

Projet Phonologie du Français Contemporain (PFC) Manuel d'utilisation du Comparateur Octobre 2002

Abderrahim Meqqori et Jacques Durand

Introduction

Le but de ce manuel est de décrire la Version 2 du Comparateur mis au point par Abderrahim Meqqori pour le projet PFC (Phonologie du Français Contemporain). Cet outil, comme son nom l'indique, permet de comparer les réalisations des mots de la liste PFC par différents locuteurs. Le logiciel intègre un outil de recherche permettant une comparaison en fonction de divers paramètres tels que l'âge, la localité ou la profession des locuteurs.

Dans sa version actuelle, l'outil est limité à la liste PFC stricto sensu (autrement dit les 94 mots précédés du chiffre de référence).

Le Comparateur est composé de trois utilitaires : le *Séquenceur*, le *Réglagefin*, et le *Compareur* lui-même. Attention, ces utilitaires requièrent Quicktime pour fonctionner. Cet outil est gratuitement téléchargeable sur le site www.apple.com.

1. Préparation des données

Pour utiliser le Comparateur, il faut en premier lieu créer un dossier dans lequel il faut copier les trois utilitaires, un fichier texte de la liste de mots sous forme brute (exemple ci-dessous) où chaque chiffre est suivi d'un point, d'un ou plusieurs espace(s) et du mot

1. roc
2. rat
3. jeune
4. mal
5. ras

Ce fichier (fourni par commodité dans le kit du Comparateur) ne doit pas comporter de titre ou d'annotations. Il doit être enregistré au format texte avec l'extension .txt.

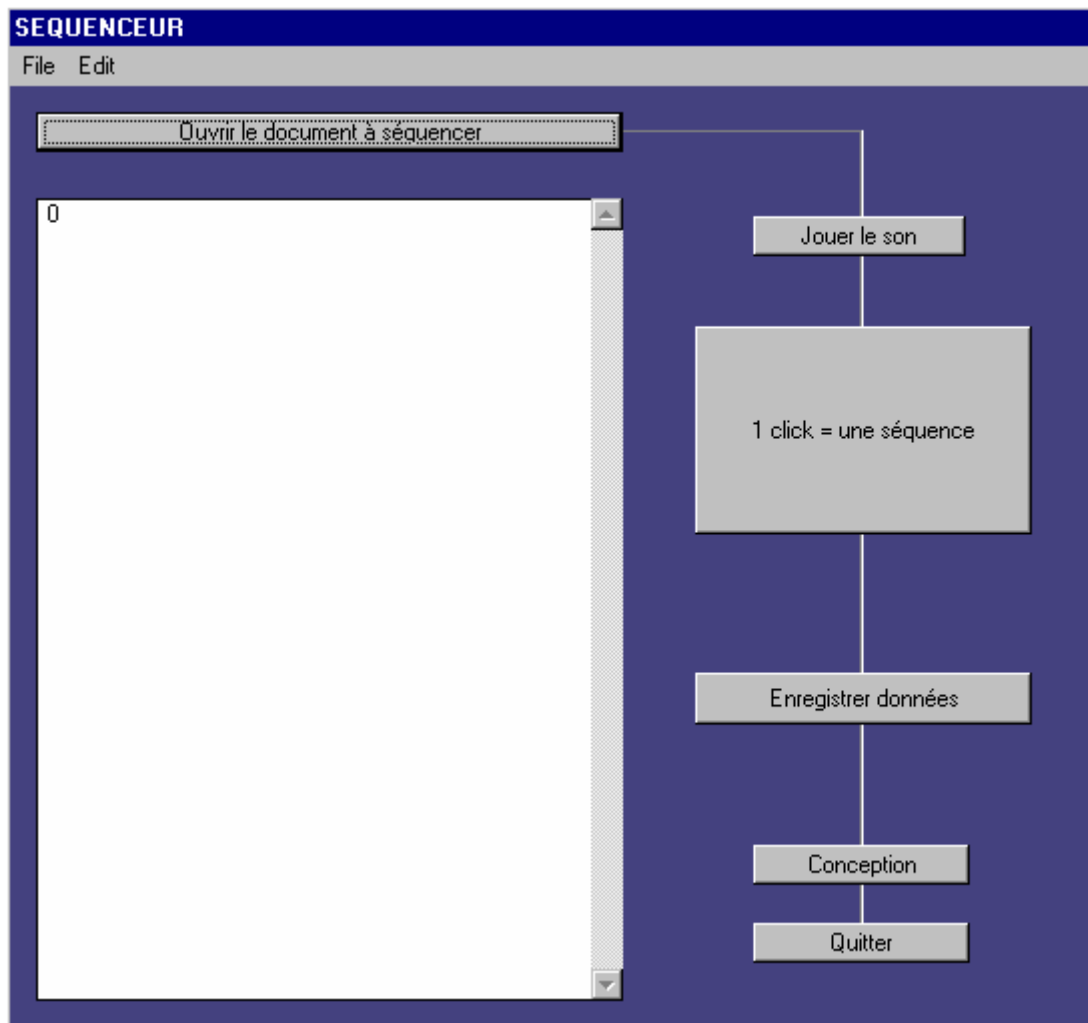
Une fois décidé le nombre de locuteurs à traiter, on copiera dans le dossier tous les fichiers .wav correspondant à la liste de mots lue par chacun des locuteurs sélectionnés. Aucune transformation préalable n'est nécessaire puisque nos utilitaires permettent de négliger tout ce qui est jugé extérieur à la liste proprement dite (introduction, commentaires, bruits, etc.).

Il faut créer un fichier contenant minimalement le nom des fichiers wav sans l'extension. Ce fichier s'appellera base.txt et contiendra jusqu'à huit informations par locuteur. Sont déjà prévus : nom, prénom, âge, profession, localité, et trois autres champs laissés au choix de l'utilisateur.

2. Utilisation du *Séquenceur*

Le *Compareur* fonctionne à l'aide de marqueurs temporels permettant de déterminer à quel moment (spécifié en fractions de secondes) commence un mot et à quel moment il se termine. Le rôle du *Séquenceur* est de créer un fichier texte contenant ces marqueurs temporels. Ce fichier est ce qui permet au *Compareur* d'aligner la réalisation de chaque mot et sa correspondance graphique.

A l'ouverture du *Séquenceur*, la fenêtre suivante s'affiche :



Le bouton *Ouvrir le document à séquencer* permet d'ouvrir le fichier son à traiter. En cliquant dessus, on fait apparaître la fenêtre traditionnelle d'ouverture de documents. Une fois le son ouvert, on peut l'écouter en cliquant sur le bouton *Jouer le son*.

Attention, l'opération de séquençage se fait en un seul mouvement et toute erreur trop importante peut entraîner des manipulations fastidieuses par la suite. L'opération consiste à faire jouer le son, à attendre que la personne prononce le premier mot et, juste au moment où l'énoncé du mot est fini, à cliquer sur le bouton *1 click = une séquence*. Le premier click détermine la fin du premier mot sachant que le début est réglé à zéro. S'il y a trop de commentaires entre le moment zéro et le premier mot, ne pas s'en préoccuper dans un premier temps car c'est ici que le *Réglagefin* interviendra. Ensuite, il faut cliquer sur le bouton *1 click = une séquence* à chaque fois que l'enquêté marque une pause entre deux mots jusqu'à la fin de la liste.

A chaque fois qu'on clique sur le bouton *1 click = une séquence*, un nombre s'ajoute au champ principal, c'est la fin du mot prononcé et le début du mot suivant.

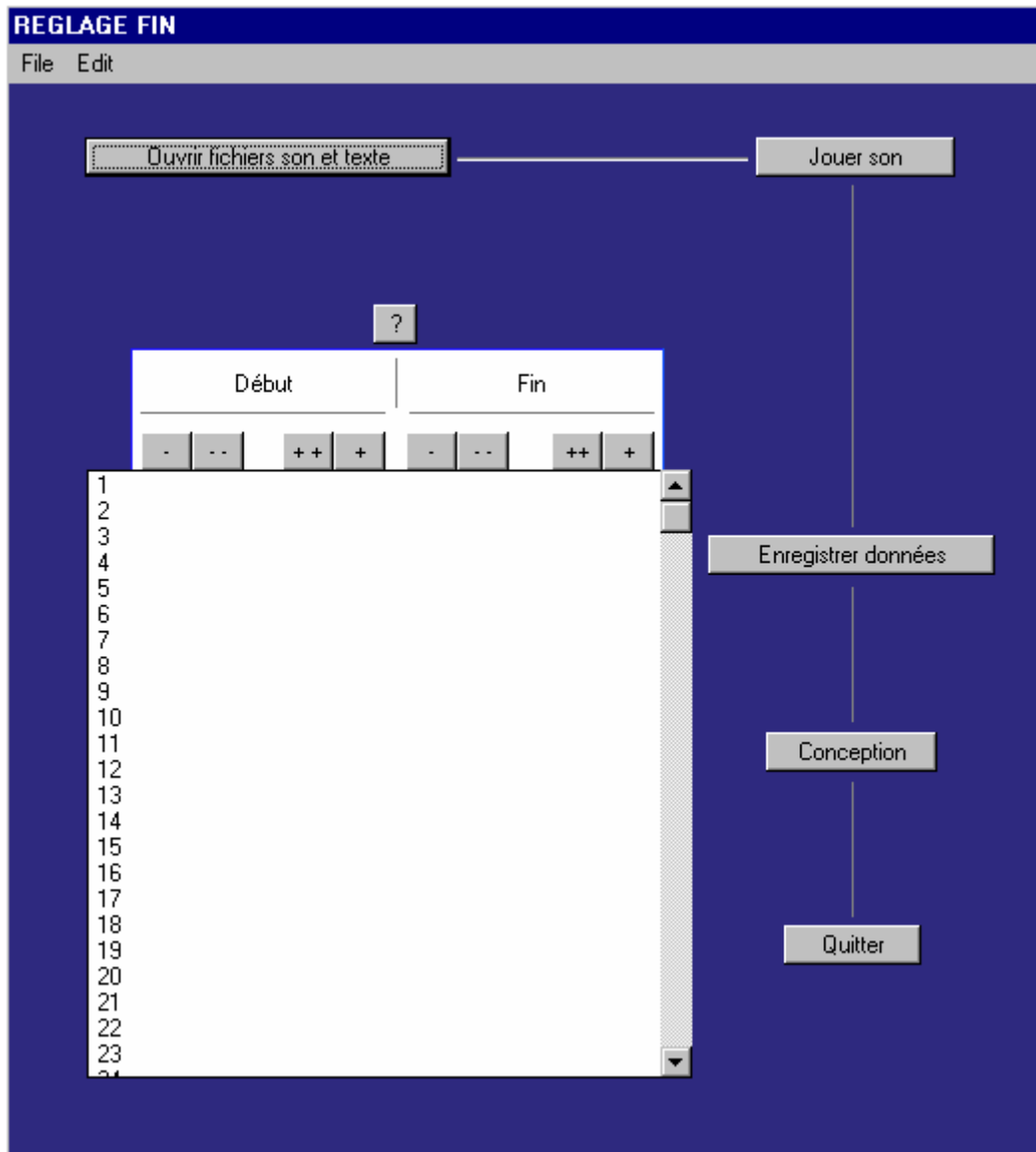
Une fois la liste terminée, sauvegarder ce travail à l'aide du bouton *Enregistrer les données*. Ce dernier permet de faire apparaître la traditionnelle fenêtre d'enregistrement de documents. Il faut donner un nom au document : ce nom devra être le même que celui du fichier son mise à part l'extension : par exemple, loc1.txt pour loc1.wav, loc2.txt pour

loc2.wav, et ainsi de suite. On l'enregistrera dans le même dossier. Le rôle du séquenceur s'arrête à ce stade, on peut donc le quitter.

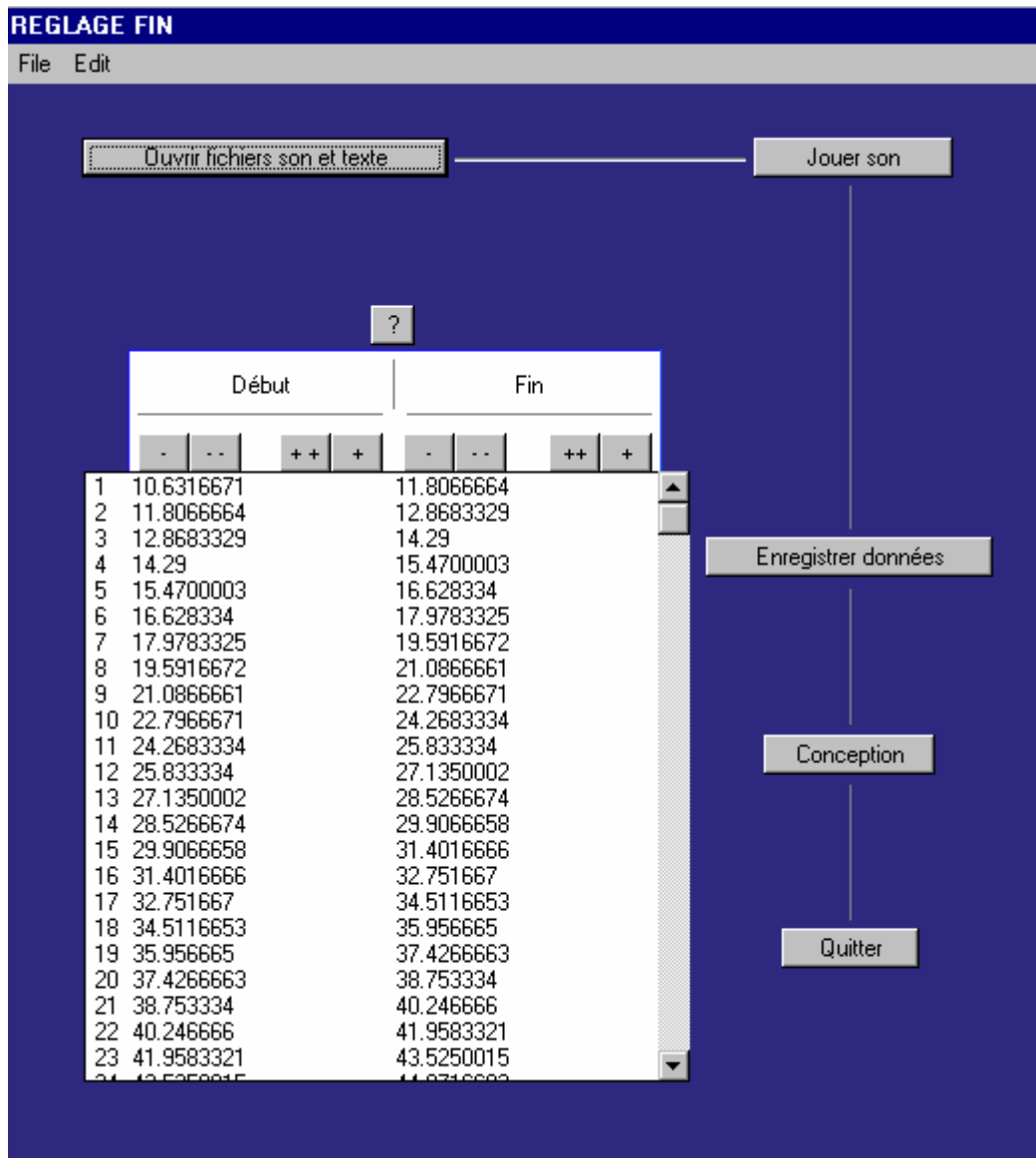
3. Utilisation du *Réglagefin*

Le *Réglagefin* permet de corriger les petites erreurs commises lors du séquençage et aussi de « cerner » au plus près les mots prononcés de manière à éviter les attentes dues aux silences précédant ou suivant les réalisations.

A son ouverture, la fenêtre suivante s'affiche :



Cliquer sur le bouton *Ouvrir fichiers son et texte*. Apparaît alors la traditionnelle fenêtre d'ouverture de documents. Attention, il faudra fournir deux fichiers l'un après l'autre en commençant impérativement par le fichier son. Lorsque la première fenêtre s'ouvre, il faut spécifier le fichier son à traiter : par exemple, loc1.wav. La deuxième fenêtre s'ouvre alors automatiquement et là c'est le fichier créé par le séquenceur qu'il faut sélectionner (à savoir, pour loc1.wav, loc1.txt, pour loc2.wav, loc2.txt etc.). La fenêtre qui s'affiche aura l'allure suivante :



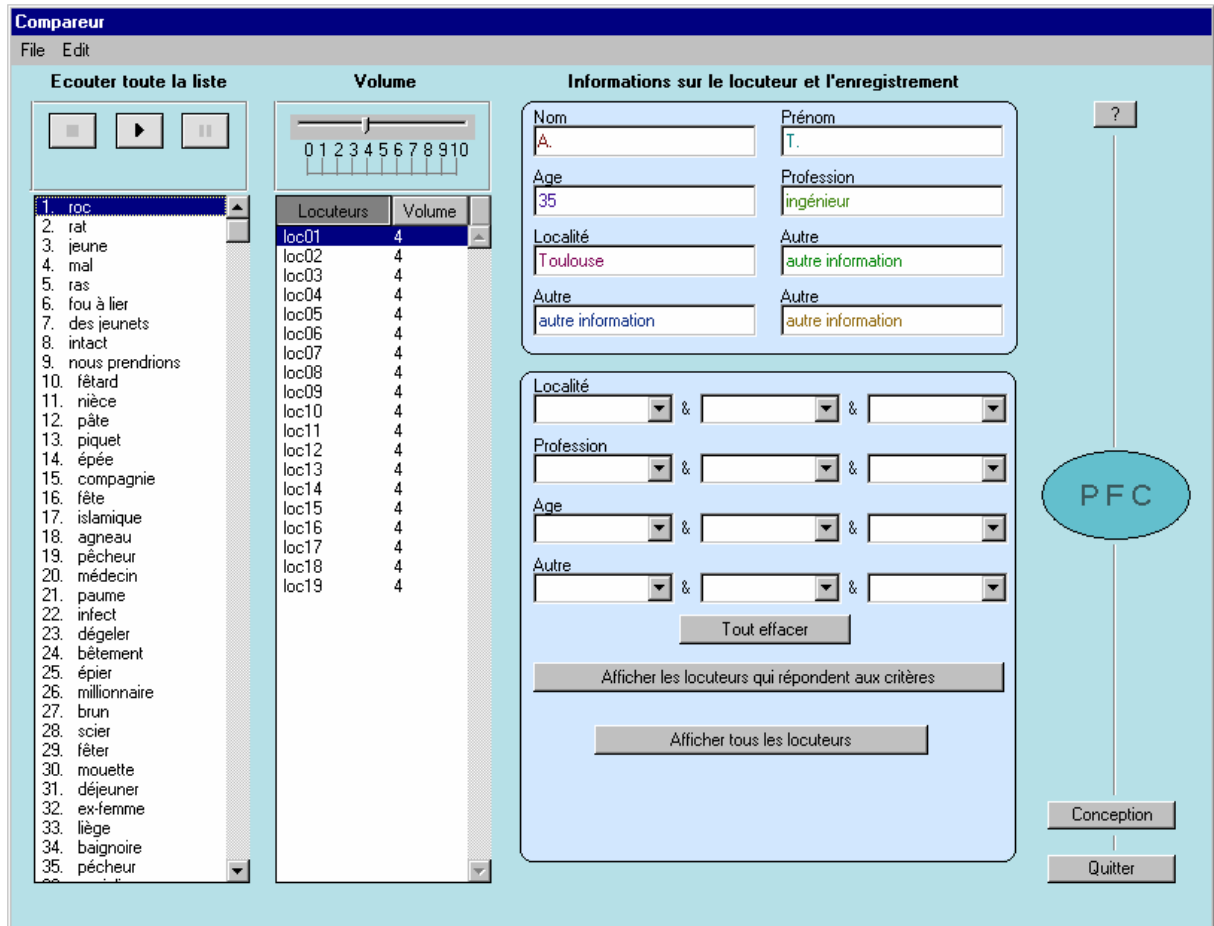
Dans le champ principal, pour chaque mot de la liste le *Réglagefin* spécifie un début et une fin en secondes. Le travail de réglage consiste à écouter chaque séquence et à ajuster les valeurs si besoin est. Pour ce faire, nous disposons de quatre boutons pour le début et de quatre boutons pour la fin. Ces derniers permettent de modifier positivement ou négativement les valeurs des frontières. Par exemple, si, au lieu d'entendre « quatre mal », on entend « tre mal », il est évident qu'on a raté dans le séquençage le début du chiffre quatre. Il faut alors soustraire de la première valeur quelques fractions de secondes. Pour ce faire, nous disposons d'un bouton « - » qui permet de retrancher à chaque clic 0.01 seconde et d'un bouton « -- » qui permet de retrancher 0.1 seconde. En revanche, si on veut allonger la fin d'une séquence, on utilisera les boutons + et ++ qui ont l'effet inverse.

4. Le comparateur proprement dit

Avant l'ouverture du programme, s'assurer de la présence des fichiers suivants dans le dossier :

- listemots.txt
- base.txt (voir §5 pour sa constitution)
- au moins un fichier .wav et le fichier .txt correspondant

A l'ouverture du comparateur, s'affiche la fenêtre suivante :



Le premier champ à gauche reproduit l'ensemble des mots de la liste qui a été saisie dans le fichier *listemots* constitué au départ. Le deuxième champ fournit la liste classée du nombre de locuteurs en fonction du fichier *base.txt* et des fichiers .wav présents dans le dossier de travail. Surplombant ce deuxième champ, on peut observer une barre de réglage qui permet d'ajuster le volume sonore pour chaque locuteur individuellement. La valeur de ce réglage qui est à 4 par défaut apparaît en face de chaque locuteur. Un ajustement individualisé peut se révéler nécessaire dans la mesure où tous les enregistrements sont disparates du point de vue acoustique. Des techniques existent cependant pour normaliser les enregistrements. Si les fichiers son ont préalablement bénéficié de telles techniques, on pourra s'éviter de telles manipulations.

La fonction du Comparateur est double. Il permet en premier lieu une écoute et une comparaison de la prononciation de mots par différents locuteurs. Une fois un mot sélectionné, un simple clic sur tel ou tel locuteur permet de l'écouter. Cet outil ne peut pas se substituer à une étude acoustique mais, d'un point de vue auditif, offre un moyen de comparaison rapide et confortable. En deuxième lieu, il permet, pour un locuteur donné, d'écouter individuellement ou en sous-groupes les mots de la liste. Une fois un locuteur sélectionné, un simple clic sur tel mot ou ensemble de mots permet d'écouter les

réalisations. Ceci est avantageux dans la mesure où les fichiers .wav de la liste de mots n'ont pas obligatoirement de fichiers textgrid associés.

Pour avoir des conditions d'écoute optimale, notre expérience suggère d'identifier le locuteur pour lequel le volume est le plus haut, d'ajuster le volume des haut-parleurs externes sur ce dernier et ensuite de monter individuellement le volume des autres locuteurs.

Information sur les locuteurs et moteur de recherche

La partie supérieure droite de la fenêtre présente les informations issues du fichier base.txt : nom, prénom, âge, profession, localité, et trois autres champs « autre information ».

Immédiatement sous ces informations, l'utilisateur dispose d'un moteur de recherche multicritères permettant de sélectionner des sous-groupes de locuteurs. Par exemple, on peut décider de ne travailler que sur les locuteurs toulousains de moins de 50 ans ou de sélectionner tous les locuteurs entre 36 et 45 ans et venant soit d'Albi soit de Toulouse.

5. Constitution du fichier base.txt

Le fichier base texte doit avoir la structure suivante, sachant que seule la première information (le nom des fichiers wav) est absolument indispensable au fonctionnement du programme :

```
loc01@A.@T.@35@ingénieur@Toulouse@autre information@autre information@autre information
loc02@B.@U.@25@étudiant@Bordeaux@autre information@autre information@autre information
loc03@C.@V.@45@commerçant@Toulouse@autre information@autre information@autre information
loc04@D.@W.@48@institutrice@Albi@autre information@autre information@autre information
loc05@E.@X.@35@ouvrier@Pézenas@autre information@autre information@autre information
```

ATTENTION, dans la version actuelle, pas de retour chariot en fin de fichier.

6. Améliorations prévues pour les futures versions du comparateur

Une version ultérieure, disons V3, est envisagée. Elle permettra notamment de comparer les mots de la liste et ceux du texte. Par la suite, le Comparateur sera intégrée dans une interface globale faisant office de base de données pour PFC.

20 janvier 2002