

# MANUEL D'UTILISATION DU 'PRÉCODEUR SCHWA'

Version 2.0

**Philippe Hambye, Jacques Durand, Julien Eychenne et Géraldine-M Mallet**

## 0. Introduction

Le précodeur schwa est un outil PFC créé et développé par Cyril Auran (Laboratoire Parole et Langage/CNRS, Université de Provence) et Julien Eychenne (ERSS/CNRS, Toulouse-Le Mirail). Cet outil vise à faciliter la tâche liée au codage des schwas potentiels dans les transcriptions PFC.

Ce précodeur permet un premier balisage des transcriptions de manière à gagner du temps lors de la saisie des codes, en précodant les sites de schwas potentiels (type de site, contextes gauche et droit). Cependant, il va de soi que le travail du codeur reste fondamental. Vu que l'outil commet inévitablement des erreurs (essentiellement dues à ses « connaissances » limitées du lexique), le codeur doit rester très attentif durant la phase de codage.

Un des avantages indéniables du précodeur est de permettre d'uniformiser les codages, surtout lorsque les conventions sont susceptibles de créer des disparités dans les pratiques.

Dans ce manuel, nous présenterons la version 2, qui est intégrée dans praat et ne fonctionne que sous Windows, ainsi que la version 1.1, qui fonctionne sur Windows et Linux. Aucun développement pour Mac n'est prévu.

## 1. Installation

Décompressez l'archive *precodeur.zip* (par exemple avec Winzip ou UltimateZip). Ceci créera un dossier « *precodeur* » contenant les fichiers suivants :

- *call\_pre-schwa\_init.praat*
- *configure\_precodeur.pl*
- *precodeur.pl*
- *lexique.pfc*
- *precodeur-tk.pl*
- *precodeur.bat*

Copiez ces fichiers dans le dossier « *outilsPFC* » que vous aurez préalablement créé à partir de la racine du disque dur (C:\) – le fichier ZIP et le dossier « *precodeur* » peuvent ensuite être supprimés. Nous rappelons que l'installation des autres outils PFC se fait au sein de ce dossier et suppose également l'installation de la version ActivePerl de Perl (voir Manuel du classeur-codeur).

Si vous souhaitez utiliser la version 1 du précodeur, il suffit d'utiliser le fichier *precodeur.bat*, qui peut être utilisé comme tout programme exécutable Windows (il suffit de double-cliquer dessus). Par commodité, on pourra décider de créer un raccourci sur le bureau.

Les instructions d'installation qui suivent ne concernent que la version 2 pour Windows. Tout d'abord, vous devez vous assurer de ce que Praat est fermé, sans quoi vous devrez réitérer la procédure intégralement. Ouvrez une fenêtre DOS (Programmes > Accessoires > Invite de commandes), et tapez les commandes suivantes (suivies de la touche [Entree]) :

*cd C:\outilsPFC*

*C:\Perl\bin\perl.exe configure\_precodeur.pl*

Le message suivant apparaît :

*Merci de patienter pendant que Praat est en train d'être configuré sur votre PC Windows.  
(ceci peut prendre quelques minutes).*

*Praat a été configuré pour lancer le pré-codage schwa chaque fois qu'un textgrid est chargé.*

*Appuyez sur [Entree] pour sortir.*

Après avoir appuyé sur la touche [Entree], le programme vous rend la main. Vous pouvez alors fermer la fenêtre DOS et utiliser le précodeur v2 depuis Praat.

## **2. Utilisation du programme**

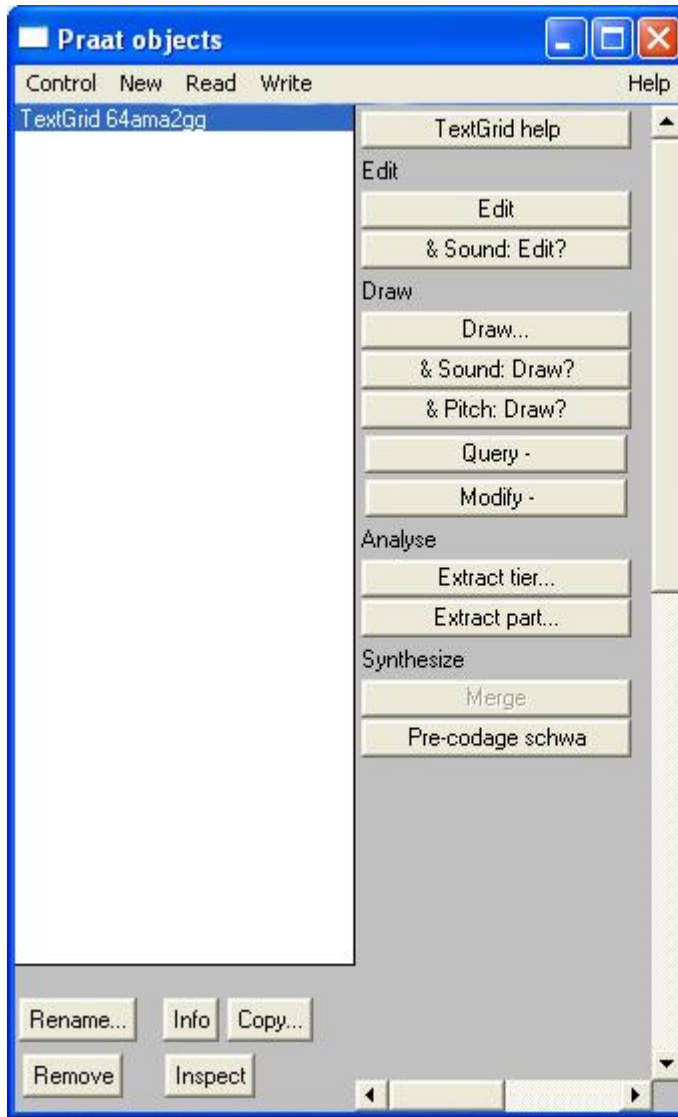
Avant d'utiliser le précodeur proprement dit, vous devez recueillir quelques informations à propos du fichier de transcription TextGrid à coder. Ouvrez celui-ci dans Praat. Repérez et notez à part le nom codé de chaque intervenant dans la transcription à coder. Ces locuteurs peuvent s'appeler AB, EADE, 11acg1 ou E, par exemple. Ne traitez pas les deux points comme faisant partie du nom d'un locuteur, et respectez la casse.

Ensuite, notez les bornes temporelles (correspondant à une *boundary* du TextGrid) de la portion de transcription que vous souhaitez coder (pour rappel 3 minutes sur les 5 minutes de transcription). Si vous utilisez la version 1, vous devez relever ces bornes avec une précision correspondant au centième de seconde, c'est-à-dire au deuxième chiffre après la virgule, sans jamais arrondir (ex. de 291,51 à 585,48).

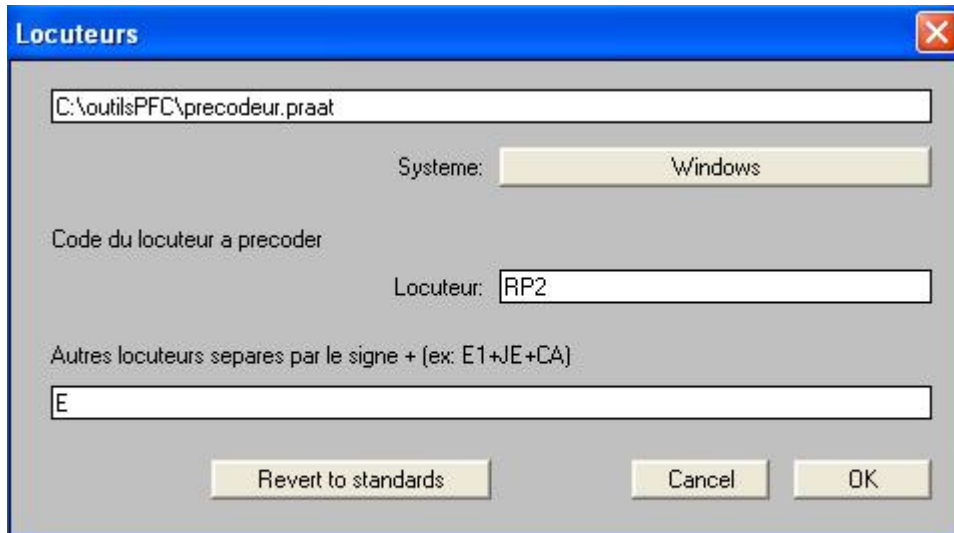
Enfin, si ce n'est déjà fait, copiez la tire de transcription en position 2 (fonction « duplicate tier » ; voir conventions établies dans le cadre du projet PFC).

Version 2

Ouvrez Praat et chargez un TextGrid comme à l'accoutumée. Un nouveau bouton nommé « précodage schwa » apparaît, comme dans la fenêtre suivante :



En cliquant sur celui-ci, une nouvelle fenêtre apparaît :



Renseignez les champs « Locuteur » et « Autres locuteurs séparés par le signe + » comme dans l'exemple ci-dessus. Si ce dernier champ contient effectivement plusieurs locuteurs, on veillera à ne pas insérer d'espace entre chaque locuteur (soit E1+JE et non E1 + JE). Cliquez sur « OK ». Cette action ouvrira le textgrid dans Praat, ainsi que la fenêtre suivante :



Cliquez sur le début de la portion à précoder et appuyez sur « Continue ». Réitérez l'opération pour la fin de la portion à précoder. Une fois le TextGrid précodé, une dernière fenêtre apparaît, demandant de sauvegarder le fichier précodé. Le fichier TextGrid original n'a pas été modifié. Le logiciel a créé un nouveau fichier TextGrid nommé sur la base du fichier original : le fichier « essai.TextGrid », par exemple, devient « essai-preschwa.TextGrid ».

### 3. Remarques sur le précodage

Les résultats du précodage sont les suivants :

- les sites de schwa potentiel sont tous précodés (sauf erreur) avec un '?' au début du code. Bien entendu, il incombe au codeur de déterminer si le schwa est réalisé ou non ;
- dans les suites de schwa ou dans certains cas de segments susceptibles de chuter, d'autres signes '?' reflètent l'impossibilité pour le précodeur de déterminer le contexte gauche ou droit du schwa. Ainsi :

*la femme de ce type // sera précodé la femme?412 de?1?2 ce?1?2 type?413 ;  
de même, parce que tu es sera précodé « parce?3?2 que?4?2 tu es »*

Notez cependant que, le précodeur ne pouvant tenir compte de toutes les variétés de français, tous les cas de figure n'ont pu être envisagés. A titre d'exemple, *c'est autre chose* sera précodé « c'est autre?422 chose?412 », en présupposant que le groupe /tr/ est réalisé ;

dans *c'est un fait*, on ne trouvera pas de codage après *fait*, et il faudra donc l'ajouter au cas où ce segment est réalisé (de même dans *en fait*).

Le précodage permet donc un gain de temps précieux mais il ne diminue pas l'attention que le codeur doit porter sur son travail. Afin de prévenir les erreurs, il devra porter une attention toute particulière aux cas de figure suivants :

- l'item *il* est toujours précodé, alors qu'il ne doit pas être codé quand la liquide n'est pas prononcée : *il?411 y avait des gens*. Il incombe donc au codeur de supprimer les codages superflus ;
- certains mots sont systématiquement précodés (ou non), alors qu'ils se terminent tantôt par une consonne, tantôt par une voyelle : *plus* (toujours codé), *fait* (jamais codé), etc. ;
- les commentaires et mots entre parenthèses sont ignorés par le précodeur, qui va chercher le mot précédent ou suivant en fonction du contexte ;
- les virgules notant une répétition sont analysées comme des frontières intonatives faibles et les mots qui les précèdent sont précodés selon cette convention ;
- certains mots ne sont pas précodés car ils ne sont pas connus de la base lexicale sur laquelle repose le précodeur : *une?411 association à Abbeville* ;
- tous les éléments des mots composés sont précodés comme des mots isolés (*là-de?212dans* ; *quatre?422-vingts*), ce qui doit être modifié manuellement dans le codage.

Bon codage...