatives par exemple, dont le découpage nécessite l'intervention manuelle d'experts, pouvant de surcroît afficher un désaccord [Koiso et al., 1998]. Ce découpage en IPU facilite non seulement la transcription mais s'avère indispensable pour les étapes de phonétisation et d'alignement avec le signal audio. [...] La procédure automatique de segmentation en IPU consiste en une détection du non voisé/voisé et d'un seuillage pour distinguer une réelle pause silencieuse d'un temps de silence dans une occlusive par exemple. (Bertrand et al. 2007)

Période intonative

Il y a segmentation en périodes intonatives si l'ensemble des règles suivantes est respecté : (1) la durée de la **pause**, ou plus précisément l'intervalle entre deux portions de F0, dépasse un seuil de l'ordre de **300 ms** ; (2) l'amplitude du **geste** – ou la différence de hauteur entre le dernier extremum de F0 et la moyenne de F0 sur toute la portion qui précède la pause – dépasse un seuil de l'ordre de 4 demi-tons ; (3) l'amplitude du **saut**, c'est-à-dire la différence de hauteur entre la dernière valeur de F0 précédant la pause et la première valeur de F0 suivant la pause (reset mélodique), dépasse un seuil de l'ordre de 3 demi-tons ; (4) il n'y a pas d'hésitation (*eah* ou syllabe démesurément allongée) à proximité immédiate de la pause. [Lacheret-Dujour et Victorri 2002]

Unités de production verbale (≈ tour de parole)

Une production verbale est la suite [d'occurrences] rattachée à un identifiant de locuteur apparaissant en début de ligne dans la transcription. [...] C'est une unité qui répond aux besoins d'une analyse automatique et qui ne correspond pas au tour de parole de la linguistique interactionnelle. Le tour de parole - tel que défini dans les analyses de la linguistique interactionnelle - est une unité difficilement identifiable automatiquement. [groupe ICOR 2006, *Glossaire*, site CORINTE, <http://icar.univ-lyon2.fr/projets/corinte/>]

Pseudo-énoncé

Unité délimitée manuellement et qui, à l'intérieur d'un tour de parole, correspond à un "énoncé" présentant une relative indépendance syntaxique, une relative autonomie intonative et une relative complétude pragmatique (voir la notion d'unité de construction de tour). [Bachy S., Hambye Ph., Simon A.C. 2006. *Tutoriel Praat, Cahier 2, Transcription et alignement du texte sur le son*. <http://valibel.fltr.ucl.ac.be/bd.htm> ; Goldman 2007]

Bertrand R., Blache P., Espesser R., Ferré G., Meunier C., Priego-Valverde B., Rauzy S. 2007. Le CID, Corpus of Interactional Data: protocoles, conventions, annotations. *Travaux Interdisciplinaires du Laboratoire Parole et Langage d'Aix-en-Provence* 25, 25-55.

Lacheret-Dujour A. & Victorri B. 2002. La période intonative comme unité d'analyse pour l'étude du français parlé: modélisation prosodique et enjeux linguistiques, *Verbum* 24/1-2, 55-73.

► Outil de segmentation et d'alignement semi-automatique

EasyAlign <http://latlcui.unige.ch/phonetique/>

Analor Contacter Anne Lacheret (Paris X): anne@lacheret.com

Éventuellement, contacter Cyril Auran ou Roxane Bertrand pour un script qui délimite les "unités inter-silence"

4. Fréquence fondamentale (stylisation)

Articulaire	fréquence des vibrations laryngées
Perceptif	hauteur (registre, étendue) mélodie (ton, contour)
Acoustique	fréquence fondamentale (f0): estimation de la fréquence laryngienne à partir du signal de parole; harmonique de plus faible rang dans le spectre complet d'un son périodique, intervalle entre deux harmoniques successifs (Lacheret-Dujour & Beaugendre 1999: 276)
Unités	<u>fréquence</u> d'un signal à un instant donné Hz : nombre de vibrations / seconde demi-tons : conversion sur une échelle perceptive; DT relatifs à 1 Hz => $F_{dt} : 40 * \log_{10} (F_{Hz}/1)$

La détection des valeurs de f0 se fait à l'aide d'un logiciel d'analyse acoustique [www.praat.org ou www.winpitch.com]. La conversion en demi-tons permet de rendre certaines mesures (étendue du registre) comparables entre différents locuteurs.

5. Stylisation: Prosogram

La **stylisation** vise éliminer tous les détails du contour mélodique qui ne jouent pas de rôle communicatif de manière à conserver les propriétés essentielles de la courbe². Le résultat d'une stylisation peut prendre la forme d'une ligne continue (le système de modélisation mélodique MOMEL interpole des valeurs pour les parties non voisées de la courbe) ou d'une suite de segments (les "tons" du prosogramme).

Si la stylisation se fonde sur un découpage du signal en syllabes ou en phones (pour ne traiter que les voyelles), une étape préalable de **segmentation** et d'alignement est requise. Si la stylisation travaille sur la courbe sans aucune information linguistique, aucun alignement préalable n'est requis. Toutefois, l'interprétation de la courbe de f0 nécessite toujours de rapporter les mouvements à des unités linguistiques.

La stylisation, qui résulte dans une version simplifiée de la courbe, peut être assortie d'une annotation en "tons" ou en "points cibles". Il s'agit d'une étape ultérieure, interprétative, qui propose une **catégorisation symbolique** des mouvements de f0 selon un modèle phonologique (nombre restreint de symboles).

Campione E., Flachaire E., Hirst D. & Véronis J. 1998. Évaluation de modèles d'étiquetage automatique de l'intonation. *Actes des 22èmes Journées d'Étude sur la Parole*. Martigny (Suisse), 99-102.

Delais-Roussarie, E., Caelen-Haumont G., Hirst D., Martin Ph. & Mertens P. 2006. Outils d'aide à l'annotation prosodique de corpus, *Bulletin PFC* 6, 7-26³.

*Hermes D. J. 2006. Stylization of Pitch Contours. In Sudhoff S. et al. (eds). *Methods in Empirical Prosody Research*, Berlin-New York: Walter de Gruyter, 29-61. [[comparaison de méthodes de stylisation](#)]

Louw J.A. & Barnard E. 2004. Automatic intonation modeling with INTSINT. *Proceedings of the Pattern Recognition Association of South Africa*. www.meraka.org.za/pubs/louwja04intsint.pdf

Martin Ph. (?) A propos de la perception et la transcription des unités prosodiques. Perpignan (référence à compléter)

Mertens P. 2004. Un outil pour la transcription de la prosodie dans les corpus oraux. *Traitement Automatique des langues* 45 (2), 109-130 <http://bach.arts.kuleuven.be/pmertens/publications.html>

► Outil pour la stylisation de la f0 et l'identification de "noyaux" syllabiques et ou vocaliques

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

Momel <http://stl.recherche.univ-lille3.fr/sitespersonnels/auran/english/index.html> (research / tools)

² Les micro variations mélodiques liées aux aspects segmentaux ou qui ne sont pas perçues par l'auditeur.

³ www.projet-pfc.net/?u_act=download&dfile=PFC_6_02delais_caelen_hirst_martin_mertens.pdf

6. Durée (débit, allongement)

Articulaire	vitesse d'articulation
Perceptif	débit (nombre d'unités par unité de temps) allongement (d'un son, d'une syllabe) rythme, tempo (en lien avec les proéminences)
Acoustique	durée
Unités	temps: ms

Temps de parole = temps d'articulation + temps de pause

Temps d'articulation soustraire la durée des pauses de la durée totale de prise de parole; en général, on considère tout silence d'une durée supérieure à 250ms comme une pause silencieuse. Les pauses peuvent être identifiées manuellement ou automatiquement (de nombreux scripts permettent d'isoler les pauses; à cet égard, chaque script ne traite pas de la même manière les prises de souffle, etc.)

Débit vitesse d'articulation mesurée au sein des macro-unités de segmentation, par ex. en nombre de syllabes / seconde (par tour de parole, énoncé...)

Débit moyen vs variations de débit Le débit moyen d'une prise de parole consiste simplement à diviser la durée totale de la prise de parole par le nombre de syllabes. Néanmoins, il peut être utile de calculer le débit moyen du temps de parole et le débit moyen du temps d'articulation. Calculer les variations du débit, c'est-à-dire la manière dont le débit évolue pour un même locuteur ou entre deux locuteurs, implique une double segmentation: en unités minimales (des sons ou des syllabes) et en macro-unités (unités séparées par des pauses, tours de parole...). À nouveau, il faut décider d'inclure ou non les pauses silencieuses aux unités macro.

Tempo Le nombre de syllabes proéminentes / seconde (calculé sur le temps de parole ou sur le temps d'articulation).

Allongement le plus souvent, allongement vocalique: augmentation significative de la durée d'une voyelle, qui peut correspondre à un phénomène d'hésitation, à un trait régional ou phonostylistique ou au marquage de la fin d'une unité intonative

Remarque: l'identification d'une syllabe ou d'une voyelle comme allongée est une opération délicate: l'allongement, outre qu'il peut être corrélé à certains types de sons (voyelles nasales vs orales), est dû à des facteurs multiples. L'allongement est rarement perçu comme tel mais souvent il participe d'une perception globale de proéminence ou de phénomène "marqué".

Grosjean, F. & Deschamps, A. (1975). Analyse contrastive des variables temporelles de l'anglais et du français: vitesse de parole et variables composantes, phénomènes d'hésitation. *Phonetica* 31, 144-184
 Uhmann S. 1992. Contextualizing relevance: on some forms and functions of speech rate changes in everyday conversation, in P. Auer & A. Di Luzio (ed.), *The contextualization of language*, Amsterdam / Philadelphia: John Benjamin, 297-336.

Voir [rythme](#).

7. Intensité

Articulaire	flux d'air, expiration, sonorité
Perceptif	niveau sonore (loudness) sonie = intensité perçue (racine carrée de la puissance acoustique, Kochanski 2006)
Acoustique	intensité (énergie contenue dans le signal de parole durant un intervalle de temps donné)
Unités	énergie, puissance, pression acoustique mesurée en dB sonie: sones ou phones

Kochanski, G. 2006. Prosody beyond Fundamental Frequency. In Sudhoff S. et al. (eds). *Methods in Empirical Prosody Research*, Berlin-New York: Walter de Gruyter, 89-121.

8. Qualité vocale

Récemment, les chercheurs en prosodie, et particulièrement ceux qui développent des systèmes de synthèse de parole, ont commencé à prendre en compte les paramètres vocaux (phonation et articulation) responsables de la perception de différents types de qualité vocale.

D'Alessandro (2006: 67) décrits les paramètres acoustiques liés à la source vocale selon quatre axes:

1. the voice register dimension: qui dépend des mécanismes sous-jacents de vibration des cordes vocales (mécanisme 1: voix modale ou de poitrine; mécanisme 2: voix de tête ou *falsetto*; mécanisme 0: *creaky voice*)
2. the noise dimension: représente la proportion relative de bruit (apériodicités) dans le signal de parole, selon le type de phonation (avec du "souffle", voix murmurée dévoisée, voix rauque...)
3. the pressed/laxed dimension: voix "étranglée" ou relâchée
4. the effort dimension: voix "forte" ou faible, dont la conséquence acoustique est l'équilibre spectral; souvent liée à la perception des prééminences.

Si l'identification auditive des aspects de la qualité vocale ne pose pas trop de problèmes (on reconnaît une voix tendue, ou murmurée), l'analyse acoustique des paramètres responsables en est ardue. Se reporter à D'Alessandro pour des consignes et des explications.

Les modifications de la qualité vocale sont souvent à interpréter dans la dimension attitudinale ou émotionnelle du discours, c'est-à-dire les fonctions dites "expressives" ou pragmatiques (ton, position, attitude, émotion, relation au discours...), y compris des informations sociolinguistiques sur le sexe, l'âge ou l'appartenance socio-culturelle du locuteur.

D'Alessandro C., Boula De Mareuil P., Prudon R. 2004. Synthèse par sélection: prosodie, dialogue et qualité vocale. *Cahiers de l'Institut de Linguistique de Louvain* 30/1-3, 153-180.

D'Alessandro C. 2006. Voice Source Parameters and Prosodic Analysis. In Sudhoff S. et al. (eds). *Methods in Empirical Prosody Research*, Berlin-New York: Walter de Gruyter, 63-87.

Scherer, K. 2003. Vocal communication of emotion: A review of research paradigms. *Speech Communication* 40, 227-256.

9. Registre et étendue tonale

Les variations de la fréquence fondamentale appréhendées à un niveau local (d'une syllabe à la suivante) s'interprètent en tons ou en contours intonatifs. À un niveau global, ces variations affectent le registre du locuteur, c'est-à-dire la hauteur moyenne de la fréquence de sa production verbale, et son étendue, c'est-à-dire la plage de fréquence qu'il exploite.

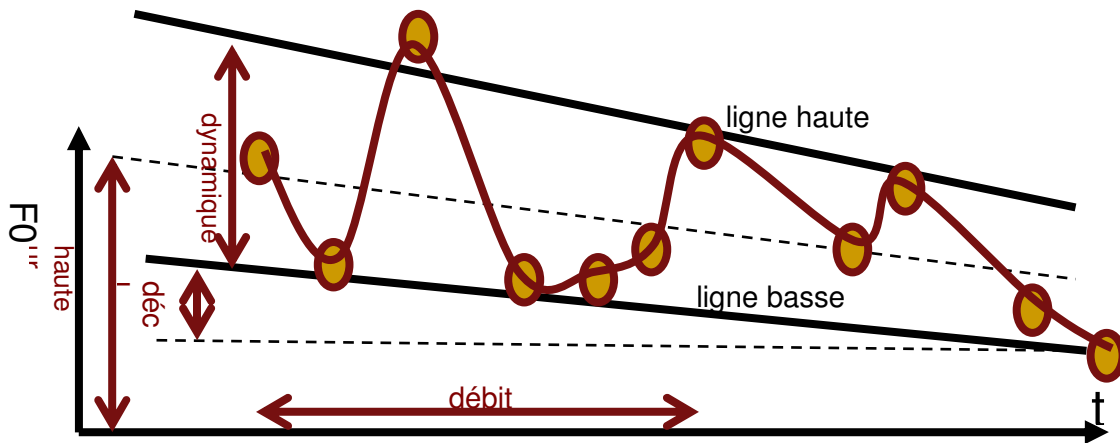


Figure (1) Registre usuel (étendue) et ligne de déclinaison (schéma tiré de Goldman 2004)

Que l'on procède manuellement ou automatiquement, il existe différentes manières de mesurer le registre et l'étendue. Ménard (2003: 42) mesure le registre et l'étendue au sein des "énoncés":

Le **registre** en tant que **hauteur de fréquence** a été calculé par le biais d'une moyenne des valeurs de fréquence fondamentale extraites au centre de la durée de chaque voyelle. Il est considéré que seuls les noyaux vocaliques sont porteurs des variations de registre. Les mesures brutes, en hertz, ont été préalablement transformées en demi-tons, échelle qui représente plus fidèlement le mode logarithmique d'intégration, par l'oreille, des fréquences exprimées en hertz.

L'**étendue** de la voix ou **plage de fréquence exploitée**, autre composante du registre, a été estimée selon deux méthodes. (1) D'abord, l'écart entre la fréquence vocalique maximale et la fréquence vocalique minimale non nulles, converties en demi-tons, pour chaque locuteur et pour chacun des énoncés, constituait l'étendue brute. (2) Afin de neutraliser l'effet d'accents à fréquence particulièrement haute ou d'abaissement finaux, l'étendue a également été déterminée en évaluant l'écart entre la fréquence moyenne des syllabes accentuées et la fréquence moyenne des syllabes inaccentuées (étendue moyenne).

Di Cristo et al. (2004: 55) n'extraient pas de la courbe des mesures brutes (en ST) mais points cibles annotés à l'aide de symboles (voir [intonation](#)) dans la courbe mélodique stylisée.

Le **niveau global** de registre (*register level*) pour un domaine donné correspond au registre moyen de référence du locuteur et aux abaissements ou rehaussements ponctuels de ce registre moyen.

Les variations de la **gamme tonale** (*register span*) correspondent aux "variations de l'intervalle de hauteur entre les cibles hautes et basses dans une séquence donnée"

Di Cristo A. et al. 2004. Outils prosodiques et analyse du discours. *Cahiers de l'Institut de Linguistique de Louvain* 30/1-3, 27-84.

Ménard L. 2003. Le registre en tant que marqueur prosodique de dialectalité: le cas du français du Québec et de la France. In M. Demers (ed), *Registre et voix sociale*, Québec, Nota bene, 35-58.

10. Proéminence: *ProsoProm*

Est proéminent un élément qui se **détache** de son entourage à cause d'une différence de durée, d'intensité, de hauteur mélodique ou de qualité vocale. La proéminence est une notion relative et graduelle (vs binaire). A priori, un élément linguistique de taille variable (syllabe, mots, énoncé) peut être proéminent. Cependant la **syllabe**, en ce qu'elle porte les variations prosodiques perçues, est souvent considérée comme le domaine approprié de la proéminence, qu'on considérera donc comme une propriété de la syllabe.

La notion de proéminence est souvent liée à l'**accentuation**; on considère alors qu'une syllabe proéminente est accentuée ou qu'elle est porteuse d'un contour intonatif fonctionnel (frontière, focus...).

L'identification manuelle ou automatique des syllabes proéminentes est une étape **indispensable** et préalable à toute analyse prosodique. En effet, les trois grandes organisations prosodiques que sont l'accentuation, l'intonation et le rythme, s'ancrent sur les syllabes proéminentes.

Il est très difficile d'établir un **accord** quant à l'identification d'une syllabe comme proéminente (Avanzi et al. 2007). L'identification d'une syllabe proéminente dépend

- de l'empan temporel écouté (plus court sera l'extrait écouté, plus nombreuses seront les syllabes perçues comme proéminentes)
- des paramètres privilégiés (durée, pause, f_0 ...)
- du "modèle linguistique" (accepte-t-on qu'un *eah* d'hésitation allongé soit considéré comme proéminent?)

ProsoProm est un outil de détection automatique des proéminences qui permet, en ajustant les seuils de détection, de repérer de manière fiable les syllabes qui se distinguent des syllabes environnantes par leur allongement ou par leur f_0 .

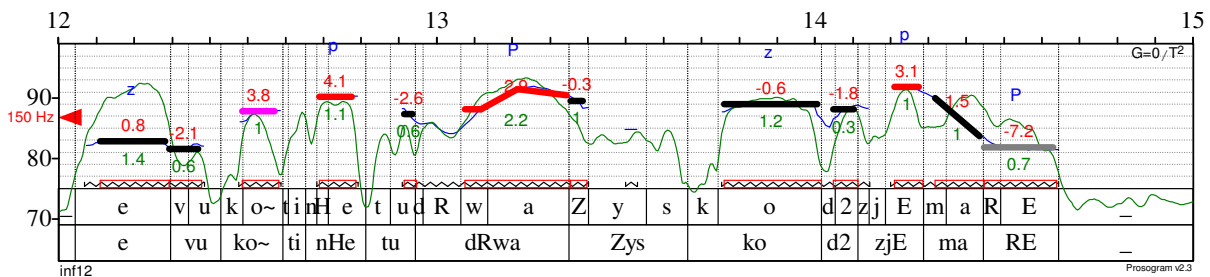


Figure (2) Prosogramme enrichi *ProsoProm* (liti-12) et vous continuez tout droit jusqu'au deuxième arrêt. Un "ton" en rouge a été identifié comme proéminent à la fois par les experts et par l'automate. Un ton gris correspond à une détection manquée de la part de l'automate, tandis qu'un ton en rose correspond à une fausse alerte (proéminence détectée par l'automate, pas par les experts). Les syllabes étiquetées "z" sont considérées comme des hésitations.

Avanzi M., Goldman J.-Ph., Lacheret A., Simon A.C., Auchlin A. 2007 (à paraître). Méthodologie et algorithmes pour la détection automatique des syllabes proéminentes dans les corpus de français parlé, *CAFLS (Cahiers de l'AFLS)*

► Outil pour la détection semi-automatique des syllabes proéminentes

ProsoProm Contacter jean-philippe.goldman@lettres.unige.ch

11. Accentuation

L'accentuation est la propriété linguistique d'une unité dite *accentogène* (syllabe, mot, constituant syntaxique) qui peut recevoir une proéminence (accent). Selon les paramètres (durée, f0, intensité) et la localisation de la proéminence (en position finale ou initiale, etc.), les accents remplissent différentes fonctions (syntaxiques, rythmiques, pragmatique, phonostylistiques...).

Tableau (3) Types d'accents en français (tableau de synthèse)

	ACCENT PRIMAIRE (contraint linguistiquement)		ACCENT SECONDAIRE		morphème intonatif
position	accent final proéminent	accent final peu proéminent	non prédictible	accent initial	accent final?
durée	très allongé (200 ms)	pas spécialement allongé	bref	bref	très allongé
pause	suivi d'une pause	pas de pause ou pause courte	pas de pause	précédé d'une pause	?
F0	ton plat ou dynamique ton fortement montant ou descendant grand intervalle (5DT) avec la syllabe précédente	ton légèrement montant	ton "plat" niveau moyen	ton "plat" niveau haut	ton amplement descendant du haut vers le grave ou en forme de cloche (montée-descente)
	NON EMPHATIQUE			EMPHATIQUE	
fonction(s)	frontière intonative majeure	frontière intonative mineure	accent rythmique (méDlterranéEN) de structuration interne	accent d'insistance	focalisation, mise à l'avant plan
Mertens (1987)	accent final		accent initial		accent final + ton HB- de focalisation
Di Cristo (1999)	accent nucléaire	accent lexical	accent rythmique	accent d'intensification	accent contrastif ou focus restreint

Une transcription des accents répartis dans la chaîne parlée implique un **modèle symbolique** qui permet de catégoriser les syllabes proéminentes selon leur forme, leur localisation et l'interprétation de leur fonction.

Di Cristo, Albert. 1999a. Vers une modélisation de l'accentuation en français (première partie), *Journal of French Language Studies* 9/II, 143-163.

Di Cristo, Albert. 1999b. Vers une modélisation de l'accentuation en français (seconde partie), *Journal of French Language Studies* 10/I, 27-45.

Mertens, P. 1993. Accentuation, intonation et morphosyntaxe, *Travaux de Linguistique* 26, 21-69

Mertens, P. 1997. De la chaîne linéaire à la séquence de tons, *T.A.L. (Traitement Automatique des Langues)* 38/1, 27-51

12. Intonation (tons et contours)

Similairement à l'accentuation, la transcription de l'intonation se réfère à un *modèle* en fonction duquel l'infinie variété des mouvements de la f0 sera réduite à un nombre restreint de tons ou de contours dits distinctifs et dotés d'une fonction particulière⁴.

Les modèles phonologiques de l'intonation du français se distinguent selon le nombre de niveaux de hauteur (2 à 4, parfois plus si on compte les niveaux relatifs), le nombre de tons ou de contours et l'importance accordée à l'alignement syllabique des contours. Quelques modèles pour une représentation **symbolique** de l'intonation du français:

- Delattre (1966) et les dix intonations de base dans une perspective d'apprentissage du français;
- Jun & Fougeron (inspiré du modèle ToBI et adapté au français; deux niveaux de hauteur H et L; accents mélodiques, accents de phrase et tons de frontière);
- Rossi (1999): intonèmes de frontière et morphèmes intonatifs;
- INTSINT (codage automatique de l'organisation tonale de surface à partir de la stylisation de la courbe) recourt à des symboles à valeur absolue *Mid, Top, Bottom* ou relative *Higher, Lower, Same, Upstepped* et *Downstepped* (Di Cristo 1998)

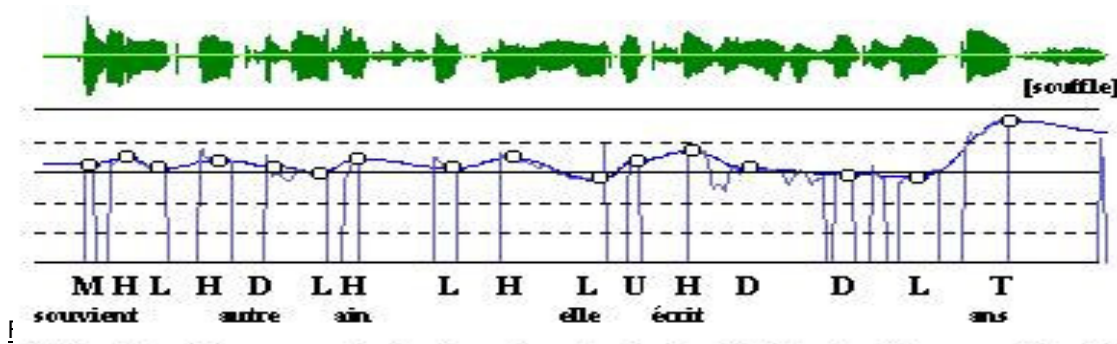


Figure (4) Transcription phonologique de surface selon le modèle INTSINT

- le modèle de l'intonation de Mertens distingue quatre niveaux de hauteurs dépendants du registre du locuteur (B-, B, H et H+) et des niveaux relatifs (/B ou \B, /H ou \H). Ces "tons" se combinent en fonction de leur localisation dans la chaîne syllabique (accent final, accent initial, syllabe atone). À chaque ton correspond une valeur sémantique de base.

entre autres de « Ainsi soit-elle » écrit il y aura bientôt vingt ans
 ... HB b bb H \h...b HB x bb HH b.....b H/H

Figure (5) Transcription intonative selon le modèle tonal de Mertens

Delattre, P. 1966. Les dix intonations de base du français, *French Review* 40, 1-14.
 Di Cristo, A. 1998. Intonation in French, in D. Hirst & A. Di Cristo (eds), *Intonation systems: a survey of twenty languages*, Cambridge: CUP, 195-219.
 Jun, S.A. & C. Fougeron. 2000. A Phonological Model of French Intonation, in A. Botinis (ed.), *Intonation. Analysis, Modelling and Technology*. Dordrecht, Boston, London: Kluwer Academic Publishers, 209-242.
 Martin Ph. 2006. Intonation du français. Parole spontanée et parole lue. *Estudios de Fonética Experimental* 15, 133-162.
 Mertens, P. 2004. Quelques allers-retours entre la prosodie et son traitement automatique. *Le français moderne* 72(1), 39-57 [version pré-finale: <http://bach.arts.kuleuven.be/pmertens/publications.html>]
 Rossi, M. 1999. *L'intonation, le système du français: description et modélisation*. Paris-Gap: Ophrys

⁴ D'où le nom parfois utilisé d'intonème, voir de morphème intonatif, en lien avec les phonèmes.

13. Rythme

Le rythme dans la parole concerne la **structuration perçue** d'une suite de **stimulations** (Carton 1973). Une figure rythmique émerge dans la perception si des proéminences sont perçues à des intervalles temporels réguliers. Le rythme est *construit* par la perception (Auer, Couper-Kuhlen & Müller 1999; Padeloup 2005).

Pour analyser le rythme perçu, on se fonde sur les **proéminences** – ou battements rythmiques – plutôt que sur la notion d'accent – qui correspond parfois à une position virtuelle dans la structure métrique vs une proéminence réalisée et perçue. L'élément proéminent est une syllabe qui se détache des syllabes environnantes par un ou plusieurs aspects (durée, hauteur mélodique, intensité) (voir [proéminence](#)).

'Stressed syllables', in our use of the term here, are those given prominence in the utterance via an increase in length or loudness and/or via a change in pitch. [...] We use the expression 'prosodic prominence' as a cover term for all kinds of syllables perceived to be stressed or accented. (Auer, Couper-Kuhlen & Müller 1999: 36-37)

13.1. Densité accentuelle

densité (accentuelle) proportion des syllabes proéminentes par rapport aux syllabes non proéminentes, par unité rythmique

Le débit correspond au nombre de syllabes réalisées par unité de temps, tandis que le tempo correspond au nombre de syllabes proéminentes réalisées par unité de temps. Le débit est articulatoire, le tempo est rythmique.

Tableau (4) rapport entre tempo et densité accentuelle (x = syllabe non proéminente; X = proéminence); la durée de chaque intervalle est représentée de manière proportionnelle à sa longueur; tous les patrons représentés dans ce tableau sont isochroniques et isométriques (voir infra)

tempo normal / densité normale	x x x X / x x x X
tempo lent / densité élevée => emphase	x X / x X
tempo rapide / densité élevée => "énergie"	x X / x X / x X / x X
tempo rapide / densité faible => "parenthèse"?	xxxX / xxxX / xxxX / xxxX

Ces configurations contextualisent certaines attitudes ou états d'activation de l'information.

13.2. Rythme: isochronie et isométrie

Un passage rythmiquement organisé se caractérise

- par son isochronie si les battements rythmiques (proéminences) sont espacés de manière régulière (intervalles de temps perçus comme identiques)⁵;
- par son isométrie si les groupes rythmiques successifs comptent le même nombre de syllabes faibles entre les syllabes fortes (proéminences)

Pour établir le tempo, on mesure la durée des intervalles rythmiques entre deux proéminences. La durée d'un intervalle rythmique dans une suite régulière varie généralement de 230 ms à 1200ms, ce qui correspond à un tempo variant de 258 à 50 proéminences / minute⁶.

Pour mesurer l'intervalle de durée entre deux proéminences, on prend comme repère l'attaque des voyelles des syllabes proéminentes (voir Couper-Kuhlen 1993; Barbosa 1994 : 73-74).

⁵ Il peut y avoir des battements "en avance" ou "en retard", voire un "battement silencieux", pour autant que l'isochronie se poursuive par après et intègre ces décalages.

⁶ C'est-à-dire de 4,3 à 0,83 battement(s) à la seconde.

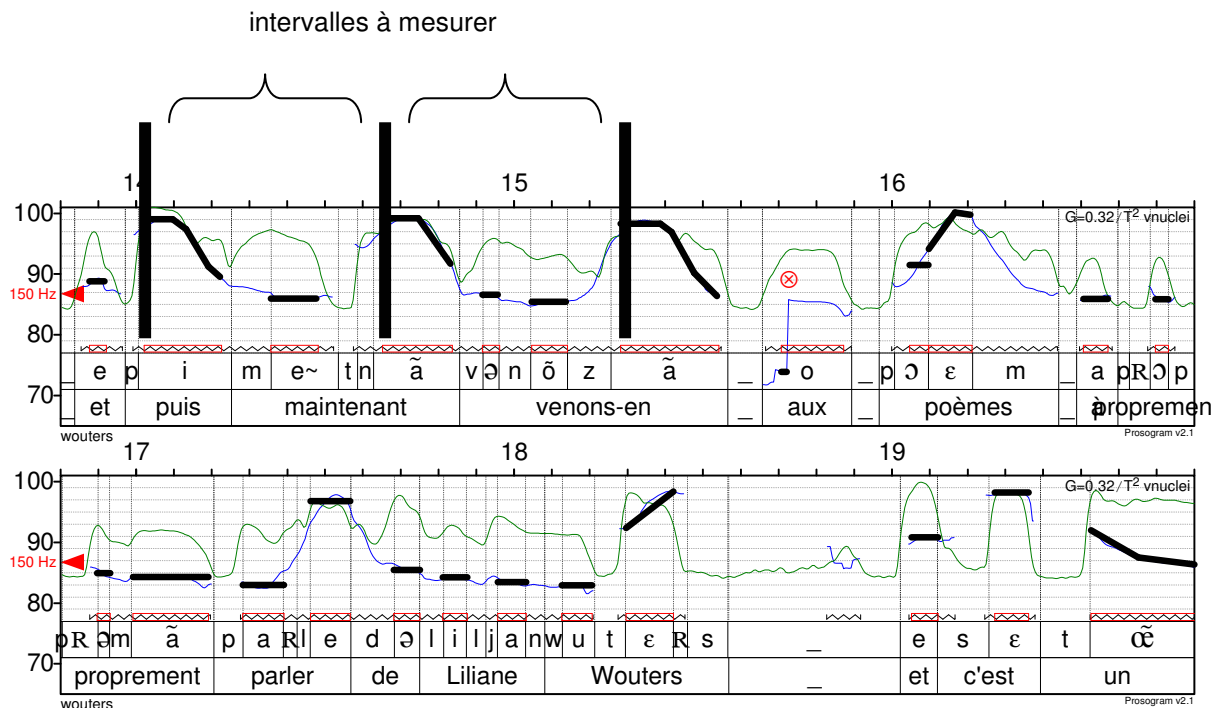


Figure (6) Prosogramme et prise de repères pour mesurer les scansions rythmique (chaque barre verticale correspond au *P-center*, soit l'attaque de la voyelle d'une syllabe préminente).

Entre deux intervalles temporels consécutifs (délimités par trois prééminences syllabiques successives):

- la variabilité perceptible (*just noticeable difference*) est de 10% pour des intervalles entre 200 et 1500 ms (Woodrow 1951 : 1224-5);
- cette variabilité peut aller jusqu'à 20% sans altérer la perception d'isochronie (et même dans certains cas excéder 20% si la régularité rythmique est doublée d'un parallélisme syntaxique ou intonatif...) (Couper-Kuhlen 1993);
- si un intervalle varie de + de 35% par rapport à la durée du précédent, aucune isochronie ne sera perçue.

13.3. Transcription

Selon les conventions de Auer et al. (1999), les passages rythmiques sont notés en caractère non proportionnel (de type *courrier*); les battements rythmiques sont alignés et analogiquement représentés par des barres obliques (/). La durée des intervalles, ou leur proportion, peut être indiquée; les syllabes prééminentes (battements rythmiques) sont transcrites en majuscules. Dans l'exemple ci-dessous, on observe à la fois une isochronie et une isométrie (une syllabe faible suivie d'une syllabe forte).

une anthologie euh réalisée par Liliane Wouters et Yves Namur

/ et PUIS	/
/ mainteNANT	/
/ venons-EN	/ (+1%)
/ aux PO-	/ (+19%)
/ èmes à proprement parLER	/ (doublé)
/ de Liliane WouTERS	/ (+24% - moyenne) et c'est un / un livre [...]

13.4. Grille d'analyse (TextGrid) pour les scansion rythmiques

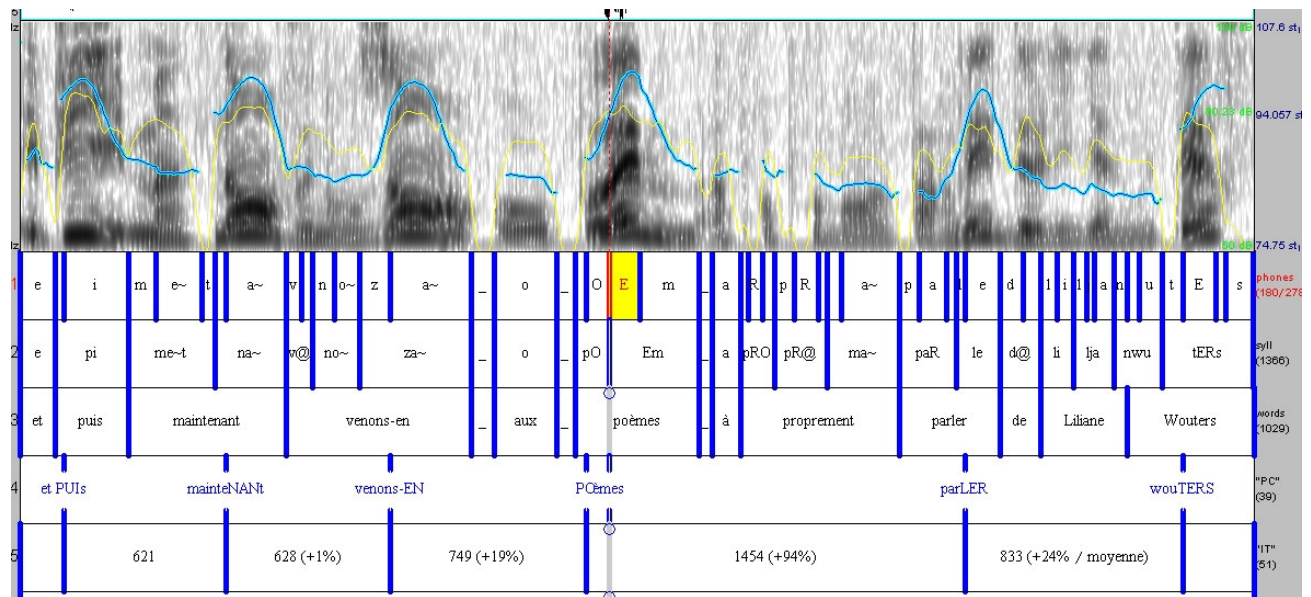


Figure (7) TextGrid pour analyser les scansion rythmiques (tires, décrites du haut vers le bas: phones, syllabes, mots graphiques, *perceptual centers* c'est-à-dire points perçus du battement rythmique et correspondant aux attaques des voyelles des syllabes proéminentes; *intervalles de durée* c'est-à-dire intervalles mesurant l'écart temporel entre deux battements rythmiques, en millisecondes)

Auer P., E. Couper-Kuhlen & F. Müller. 1999. *Language in time*. NY / Oxford: Oxford University Press.
 Barbosa P. 1994. *Caractérisation et génération automatique de la structuration rythmique du français*. Thèse de doctorat, Institut national polytechnique de Grenoble
 Couper-Kuhlen E. 1993. *English speech rhythm: form and function in everyday verbal interaction*. Amsterdam-Philadelphia: John Benjamin.
 Padeloup, V. 2005. Figures et fond dans la scène prosodique: leur résistance face aux variations du débit de parole. *Interface Discours-Prosodie* (2005 septembre 8-9 : Aix-en-Provence, France).
 Uhmman, S. 1992. Contextualizing relevance: on some forms and functions of speech rate changes in everyday conversation, in P. Auer & A. Di Luzio (ed.), *The contextualization of language*, Amsterdam - Philadelphia: John Benjamin, 297-336.

14. Mesures acoustiques automatiques: *Voice Report*⁷

Voice report on corpus xxx (à partir d'un corpus aligné phones + syllabes et stylisés par prosogramme)

Number of nuclei: 368

Speech duration : 106.965 s (including pauses)

Articulation dur: 73.306 s (excluding pauses)

Articulat. ratio: 68.5% (articulation / speech)

Speech rate: 3.4 nuc/s (over speech duration)

Articulation rate: 5.0 nuc/s (over articulation duration)

Nuclei duration : 31.870 s (only nuclei)

Nuclei ratio: 43.5% (nuclei / articulation)

Syllable duration: mean=0.199 s / stddev=0.092 s

Nuclei duration : mean=0.087 s / stddev=0.061 s

About nuclei F0

F0 mean: 74.99 ST (above 1Hz)

F0 stddev: 12.41 ST (standard deviation)

Estimated quantiles: (in semitones above 1Hz)

. 1% 0 (min)

>. 5% 71.3

. 50% 74.3 (median)

>. 95% 86.9

. 99% 91.8 (max)

Estimated spreading:

. max-min 91.8 ST (range)

>. 95%-5% 15.6 ST (narrow range)

>. median-5% 12.6 ST (left range)

>. 95%-median 3.0 ST (right range)

About dynamics

Static : 77.7% (286)

Rising : 6.8% (25) with 1.6% (6) complex

Falling: 15.5% (57) with 0.5% (2) complex

>Complex:

>Hat : 1.2% (1/82)

>Hacek: 8.5% (7/82)

By nuc (=syll):

Neat intra-mvt: 4.2 ST (mean mvt of dyn tones, i.e. excluding static tones)

Intra-mvt: 0.9 ST (mean mvt of ALL tones, i.e. including static tones)

Inter-mvt: 3.1 ST (mean mvt between tones)

Agitation: 4.0 ST (inter-mvt + intra-mvt)

About prom

prom 35.3 (130/368)

nonProm 35.3 (238/368)

⁷ Outil en cours de développement: jean-philippe.goldman@lettres.unige.ch