

Pour une segmentation prosodique de la parole spontanée
Exemple sur le parler régional de Tlemcen, Étude à base de Praat

BEZZEGHOUD Sihem

Université Aboubekr BELKAID, DYNLAMED, TLEMEN

Doctorante en sciences du langage dirigée

Par le Pr B.BENMOUSSAT¹ et Pr M.SENOUCI²

Résumé

Dans cet article nous allons tenter de résoudre un problème épineux auquel les linguistes et les phonéticiens font face lors du traitement du son et du signal à travers l'élaboration d'un outil à base du logiciel PRAAT qui permet la segmentation automatique de plusieurs unités du corpus ainsi que l'annotation et la transcription du signal. Ce dernier est composé d'un script qui aide l'application de plusieurs tiers afin d'aboutir à une segmentation de plusieurs unités de son en passant par la constitution de fichiers Text Grid : l'unité principale pour toute analyse prosodique sur la base de Praat, ce qui facilitera la représentation graphique du signal de parole et par conséquent la création des courbes intonatives et la segmentation en tiers afin de permettre l'inclusion des transcriptions phonétique propres à chaque énoncée.

Cet article décrit cet outil, son mode de conception et de fonctionnement et présente les résultats d'évaluation et la performance de cet outil à travers la segmentation automatique du corpus recueilli sur le parler de la région de Tlemcen.

Mots clés : Prosodie, corpus parole spontanée, scripte, segmentation automatique, parler régional de Tlemcen.

Abstract

In this paper we try to resolve a great problem that linguists and phoneticians are facing through the conception of a Praat based tool for the automatic segmentation in prosodic units of speech corpora, signal annotation and transcription in addition to the description of the sets this tool is made up of starting by the constitution of Text Grids : the basic unit of each prosodic analysis using Praat which will facilitate the graphic representation of sound signals . As a result the realisation of boundaries and units segmentations into tiers to allow including phonetic transcriptions .

This paper briefly describes the work flow of each set of this tool and the way it has been successfully used on the corpora of Tlemcen and its surroundings .It also presents the results of an evaluation of performance of this tool in an automatic annotation task on a large choosed corpora .

Key words

Prosody, spontaneous speech corpora, script,automatic segmentation, Tlemcen regional spoken language .

Introduction

Le traitement de la parole spontanée ne peut se limiter de nos jours à une analyse purement segmentale, il peut en plus nécessiter la prise en considération des éléments et des phénomènes prosodiques. Les recherches sur le traitement de la prosodie ont permis d'améliorer considérablement la qualité de la synthèse et du traitement de la prosodie grâce aux techniques modernes d'analyse numérique du signal. L'analyse prosodique nécessite l'élaboration et l'organisation d'un corpus de paroles recueillies orienté dans cette optique. Toutefois, le problème du naturel de la parole reste un domaine d'un grand intérêt pour la recherche tant au niveau linguistique qu'au niveau du traitement du signal. L'étude des techniques numériques est un niveau moderne du traitement du signal qui permet d'analyser les paramètres prosodiques. En effet, la prosodie englobe les variations d'intonation le rythme, l'accentuation, l'intensité de la parole naturelle. Les paramètres prosodiques mesurables sont la fréquence fondamentale F_0 , la durée et l'intensité. La segmentation d'un corpus de parole demande beaucoup de temps si il doit être effectué d'une façon naturelle et manuelle, spécialement si le corpus est important. Cependant, une programmation informatique sur la base d'un outil spécifique destiné à la segmentation des énoncés est plus que nécessaire quand le corpus s'avère être volumineux ou hétérogène. Dans ce cadre notre problématique est de prouver que les énoncés sélectionnés du corpus recueilli mettent en relief les aspects prosodiques tels que l'intonation, l'accentuation et le rythme après sélection en première phase et en deuxième phase. Ceci aboutira impérativement après l'élaboration d'un script qui permettra cela et de combiner comparaison systématique et analyse détaillée des énoncés. Ce programme permettra aussi de générer un nombre considérable de données de sortie et d'analyser un large nombre de fichiers audio ce qui facilitera l'analyse du signal après reversion des fichiers .wave en fichiers TextGrid, en d'autres termes ces énoncés seront exposés à un outil d'analyse prosodique personnalisé en utilisant Praat qui permettra la constitution des fichiers Text Grids et par la suite différents niveaux de segmentation, d'annotation et de traitement et d'analyse du signal.

A-Matériel et Méthodologie

La carence des études prosodiques consacrées au parler de la région Tlemcen n'a pas été un facteur décourageant pour essayer de nous lancer dans cette optique, malgré que le manque de documents a été un réel handicap pour nous. Dans un tel contexte d'indigence et de manque de référence propre à la prosodie de notre parler : le risque de ne pas pouvoir trouver une explication à un certain nombre de phénomènes reste inévitable.

Toute production langagière ou dialecte, toute énoncée possède inévitablement une dimension prosodique ; des éléments suprasegmentales, syllabe, accent, intonation, durée. Et dans ce cadre, la conception de notre outil et l'élaboration du script va être essentiellement orientée sur l'annotation des limites prosodiques 'boundaries'. L'annotation automatique contrairement à l'étiquetage manuel est en fait un gain de temps surtout quand il s'agit de traiter et d'analyser des corpus larges et colossaux et la transcription du signal l'analyse intonative et l'accentuation des énoncées sélectionnées en plus du calcul de la durée des phonèmes suivants: [q], [k],[S],[tS] et [?] par rapport aux voyelles adjacentes .

A.1-Description du Matériel utilisé

A.1.1 Description et utilité de PRAAT

Plusieurs outils informatiques d'aide à la transcription et à l'annotation de données sonores ont été élaborés parmi les outils permettant à la fois la transcription, l'annotation et l'analyse des corpus constitués de fichiers de parole, Parmi tous les logiciels qui facilitent la manipulation de ces données c'est-à-dire le traitement des fichiers audio, nous avons choisi PRAAT, qui a été développé par Paul Boersma et David Weenink à l'Institut des Sciences Phonétiques de l'Université d'AMSTERDAM .

L'intérêt que lui porte les linguistes est vaste pour cause de sa grande précision dans l'analyse audio et la compatibilité de ses fichiers produits avec d'autres logiciels en les convertissant à d'autres formats, il permet une analyse du son dans un environnement ergonomique avec précision ,c'est donc un outil d'analyse et de modélisation en phonétique et en phonologie largement utilisé dans diverses domaines des sciences du langage .

Le programme PRAAT est un outil complexe dans sa conception et dans ses applications mais il reste désormais le plus perfectionné et adaptable au traitement et à l'analyse des fichiers audio. Son téléchargement est gratuit et fonctionne sous toutes les plateformes courantes tel que Windows. Après l'élaboration d'un script adapté il facilite la transcription du signal, qui s'effectue sur la base de deux fichiers : le fichier audio en premier converti en fichier textGrid, ce dernier constitue finalement le fichier de transcription propriétaire au logiciel. Il facilite aussi l'alignement des transcriptions avec les fichiers audio appropriés, leur approfondissement et lecture en temps réel. Ce logiciel semble répondre à nos attentes et nous initier à la programmation dans le but de traiter le corpus recueilli.

Dans cet article nous nous proposons de présenter plus en détails certaines des fonctionnalités de ce programme. Comme tout logiciel, il se présente sous un ensemble de programmes couvrant nos besoins dans le domaine de l'analyse prosodique, fonctionnant sur plusieurs plateformes, chaque fonction est un petit programme qui réalise une ou plusieurs actions bien précises pour transcrire orthographiquement et annoter les données, nous nous concentrons sur quelques fonctionnalités qui peuvent être intéressantes pour les tâches de transcription et d'annotation des corpus oraux, comme les tâches de saisie des données audio, l'enregistrement de fichiers audio qui peuvent ensuite être analysés, la transcription, l'étiquetage et la segmentation de fichiers, l'analyse phonétique et acoustique au niveau segmental : le logiciel permet de mener des analyses spectrographiques et d'autres analyses prosodiques F_0 , durée et intensité et modification par stylisation des courbes de fréquence fondamentale et d'intensité construction de réseau de neurones et la possibilité d'écriture de script pour effectuer plus rapidement certaines tâches d'extraction d'information comme nous prévoyons de le faire.

Ce logiciel nous convient donc parfaitement dans le sens où il nous aide à traiter les fichiers du format .wave, établir l'étiquetage des phonèmes ciblés et de déterminer l'identification des voyelles et consonnes avoisinantes ce qui me permettra par la suite de recueillir un maximum d'informations afin de comparer la durée de chaque phonème dans différents environnements. En d'autres termes établir des traitements au niveau segmental et suprasegmental.

A.1.2 Méthodologie d'élaboration du Script

Le script est un langage informatique élaboré dans le cadre d'un ensemble commandes programmées pour faire exécuter des fonctions à PRAAT, notre script est organisé comme suit :Organisé en deux phases, *la phase initiale* représentée sous forme de warm up (introduction) contenant un ensemble d'instructions qui permettront une meilleure orientation de l'analyse des données en indiquant l'emplacement des fichiers d'entrée et de sortie et essentiellement permettant d'identifier et de définir le phonème à analyser.

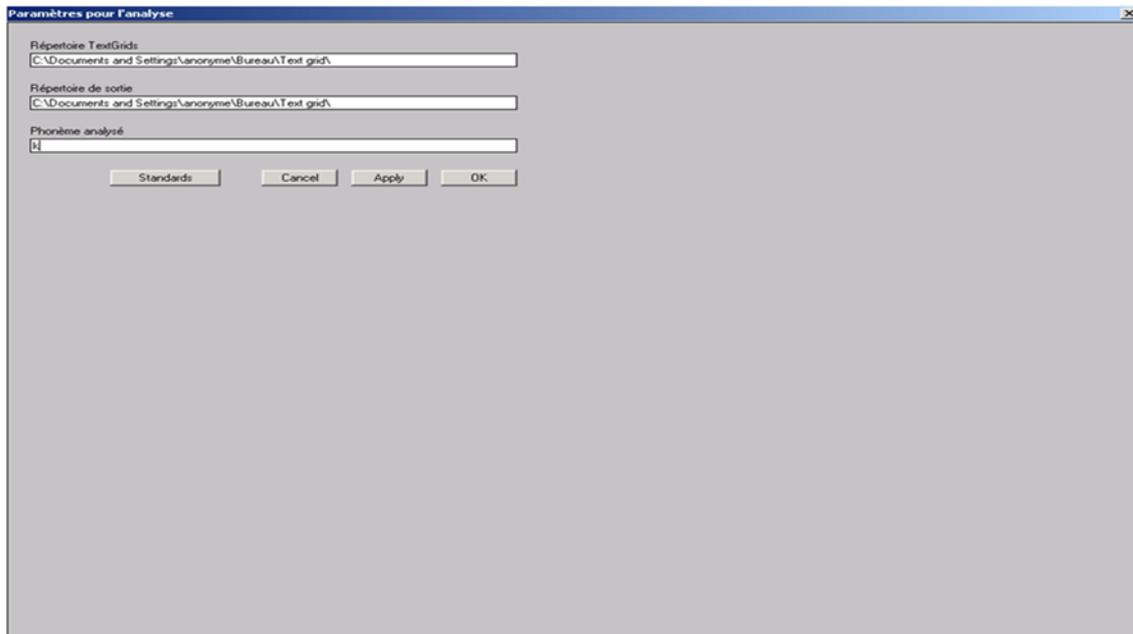


Figure 1 Fenêtre ouverte sur la partie initiale du script

Permettant l'identification du phonème

En d'autres termes, dans cette phase le logiciel est sensé localiser le répertoire dans lequel les fichiers se trouvent les identifier un à un selon le type et la catégorie à laquelle ils appartiennent et les classer selon leurs différences tout en respectant la mention rural et urbain, ce qui permettra en phase finale de ressortir correctement les caractéristiques prosodiques de tous les fichiers audio, leur approfondissement et leur analyse à une échelle comparative. C'est à dire combiner comparaison systématique et analyses détaillées des fichiers. Cette étape est très importante, le filtrage des fichiers audio facilite une lecture automatique correcte, et par la suite une répartition exacte des labels et de l'étiquetage correspondant à l'enregistrement approprié.

La seconde partie est présentée sous la forme de **deux boucles** précédées par une instruction qui permet de créer **la liste des fichiers text Grid** et en définissant leur nombre. Le fichier de sortie du type .txt nommé ‘résultats’ dans lequel les résultats sont affichés .Ce fichier texte est manuellement converti en fichier Excel afin de permettre une meilleure lecture.

A partir du début de la première boucle l’instruction donnée est de parcourir l’ensemble des fichiers de **1** jusqu’à **n** nombre de fichiers, sélectionner la liste, choisir un fichier à traiter selon le phonème qu’on veut analyser, aboutir à sa lecture et enfin afficher l’information. Cette opération est suivie par une pause avant de passer à l’étape suivante qui est supposée déterminer le nombre d’intervalles à traiter. La deuxième boucle est initialisée par une instruction qui consiste à parcourir toutes les intervalles, identifier les labels et si l’étiquette correspond au phonème choisi, l’étiquette des voyelles ou consonnes adjacentes est récupérée. Une condition a été introduite : si le fichier ne connaît pas de label après le phonème en question la mention ‘fin de fichier ’ est introduite. Dans le cas échéant on passe à l’intervalle suivant. Pour chaque fichier text Grid traité, le script est automatiquement affiché les informations relatives à la durée dans le fichier de sortie .txt appelé ‘résultat’ converti en fichier Excel.

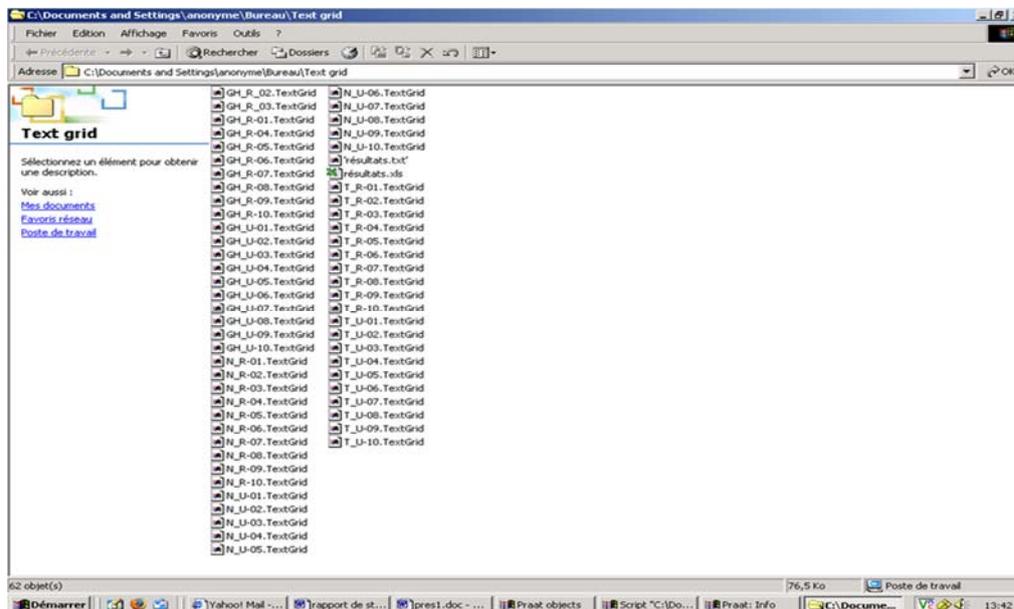


Figure 2 :Fenêtre ouverte sur les fichiers traités convertis en text grid en plus du fichier résultat

A-2- Méthodologie (élaboration du corpus)

En ce qui concerne le corpus utilisé, vu l'absence de toute étude préalable sur le parler arabe de la région de Tlemcen, il était nécessaire pour nous de constituer notre propre corpus nous-même. Autour de ce noyau l'élaboration du corpus qui a servi à l'aboutissement de cette étude prosodique a été réalisée dans la région de Tlemcen dans des zones rurales et urbaines. La taille du corpus dans le sens strict du terme a aussi été un critère exclusif à prendre en considération qui a été justement déterminée en fonction de nos besoins d'analyses linguistiques et prosodiques. Les locuteurs et auditeurs (hommes et femmes) sont entièrement natifs de la région, âgés de 18 à 85 ans ayant vécu et évolué dans la région depuis leur tendre enfance pour tous les âges confondus. Si il leur était arrivé au courant de leur vie de se déplacer en dehors de la région c'était pour des périodes de courtes durées pour des raisons professionnelles ou autres.

Le corpus que nous ne sommes entrepris de constituer sur la base d'enregistrements audio établis à l'aide d'enregistreur numérique et dictaphone. Les enregistrements ont été établies dans le cadre de conversations spontanées dans leur environnement naturel, une dynamique avait été donc établie afin d'instaurer un véritable dialogue entre les participants et l'utilisateur de l'outil d'enregistrement sans pour autant être esclave de l'outil. On a utilisé un ensemble de trois heures de conversations non interrompues en utilisant un enregistrement numérique en utilisant un dictaphone qui était dissimulé. Le nombre des sujets avoisinait la cinquantaine mais seuls les meilleurs fichiers ont été choisis selon les conditions d'enregistrement. D'autres enregistrements étaient établis sous la forme de lecture ou d'interviews. En phase initiale les thèmes étaient centrés sur leurs activités quotidiennes afin de leur permettre de mieux se concentrer sur notre objectif principal, c'est-à-dire : ressortir les particularités vernaculaires et paramètres prosodiques.

Nous avons utilisé une soixantaine de fichiers avec une durée variable dans l'ensemble du corpus mais seuls ceux regroupant le plus de paramètres prosodiques avaient été sélectionnées. Nous avons voulu constituer une plateforme linguistique afin que ces échantillons linguistiques recueillis permettent une analyse segmentale et supra segmentale et soient vraiment adaptables au script élaboré.

B-Résultats

B-1 Annotation et étiquetage

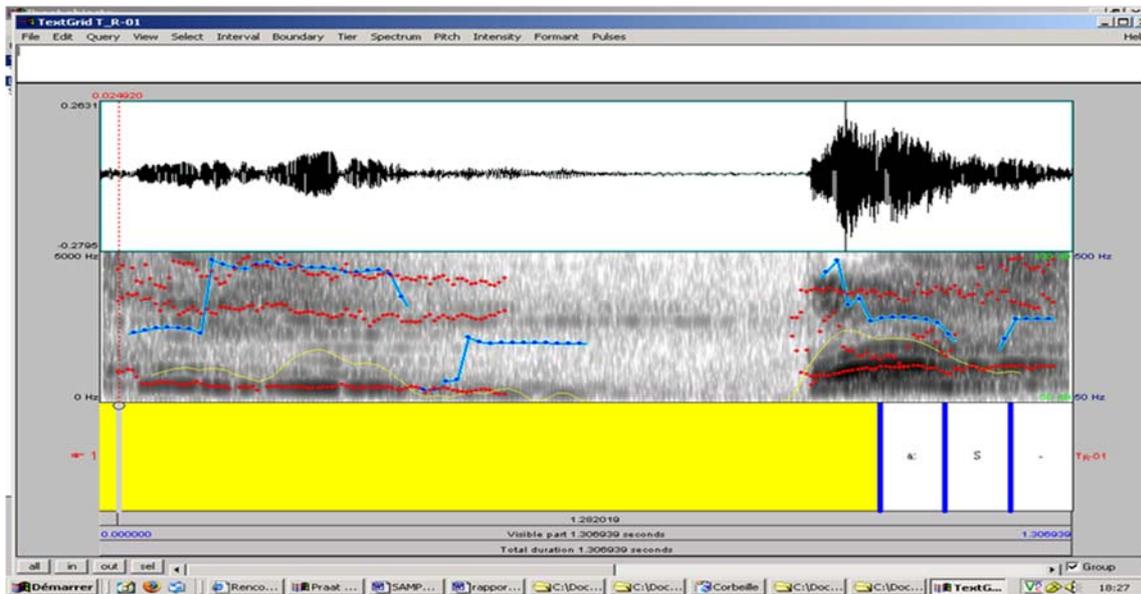


Figure 3 : Fenêtre ouverte du fichier Texte Grid T_R-01

L'annotation automatique à travers script et un fichier texte Grid n'est pas seulement un gain de temps mais d'une précision extrême. L'étiquetage automatique est une prometteuse solution alternative à tous les problèmes d'identification perceptuelle des unités et dans notre étude la position exacte du phonème en question.

B-2 Représentation intonative du signal

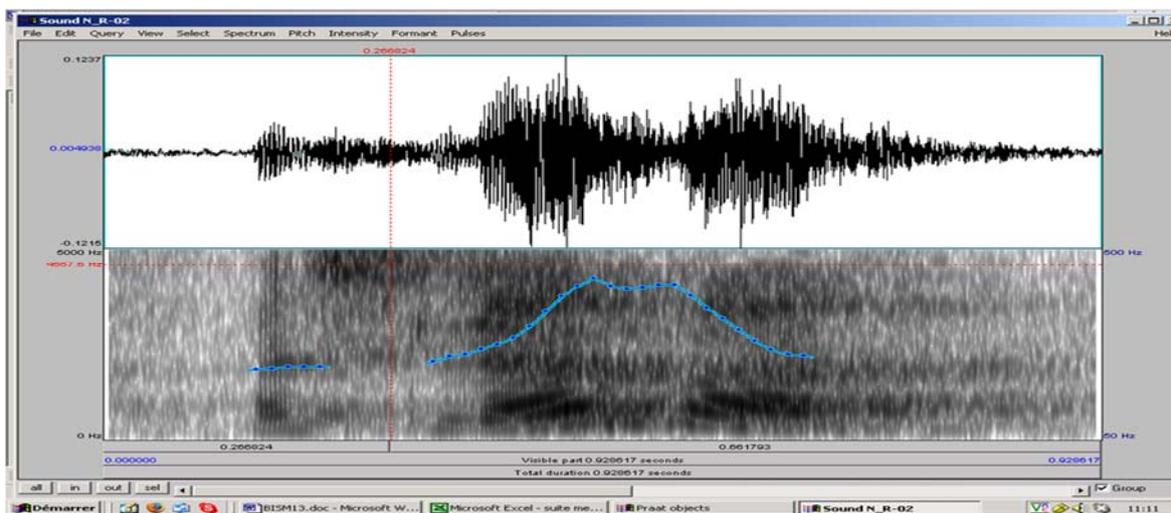


Figure 4 : Fenêtre ouverte sur la représentation intonative du fichier (N_R-02.wav)

La création des fichiers text grid à travers le script permet un scan complet du signal et une représentation intonative graphique de l'intonation juxtaposée permettant la mise en relief du positionnement spécifique de F0 dans le nucleus de la syllabe accentuée.

B-3 calcul de la durée

Nombre	Nom &	Phonème	Voy/Consn	Voy/Consn	Durée
De fichiers traités	Type de fichier	analysé	Précédente	Suivante	Calculée
1	N_U-01.TextGrid		q	- a	0,0764
2	N_U-02.TextGrid		q	t a	0,039
3	N_U-03.TextGrid		q	- a	0,0274
4	N_U-04.TextGrid		q	- a	0,0379
5	N_U-05.TextGrid		q	- i	0,0416

Figure 5 Fichier de sortie représentant le calcul de la durée des phonèmes**C-Analyses et Discussions****C-1 Analyses**

Il existe plusieurs niveaux d'analyses de la prosodie et dans chacun de ces niveaux il y a des tâches spécifiques qui révèlent d'avantage sur le niveau phonétique et phonologique.

La prosodie d'une analyse est formée de deux systèmes : le système accentuel qui participe à la constitution du rythme et le système mélodique qui va gérer le ton d'un mot ou l'intonation d'une énoncée. L'accent appartient à la prosodie lexicale, lorsque ses fonctions s'exercent dans le cadre du mot ou du morphème et à la prosodie supra lexicale lorsqu'ils s'exercent au-delà du mot et donc il faut distinguer entre l'accentuation lexicale et l'accentuation rythmique.

Le rythme et l'accentuation sont intimement liés dans le sens où le rythme se constitue sur l'accent. Cependant pour qu'il y ait un rythme dans une énoncée quelconque ; il faut distinguer les temps forts des temps faibles.

Et dans la limite du corpus recueilli ; nous avons sélectionné quelques énoncés que nous avons tenté de segmenter en groupes intonatifs et mots prosodiques sur la base du nombre de syllabes dont ils sont constitués du moment que la syllabe est considérée comme étant le noyau dur de la prosodie. Puis, nous avons essayé de mettre en relief les syntagmes proéminents qui se distinguent.

C-2-Discussions

Parler d'intonation dans une langue ou un dialecte est très difficile dans le sens où cette dernière est universelle à tous les dialectes et langues et en même temps avec des spécificités particulières et ceci s'applique sur tous les langages humains. Mais si en plus aucune étude prosodique préalable n'avait été faite pour nous éclaircir (comme c'est le cas des dialectes Algérien et particulièrement en ce qui concerne le parler régional de Tlemcen) la difficulté qui s'annonçait était encore plus importante. Néanmoins, nous avons pu soustraire des informations qui nous ont paru assez intéressantes pour des recherches initiales orientées dans ce cadre.

Les résultats d'analyses prosodiques des enregistrements recueillis ont été classés comme suit :

➤ Sur le plan matériel

Nous avons constaté que la durée des voyelles et consonnes adjacentes était variables dans l'environnement des phonèmes en question. Cette variabilité concerne particulièrement le [u] et le [a] qui étaient réalisées en position finale sous la forme transcrite de [ə] .

Ceci va faire l'objet d'une comparaison plus intense dans nos prochaines études doctorales en exposant les sujets à un éventail de discours plus varié et surtout en utilisant des outils de perception plus pointu.

➤ Sur le plan lexical

Concernant l'accentuation, nous avons constaté que le degré de proéminence de l'accent varie : l'accentuation du groupe prosodique ne correspond pas à l'accentuation d'un seul terme ceci veut dire que l'accent rythmique initial appuie l'accent final.

Concernant l'intonation, Le degré de l'intonation varie autour d'une syllabe ou dans un environnement de syllabes successives elle dépend et est varié entre les locuteurs selon le style et les tempéraments de nos participants.

Conclusion

Etant un outil d'analyse intégré, le script élaboré résout un dilemme important entre analyse linguistique et prosodique et permet une approche comparative et systématique en plus d'un traitement suprasegmental approfondi du corpus recueilli. Il minimise aussi le temps d'analyse en automatisant les tâches qui ne requièrent pas un jugement humain et facilite l'intervention humaine qui peut être sujette à l'erreur en délivrant des résultats de haute précision et permettant une analyse prosodique pointue. Il s'est avéré être donc très utile pour notre annotation automatique à l'échelle prosodique d'un large corpus. Un gain de temps considérable, qui nous a plus permis de focaliser sur une analyse intégrale et d'éviter de perdre trop de temps dans le traitement manuel.

L'analyse prosodique nous a démontré qu'une étude approfondie dans le cadre de la prosodie peut être effectuée sur les dialectes régionaux. Ce que nous avons constaté est que l'accentuation du groupe prosodique ne correspond pas à l'accentuation d'un seul terme ceci veut dire que l'accent rythmique initial appuie l'accent final. L'intonation varie selon le nombre et l'environnement de la ou des syllabes et d'après le style utilisé par les participants et leur tempérament.

La durée des voyelles et consonnes adjacentes étaient variables, dans l'environnement de [q] : le [u] et le [a] elles étaient le plus souvent réduites en [ə] en position finale.

En résumé, nous pouvons conclure que l'analyse prosodique de notre corpus nous a conduit au résultat suivant :

- L'accentuation du groupe prosodique ne correspond pas à l'accentuation d'un seul terme ceci veut dire que l'accent rythmique initial appuie l'accent final.
- L'intonation varie selon le nombre et l'environnement de la ou des syllabes et d'après le style utilisé par les participants et leur tempérament.
- la durée des voyelles et consonnes adjacentes étaient variables, dans l'environnement de [q] : le [u] et le [a] sont réduites en [ə] en position finale.

Références bibliographiques

[1] Bel,B et Blache ,P., Le centre de Ressources pour la description de l'Oral (CRDO) « Travaux interdisciplinaires du laboratoire parole et langage d'Aix en Provence (TIPA) »,25 :13-18,2006.

[2]Bruce ,G et Touati,P., « On the analysis of prosody in spontaneous speech with exemplification from Swedish and French,Speech Communication 11 :453-458,1992.

[3] Dicristo,A., « La prosodie de la parole » :3-13,deboeck ,2012.

[4]Hirst,D.J et Espesser ,R., « Automatic modelling of fundamental frequency using a quadratic spline function »Travaux de l'institut de phonétique d'Aix ,15-75,1993.

[5]Hirst,D.J « A praat plugin for Momel and INSTINT with improved algorithms for modelling and coding intonation ».In proceedings of the XVth International Conference of Phonetic Sciences 1233-1236,Saabrucken ,2007.

[6] Hirst,D.J et Dicristo,A., «A survey of intonation systems»,1-44 Cambridge University,1998.

[7] Jacques,M.P « Pourquoi la linguistique de Corpus », « La linguistique de Corpus »Presse Universitaire de Rennes :21-30,2005.

Liens électroniques

[1] <http://www.fon.hum.uva.nl/praat/>

[2] <http://latlcui.unige.ch/phonetique/easyalign/tutorielpraat.pdf>

[3] http://icar.univ-lyon2.fr/projets/corinte/documents/PRAAT_BalthasarValero.pdf

[4] <http://llacan.vjf.cnrs.fr/fichiers/manuels/Praat/PRAAT%20TutorialFR.pdf>

[5] <http://www.lacheret.com/Xinha/UPLOAD/TP-PRAAT-2011.pdf>