



HAL
open science

Annotation des proéminences pour la segmentation de corpus oraux : l'expérience du projet SegCor

François Delafontaine, Biagio Ursi, Luisa Acosta Córdoba

► To cite this version:

François Delafontaine, Biagio Ursi, Luisa Acosta Córdoba. Annotation des proéminences pour la segmentation de corpus oraux : l'expérience du projet SegCor. SHS Web of Conferences, 2018, 6e Congrès Mondial de Linguistique Française, 46. hal-01959747

HAL Id: hal-01959747

<https://hal.science/hal-01959747>

Submitted on 18 Dec 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Annotation des proéminences pour la segmentation de corpus oraux : l'expérience du projet SegCor

François Delafontaine^{1*}, Biagio Ursi² et Luisa Acosta³

¹UMR 7270 LLL, Université d'Orléans

²UMR 5191 ICAR, École Normale Supérieure de Lyon

³UMR 5191 ICAR, Université Lumière Lyon 2

Résumé. Dans le but d'annoter des unités syntaxiques, macrosyntaxiques et interactionnelles, et pour confirmer le recours aux indices prosodiques faits par les annotateurs, nous avons voulu disposer d'une annotation prosodique sous la forme de proéminences et reprenant le protocole établi par le projet Rhapsodie. Après avoir préparé nos données, nous avons procédé à l'annotation manuelle par trois annotateurs, ainsi qu'à l'annotation automatique à l'aide de l'outil ANALOR. Cette expérience nous a amené à réviser le protocole concernant la hauteur mélodique et l'attaque syllabique, tandis que les résultats obtenus soulignent le potentiel de ces révisions. Nous avons constaté la nécessité des sessions d'entraînement entre les annotateurs non experts avant l'annotation, ainsi que la nécessité d'entraînement de l'outil d'annotation automatique, qui a fait preuve de robustesse sur un enregistrement audio jugé de mauvaise qualité.

Abstract. Annotation of prominences for the segmentation of oral corpora: the Segcor project experiment. In pursuing the annotation of syntactic, macrosyntactic and interactional units, and in order to take into account the mobilisation of prosodic clues by annotators, we have conducted a prosodic annotation experiment, dealing with prominences and following the guidelines established by the Rhapsodie project. After preparing our data, we have proceeded with manual annotation by three annotators and automatic annotation using the ANALOR software. The experiment led us to revise the guidelines concerning the pitch and the syllabic onset. The results underline the potential of those revisions. Before the annotation task, training sessions between non-expert annotators seem to be required. As for the automatic annotation tool, its training is necessary and its reliability seems to be consistent on a poor quality audio recording.

1 La segmentation de corpus oral

L'exploitation des corpus oraux, malgré les traitements et transcriptions dont ils font l'objet, demeure limitée par l'absence de segmentations en unités linguistiques aux niveaux morphosyntaxique, macrosyntaxique ou encore interactionnel. Le projet SegCor, cadre de cette contribution, se donne pour objectif la segmentation de grands corpus oraux en de telles unités. Cet objectif passe par la prise en compte d'indices fiables, y compris les indices prosodiques, c'est-à-dire les spécificités acoustiques et perceptives de la parole et les paramètres qui y sont associés. Bien que la segmentation en unités prosodiques n'ait pas fait partie du projet, la nécessité de fixer les indices prosodiques a donné lieu à un codage exploratoire. Ce codage est destiné à être mobilisé comme outil dans la délimitation de divers types d'unités relevant de plusieurs dimensions analytiques. Plus précisément, dans nos intentions, le rôle de l'annotation prosodique a été questionné quant à la possibilité d'évaluer les choix de segmentation effectués aux autres niveaux.

Notre recours à la prosodie en ces termes s'inscrit dans l'usage qu'en font plusieurs linguistes s'intéressant à l'oral. Ainsi, en raison de l'absence de marques segmentales (dans les enregistrements) ou graphiques (dans leur transcription), l'appel aux critères dits suprasegmentaux est un moyen récurrent pour la segmentation en unités et le rattachement (ou détachement) de certains éléments. Déjà en 1961, Morier faisait appel à la prosodie pour définir la parataxe dans les termes suivants : « juxtaposition de deux propositions entre lesquelles le lien de dépendance n'est qu'implicite, la courbe mélodique commune dispensant de l'usage d'un outil de coordination ou de subordination » (Morier 1961, cité par Béguelin 2010 : 3). L'étude de la langue parlée à l'aide de grands corpus a entraîné la critique de la phrase en tant qu'élément fondamental du langage et de la pensée (Bruneau & Heulluy 1935, cités par Groupe de Fribourg 2012 : 4) et, parallèlement, la mise en valeur de la prosodie en tant que dimension analytique primordiale. Du côté de la macrosyntaxe, l'apport de la prosodie est explicite dans la définition de l'unité de base de l'approche aixoise, le noyau :

* francois_delafontaine@outlook.com

« [c]aractérisé par une intonation autonome et une mélodie conclusive, il donne l'effet d'énoncé complet, doté d'autonomie » (Blanche-Benveniste 2010 : 159). Ces termes sont également présents dans la macrosyntaxe fribourgeoise (Groupe de Fribourg, 2012). En linguistique interactionnelle, d'autres éléments prosodiques sont ajoutés au caractère reconnaissable et solidaire de la courbe mélodique pour le repérage d'une unité de construction du tour : la variation de la fréquence fondamentale (F0), le rattachement prosodique et l'accélération du débit de parole en début d'unité ; l'allongement, la voix craquée sur les dernières syllabes, qui peuvent être suivies de pauses (Selting 1996 : 368). Malgré cette mise en valeur, l'annotation systématique d'indices prosodiques est une difficulté permanente dans l'étude de la langue parlée, notamment en raison de la complexité des processus de production et perception des signaux acoustiques. En bref, la linguistique de l'oral ne peut pas faire l'économie des indices prosodiques, à tous les niveaux d'analyse considérés, mais se heurte à la difficulté d'y associer des critères fiables.

Ces considérations ne sont pas nouvelles : le projet Rhapsodie (Lacheret-Dujour *et al.* 2014) y a répondu par un protocole d'annotation des proéminences prosodiques (voir § 2) et l'outil d'annotation automatique ANALOR (Avanzi *et al.* 2008, Avanzi 2012). Étant une dimension analytique complexe et émergente, la prosodie ne pourrait pas être réduite à un paramétrage d'indices acoustiques généraux (fréquence, intensité, timbre, durée etc.) – tel que le logiciel ANALOR peut le faire. Elle devrait aussi être appréciée de manière locale et située par des locuteurs/récepteurs qui sont capables d'indexer les valeurs linguistiques de ces paramètres. Ainsi, le protocole cherche à traiter des réalisations dans le contexte naturel de production de la parole en interaction. L'intérêt de l'outil ANALOR est alors de tester l'applicabilité d'une systématisation d'annotation par seuils de paramétrages, en tant qu'ancrage objectivant d'annotation s'appuyant sur des facteurs acoustiques. Les deux méthodes d'annotation, pertinents pour des raisons différentes, font donc l'objet de ce travail.

Nous avons repris ici protocole et outil Rhapsodie pour les appliquer à nos données. À l'issue de cette expérience, nous nous sommes posé trois questions centrales, auxquelles nous essaierons de répondre dans cette contribution :

- Munis du protocole Rhapsodie, quels sont les critères effectivement utilisés par les annotateurs ?
- Quelle est la fiabilité de l'annotation obtenue ?
- L'annotation des proéminences peut-elle servir d'indice prosodique pour la segmentation d'unités de haut niveau ?

Pour répondre à la première question, nous comparerons les conventions fournies pour l'annotation manuelle avec les critères utilisés par l'outil d'annotation automatique, puis les critères utilisés par nos annotateurs. Quant à la deuxième question, nous testerons l'accord inter-annotateurs, y compris pour l'annotation automatique. Le dernier point nous servira essentiellement de conclusion et de synthèse pour rassembler les éléments de réponse à nos questions initiales et envisager des perspectives de recherche.

Concernant le corpus annoté, il faut préciser que, contrairement à d'autres projets francophones portant principalement sur la prosodie ou la phonétique, notre corpus n'a pas été conçu exclusivement pour l'analyse acoustique. En effet, les enregistrements ont été effectués dans leur contexte ordinaire en respectant les conditions écologiques de la production de la langue parlée (Baude *et al.* 2006 : 50). Le corpus annoté est constitué par la partie française du corpus pilote du projet SegCor¹, composé d'enregistrements tirés des bases de données CLAPI² et ESLO³. Il contient 100 minutes d'enregistrements documentant des situations de monologue et dialogue, ainsi que plusieurs types d'interactions : appels téléphoniques, conversations ordinaires en face à face, échange médiatique, conférence ou encore lecture à un enfant. Les conditions d'enregistrement sur le terrain offrent une qualité sonore variable, avec bruits de fond, qualité et distance du microphone (dont téléphone) différentes et chevauchements de parole : un enregistrement à la qualité sonore trop mauvaise, enregistré dans un petit commerce, n'a pas été annoté.

Les annotations ont été réalisées avec l'outil Praat (Boersma & Weenink 2017), soit 90 minutes annotées automatiquement et manuellement par deux annotateurs (24'038 syllabes) ; deux transcriptions, soit 20 minutes d'enregistrement, ont un troisième annotateur (5'999 syllabes). Le tableau 1 détaille pour chaque enregistrement la base de données d'origine, sa durée en secondes, son nombre de syllabes, le nombre de locuteurs annotés et sa qualité globale. On notera « Moyenne » une transcription avec chevauchements, rires et distance au microphone variable ; on notera « Mauvaise » une transcription avec bruit de fond ambiant ; la qualité reste indicative.

Tableau 1. Détails du corpus annoté.

Origine	Nom	Durée	Syllabes	Locuteurs	Qualité
CLAPI	PhoneCall	577	2567	2	Moyenne
CLAPI	PreparingMeal	658	1227	2	Mauvaise
CLAPI	SocialMeeting	605	2672	7	Moyenne
CLAPI	TableTalk	618	2722	3	Moyenne

ESLO	Conf	619	3277	1	Bonne
ESLO	Ent	572	2592	2	Bonne
ESLO	LivreEnf	569	2442	2	Bonne
ESLO	Media	602	3177	4	Bonne
ESLO	TableRonde	682	3362	7	Bonne

2 Critères d'annotation des prééminences

Nous avons choisi, pour notre annotation prosodique, de passer par le concept de proéminence. Nous utiliserons ici le terme de « proéminence » pour désigner toute syllabe rendue saillante dans son contexte (Lacheret-Dujour & Beaugendre 1999). Quant aux critères :

D'un point de vue phonétique, les principaux traits activés sont la durée, la fréquence fondamentale (F0), l'intensité et la qualité vocale. Ces paramètres trouvent leurs correspondants perceptifs dans l'allongement syllabique, les modulations mélodiques (changement de niveau de hauteur, dynamique des contours, pauses silencieuses), les pics de sonie et les variations de timbre (réduction vocalique, emphase spectrale, *creaky voice*, etc.). (Avanzi 2012 : 63)

Ces critères sont nombreux, se combinent et ne sont pas les seuls critères pertinents pour la perception du phénomène (Martin 2006, Simon & Christodoulides 2016). Nous nous inscrivons donc dans une approche perceptive⁴ déjà largement explorée dans le cadre du projet PFC (Lacheret-Dujour & Victorri 2002, Poiré 2006, Morel *et al.* 2006) et du corpus C-PROM (Avanzi *et al.* 2007, 2010), aboutissant avec Rhapsodie à un protocole d'annotation manuelle des proéminences et à un outil d'annotation automatique.

2.1 Critères d'annotation manuelle

Le protocole Rhapsodie⁵ souligne la multiplicité des critères et fournit pour seule consigne d'annoter le contraste. Parmi les critères listés se trouvent la fréquence, la pause, la durée, la qualité vocale et leurs combinaisons.

Les consignes d'annotation elles-mêmes sont conçues pour éviter le sur/sous-codage : écouter au préalable quelques secondes du fichier, puis annoter par empan de 3 à 5 secondes en n'annotant qu'un locuteur à la fois et sur une unité linguistique sémantiquement et syntaxiquement « achevée ». Le protocole suggère d'éviter de coder pendant plus de 30 minutes. Les proéminences elles-mêmes sont annotées en deux étapes : d'abord les proéminences « fortes », se détachant de l'environnement aux premières écoutes ; puis, en utilisant ces proéminences comme frontières pour des intervalles plus petits, les proéminences « faibles ». Chacune des étapes est limitée à trois écoutes maximum. Le degré de proéminence, forte ou faible, doit être attribué relativement à ce qui suit et non pas à ce qui précède. Quelques remarques découlent de ces conventions.

Tout d'abord, elles ont été conçues pour des annotateurs non experts⁶ (Avanzi *et al.* 2007). Le protocole prévient de ne pas tenir compte seulement de la hauteur, mais un annotateur non expert ne saura alors plus à quels critères se fier. Ce problème était déjà présent dans l'expérience de Poiré (2006) qui concluait que les codeurs, pourtant experts, ne partageaient pas la même définition de la proéminence. Il peut expliquer le faible score obtenu chez les annotateurs non experts de Rhapsodie (entre 0.23 et 0.52, score Kappa-Fleiss (1971)). Dans notre cas, nous avons réduit les critères à la hauteur et à la durée, mais ces critères demeurent flous, spécialement pour les non-experts.

Une autre remarque importante est que les consignes Rhapsodie exigent que la fenêtre de 3 secondes corresponde à une « macro-unité » ou « unité de sens », ainsi qu'à la production d'un seul locuteur (Avanzi *et al.* s.d.). Cette consigne présuppose que de telles unités sont données à priori, ce qui contredit, à notre sens, l'objectif d'une annotation prosodique en vue d'évaluer ensuite les choix d'annotation de ces mêmes unités. Nous avons réduit cette consigne à l'obligation de ne pas couper un empan à l'intérieur d'une unité lexicale.

En dernière remarque, le protocole fait de la proéminence forte un phénomène local, tel que toute proéminence perçue à l'échelle locale en première écoute sera considérée comme forte. La consigne prévoit, en cas de doute dans cette étape, la possibilité d'annoter une proéminence faible. Ce principe d'annotation en deux étapes est au cœur du protocole Rhapsodie et influence grandement le résultat obtenu.

2.2 Critères d'annotation automatique

L'annotation automatique des proéminences avec le logiciel ANALOR (Avanzi *et al.* 2008, Avanzi 2012), entraîné sur des données annotées manuellement suivant un protocole similaire, repose sur les critères de la fréquence F0, de la durée et de la pause.

Le programme utilise pour empan un contexte de 3 syllabes avant et après la syllabe mesurée (Avanzi 2012 : 76), soit 7 syllabes. La moyenne des durées et des fréquences fondamentales F0 pour chaque syllabe permet de calculer la proéminence en termes relatifs, par seuils. Le programme exploite une segmentation phonémique, par phonèmes, pour calculer les points de F0 et le poids de la durée (suivant le nombre de phonèmes dans la syllabe). Enfin, et toujours en usant des phonèmes, le programme intègre également les glissandos⁷, de sorte qu'une syllabe est proéminente si :

- son pic de F0 est supérieur de 1.5 demi-ton à la moyenne du contexte
- sa durée est supérieure de 1.5 fois à la moyenne du contexte
- son noyau syllabique a une montée intrasyllabique de 3 demi-tons

Un demi-ton valant 1.06 fois la moyenne environ : par exemple, pour une fréquence de 300 Hz, la syllabe est proéminente passé 318 Hz. Une troisième mesure est la présence d'une pause dès 0.25s. Le dépassement d'un seuil pour n'importe quel critère suffit à activer la détection d'une proéminence. L'algorithme précis est présenté par Avanzi, Lacheret-Dujour, Obin et Victorri (2011 : 61) et révisé par Avanzi (2012).

Là aussi, nous devons faire plusieurs remarques.

La plus importante est que les seuils donnés ci-dessus, pour la fréquence, la durée et le glissando, sont les seuils par défaut d'ANALOR, indépendamment de l'entraînement sur les données. En ce sens, ANALOR devrait d'abord être entraîné (à l'aide des outils intégrés au programme) sur de nouvelles données, et ensuite exploité pour l'annotation avec les nouveaux seuils obtenus. Nous n'avons pas entraîné le programme sur nos données, et les seuils utilisés sont donc ceux par défaut, y compris pour le poids relatif à donner à chaque seuil pour le calcul des proéminences. À noter que ces seuils peuvent être modifiés manuellement dans le programme.

Il faut préciser par ailleurs que l'annotation automatique requiert des transcriptions dans un certain format, et ce format a eu des conséquences sur la présentation des données. Notamment : (a) les disfluences doivent avoir été annotées à l'avance, (b) la transcription doit être rassemblée sur une seule « tire » (ligne de transcription) contenant la transcription orthographique et (c) l'identifiant des locuteurs eux-mêmes doit être indiqué sur une autre tire. Ce rassemblement a supprimé les chevauchements, n'y laissant qu'un seul locuteur annoté : celui qui parlait avant le début du chevauchement, sauf si un autre locuteur est plus audible.

Enfin, le programme prévoit le cas où il n'y a pas de tire phonémique. Sans phonème, ANALOR ne peut pas trouver de noyau vocalique, ce qui impacte le calcul du point de F0 et du glissando. Le programme compense alors, mais, inévitablement, il perd en précision.

2.3 Comparaison des critères

L'annotation manuelle (perceptive, par l'oreille humaine) et l'annotation automatique (acoustique, sur instrument) ne reposent donc pas sur les mêmes critères. Les deux types d'annotation reposent sur les mêmes paramètres, ont des critères multiples, et un seul critère suffit pour créer une saillance et donc réaliser une proéminence. Au niveau des critères, toutes deux partagent la hauteur (variation de F0) et la durée.

Néanmoins, l'annotation automatique utilise des seuils fixes (en valeurs relatives), ajoute le critère du glissando, éventuellement de la pause, et ne considère aucun autre critère. L'annotation manuelle ne donne pas de seuils et ouvre la liste des critères potentiels. Dans l'approche où nous nous inscrivons, l'annotation manuelle prime et sert de référence pour évaluer l'annotation automatique (Lacheret 2016 : 68). Dans notre cas, nous comparerons notre annotation manuelle aux seuils préalablement établis ci-dessus.

Cette même approche s'intéresse à la perception des proéminences et, de ce fait, ne fixe pas de référence ou d'annotation « juste » : les critères sont révisés selon ce que les gens ont annoté, et non l'inverse. La comparaison entre annotateurs n'évalue pas les annotateurs mais la fiabilité du protocole d'annotation.

3 Observations sur l'annotation

Avant de pouvoir observer des résultats de notre expérience, nous devons traiter du travail réalisé en amont et qui a pu impacter les résultats. L'annotation de notre corpus a impliqué une phase de pré-traitement avec, notamment, la segmentation syllabique et l'annotation des disfluences, puis une phase d'annotation manuelle et une seconde phase d'annotation automatique.

Praat présente l'annotation sous forme de « tires » ou lignes de transcription. Dans les figures contenant un extrait, nous donnerons la cote de cet extrait avec le nom de la base de données d'origine (CLAPI ou ESLO), le nom de l'enregistrement (voir tableau 1) et entre parenthèses le début et la fin de l'extrait en secondes. Chaque ligne sous le signal sonore est une tire où les lignes bleues verticales sont les frontières temporelles des segments.

3.1 Pré-traitement

Les frontières des transcriptions d'origine n'étaient pas toujours alignées correctement sur le son. D'autre part, certaines frontières manquaient pour le repérage des pauses (dès 0.1s). Une première étape a donc consisté à créer une nouvelle tire alignant le texte de la transcription sur les pauses du signal. Lors de cette étape, pour permettre le traitement automatique du texte, nous avons éliminé des transcriptions tous les symboles non orthographiques, comme par exemple les antiquotes pour signaler les élisions non standards, la notation de certains phénomènes vocaux comme les soupirs, les aspirations, les baisses du volume de la voix, etc.⁸

L'étape suivante était la segmentation syllabique. Nous avons choisi de ne pas utiliser EasyAlign (Goldman 2011) ou JTrans (Cerisara, Mella & Fohr 2009) en raison du temps nécessaire pour corriger les tires syllabiques et phonémiques créées. Les deux programmes créent des pauses entre les « mots », syllabes et phonèmes qu'il faut ensuite éliminer manuellement ou avec un script, en distinguant alors ces pauses de celles pré-annotées. Dans le cas d'EasyAlign, le programme peut échouer à segmenter certains passages et les laissera donc vides. Enfin, les frontières produites doivent être réalignées manuellement sur le son. Nous avons calculé que la segmentation syllabique pouvait nous prendre jusqu'à 8 heures pour 10 minutes de transcription.

À la place, nous avons choisi de « tokeniser » au préalable notre transcription avec l'outil d'annotation morpho-syntaxique DisMo (Christodoulides *et al.* 2015). Cet outil ne tient pas compte du son, mais a été entraîné sur des transcriptions de l'oral. D'un point de vue qualitatif, DisMo a fourni une précision dans la segmentation qui est comparable à EasyAlign ou Jtrans. En prenant en compte le temps nécessaire pour la correction, ce logiciel s'est révélé performant. Nous avons réaligné manuellement les frontières de tokens, puis utilisé EasyAlign pour « phonétiser » le texte (selon la convention SAMPA) avant de créer manuellement, depuis les tokens, la tire syllabique, en rajoutant les frontières des tokens multisyllabiques, en corrigeant les frontières entre deux tokens et en corrigeant les phonèmes de syllabes. Cette méthode a été bénéfique uniquement grâce à l'absence de correction automatique des micro-pauses de JTrans, et parce que nous avons décidé de ne pas annoter de tire phonémique : nos annotations ne segmentent donc pas les phonèmes. Nous avons calculé que la segmentation syllabique nous aurait pris environ 4 heures pour 10 minutes de transcription.

Une fois la tire syllabique corrigée, nous avons encore annoté préalablement et sur une tire à part les disfluences, soit : (a) les allongements excessifs, (b) les répétitions, (c) les abandons-reformulations et (d) les « euh » d'hésitation. Les disfluences sont signalées par un « H » dans nos annotations.

Une fois ce pré-traitement terminé, chaque annotateur a annoté les proéminences séparément, en utilisant « P » pour les proéminences fortes et « p » pour les proéminences faibles. Chaque annotateur (FD, BU, LU) dispose de sa propre tire, y compris l'outil d'annotation automatique (ANA).

3.2 Perception de la mélodie

L'annotation manuelle a rapidement donné lieu à de nouvelles observations et une session a été menée pour comparer les annotations et chercher à déterminer les critères effectivement utilisés par les annotateurs. Ainsi, pour le critère de la hauteur, nous présentons des éléments de discussion à partir de deux extraits du corpus :

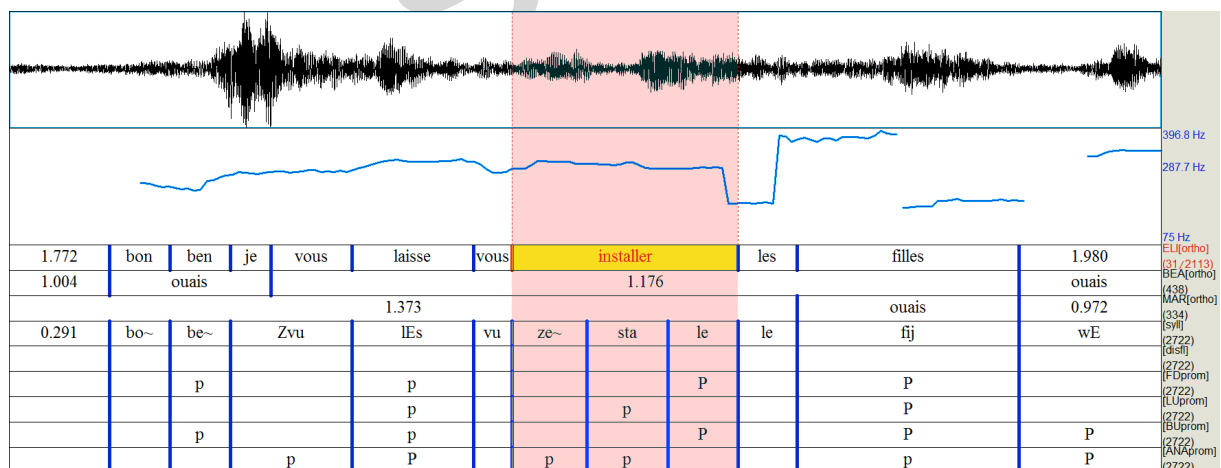


Fig. 1. CLAPI_TableTalk (18-20)

Nous observons à la figure 1 que deux annotateurs (FD et BU) ont repéré une proéminence forte. Or, cette dernière n'est pas marquée intonativement, ce qui peut expliquer que LU ne l'ait pas repérée. Pour autant, les annotateurs non experts diront s'être fiés à la courbe mélodique, à l'intonation, mais nous n'avons aucune mesure objective pour conforter cette analyse. La même chose se répète en figure 2 :

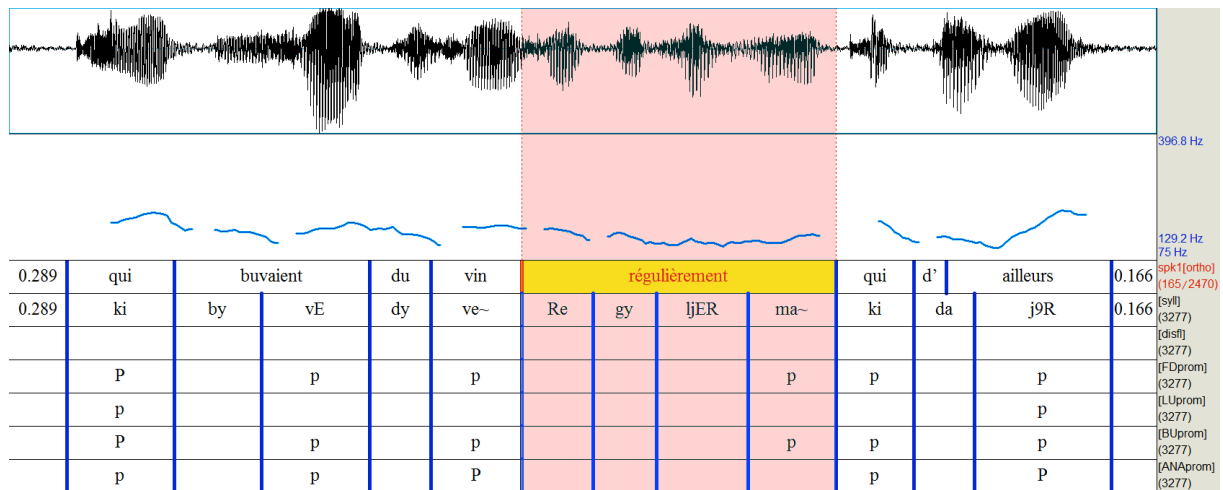


Fig. 2. ESLO_Conf (654-657)

À l'écoute la proéminence sur /ma~/ (dernière syllabe du mot « régulièrement ») est moins évidente. Le pitch fourni par Praat est lui relativement plat, là où il fournit des pics visibles pour /ki/, et manifestement pour le /j9R/ de « d'ailleurs » où, au contraire, l'annotation automatique trouve une proéminence forte due à plusieurs seuils atteints (fréquence, durée, présence d'une pause), là où tous les annotateurs n'ont vu qu'une proéminence faible.

Pour expliquer ces choix d'annotation, nous décrirons les cas (1) et (2) comme des « maintiens de la fréquence ». Essentiellement, nous faisons intervenir ici la subjectivité de l'annotateur en supposant que, dans ces deux cas, il s'attendrait à une baisse de la mélodie et percevrait donc, quand celle-ci se maintient, une hausse. Inversement, dans le cas de « d'ailleurs » en (2) les annotateurs s'attendent probablement par défaut à une hausse ou du moins à un maintien de la mélodie, de sorte que l'écart perçu est moindre pour eux.

Nous avons testé cette hypothèse en changeant les consignes de nos annotateurs pour nous aligner sur ce comportement. D'une part, nous avons détaché la proéminence de son contexte local : l'annotateur devait se fier à une moyenne « globale » de la voix de chaque locuteur pour déterminer les proéminences faibles ou fortes. Nous n'avons cependant eu aucun moyen de définir cette moyenne et, au final, il semble que cette consigne n'ait pas eu d'effet. D'autre part, nous avons demandé à nos annotateurs, afin d'évaluer les attentes dans la fenêtre temporelle annotée, de faire une première écoute, puis de répéter le segment par eux-mêmes avec une « voix monotone » : le but étant de ne conserver que l'accentuation propre à la langue. L'annotateur comparait ensuite cette production artificielle avec son écoute pour déterminer les proéminences. Cette méthode, quand même imprécise, a néanmoins permis de régler la majorité des indécisions lors des séances de test.

Essentiellement, donc, et relativement à la fréquence, nos annotateurs ont annoté comme proéminent une « hausse ou maintien de la fréquence » relativement à une production supposée neutre ou « monotone ».

3.3 Perception de la durée

Concernant le critère de durée, nous avons observé qu'il semblait négligeable chez nos annotateurs, comme le montrent les deux extraits ci-après :

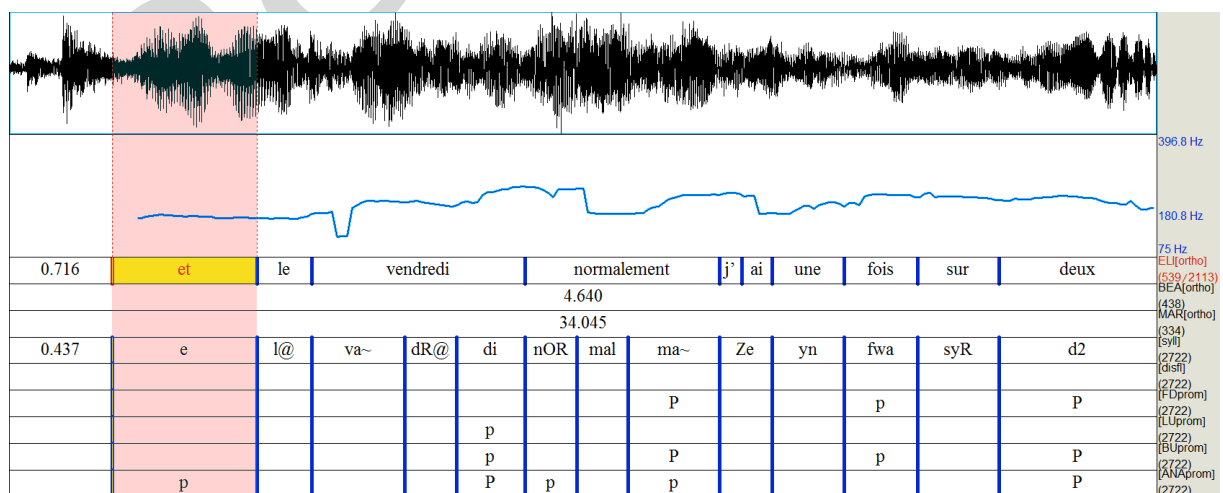


Fig. 3. CLAPI_TableTalk (232-235)

En excluant que ce /E/ de « et le vendredi » soit une disfluente, nous disposons là d'une syllabe d'une durée de 0.28s au milieu d'un contexte (sur 3 syllabes) de 0.14s (0.18s pour le /va~/ de « vendredi ») chez un locuteur où les

« et » varient plutôt entre les 0.08s et 0.16s. ANALOR le repère et l'annote conformément, mais aucun des trois locuteurs n'y voit une proéminence. Cela correspond assez bien au constat déjà réalisé par Morel et ses collaborateurs (2006) où, passé une certaine durée, l'accord sur la perception d'une proéminence diminue. Le cas en (3) n'est pas isolé : il se répète à la seconde 335 dans la transcription, avec une durée de 0.36s : aucun annotateur n'a mis de proéminence.

Tous les cas rencontrés n'ont pas des durées aussi excessives. Dans le cas de (4) ci-dessous, la syllabe /ma~/ de « vraiment » a une durée de 0.21s autour de syllabes d'une moyenne de 0.12s, soit un seuil entre 1.6 et 1.75. Le locuteur prononce un second « vraiment » à la seconde 478 qui dure lui 0.14s. Il y a donc allongement, et ANALOR se fie bien à la durée pour annoter une proéminence faible mais une fois encore, aucun des trois annotateurs n'a placé de proéminence sur cette syllabe. Autrement dit, la durée seule ne semble pas un critère suffisant pour nos annotateurs non experts.

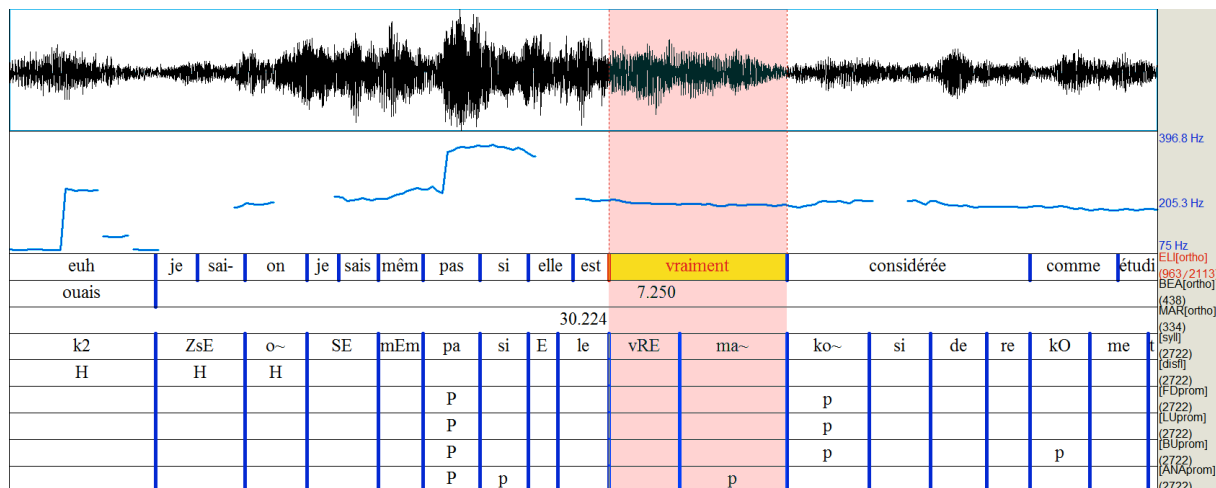


Fig. 4. CLAPI_TableTalk (338-340)

Inversement, dans l'annotation présentée ci-dessus, tous les locuteurs ont placé la proéminence sur la première syllabe de « considérée » (/ko~/), à part l'annotation automatique, ce qui nous amène à notre dernière observation.

3.4 Perception de l'attaque syllabique

Nous avons très tôt observé que les consonnes occlusives, notamment sourdes, en début de syllabe tendaient à provoquer la perception d'une proéminence. Nous avons alors noté cela parmi les pièges pour l'annotateur, à côté par exemple de l'amplitude : une syllabe prononcée plus fortement mais sans hausse de fréquence.

Il est cependant apparu que, pour une série de proéminences, les annotateurs ne pouvaient pas justifier leur choix par la fréquence ou par la durée. Le cas de (4) est exemplaire, avec un pitch plat, aucune attente probable de proéminence mais, pour les annotateurs, un /k/ particulièrement saillant pour « considérée ». Un autre cas serait le suivant :

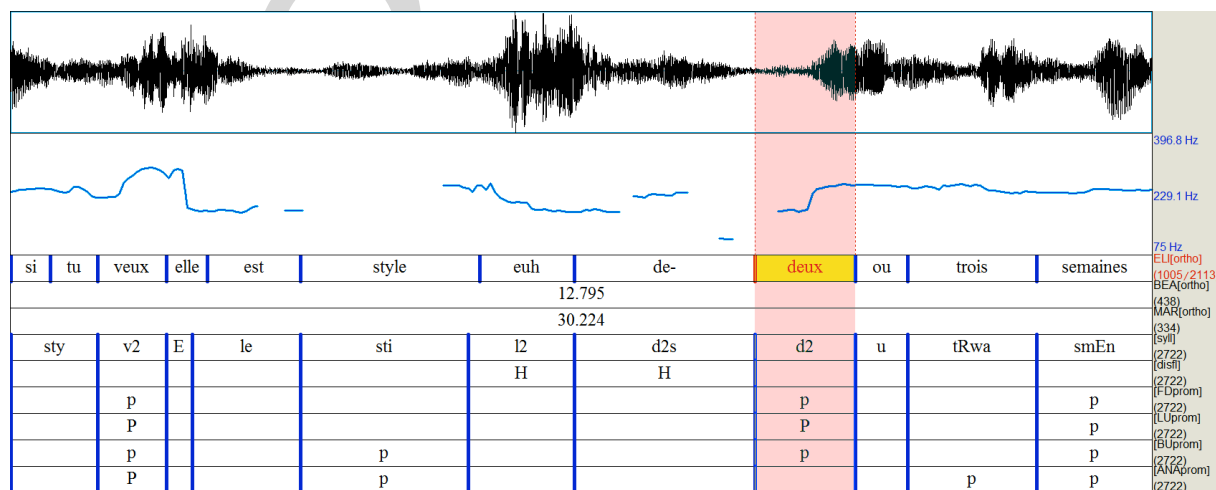


Fig. 5. CLAPI_TableTalk (347-350)

Le pitch de Praat est moins lisible, et les disfluences dans le contexte de gauche ont pu influencer le résultat d'ANALOR, mais les annotateurs n'ont pas perçu de différence de hauteur notable, malgré quoi ils ont trouvé une proéminence, dont une forte. La raison donnée par les annotateurs, visible sur l'oscillogramme, est la durée de la consonne /d/.

Ce cas ne se limite pas aux seules consonnes : en (6), il y a bien un passage de /tE/ à /e~/ avec une hausse de 2 demi-tons, mais la moyenne des points de fréquence dans le contexte élimine ce seuil. Concernant les annotateurs, ceux-ci n'ont pas utilisé la hausse de la mélodie comme argument. La force de la proéminence, qui n'était plus locale selon nos consignes, a été justifiée plutôt par les annotateurs par une particularité de /e~/ décrite comme « un retard sur le premier phonème de la syllabe », en l'occurrence ici le seul phonème de la syllabe.

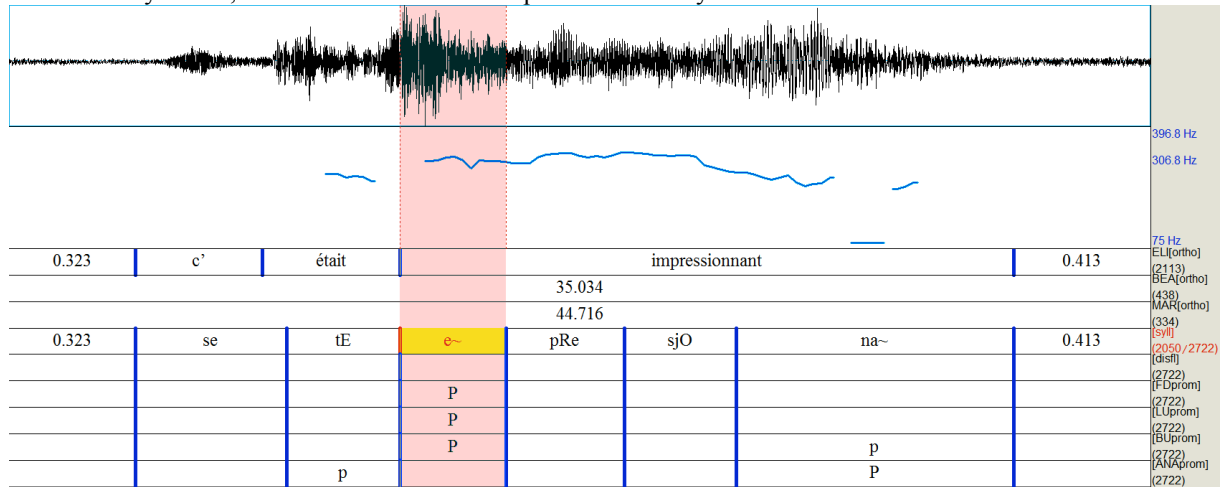


Fig. 6. CLAPI_TableTalk (480-481)

Ce critère perceptif pour le repérage des proéminences a été intégré à nos conventions d'annotation : il a semblé suffisamment fréquent et fiable malgré notre difficulté à le décrire. Il ne s'agit pas d'un critère novateur : Mertens mentionnait déjà, au sujet des accents initiaux : « allongement de la tenue consonantique (...) Si elle [la syllabe] commence par une voyelle (...) et, surtout, l'insertion d'un coup de glotte » (1987 : 85-86). Nous parlerons pour notre part d'allongement de l'attaque syllabique.

Il y a évidemment des cas où l'accord entre annotateurs diffère :

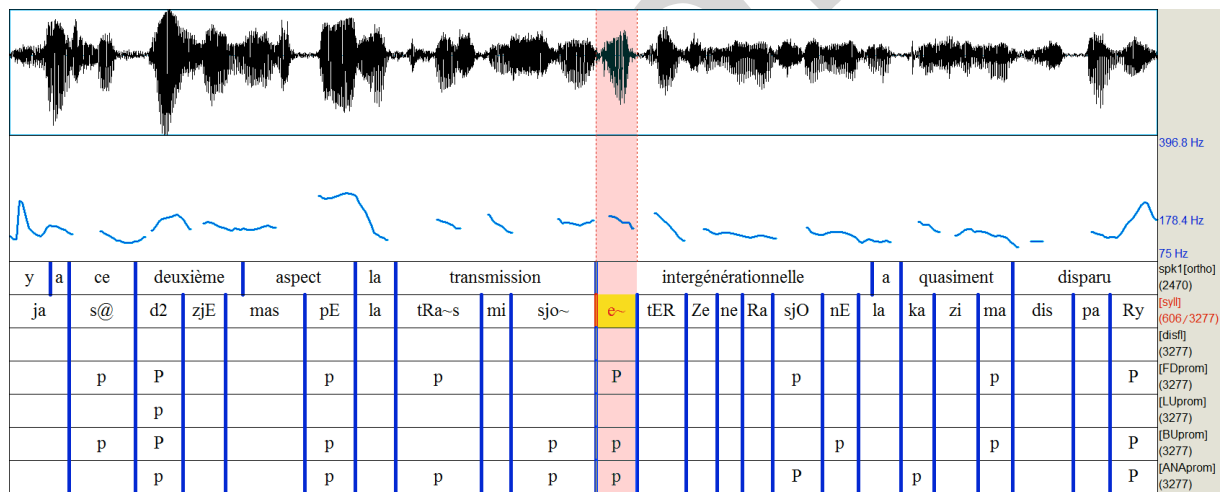


Fig. 7. ESLO_Conf (721-723)

Dans cet enregistrement, nous avons deux occurrences de « intergénérationnelle » du même locuteur où, pour la première, devant disfluence et avec un pic mélodique, tous les annotateurs perçoivent une proéminence sur la syllabe initiale alors que pour la seconde, en (7), les annotateurs ne s'entendent pas. C'est aussi vrai pour « transmission » où un seul annotateur semble percevoir ce phénomène.

4 Fiabilité de l'annotation

Une fois le corpus annoté, nous avons calculé des scores Kappa-Fleiss entre chaque annotateur, annotation automatique incluse.

Dans les tableaux suivants, nous donnons en colonnes les différents enregistrements et en lignes les paires d'annotateurs testées. Dans le cas de LU, nous avons marqué avec « x » tous les enregistrements où cet annotateur n'a pas annoté. Le total pour chaque transcription correspond à un score Kappa calculé sur l'ensemble des annotateurs de cette transcription, incluant l'annotation automatique. Le total pour l'ensemble des transcriptions est un score Kappa excluant LU : nous estimons problématique de comparer le total sur deux enregistrements avec le total des autres annotateurs, obtenu sur neuf enregistrements.

Tableau 2. Accord inter-annotateur (Kappa-Fleiss)

Proéminences	Phone Call	Prepar. Meal	Social Meeting	Table Talk	Conf	Ent	Livre Enf	Media	Table Ronde	Moyenne
FD-BU	0.76	0.82	0.83	0.66	0.82	0.88	0.85	0.9	0.85	0.83
FD-LU	x	x	x	0.41	0.4	x	x	x	x	x
FD-ANA	0.33	0.34	0.35	0.26	0.43	0.41	0.41	0.47	0.46	0.4
BU-LU	x	x	x	0.43	0.39	x	x	x	x	x
BU-ANA	0.35	0.36	0.34	0.35	0.5	0.42	0.44	0.49	0.51	0.43
LU-ANA	x	x	x	0.24	0.27	x	x	x	x	x
Total	0.47	0.51	0.5	0.39	0.48	0.56	0.56	0.61	0.6	0.54

4.1 Accord entre annotations manuelles

Le résultat principal de cette expérience est un accord entre les deux annotateurs FD et BU de 0.83. Nous pouvons comparer ce résultat avec le kappa obtenu dans le cadre de C-PROM (Avanzi *et al.* 2010), de 0.86, et celui obtenu dans le cadre du projet Rhapsodie (Lacheret *et al.* 2014), entre 0.78 et 0.81, score sur l'annotation manuelle à chaque fois.

Ce score, relativement dans la moyenne, doit être accompagné d'une remarque. Nous avons calculé le score sur trois catégories : proéminence forte, faible ou rien. Dans le cas de C-PROM, les deux types de proéminence ont été réduites à une seule (là encore, Avanzi *et al.* 2010). Nous avons donc refait nos scores en suivant cette méthode et en réduisant les proéminences à une seule catégorie :

Tableau 3. Réduction à un seul type de proéminence

Proéminences	Phone Call	Prepar. Meal	Social Meeting	Table Talk	Conf	Ent	Livre Enf	Media	Table Ronde	Moyenne
FD-BU	0.84	0.89	0.89	0.7	0.87	0.92	0.9	0.94	0.92	0.88
FD-LU	x	x	x	0.52	0.49	x	x	x	x	x
FD-ANA	0.42	0.49	0.43	0.31	0.62	0.5	0.54	0.58	0.58	0.51
BU-LU	x	x	x	0.54	0.48	x	x	x	x	x
BU-ANA	0.42	0.49	0.41	0.43	0.67	0.49	0.58	0.6	0.62	0.54
LU-ANA	x	x	x	0.3	0.42	x	x	x	x	x
Total	0.55	0.62	0.57	0.46	0.61	0.63	0.67	0.7	0.7	0.64

Notre score final augmente alors à 0.88, de façon assez égale sur l'ensemble des enregistrements. Nous avons du mal à expliquer ce score : comme pour C-PROM, les annotateurs avaient eu des séances d'entraînement et de discussion pour s'entendre sur les critères, et c'est là sans doute la raison principale, mais les changements importants dans le protocole et le manque d'expertise des annotateurs rendent le résultat surprenant.

4.2 Accord avec l'annotation automatique

La seconde surprise vient du programme ANALOR. À 0.4 ou 0.51, il est bien inférieur aux attentes des travaux précédents, mais les explications sont multiples :

- nous n'avons pas de tire phonémique ;
- nous avons utilisé les seuils par défaut ;
- nous avons changé le protocole d'annotation manuelle.

À quoi s'ajoutent encore deux causes pour ce faible score, qui ne sont pas négligeables. La première est le pitch extrait de Praat et fourni à ANALOR : nous avons là aussi utilisé les options par défaut, d'où un pitch potentiellement de plus faible qualité. La seconde a trait à la qualité même des enregistrements qui, comme dit plus haut (point 1), était variable.

Or la comparaison des tableaux 1 et 2-3 produit une troisième surprise : l'enregistrement « Preparing Meal » a une mauvaise qualité sonore. Il s'agit d'un enregistrement de repas, avec de nombreux bruits de fond dont, pendant 215 secondes, un bruit de machine en continu qui sature le signal sonore. Or, malgré ces conditions, ANALOR obtient un score comparable aux autres enregistrements du corpus CLAPI. Cet enregistrement avait été conservé et annoté exactement pour voir comment l'annotation automatique s'en sortirait et il semble que le pitch extrait demeure suffisamment fiable.

Inversement, on pourra arguer que les enregistrements considérés de bonne qualité ont obtenu des scores plus élevés, qui suggèrent que la qualité sonore demeure un critère important. Nous n'approfondirons pas ici la question.

Le faible score obtenu ne remet pas en question l'outil ANALOR : il appelle une étude qualitative pour comprendre les causes de cette forte divergence et pour en déduire si les critères acoustiques doivent être révisés, ou si les changements réalisés dans le protocole doivent être remis en question ; ou encore s'il s'agit simplement d'une mauvaise manipulation de l'outil et de données insuffisantes. Nous dirions sur ce dernier point que l'outil s'est montré plutôt robuste face à l'absence de segmentation en phonèmes, absence qui rend difficile le calcul des glissandos.

4.3 Le troisième annotateur

Enfin, l'accord inter-annotateurs entre LU et aussi bien les autres annotateurs que l'annotation automatique s'avère remarquablement inférieur : 0.39 à 0.43 pour l'annotation manuelle et 0.24 à 0.27 pour l'annotation automatique. La réduction des prééminences à un seul type ne change pas ce constat et nous ne donnerons rapidement que trois pistes pour expliquer ce résultat. La première est que LU n'est pas un locuteur natif. La seconde est que LU a de l'expérience en annotation prosodique qui ne correspond pas à notre protocole. La troisième est que LU n'a pas pu participer aux séances de test. Au final, son score rappelle celui des annotateurs non experts de Rhapsodie (point 2.1) et pose problème par rapport à l'applicabilité de nos conventions révisées d'annotation manuelle des prééminences.

De façon plus générale, nous pouvons observer les causes pratiques de la divergence au travers d'un décompte des occurrences. La première ligne contient le total de syllabes, puis les lignes suivantes donnent le locuteur et le nombre d'occurrences des prééminences faibles « p » et fortes « P ».

Tableau 4. Décompte des occurrences.

Décompte	Phone Call	Prepar. Meal	Social Meeting	Table Talk	Conf	Ent	Livre Enf	Media	Table Ronde	Total
Syllabes	2567	1227	2672	2722	3277	2592	2442	3177	3362	24038
FD-p	409	250	432	321	685	454	504	616	565	4236
FD-P	310	205	291	265	280	269	350	374	369	2713
BU-p	453	267	475	478	738	474	549	661	602	4697
BU-P	307	194	268	310	295	258	347	339	309	2627
LU-p	x	x	x	384	375	x	x	x	x	x
LU-P	x	x	x	187	129	x	x	x	x	x
ANA-p	615	276	649	643	616	567	666	693	651	5376

ANA-P	341	165	336	370	532	359	387	523	390	3403
-------	-----	-----	-----	-----	-----	-----	-----	-----	-----	------

Le total nous intéresse principalement, puisqu'il montre qu'ANALOR sur-annote, par rapport aux deux annotateurs FD et BU, les proéminences d'environ 6% : 1'000 à 1'700 occurrences de différences, dont environ 700 pour les proéminences fortes. Cela suggère simplement que les seuils fournis sont trop faibles. Concernant FD et BU eux-mêmes, ils ne semblent pas diverger grandement, bien que BU tende à annoter plus de proéminences faibles.

Dans le cas de LU, la comparaison est difficile du fait du faible nombre d'occurrences. Cela dit, on observera un nombre de proéminences sensiblement plus réduit : jusqu'à un tiers moins de proéminences faibles et moitié moins de proéminences fortes.

5 Conclusion

L'expérience de l'annotation en proéminences dans le cadre du projet SegCor amène donc à des constats mitigés.

Concernant notre première question, nous sommes amenés à accepter trois résultats : (a) la perception de la fréquence dépend des attentes du sujet parlant ; (b) la durée syllabique semble peu pertinente ; (c) les annotateurs se fient à l'attaque syllabique pour repérer les proéminences. Nous avons modifié le protocole de Rhapsodie en conséquence :

- en comparant l'empan annoté à une « voix monotone » pour chaque locuteur ;
- en ajoutant l'attaque syllabique aux critères de la durée et de la fréquence.

Le premier point a systématiquement été appliqué par nos annotateurs afin d'« objectiver » leur perception des niveaux de proéminence, en tenant compte de leur double compétence, à la fois en tant que locuteurs (capables de reproduire une courbe mélodique non marquée) et en tant que récepteurs (capables d'apprécier les traits saillants des proéminences dans le signal). Le deuxième point a fait l'objet de plusieurs commentaires au cours des séances test et a émergé comme ayant un impact dans les choix opérés par nos annotateurs. Il semble qu'en se concentrant sur les trois critères énumérés ci-dessus, et en les définissant ainsi, nous obtenons un accord inter-annotateur remarquable, mais ne correspondant plus forcément aux proéminences des travaux précédents.

Concernant notre deuxième question, nous devons aussitôt remettre en question cet accord en tenant compte du troisième annotateur, pour lequel le protocole s'avère toujours insuffisant. Nous insisterons particulièrement sur l'importance d'au moins une séance d'entraînement en commun pour comparer les expériences et s'entendre sur des cas pratiques. Quant à l'annotation automatique, elle s'avère plus robuste que prévu sur des enregistrements de mauvaise qualité, mais l'absence de tire phonémique et d'entraînement de l'outil réduisent considérablement l'accord.

Nous devons insister à ce sujet sur la signification des scores : ils ne mesurent que le consensus entre les différents annotateurs, non leur validité par rapport à une référence qui n'existe pas. L'accord inter-annotateurs évalue avant tout la fiabilité du protocole pour cerner la proéminence en tant que phénomène effectivement perçu et exploité par les sujets parlants pour, entre autres, la démarcation d'unités de différents niveaux. Un désaccord ne signale pas que les annotateurs auraient tort, mais que le protocole et les critères fournis sont remis en cause. La réévaluation des critères doit passer, à son tour, par l'analyse qualitative des divergences là où elles se produisent entre annotateurs.

Quant à l'exploitation de ces données, à ce stade, elle s'avère difficile, qu'il s'agisse de repérer des groupes intonatifs ou de déterminer le statut de mots-phrases (Tesnière 1959) tels que *oui, non, si, voilà*, qui peuvent souvent être considérés comme des unités non autonomes et problématiques⁹. Les séquences comme « oui mais », « non mais », où le mot-phrase est lié au « mais » et les deux forment un seul mot prosodique posent également problème pour l'analyse macrosyntaxique (Lacheret-Dujour *et al.* 2011 : 73-74). En linguistique interactionnelle, par ailleurs, on peut se pencher sur le positionnement de certaines formes linguistiques (comme les marqueurs de réception) ou sur les profils prosodiques de certains formats (comme les hétéro-répétitions), et en fournir des caractérisations actionnelles (Persson 2017). Dans les deux perspectives, il s'agit alors d'envisager les proéminences comme relevant d'un niveau d'analyse spécifique, contribuant à une analyse multidimensionnelle de la parole en interaction, plutôt que comme un support pour l'annotation à d'autres niveaux. Pour l'instant, la seule annotation des proéminences ne semble pas apporter de solides arguments aux choix opérés dans la segmentation d'unités de haut niveau (interactionnelles et macrosyntaxiques), elle devrait s'inscrire dans une analyse prosodique et phonétique plus large. Dans le cadre du projet SegCor, nous nous proposons de revenir sur cette annotation une fois qu'une liste de cas critiques sera finalisée pour la segmentation du corpus pilote, à partir des guidelines en cours d'élaboration. Les indications des guidelines pourront alors fournir des éléments utiles pour repérer des « unités maximales » pouvant être éventuellement caractérisées en tant que telles au niveau prosodique.

Dans notre contribution, nous avons décrit la préparation du corpus à annoter et illustré les résultats de l'expérience d'annotation des proéminences à l'aide des accords inter-annotateurs. Ensuite, nous avons souligné les aspects pratiques qui nous ont amené à présenter de nouveaux éléments sur lesquels s'appuyer pour une annotation des proéminences et à proposer des modifications concernant le protocole de codage Rhapsodie. Quant au rôle de l'annotation des proéminences (et prosodique en général) pour la segmentation d'unités de haut niveau, les éléments évoqués ont permis d'envisager quelques pistes de recherche à partir des cas d'annotation problématique rencontrés dans notre projet et dans des travaux précédents.

Références bibliographiques

- Avanzi, M. (2012). *L'interface prosodie/syntaxe en français : Dislocations, incisives et asyndètes*. Berne : Peter Lang.
- Avanzi, M., Goldman, J.-P., Lacheret, A., Simon, A. C. & Auchlin, A. (2007). Méthodologie et algorithmes pour la détection automatique des syllabes proéminentes dans les corpus de français parlé. *Cahiers of French Language Studies*, 13(2), 2-30.
- Avanzi, M., Lacheret-Dujour, A., Obin, N. & Victorri, B. (2011). Vers une modélisation continue de la structure prosodique : le cas des proéminences syllabiques. *Journal of French Language Studies*, 21, 53-71.
- Avanzi, M., Lacheret, A., Sauvage-Vincent, L. & Victorri, B. (s.d.). *Protocole de codage prosodique. Annotation des proéminences, disfluences et périodes dans le projet Rhapsodie*, <<http://projet-rhapsodie.fr/tuto/Codage%20Prosodique.pdf>>.
- Avanzi, M., Lacheret, A. & Victorri, B. (2008). ANALOR. A Tool for Semi-Automatic Annotation of French Prosodic Structure. In *Proceedings of the 4th International Conference on Speech Prosody*, Campinas, Brazil, 119-122, <http://www.isca-speech.org/archive/sp2008/papers/sp08_119.pdf>.
- Avanzi, M., Simon, A.-C., Goldman, J.-P. & Auchlin, A. (2010). C-PROM. An annotated corpus for French prominence study. *Speech prosody, 5th international conference*, Chicago, <<https://www.researchgate.net/publication/236971928>>.
- Baude, O., Blanche-Benveniste, C., Calas, M.-F., Cappeau, P., Cordereix, P., Goury, L., Jacobson, M., de Lamberterie, I., Marchello-Nizia, C. & Mondada, L. (2006). *Corpus oraux : guide des bonnes pratiques*. Paris : CNRS Éditions & Presses Universitaires d'Orléans.
- Béguelin, M.-J. (2010). Noyaux prédicatifs juxtaposés. In M.-J. Béguelin, M. Avanzi & G. Corminbœuf (éds), *La Parataxe. Tome 1 : Entre dépendance et intégration*. Berne : Peter Lang, 3-33.
- Benzitoun, C., Sabio, F., Kahane, S. & Pietrandrea, P. (2012). *Protocole de codage macrosyntaxique du projet Rhapsodie*, <http://www.projet-rhapsodie.fr/plus/publications/doc_download/188-tutoriel-de-codage-syntaxique.html>.
- Blanche-Benveniste, C. (2010). *Le français. Usages de la langue parlée* [Avec la collaboration de Ph. Martin pour l'étude de la prosodie]. Leuven : Peeters.
- Blanche-Benveniste, C. (2013 [1979]). À propos de la langue écrite. *TRANEL – Travaux neuchâtelois de linguistique*, 58, 21-30.
- Boersma, P. & Weenink, D. (2017). Praat: doing phonetics by computer, <praat.org>.
- Cerisara, C., Mella, O. & Fohr, D. (2009). JTrans, an open-source software for semi-automatic text-to-speech alignment. *Proceedings of the 10th Annual Conference of the International Speech Communication Association – Interspeech 2009*, <http://www.isca-speech.org/archive/archives_papers/interspeech_2009/papers/i09_1823.pdf>.
- Christodoulides, G., Barreca, G. & Avanzi, M. (2015). DisMo : un annotateur multi-niveaux pour les corpus oraux. In *Actes de la 22ème conférence Traitement Automatique des Langues Naturelles*, <http://www.atala.org/taln_archives/TALN/TALN-2015/taln-2015-demo-012.pdf>.
- Fleiss, J. L. & Cohen, J. (1971). The Equivalence of weighted Kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement*, 33, 613-619.
- Goldman, J.-Ph. (2011). EasyAlign : an automatic phonetic alignment tool under Praat. In *Proceedings of the 12th Annual Conference of the International Speech Communication Association – Interspeech 2011*, <http://www.isca-speech.org/archive/archives_papers/interspeech_2011/i11_3233.pdf>.
- Groupe de Fribourg (2012). *Grammaire de la période*. Berne : Peter Lang.
- Lacheret-Dujour, A. (2016). Modélisation de la structure prosodique du français parlé basée sur l'usage : l'expérience Rhapsodie. *Langue française*, 191, 67-82.
- Lacheret-Dujour, A. & Beaugendre, F. (1999). *La prosodie du français*. Paris : Éditions du CNRS.
- Lacheret, A., Kahane, S., Beliao, J., Dister, A., Gerdes, K., Goldman, J.-P., Obin, N., Pietrandrea, P., Tchobanov, A. (2014). Rhapsodie : un Treebank annoté pour l'étude de l'interface syntaxe-prosodie en français parlé. In F. Neveu *et al.* (éds), *4e Congrès Mondial de Linguistique Française – CMLF 2014*, Les Ulis : EDP Sciences, 2675-2689.
- Lacheret-Dujour, A., Kahane, S., Pietrandrea, P., Avanzi, M. & Victorri, B. (2011). Oui mais elle est où la coupure, là ? Quand syntaxe et prosodie s'entraident ou se complètent. *Langue Française*, 170, 61-80.
- Lacheret, A. & Victorri, B. (2002). La période intonative comme unité d'analyse pour l'étude du français parlé : modélisation prosodique et enjeux linguistiques. *Verbum*, XXIV, 55-72.
- Martin, Ph. (2006). La transcription des proéminences accentuelles : mission impossible ? *Bulletin PFC*, 6, 81-87.
- Mertens, P. (1987). *L'intonation du français : de la description linguistique à la reconnaissance automatique*. Thèse de doctorat, Katholieke Universiteit Leuven.
- Morel, M., Lacheret-Dujour, A., Lyche, C. & Poiré, F. (2006). Vous avez dit proéminence ? In *Actes des 26^e Journées d'Étude sur la Parole – JEP 2006*, 183-186, <http://www.afcp-parole.org/doc/Archives_JEP/2006_XXVIe_JEP_Dinard/JEP06_ACTES.PDF>.

- Obin, N., Beliao, J., Veaux, C. & Lacheret, A. (2014). SLAM : Automatic Stylization and Labelling of Speech Melody. In *Proceedings of the 7th International Conference on Speech Prosody*, 246-250, <hal-00997238>.
- Persson, R. (2017). La prosodie comme ressource pour l'organisation de l'interaction : état des lieux et illustrations. *Revue française de linguistique appliquée*, XXII (2), 33-52.
- Poiré, F. (2006). La perception des proéminences et le codage prosodique. *Bulletin PFC*, 6, 69-79.
- Selting, M. (1996). On the interplay of syntax and prosody in the constitution of turn constructional units and turns in conversation. *Pragmatics*, 6 (3), 371-388.
- Simon, A.-C., Avanzi, M. & Goldman, J.-Ph. (2008). La détection des proéminences syllabiques. Un aller-retour entre l'annotation manuelle et le traitement automatique. In J. Durand, B. Habert & B. Laks (éds), *1er Congrès Mondial de Linguistique Française – CMLF 2008*, 1673-1686.
- Simon, A.-C. & Christodoulides, G. (2016). Frontières prosodiques perçues : corrélats acoustiques et indices syntaxiques. *Langue française*, 191, 83-106.
- Tesnière, L. (1959). *Éléments de syntaxe structurale*. Paris : Klincksieck.

¹ Dans le projet franco-allemand SegCor, la partie allemande du corpus pilote est issue de la base FOLK (<http://agd.ids-mannheim.de/folk.shtml>), développée par l'Institut für Deutsche Sprache (IDS) de Mannheim. Les aspects contrastifs ne font pas l'objet de cette contribution, mais ont été pris en compte dès le début du projet.

² <http://clapi.icar.cnrs.fr/>

³ <http://eslo.huma-num.fr/>

⁴ Anne Lacheret-Dujour (2016 : 69) parle d'approche fonctionnelle « mettant l'auditeur au centre du dispositif ».

⁵ <http://www.projet-rhapsodie.fr/tuto/Codage%20Prosodique.pdf>

⁶ Nous considérons annotateur non expert un locuteur qui n'a pas de formation spécifique en phonétique et phonologie du français.

⁷ Le glissando est une montée ou descente de la F0 dans le noyau vocalique (intrasyllabique).

⁸ C'est le cas des données issues de la base CLAPI, elles ont été transcrites selon la convention ICOR, consultable à l'adresse suivante : <http://icar.univ-lyon2.fr/projets/corinte/documents/2013_Conv_ICOR_250313.pdf>.

⁹ Ces mots sont parfois annotés comme « unités illocutoires associées » et signalés comme « marqueurs de discours » dans le protocole de codage macrosyntaxique Rhapsodie (Benzitoun *et al.* 2012).