

Un système d'annotation automatique de la structure prosodique

Philippe Martin
LLF, UFRL, Université Paris Cité
Place Paul Ricoeur, 75013 Paris, France
philippe.martin@utoronto.ca

RESUME

On présente un système d'annotation prosodique permettant de visualiser les structures prosodiques générées par des règles de dépendance appliquées sur des événements prosodiques annotés automatiquement.

Les événements prosodiques sont définis par des cibles tonales dans la notation ToBI, ou par des contours mélodiques, montants ou descendants, atteignant la hauteur la plus basse ou la plus haute de la phrase, et au-dessus ou au-dessous du seuil de glissando (c'est-à-dire perçus comme un changement mélodique ou un ton statique), mais d'autres définitions peuvent être utilisées au gré de l'utilisateur.

À partir de ces définitions, les contours ou les cibles tonales alignés sur les voyelles des syllabes accentuées localisées sont affichés automatiquement. Des règles de dépendance définies par l'utilisateur opèrent sur ces événements prosodiques pour déterminer et afficher la structure prosodique correspondante, permettant une comparaison visuelle avec la structure morphosyntaxique et conduisant à une meilleure compréhension de la manière dont les structures prosodiques peuvent amorcer le décodage syntaxique par l'auditeur.

Abstract

An automatic prosodic annotation system

A prosodic annotation system is presented that enables the prosodic structure generated by dependency rules applied to automatically annotated prosodic events to be visualized.

Prosodic events are defined by tonal targets in ToBI notation, or by melodic contours, rising or falling, reaching the lowest or highest pitch in the sentence, and above or below the glissando threshold (i.e. perceived as a melodic change or static tone), but other definitions can be used at the user's discretion.

Based on these definitions, vowel-aligned contours or tonal targets of localized stressed syllables are displayed automatically. User-defined dependency rules operate on these prosodic events to determine and display the corresponding prosodic structure, enabling visual comparison with the morphosyntactic structure and leading to a better understanding of how prosodic structures can initiate syntactic bootstrapping by the listener.

MOTS-CLES : Structure prosodique, cible tonale, contour mélodique, annotation prosodique.

KEYWORDS: Prosodic structure, tonal target, melodic contour, dependency rules, prosodic annotation.

1 Introduction

Il est généralement admis que la structure prosodique d'une phrase résulte de la fusion des groupes accentuels AP (*accent phrases*) en syntagmes intonatifs intermédiaires ip (*intermediate intonation*

phrases), des ip en syntagmes intonatifs IP (*Intonation Phrases*) et, finalement, des IP en structure prosodique PS (*Prosodic Structure*). À ce jour, plusieurs systèmes d'annotation prosodique automatique ont été développés, basés sur le modèle autosegmental-métrique et l'annotation ToBI (Wightman and Ostendorf 1994, Syrdal et al. 2001, Rosenberg 2010). Des implémentations récentes utilisent des algorithmes d'apprentissage profond (Zhai and Hasegawa-Johnson 2023), dont l'efficacité repose sur la fiabilité de l'annotation ToBI d'un corpus d'apprentissage.

On décrit ici un système d'annotation automatique basé sur un modèle sensiblement différent, impliquant l'autonomie de la structure prosodique par rapport à la morphosyntaxe, et autorisant la transcription des événements prosodiques par contours aussi bien que par cibles tonales. Pour décrire les événements prosodiques, par hypothèse localisés essentiellement sur les voyelles des syllabes accentuées, ce système intègre la valeur de glissando pour mieux lier la perception de l'auditeur à la réalité acoustique. Ce choix contraste avec les systèmes dominants dans lesquels la transcription des événements prosodiques est plus étroitement liée à la courbe de la fréquence fondamentale et pour lesquels, pour les langues à accent lexical comme l'italien, seuls les tons de frontière et non les accents de hauteur jouent un rôle dans l'indication de la structure prosodique (Selkirk 1984).

Sur la base de l'autonomie de la structure prosodique, dans les unités desquelles viendraient s'insérer les unités syntaxiques, on est conduit à abandonner, voire à inverser, la vision selon laquelle la structure prosodique résulterait d'une mise en correspondance avec la structure morphosyntaxique de la phrase. Dès lors, on est conduit à admettre que 1) il doit exister des marqueurs prosodiques qui indiquent les fusions successives des AP en groupes intonatifs plus grands indépendamment des autres structures morphosyntaxiques ou informationnelles, et 2) il doit exister des règles qui régissent la distribution de ces marqueurs prosodiques pour indiquer sans ambiguïté une structure prosodique donnée. Un choix théorique porte alors sur 1) la sélection de caractéristiques dans le matériel prosodique qui instancient de manière adéquate les marqueurs prosodiques, et 2) le type de grammaire prosodique approprié.

2 Sélection des paramètres acoustiques

Un système d'annotation prosodique automatique doit s'appuyer sur des caractéristiques acoustiques et donc sur la fiabilité de l'algorithme d'analyse de la parole pour obtenir l'intensité syllabique, la durée des voyelles et, en particulier, la hauteur mélodique. Excepté pour les enregistrements effectués dans des salles insonorisées ou dans des conditions équivalentes, l'analyse de la parole lue et spontanée nécessite des algorithmes fiables de suivi mélodique, qui en pratique peuvent se révéler déficients pour beaucoup d'enregistrements, en particulier spontanés.

Ces systèmes peuvent être complétés par l'affichage superposé, sur la même échelle de fréquence, d'un spectrogramme à bande étroite. L'utilisateur peut alors être guidé par la première ou la deuxième harmonique affichée sur le spectrogramme pour saisir manuellement à l'écran un segment de courbe de hauteur corrigée, grâce à une fonction de dessin intégrée au logiciel venant au secours d'une mesure de F0 défaillante. On dispose ainsi d'un outil permettant l'annotation prosodique (et aussi des formants) quelle que soit la qualité de l'enregistrement, même ceux très dégradés du début du 19^{ème} siècle.

Le choix des caractéristiques acoustiques devrait aussi refléter la perception humaine mieux que la courbe de fréquence fondamentale souvent prise telle quelle comme source de données. Au lieu de décrire les caractéristiques prosodiques des voyelles par les changements de durée, d'intensité et de fréquence fondamentale évalués séparément, on adopte le paramètre de glissando, un paramètre combinant ces trois entités acoustiques en tenant en compte avec une approximation acceptable la

perception d'un événement prosodique en termes de changement mélodique, perçu comme tel ou perçu comme un ton statique, selon sa valeur par rapport à un seuil (Rossi 1971).

La valeur du glissando est obtenue en rapportant la variation de hauteur en demi-tons à leur durée :

$\text{Glissando} = (\text{DT}_2 - \text{DT}_1) / (t_2 - t_1)$ avec DT (Demi-ton) = $12 * (\log(F0_t/100.0)) / \log(2.0)$ ($F0_t$ étant la valeur de la fréquence fondamentale à l'instant t).

Le seuil de glissando en demi-tons est donné par $\text{coeff} / (t_2 - t_1)^2$, coeff étant un paramètre ajustable variant de 0,16 à 0,32 de manière à prendre en compte le degré éventuel de non-linéarité de la courbe de fréquence fondamentale de la voyelle ainsi que la sensibilité des auditeurs aux changements de hauteur.

Le choix des contours mélodiques à l'endroit des voyelles accentuées comme marqueurs prosodiques est renforcé par des recherches neuro-perceptives récentes, qui suggèrent que les accents de hauteur sont encodés comme des catégories discrètes et contrastives dans le cerveau (Llanos et al. 2021). Le positionnement des syllabes accentuées est semi-automatique pour le français, langue à accent rythmique, et ne peut être totalement déterminée à partir du texte, car dépendant entre autres du débit de parole (Martin 2018). Par défaut, ce sont les mots grammaticaux, dont la catégorie est établie à partir d'un lexique (Lexique 3), suivi d'un correcteur trigramme, qui reçoivent l'accent. Pour l'italien par contre, langue à accent lexical, la position des syllabes accentuées relève directement d'un lexique, mais peut être obtenue également à partir d'une analyse morphologique des mots (Martin 1990).

Pour l'annotation prosodique de la phrase française, les événements prosodiques situés sur les voyelles des syllabes accentuées sont classés automatiquement par le logiciel à partir de la mesure de trois paramètres : a) $F0$ montant – $F0$ descendant ; b) en dessous ou au-dessus du seuil de glissando et c) atteignant la valeur de $F0$ la plus basse ou la plus élevée parmi toutes les autres voyelles accentuées de la phrase.

Les classes de contours sont alors :

- a. **Cdec↓**, (L*L%) Contour terminal déclaratif atteignant la valeur mélodique ($F0$) la plus basse dans la phrase parmi les autres fins de voyelles accentuées
- b. **Cint↑**, (H*H%) Contour terminal interrogatif atteignant la valeur mélodique ($F0$) la plus haute dans la phrase parmi les autres fins de voyelles accentuées
- c. **Cris↗**, (LH*) Variation mélodique montante supérieure au seuil de glissando (cf. continuation majeure en français)
- d. **Cfap#↘**, (H*L#) Variation mélodique descendante supérieure au seuil de glissando suivie d'une pause d'au moins 250 ms (contour de la dictée)
- e. **Cfal↘**, (HL*) Variation mélodique descendante supérieure au seuil de glissando (cf. continuation mineure en français)
- f. **Cneu—**, (H*) Montée ou descente mélodique inférieure au seuil de glissando (contour neutralisé)

Pour l'italien, il existe un contour complexe g) **Ccom √**, (LL*H), descendant en dessous du seuil de glissando sur la voyelle accentuée, et généralement montant au-dessus du seuil de glissando sur la voyelle de la syllabe finale (et éventuellement sur la consonne voisée finale). Ce contour mélodique complexe est réalisé sur une même voyelle d'une syllabe finale et accentuée.

En plus de ces 7 catégories adaptées au français et à l'italien, le logiciel permet de définir un total de 14 catégories distinctes s'adaptant à la langue et aux conditions de l'analyse (par exemple en ajoutant un contour d'insistance, d'hésitation, etc.).

À partir de ces définitions et de l'analyse acoustique (durées vocalique, courbes de F0 et d'intensité), les contours ou les cibles tonales ToBI sont automatiquement affichés avec leurs caractéristiques graphiques sélectionnées par l'utilisateur (couleur, largeur du trait...), comme le montre la figure 1. Cette classification est effectuée en fonction des limites des voyelles accentuées sur l'échelle temporelle, définies soit manuellement par l'opérateur, soit à l'aide d'un moteur de segmentation automatique intégré au logiciel (disponible dans plus de 42 langues...).

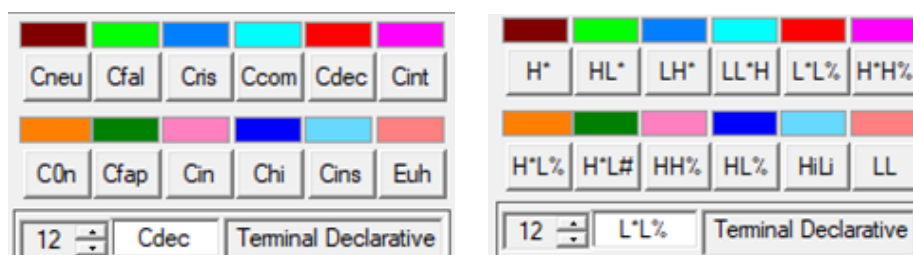


FIGURE 1. Définitions programmables des événements prosodiques en contours (à gauche) ou en notation ToBI (à droite).

3 Choix d'une grammaire

En utilisant le système d'annotation ToBI, une grammaire prosodique dans le cadre autosegmental-métrique inclut un ensemble de règles décrivant toutes les séquences bien formées d'événements prosodiques dans une phrase. En pratique, ces règles sont découvertes et énoncées à partir de l'analyse de diverses configurations syntaxiques qui sous-tendent la primauté de la morphosyntaxe sur la prosodie de la phrase (Delais, Post et Yoo 2020).

Cependant, une autre vision peut être envisagée dans laquelle la structure prosodique est considérée comme autonome par rapport à la morphosyntaxe et à toute autre structure de la phrase. On peut alors considérer que la structure prosodique, c'est-à-dire les regroupements successifs d'AP en ip, d'ip en IP et d'IP en SP, sont indiqués par des marqueurs prosodiques spécifiques perceptivement saillants tels que décrits plus haut.

Afin d'intégrer les marqueurs prosodiques dans une grammaire, le choix s'est porté sur une grammaire de dépendance, basée sur une réinterprétation du concept de continuité, présent par exemple chez (Delattre 1966). En effet, en tant que marqueur prosodique, la continuité mineure indique à l'auditeur que la phrase se poursuivra jusqu'à l'apparition d'une continuité majeure, et la continuité majeure indique que la phrase se poursuivra jusqu'à l'apparition d'un contour terminal.

En termes de relations de dépendance sur l'échelle temporelle, l'occurrence de la continuité mineure dépend de l'avènement attendu de la continuité majeure plus loin dans la phrase. De même, la continuité majeure dépend de l'occurrence attendue d'un contour terminal dans le futur de la phrase. En termes de regroupements des groupes accentuels, les séquences de groupes accentuels dont le dernier se termine par une continuité mineure sont fusionnées avec tous les groupes accentuels déjà regroupés avec une continuité majeure, et tous les groupes accentuels se terminant par une continuité majeure sont fusionnées avec tous les groupes accentuels déjà regroupés se terminant par un contour terminal.

4 Grammaire de dépendance prosodique

Cfal	2	->	Cris
Contour	Rank	Direct...	Contour
Cneu	1	->	Cdec
Cneu	1	->	Cint
Cneu	1	->	C0n
Cneu	1	->	Cin
Cneu	1	->	Cfal
Cneu	1	->	Cris
Cneu	1	->	Cfap
Cfal	2	->	Cris
Cfal	2	->	Cfap
Cfal	2	->	Cint
Cfap	3	->	Cint
Cris	3	->	Cdec
Cris	3	->	Cint
Cfap	3	->	Cdec
Cin	4	<-	Cint
C0n	4	<-	Cdec
Cint	6	->	
Cdec	6	->	

Ccom	4	->	Cint
Contour	Rank	Direct...	Contour
Cneu	1	->	Ccom
Cneu	1	->	Cdec
Cneu	1	->	Cint
Cneu	1	->	C0n
Cneu	1	->	Cin
Cneu	1	->	Cfal
Cneu	1	->	Cris
Cfal	2	<-	Cris
Cris	3	->	Ccom
Cris	3	->	Cdec
Cris	3	->	Cint
Ccom	4	->	Cdec
Ccom	4	->	Cint
C0n	5	<-	Cdec
Cin	5	<-	Cint
Cdec	6	->	
Cint	6	->	

204

la représentation graphique de la structure prosodique, il reste isolé, ce qui se traduit par une structure prosodique incomplète. Un exemple est donné figure 5.

La structure résultante est automatiquement affichée, avec une représentation arborescente utilisant des branches orthogonales orientées selon la direction des dépendances (Figure 3 et 4).

Chaque règle de dépendance est modifiable par l'utilisateur avec le rang de l'événement prosodique et la direction gauche ou droite de la dépendance. Par exemple, en français, **Cfal**↘ dépend de **Cris**↗ à droite, alors qu'en italien, **Cfal**↘ dépend de **Cris**↗ à gauche. Le code couleur correspond aux catégories d'événements prosodiques définies par l'utilisateur dans le tableau de la figure 2.

5 Exemples en français et en italien

Le français et l'italien permettent une comparaison entre une langue à accent rythmique et une langue à accent lexical, en particulier en ce qui concerne l'apparition de contours mélodiques "complexes" placés sur la voyelle accentuée du groupe accentuel, mais en dessous du seuil de glissando, et remontant au-dessus de ce seuil sur la voyelle de la syllabe finale (et éventuellement sur la consonne voisée finale). Ce contour mélodique, qui n'est pas normalement présent en français, est automatiquement identifié par le système, étiqueté et colorié selon les propriétés graphiques sélectionnées par l'utilisateur. L'analyse acoustique dispose de plusieurs algorithmes de détection de F0 (peigne spectral, AMDF, autocorrélation), complété par une estimation linéaire aux moindres carrés pour les sections des voyelles accentuées.

5.1 Exemple en italien

Un premier exemple de parole lue en italien (corpus SIWIS 2016) montre une structure prosodique reflétant partiellement la syntaxe (les voyelles accentuées sont en majuscules grasses, figure 3)

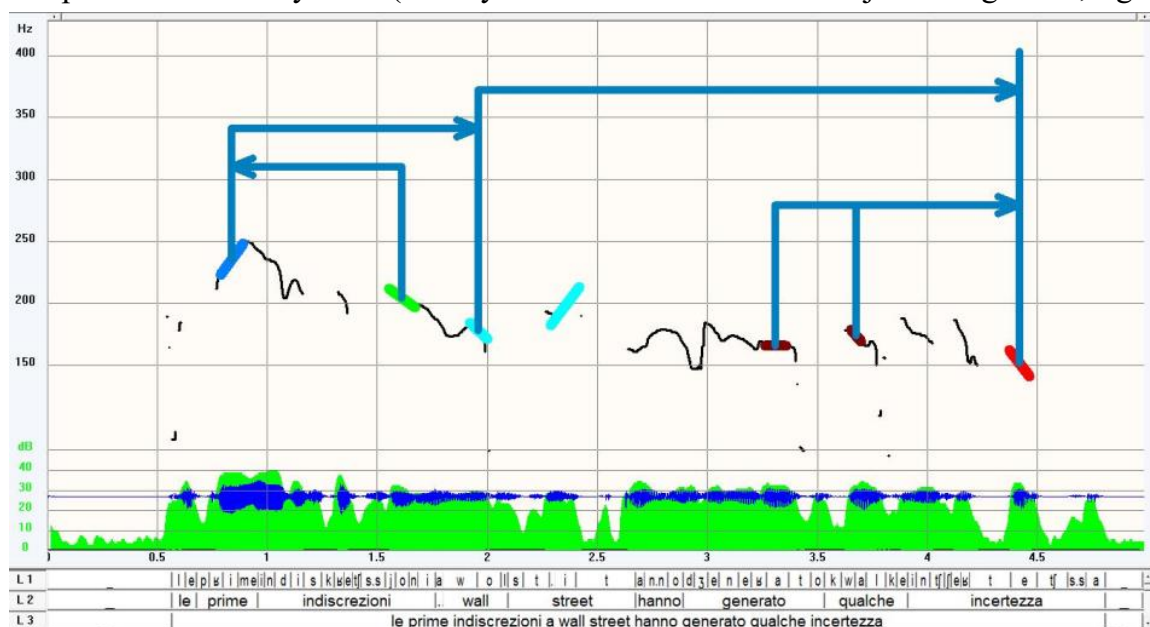


FIGURE 3. Contours colorés des voyelles de syllabes accentuées successives : **Cris**↗ au-dessus du seuil de glissando (GT), **Cfal**↘ au-dessus de GT, **Ccom**↗ complexe non final, **Cneu**— en dessous de GT, **Cneu**— en dessous de GT, et **Cdec**↓ terminal. (**Cris**↗ est représenté en bleu, **Cfal**↘ en vert, **Ccom**↗ en turquoise, **Cneu**— en brun et **Cdec**↓ en rouge). [Le *prIme*] **Cris**↗ [*indiscreziOni*] **Cfal**↘ [*a Wall StrEEt*] **Ccom**↗ [*hanno generAto*] **Cneu**— [*quAlche*] **Cneu**— [*incertEzza*] **Cdec**↓ (SIWIS it_a1_08_123).

{[(Le *pr*Ime) (indiscreziOni)] [(a Wall StrEEt)]} {(hanno generAto) (quAlche) (incertEzza)}

"Les premières rumeurs à Wall Street ont généré une certaine incertitude" regroupant les deux premiers groupes accentuels avec la séquence **Cris**↗ - **Cfal**↘ avec une dépendance "à gauche", suivie de la fusion du groupe avec le premier AP présentant un contour complexe **Ccom** √.

Les sections surlignées de la courbe mélodique de cet exemple montrent les variations mélodiques successives sur les voyelles accentuées, annotées par des variations linéaires classées automatiquement par le logiciel selon les catégories définies par l'utilisateur et décrites ci-dessus.

5.2 Exemple en français

La figure 4 affiche la séquence de contours prosodiques et la structure prosodique correspondante d'un segment de parole spontanée "et du coup tu sais j'aurai au moins j'aurai de peau mieux ça sera quoi " (corpus ORFEO 2017).

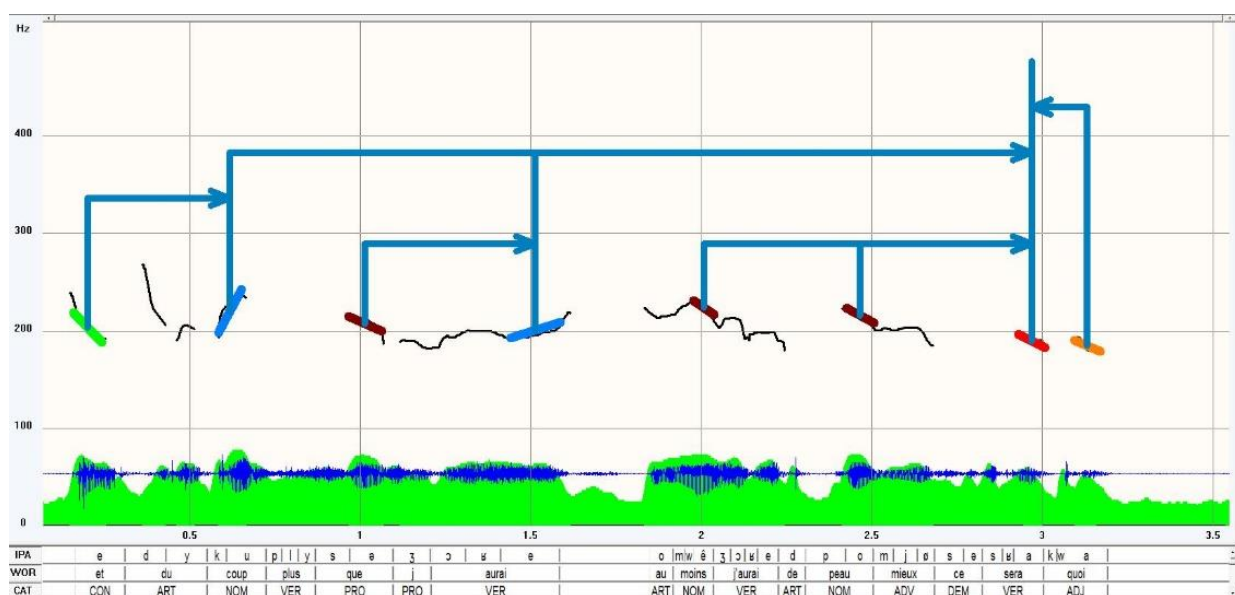


FIGURE 4. Représentation de la structure prosodique de l'exemple (*Et*) **Cfal**↘ (*du cOUp*) **Cris**↗ [*plus quE*] **Cneu**— (*j'aurAI*) **Cris**↗ (*au mouINS*) **Cneu**— (*j'aurai de pAEU*) **Cneu**— (*mieux ça sera*) **Cdec**↓ (*quOI*) **Cneu**← (ORFEO 07madmc110912).

Le syntagme accentuel [*tu sAIs*] est aligné sur une parenthèse prosodique terminée par un contour conclusif terminal **Cdec**↓ et constitue une structure prosodique indépendante à un seul groupe accentuel. Le dernier groupe accentuel (*quOI*) porte un contour neutralisé **Cneu**— indiquant une dépendance « à gauche » vers le contour terminal (cf. rhème dans une configuration rhème-thème).

Le calculateur de structure prosodique fournit une représentation graphique prenant en compte a) les événements prosodiques et b) les règles de dépendance opérant sur les événements prosodiques, événements et règles programmables et définies par l'utilisateur. Le système est interactif, et constitue un outil efficace permettant d'expérimenter l'adéquation de la description des événements prosodiques et la validité de la grammaire prosodique sur de nombreux exemples facilement analysables, quelle que soit la qualité des enregistrements. Il est également possible d'écouter l'effet d'un changement de contour effectué par une commande graphique de l'utilisateur grâce au morphing prosodique, qui utilise un algorithme Psola intégré.

On peut par exemple vérifier sur la figure 5 l'absence de grammaticalité en français de **Cfal**↘ chute du contour mélodique au-dessus du seuil de glissando suivi de **Cdec**↓, ou l'absence de grammaticalité du même contour **Cfal**↘ en position initiale dans une structure prosodique en italien. La figure 5 affiche une séquence de contours invalide pour le groupe accentuel [*pour cette minUte*]

annotée avec un **Cfal** non grammatical au lieu du **Cneu** correct. Le groupe accentuel n'est donc pas intégré à la structure prosodique de la phrase.

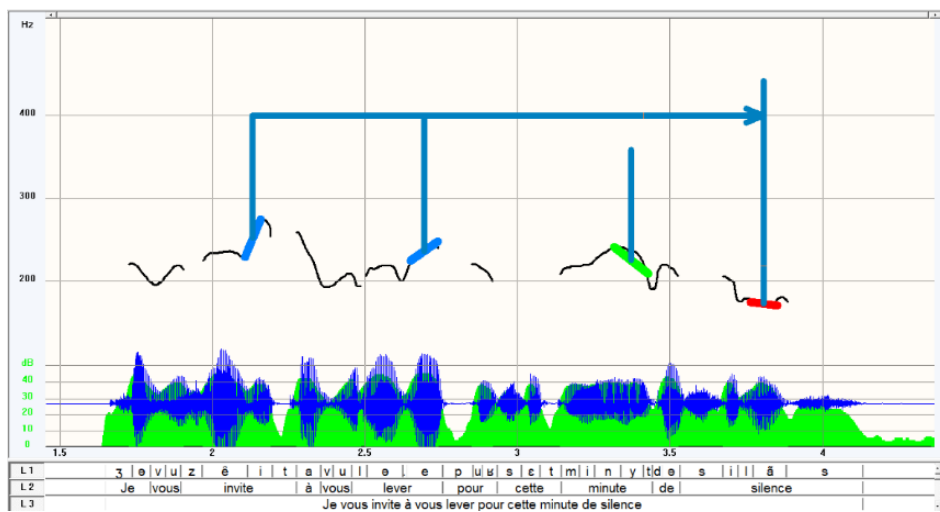


FIGURE 5. **[Je vous invIte] Cris [à vous levER] Cris [pour cette minUte] Cfal [de silENce] Cdec* (SIWIS fr_b_29_001) avec *[pour cette minUte]* annoté **Cfal** au lieu de **Cneu**, menant à la non intégration dans la structure prosodique.

Les caractéristiques de chaque contour annoté, la gamme de fréquence fondamentale, la durée et l'intensité sont affichées dans un tableau et peuvent être sauvegardées au format Excel ou équivalent d'un simple clic de souris (figure 6).

Name	Width	T 1 [s]	T 2 [s]	F0 1 [Hz]	F0 2 [Hz]	Duration [s]	Range [Hz]	F0 1 [ST]	F0 2 [ST]	Int 1 [dB]	Int 2 [dB]	Diff Int [dB]	Glissando [st/s]	Glissando ratio	Text
Cris	12	0.781	0.904	219	248	0.123	29	13	15	36	36	1	17//10	1.700	
Cfal	12	1.554	1.685	216	198	0.131	-18	13	11	27	27	2	13//9	1.444	
Ccom	12	1.929	1.983	183	175	0.054	-8	10	9	30	30	-2	14//54	0.259	
Ccom	12	1.983	2.308	175	186	0.324	11	9	10	28	28	1	4//1	4.000	
Ccom	12	2.308	2.385	186	209	0.077	23	10	12	29	29	-3	28//26	1.076	
Cneu	12	3.247	3.363	166	165	0.116	-1	8	8	31	31	1	1//11	0.090	
Cneu	12	3.657	3.715	175	165	0.058	-10	9	8	32	32	2	19//47	0.404	
Cdec	12	4.400	4.450	158	144	0.050	-14	7	6	34	34	-5	36//63	0.571	

Name	Width	T 1 [s]	T 2 [s]	F0 1 [Hz]	F0 2 [Hz]	Duration [s]	Range [Hz]	F0 1 [ST]	F0 2 [ST]	Int 1 [dB]	Int 2 [dB]	Diff Int [dB]	Glissando [st/s]	Glissando ratio	Text
Cfal	12	0.157	0.254	206	188	0.096	-18	12	10	27	27	6	21//17	1.235	e
Cris	12	0.593	0.656	219	253	0.063	34	13	16	34	34	6	43//39	1.102	u
Euh	12	0.980	1.059	214	208	0.078	-6	13	12	35	35	-2	7//25	0.280	ø
Cris	12	1.485	1.612	192	214	0.127	22	11	13	32	32	-15	29//9	3.222	ε
Cneu	12	1.970	2.027	232	221	0.057	-11	14	13	37	37	0	13//48	0.270	é
Cneu	12	2.436	2.499	221	208	0.063	-13	13	12	37	37	-5	20//39	0.512	o
Cdec	12	2.923	2.983	201	188	0.060	-13	12	10	27	27	-1	19//43	0.441	a
C0n	12	3.104	3.141	193	190	0.036	-3	11	11	22	22	4	9//121	0.074	a

FIGURE 6. Liste des contours annotés au format Excel pour les exemples des figures 3 (en haut) et 4 (en bas), avec leurs valeurs de F0, de durée, d'intensité et de glissando de F0 pour les voyelles accentuées.

6 Conclusion

Un outil d'annotation prosodique efficace est essentiel pour mieux comprendre le rôle de la structure prosodique en relation avec la morphosyntaxe et l'accès au sens par l'auditeur. Il devient également crucial lorsqu'un grand nombre d'exemples annotés est nécessaire pour entraîner des systèmes d'apprentissage profond. L'outil est intégré dans le logiciel d'analyse de la parole WinPitch (2024) et peut être téléchargé gratuitement sur www.winpitch.com.

Références

- DELAIS-ROUSSARIE E, POST B., YOU H-Y. (2020) Unités prosodiques et grammaire intonative du français : vers une nouvelle approche, *Actes de la 6e conférence conjointe Journées d'Études sur la Parole*, Nancy, France, 08-19 juin 2020.
- DELATTRE P. (1966) Les dix intonations de base du français, *French Review* (40) 1-14.
- LLANOS F., SNEED GERMAN J., NIKE GNANATEJA G., CHANDRASEKARAN B. (2021) The neural processing of pitch accents in continuous speech, *Neuropsychologia*, 2021, 158.
- LEXIQUE 3 <http://www.lexique.org/shiny/lexique>
- MARTIN PH. (1990) Positionnement automatique de l'accent lexical de l'Italien, *Actes des XVIIIèmes Journées d'Études sur la parole*, Montréal 1990, 149-152.
- MARTIN PH. (2018) *Intonation, structure prosodique et ondes cérébrales*, London : ISTE, 322 p.
- ORFEO (2017) Outils et Recherches sur le Français Écrit et Oral. <http://www.projet-orfeo.fr/>
- ROSENBERG R. (2010) AuToBI - A tool for automatic ToBI annotation, *Proc. Interspeech September 2010*, doi: 10.21437/.2010-71.
- SELKIRK E. O. (1984) *Phonology and syntax: the relation between sound and structure*. Cambridge, Mass., MIT Press.
- ROSSI M. (1971) Le seuil de glissando ou seuil de perception des variations tonales pour la parole, *Phonetica* (23) 1-33.
- SYRDAL A, HIRSCHBERG J, MCGORY J. and BECKMAN M. (2001) Automatic ToBI prediction and alignment to speed manual labeling of prosody. *Speech Communication*, 33(1-2), 135-151.
- SIWIS Corpus (2016) Yamagishi, J. et al. The SIWIS French Speech Synthesis Database, <https://doi.org/10.7488/ds/1705>.
- WIGHTMAN C. and OSTENDORF M. (1994) Automatic labeling of prosodic patterns. *IEEE Transactions on Speech and Audio Processing*, 2(4), pp. 469-481.
- WINPITCH (2024) www.winpitch.com
- ZHAI W., HASEGAWA-JOHNSON M. (2023) Wav2ToBI: a new approach to automatic ToBI transcription. *Proc. INTERSPEECH 2023*, 2748-2752, doi: 10.21437/Interspeech.2023-477.