GUIDE DU TRANSCRIPTEUR ET DU RELECTEUR DES ESLOs

Version 4

Mai 2013

Table des matières

Table des figures

Présentation

I. Objectifs du guide

Dans ce guide, vous trouverez l'ensemble des informations pratiques sur l'utilisation du logiciel d'aide à la transcription *Transcriber*, les conventions de transcription utilisées ainsi que la description des étapes à suivre pour faire une transcription et une relecture des enregistrements des ESLOs. Ce guide doit vous servir de référence ; nous vous conseillons de l'avoir à portée de main.

Par ailleurs, ce guide est accompagné d'un document Lexique-Eslo qui recense un certain nombre de graphies particulières, de difficultés orthographiques, etc. Ce Lexique, fourni en version électronique, est mis à jour régulièrement. Dès qu'une modification est apportée, nous vous envoyons la dernière version par mail.

II. Quelques mots sur les ESLOs

ESLO1, Enquête Socio-Linguistique à Orléans a été conduite à partir de 1968 par des universitaires britanniques avec une visée didactique : l'enseignement du français langue étrangère dans le système public d'éducation anglais. Elle comprend environ 200 interviews, toutes référencées (caractérisation sociologique des témoins, identification de l'enquêteur, date et lieu de passation de l'entretien), soit au total plus de 300 heures de parole incluant pour moitié des interviews en face à face et pour moitié une gamme d'enregistrements variés (conversations téléphoniques, réunions publiques, transactions commerciales, repas de famille, entretiens médico-pédagogiques, etc.).

En partant des acquis d'ESLO1, une nouvelle enquête a été mise en chantier par le LLL : ESLO2. Il s'agit, à quarante années de distance, de constituer un corpus comparable dans la produit attendu et dans les modalités de la collecte : l'objectif a été fixé à 400 heures environ de documents sonores répondant à une approche variationniste et conçus à travers des programmes spécifiques.

III. Disponibilité du corpus

L'ensemble des enregistrements et des transcriptions, mais aussi les informations concernant les conditions de recueil de l'enregistrement, les locuteurs, les transcriptions (ce que l'ont appelle les métadonnées) sont mis en base de données et rendus disponibles à l'adresse suivante : **http://eslo.tge-adonis.fr**

Ce site permet en outre de consulter les catalogues (par enregistrements, locuteurs et transcriptions), de créer des sous-corpus, de télécharger les données (sons et transcriptions) et métadonnées, de lire les transcriptions synchronisées au son, de faire des requêtes (par chaine de caractères).

Utilisation du logiciel d'aide à la transcription : *Transcriber*

I. Quelques principes sur le logiciel

L'outil choisi pour transcrire est **TRANSCRIBER** : logiciel d'aide à l'annotation de signaux de parole. Ce choix repose notamment sur son interface graphique simple permettant à un utilisateur non informaticien de segmenter des enregistrements de longue durée, de les transcrire et de marquer les tours de parole, la segmentation thématique et les conditions acoustiques, mais aussi pour sa robustesse face à de grands corpus.

II. Installation du logiciel

1. Téléchargement

Le logiciel Transcriber se télécharge sur internet à l'adresse ci-dessous. Il convient de télécharger la dernière version pour Windows, soit la version 1.5.1.

http://trans.sourceforge.net/en/presentation.php

2. Paramétrages

Un certain nombre de paramètres sont à intégrer dans Transcriber avant la première utilisation. Par la suite, il n'est plus nécessaire d'y revenir[1].

Dans l'onglet **OPTIONS > GENERAL** :

- « **Nom du transcripteur** » : mettre son nom sous la forme Prénom NOM
- « **Enregistre l'activité dans** » : écrire [**C:\time_Prenom.txt**], par exemple [C:\time_Celine.txt]. Cette procédure permettra d'évaluer le temps de transcription pour les enregistrements ESLO. Se crée alors un fichier dans C.
- « Langue » : français

ATTENTION : Une fois ces paramètres entrés dans Transcriber, pour qu'ils soient enregistrés, il faut retourner dans OPTIONS et cliquer sur « **Enregistrer la configuration** ».

[1] Il n'est plus besoin d'y revenir si vous utilisez toujours le même ordinateur. Si vous êtes amené à utiliser des ordinateurs différents, cette procédure est à appliquer à chaque ordinateur.

3. Installation du dictionnaire

Un dictionnaire interne à Transcriber est disponible. Pour le télécharger, suivre la procédure présentée ci-dessous :

- Aller sur la page : http://aspell.net/win32/
- Cliquer sur **Full installer** (Released Dec 22, 2002), ce qui permet de télécharger le fichier [Aspell-0-50-3-3-Setup.exe]. En faire l'installation.
- Revenir sur la page : http://aspell.net/win32/
- Dans la section Precompiled dictionaries, télécharger le dictionnaire français [aspell-fr-0.50-3-3.exe]. En faire l'installation.

III. Manuel d'utilisation simplifié[2]

Un manuel complet existe en version française à l'adresse suivante :

http://trans.sourceforge.net/en/transguidFR.php

Vous trouverez ci-dessous les manipulations qui nous semblent les plus utiles.

1. Créer et nommer une section (ou Report)

Les sections sont les parties de la transcription qui correspondent à une question de la trame d'entretien. Pour dénommer les sections (ou report), il faut se référer aux codages utilisés dans la trame d'entretien, chaque question de la trame correspondant à un code sous la forme : **S + chiffre** (dans l'exemple ci-dessous : S11). Dans Transcriber, les sections apparaissent sur fond rose, au milieu de la page et portent le nom « Report ».

Segmentation > Créer une section (ou CTRL r)

Catégorie : Report

Nouveau sujet (si la question n'a jamais été abordée)

Ou sélectionner le sujet dans la fenêtre

[2] Pour plus de précisions, voir le document rédigé par Jean-Yves ANTOINE.

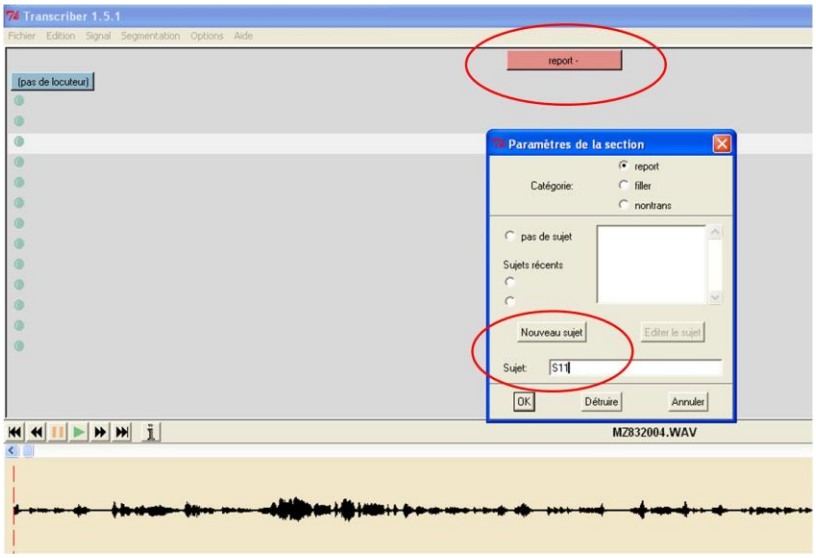

Figure 1: Créer une section (Report)

IV. Créer et nommer un tour de parole

L'ensemble de ce que dit un locuteur s'appelle un tour de parole. Un tour de parole doit donc être attribué à un locuteur. Tout locuteur (chercheur ou témoin) est codé, les codes vous seront fournis avec les fichiers sons, voir aussi la section « Identification des locuteurs » (p.21). Pour ce faire, suivre la démarche ci-dessous :

Segmentation > Créer un tour (ou CTRL t)

Créer un locuteur (si le locuteur n'a as encore été créé)

Ou choisir le locuteur dans la fenêtre

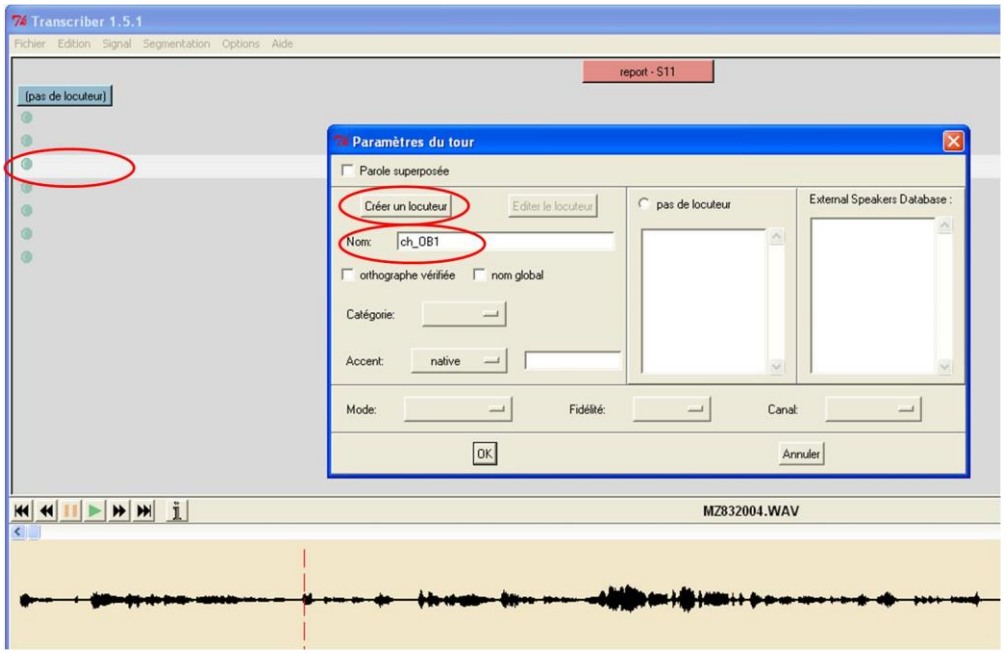

Figure 2: Créer un tour de parole

1. Créer des segments dans un tour de parole

A l'intérieur d'un tour de parole, si ce tour est long, on doit créer des segments (voir plus loin, section « Segmentation des segments »). Pour ce faire :

Taper sur « Entrée » au fil du déroulement du son.

Il se créera alors autant de points bleus à l'initiale.

2. Chevauchements

La parole entre deux locuteurs peut se chevaucher (deux locuteurs ou davantage qui parlent en même temps). Pour faire apparaitre les chevauchements (ou parole superposée), il faut caractériser les locuteurs impliqués dans le tour de parole. Dans la transcription, on verra alors apparaitre autant de lignes de transcription que nécessaire dans le tour de parole à deux locuteurs.

Voici la démarche pour créer des chevauchements, avec deux locuteurs :

Segmentation > Créer un tour (ou CTRL t)

Cocher « Parole superposée » (en haut à gauche)

Déterminer qui est le premier locuteur

Puis déterminer qui est le deuxième locuteur

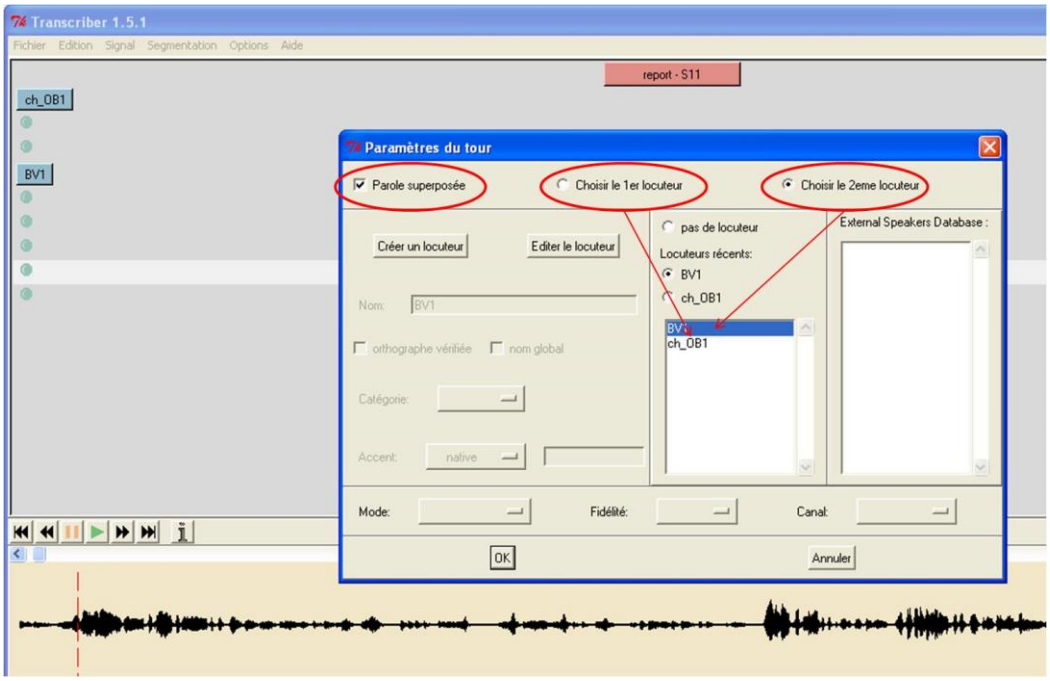

Figure 3: Créer un tour de parole avec chevauchement

Dans les cas où il y aurait plus de deux locuteurs, il faut noter en chevauchement les deux locuteurs principaux (ceux qu'on entend le mieux), et utiliser la **balise [conv]** « Conversation de fond ».

3. Raccourcis clavier

Voici quelques raccourcis claviers qui pourront vous être utiles pour les actions les plus fréquentes.

Créer un segment : **Entrée**

Play / Pause : **Tabulation**

Créer une section : **CTRL r**

Créer un tour de parole : **CTRL t**

Paramètres du tour de parole : **CTRL ALT t**

Insérer une balise : **CTRL d**

Annuler un tour de parole : **CTRL SUPPR (ou CTRL DELETE)**

Enregistrer : **CTRL s**

Procédures de transcription, relecture, validation

I. Étapes préalables à toute transcription

Dans cette section, vous trouverez les étapes à suivre avant de commencer toute transcription. Nous terminons par une partie récapitulant les documents à rendre pour chaque transcription.

♦ *Paramètres de l'épisode (= transcription)*

Avant de commencer une transcription, il y a des informations à paramétrer dans **Fichier > Paramètres de l'épisode**.

Il s'agit de deux informations :

- Prénom NOM du transcripteur
- Choisir la langue principale : français

Ainsi, pour chaque transcription, le nom du transcripteur est inscrit dans le document et peut être récupéré.

♦ *Nom du fichier de transcription*

Le fichier transcription doit être enregistré sous le même nom que celui porté par le fichier son, auquel il suffit de rajouter la version de transcription (A, B ou C). Par exemple, le fichier son se nomme [ESLO2_ENT_1001], le fichier transcription devra être appelé **[ESLO2_ENT_1001_A]**, le fichier relecture **[ESLO2_ENT_1001_B].**

♦ *Codes témoin et chercheur*

Pour les entretiens, on a généralement un témoin (interviewé) et un chercheur (le membre de l'équipe qui assure l'interview). Ces deux locuteurs sont codés, le code étant déterminé par l'inscription de l'enregistrement dans la base de données.

Le code du témoin est attribué de manière aléatoire par la base, et correspond à deux lettres en majuscule suivies d'un chiffre (ex : BV1, RL2).

Chaque chercheur a un code qu'il conserve pour toute transcription et qui prend la forme suivante : ch_Initiales du chercheur + chiffre (ex : ch_OB1, ch_CD2).

⇨ Ces codages vous seront donnés en même temps que le fichier son.

Pour les autres modules, on peut avoir une organisation différente, avec plus de locuteurs, avec des locuteurs sur lesquels on n'a pas d'informations, etc.

⇨ Nous vous donnerons systématiquement les codes des locuteurs principaux.

Il est possible que des locuteurs apparaissent au fil de l'enregistrement, alors qu'ils n'étaient pas prévus.

Plusieurs cas de figure :

- **Le locuteur a un lien de parenté/amitié** avec l'un des locuteurs de l'enregistrement, un code explicite lui est alors attribué, qui prend la forme suivante (exemple avec un témoin fictif codé IG297) :
    - Pour le **mari** du témoin IG297 > IG297MAR
    - Pour la **femme** du témoin IG297 > IG297FEM
    - Pour son **père** > IG297PER
    - Pour sa **mère** > IG297MER
    - Pour son **grand-père** > IG297GRP
    - Pour sa **grand-mère** > IG297GRM
    - Pour sa **fille** > IG297FIE
    - Pour son **fils** > IG297FIL
    - Pour sa **petite fille** > IG297PFIE
    - Pour son **petit fils** > IG297PFIL
    - Pour sa **belle fille** > IG297BFIE
    - Pour son **beau fils** > IG297BFIL
    - Pour son **ami** > IG297AMI
    - …

- **Le locuteur est un tiers n'ayant pas de lien de parenté avec le locuteur témoin mais ayant cependant un rôle dans la conversation**. Il est alors considéré comme un locuteur intervenant au cours de la conversation. Il parle et fait partie de l'intervention donc il sera noté sous la forme :
    - **numéro de l'entretien + LOC,** ex. « 083LOC »

  !!! S'il existe plusieurs intervenants, il conviendra de les différencier sous la forme → « 083LOC1 », « 083LOC2 », « 083LOC3 », …

- **Le locuteur est un tiers n'ayant pas de lien de parenté avec le locuteur témoin et qui est, de plus, une personne non sollicitée** ; on ne possède alors aucune information à leur sujet. On considère alors qu'il n'est pas nécessaire de définir le statut de la personne qui parle. Ce locuteur sera noté sous la forme :
    - **Pour ESLO1 : numéro de l'entretien + INC** (pour inconnu), ex. « 083INC »
    - **Pour ESLO2 : ENT_numéro de l'entretien + INC** (pour inconnu), ex. « ENT_29INC »

  !!! S'il existe plusieurs intervenants, il conviendra de les différencier sous la forme → « 083INC1 », « 083INC2 », « 083INC3 », …

!!! Dès lors que vous créez un code locuteur correspondant à l'un de ces trois cas particuliers, il faut l'indiquer dans le fichier « Remarques » (*voir infra*).

♦ *Première écoute de familiarisation (facultative)*

Avant de se lancer dans la transcription, écouter quelques extraits (environ 5-10 minutes) du fichier son afin de se familiariser avec les voix des locuteurs, leurs façons de parler, etc. Cette étape permet de se préparer et de mieux appréhender la phase de transcription.

♦ *Problèmes et remarques (au fil de la transcription)*

Au fil de la transcription, ou à un moment précis de la transcription, vous pouvez rencontrer des problèmes, liés ou bien à l'acoustique de l'enregistrement, à un fond sonore ou à tout autre type de difficulté. Vous pouvez avoir créé des codes locuteurs non prévus. Par ailleurs, vous pouvez avoir des remarques à faire sur une transcription, par exemple sur une attitude particulière du témoin ou du chercheur (tic de langage, accent, etc.), sur un passage de l'entretien qui vous semble particulièrement intéressant, etc.

Afin d'informer ces problèmes et remarques, vous utiliserez un document à remplir intitulé « **Fichier_Remarques** »[3]. Vous renommerez ce fichier de la façon suivante : nom du fichier de transcription suivi de « Remarques ». Par exemple : **[ESLO2_ENT_1001_A_Remarques]**.

♦ *Documents à rendre : récapitulatif*

Pour chaque transcription, deux documents devront être rendus, les deux sous format électronique :

- le fichier de transcription (Transcriber) Ex. de dénomination : **[ESLO2_ENT_1001_A]**
- le fichier « Problèmes et remarques » (facultatif) Ex. : **[ESLO2_ENT_1001_A_Remarques]**

Ces documents doivent être transmis par mail à C. Dugua (celine.dugua@univ-orleans.fr) et L. Kanaan-Caillol (layal.kanaan@univ-orleans.fr).

!!! Lors des envois par mail, indiquez dans l'objet, le nom de la transcription (dans cet exemple : [ESLO2_ENT_1001_A]).

[3] Voir Annexes

II. Qu'est-ce que transcrire, relire, valider ?

La transcription des ESLOs se fait en trois étapes : une transcription brute (version A), une relecture (version B), une validation (version C). Chaque version est réalisée par un transcripteur différent.

Transcrire revient à mettre en œuvre 4 types de tâches, qui peuvent être considérées comme quatre objectifs.

1. Les quatre tâches inhérentes à toute transcription/relecture/validation

♦ *Tâche de segmentation*

La segmentation consiste en l'alignement de la transcription avec le son. La segmentation comprend plusieurs niveaux :

- Découpage en sections ou reports, avec codage de ces sections
- Découpage en tours de parole, avec attribution de la parole aux locuteurs (usage de codes, voir « anonymisation »)
- Découpage en segments à l'intérieur des cours de parole : il peut s'agir de parole ou de temps de pauses
- Réalisation des chevauchements

♦ *Tâche d'écoute*

Transcrire passe forcément par un temps d'écoute qui consiste à réaliser la correspondance entre ce qui est dit, et ce que l'on entend (ou perçoit). Cette écoute peut être plus ou moins fine et plus ou moins problématique.

Des outils (balises) peuvent être utilisés dans les cas où le transcripteur ne comprend pas ce qui est dit ou pour signaler qu'il a des doutes sur ce qu'il entend.

♦ *Tâche de transcription (ou passage à l'écrit)*

La transcription (ou passage à l'écrit) est la mise en mots (écrits) de la parole. Elle doit respecter les règles orthographiques mais suit également un certain nombre de conventions (voir plus loin, ainsi que le Lexique-Eslo). La transcription doit alors répondre aux deux objectifs suivants :

- Respect de l'orthographe
- Respect des conventions de transcription ESLOs

♦ *Tâche d'anonymisation*

L'anonymisation consiste à rendre anonyme la transcription, elle se fait à deux niveaux :

- Anonymisation des locuteurs connus (les témoins) par le biais de codes établis locuteurs que nous vous transmettons.
- Anonymisation de certains éléments de parole :
    - Les noms de personne dont on parle et qui ne sont pas des personnalités : **NPERS**
    - Les paroles très personnelles qui pourraient porter préjudice à la personne dont on parle (passages dits « délicats ») : **NANON**

2. Répartition des tâches dans les trois versions

Globalement, toutes ces tâches sont mises en œuvre dans une transcription, quelle que soit la version (A, B ou C). Toutefois, nous proposons la procédure ci-dessous qui hiérarchise les tâches en fonction de la version. Cela permet de limiter le nombre de compétences à mettre en œuvre pour une version donnée mais aussi de s'assurer que tous les objectifs sont bien atteints.

♦ *Niveau brut - Version A*

La transcription version A est une transcription dite « brute ». Les objectifs peuvent être hiérarchisés ainsi :

1) **Segmentation** en sections et codage, en tours de parole et attribution de la parole aux locuteurs (codés), en segments, réalisation des chevauchements.
2) **Transcription brute avec conventions** : réaliser une transcription qui respecte l'orthographe française et les conventions de transcription (relire sa transcription une fois avant de la rendre). Mais ne pas affiner la transcription : pas de recherche sur dictionnaire, internet, grammaire.
3) **Anonymisation** : deux niveaux :
    - coder les locuteurs (cf codes donnés avec le son) : être très vigilant sur la typographie de ces codes. A la fin de la transcription, vérifier les codes des locuteurs dans les « Paramètres du tour ».
    - coder les noms de personne (noms de famille) dont on parle et qui ne sont pas des personnalités en utilisant **NPERS**.
4) **Écoute** : l'écoute se fait a minima. Ne pas passer du temps à réécouter des segments,, il est préférable d'utiliser les balises dans les cas de difficultés de compréhension.

!!! Dans la version A, le transcripteur ne doit pas réécouter plus de deux fois un segment.

A la fin de la version A, les éléments suivants sont considérés comme finalisés, c'est-à-dire qu'ils ne seront plus vérifiés dans les versions ultérieures :

- les emplacements et codes des Reports (sessions),
- les codes des locuteurs : typographie + attribution de la parole.

♦ *Niveau relu - Version B*

La relecture (version B) consiste à vérifier la première version de transcription sur trois points essentiellement, qui peuvent être hiérarchisés ainsi :

1) **Transcription** : vérifier le respect de l'orthographe et des conventions.

Méthode : exporter le fichier en .txt sous word et passer le correcteur orthographique. Dès qu'il y a des éléments soulignés en rouge ou vert, faire un Rechercher (CTRL F) sous *Transcriber* et remplacer par la forme adéquate.

Vous documenter : Vérifier dans le Lexique-ESLO, les dictionnaires, les grammaires, internet, manuels de conjugaison tous les mots, toutes les formes sur lesquelles vous avez un doute.

2) **Écoute** : affiner la version A en écoutant plusieurs fois les passages peu audibles ou difficilement compréhensibles.

On trouve essentiellement deux cas de problèmes liés à l'écoute :

- la transcriptrice de la version précédente n'a pas bien entendu un segment de parole (usage de la balise Pi = prononciation inintelligible) et vous entendez quelque chose de distinct
- vous entendez autre chose que ce qu'a noté la transcriptrice précédente

Dans les deux cas, vous transcrivez ce que vous entendez.

3) **Segmentation** : se limiter à vérifier les segments et surtout les chevauchements avec l'ordre des locuteurs dans les chevauchements (il s'agira surtout d'affiner les chevauchements par rapport à la version A)

> Parmi les trois objectifs de la version B, seul celui concernant la segmentation (et sa déclinaison) ne sera plus vérifié par la suite.
>
> Mais il s'agit surtout de corriger l'orthographe en vous documentant, de vérifier les conventions et d'affiner l'écoute afin de limiter les balises de non compréhension.

♦ *Niveau validé - Version C*

L'étape de validation consiste à relire la version B des transcriptions qui était déjà une relecture de la première version (version A). Après avoir testé différentes méthodes, celle qui présente le meilleur rapport temps/correction consiste à corriger directement sous *Transcriber* en écoutant le son simultanément.

La validation présente un triple objectif, hiérarchisé ainsi :

1) **Transcription** : vérifier le respect de l'orthographe et des conventions, en s'aidant des différents outils mis à disposition.
2) **Écoute** : affiner les difficultés d'écoute
3) **Anonymisation** : sur deux aspects :

- Vérifier les **NPERS** : il ne faut pas qu'il y ait eu d'oublis, et s'assurer que des noms passés en NPERS ne sont pas des « personnalités ». Le cas échéant, rétablir le nom.
- Vérifier s'il y a des passages où, sans citer de noms de famille, il peut s'avérer problématique/préjudiciable de laisser la fonction ou le prénom parce que la personne parle de faits privés (ex : le député, le maire…). Dans le cas où la transcriptrice en trouverait, il est recommandé d'en discuter avec C. Dugua ou L. Kanaan-Caillol ou O. Baude et de décider, le cas échéant, ce qu'on passe en **NANON** (cela peut être un segment long).

3. Schéma récapitulatif de la répartition des tâches dans chacune des trois versions de transcription

Le schéma synthétique ci-dessous représente la part que prennent les quatre tâches dans chacune des trois versions de transcription. Ces parts peuvent être considérées comme du « temps passé ».

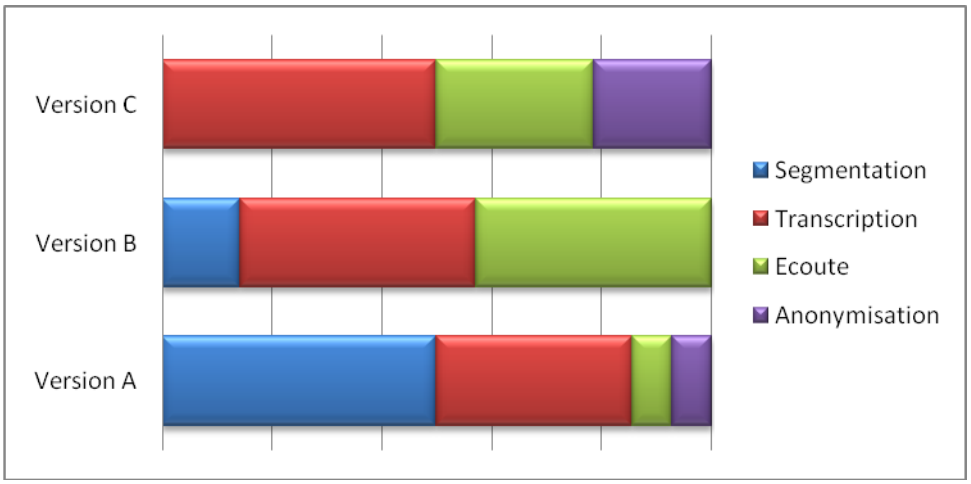

Figure 4: Schéma de la répartition des tâches en fonction de la version de transcription

Conventions de transcription

La transcription se fait selon certains principes et en suivant des conventions que nous allons préciser dans les sections suivantes.

I. Principes et règles de base des conventions de transcription

La transcription des corpus ESLO répond à deux principes, qui peuvent s'avérer, dans certains cas, contradictoires :

- le respect de l'orthographe,
- le respect de ce qui a été dit, notamment en termes de structure grammaticale.

Voici un exemple de contradiction entre ces deux principes : « les lettres que j'ai écrits ». On ne met pas la marque du féminin parce qu'on ne l'entend pas (/λελΕτ{κ↔Ζεεκ{ι/). En revanche, on met le pluriel parce qu'à l'oral on ne peut pas discriminer si le pluriel a été « mis » ou pas…

Le dictionnaire qui servira de référence est le TLFi[4], disponible en ligne :

http://atilf.atilf.fr/tlf.htm

Vous pouvez aussi utiliser le dictionnaire interne à Transcriber (**Edition > Correction orthographique**).

Ci-dessous, quatre règles de base à respecter, certains de ces éléments étant repris de manière plus détaillée par la suite :

- Ne pas utiliser la majuscule en début de phrase ni en début de tour de parole sauf pour les noms propres
- Excepté le point d'interrogation, aucun signe de ponctuation ne devra figurer dans les transcriptions. **Attention : laisser un espace devant le point d'interrogation**.
- Dans une même transcription, l'écriture des mots doit être homogène, aussi bien pour les noms propres que pour les noms communs. Ex : *et cætera* ou *et cetera* ; orthographe des noms propres
- Indiquer par le signe « & » devant un mot sans espace, tout mot dont l'orthographe n'est pas attestée dans un dictionnaire (voir partie « Graphies incertaines », p.31)

Figure 5: Principes de base des conventions de transcription

[4] Il n'est plus possible d'utiliser le Petit Robert mis en ligne par le portal de la BU pour l'instant. Si cette ressource devenait à nouveau disponible, nous la réutiliserions comme référence.

II. Principes de segmentation

1. Trois niveaux de segmentation

Trois niveaux de segmentation sont possibles avec Transcriber :

- En sections
- En locuteurs
- En segments

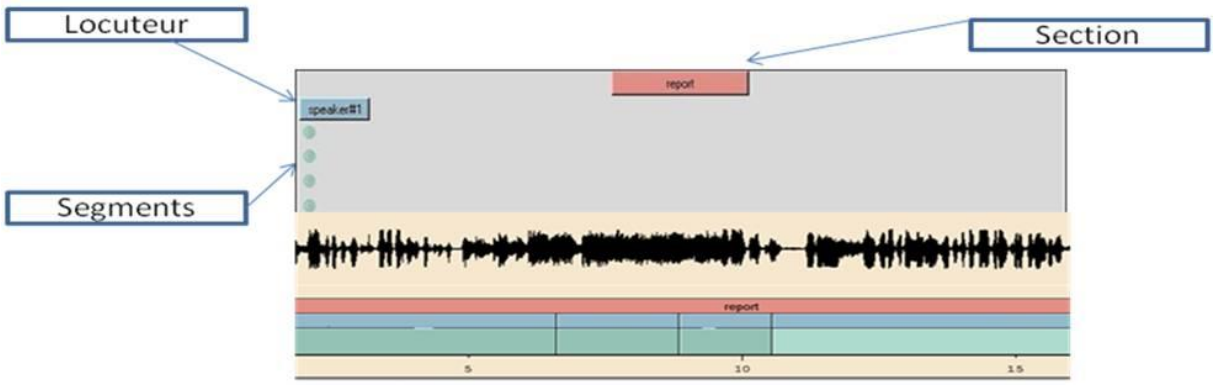

Figure 6: Trois niveaux de segmentation: section, locuteur, segment

2. Règles de segmentation

Quelques règles de segmentation seront à respecter

♦ ***Segmentation des sections :***

Chaque section (ou *Report*) a un code qui renvoie à chaque question ou thème de la trame d'entretien (voir document « Trame d'entretien ») : S12 (section 1, question 2 : Le logement). Ainsi chaque *report* dans Transcriber englobera la question et la réponse.

Exemple :

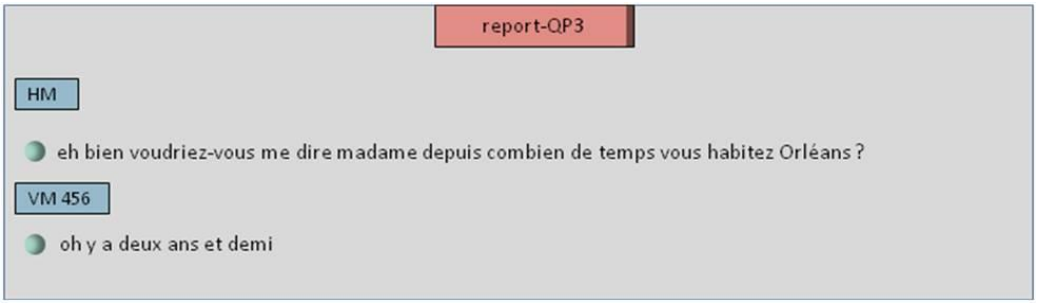

Figure 7: Exemple de section

Dans le cas où vous rencontreriez des thématiques que vous ne parvenez à rapprocher d'aucune question de la trame d'entretien, intégrez-là à la thématique précédente.

♦ *Identification des locuteurs*

Locuteurs identifiés

Nous distinguons ici ESLO1 et ESLO2.

ESLO1

Pour l'identification des locuteurs, il faut se référer au document : « Catalogue des enregistrements des enquêtes sociolinguistiques à Orléans 1968 – 71 ».

Ainsi, pour chaque enregistrement d'ESLO1 vous disposez d'un code pour le témoin (ensemble de chiffres et de lettres) et d'un code pour le chercheur (ses initiales).

Prenons pour exemple l'enregistrement numéro 013 :

- Le témoin : MD461
- Le chercheur : BV

ESLO2

Entretien

Pour les entretiens d'ESLO2, vous utiliserez les codes qui sont attribués aux locuteurs par la base de données. Le code témoin se compose de deux lettres et un chiffre, le code chercheur de la suite « ch_ » suivie des initiales du chercheur et d'un chiffre. Pour un chercheur donné, son code sera toujours le même, quelle que soit la transcription.

Par exemple, pour l'enregistrement ESLO2_ENT_1001, les locuteurs sont :

- Témoin : BV1
- Chercheur : ch_OB1

Autres modules

Pour les autres modules d'ESLO2, nous vous fournirons les codes locuteurs.

Hésitations entre deux locuteurs

Dans certains cas, il peut s'avérer difficile d'attribuer le tour de parole à l'un ou l'autre des locuteurs. Vous devrez prendre une décision, et choisir l'un des locuteurs possibles.

Locuteur non identifié

Voir paragraphe *Codes témoin et chercheur*, p. 12

Rappel : indiquez dans la fiche Remarques toutes les fois où vous créez un code de locuteur non identifié.

♦ *Segmentation des segments*

Règles de segmentation des segments

La segmentation en unités doit se faire en respectant certaines règles :

- Les segments ne doivent pas être trop longs : ils ne doivent pas dépasser 15 secondes (comme c'est le cas dans l'exemple ci-dessous pris dans ESLO1) et, dans la mesure du possible, ne doivent pas dépasser deux lignes dans Transcriber
- Veillez à avoir des unités syntaxiques et sémantiques cohérentes
- Ne pas segmenter mot à mot mais par groupe pertinent

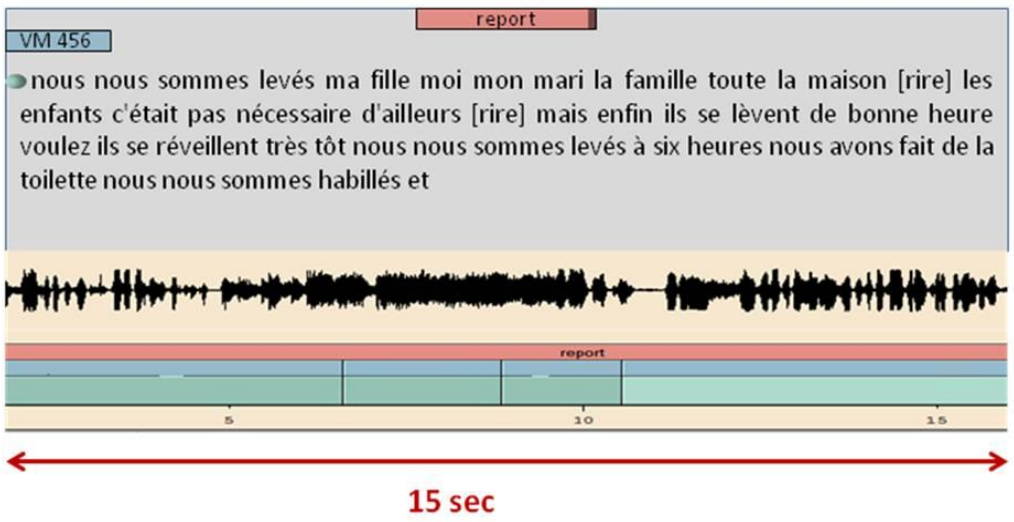

Figure 8: Exemple de tour de parole trop long

Nous lui préférerons cette segmentation :

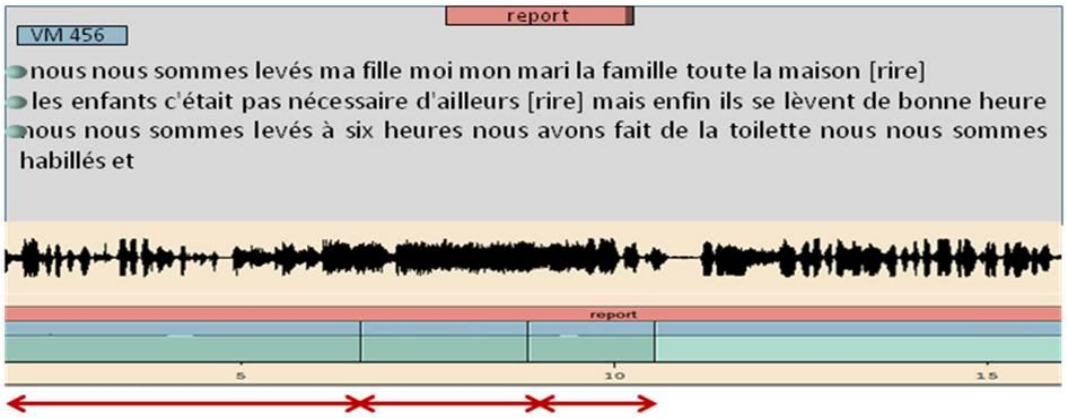

Figure 9: Exemple de tour de parole découpé en segments

Pauses

Un des phénomènes de l'oral qui nécessite une segmentation particulière est la pause, elle est à noter par un segment vide. Cette segmentation permettra d'avoir précisément la durée de la pause.

On attribue les pauses à un locuteur. Exceptionnellement seulement, lorsqu'une pause longue n'est pas attribuable à un locuteur, on met « **NO SPEAKER** » comme locuteur.

Si la pause apparait entre deux tours de parole ou entre une question et la réponse, on met la pause plutôt au premier locuteur.

♦ *Chevauchements*

Chevauchement de paroles

Lorsque deux locuteurs parlent en même temps, vous utiliserez la segmentation que propose *Transcriber* (paramètres du tour > parole superposée). Il faudra seulement mettre la portion de parole prononcée simultanément et non l'ensemble de l'intervention des locuteurs.

Figure 10: Exemple de chevauchement de paroles

Cas particulier : Chevauchement avec des marques d'acquiescement

Dans le cas de chevauchement avec des marques d'acquiescement (back chanel), on les notera sans tenir compte du moment précis de leur réalisation (sans interrompre la prise de parole). En revanche, si elles constituent un tour de parole, elles sont notées telles quelles.

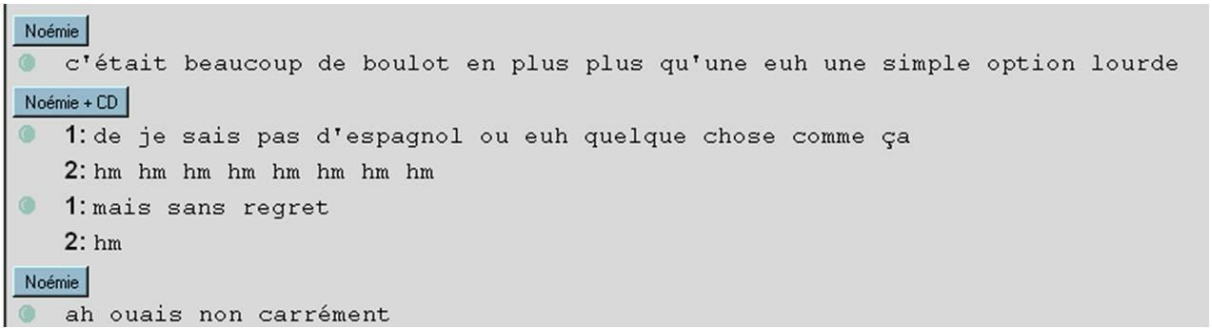

Figure 11: Exemple de chevauchements avec des marques d'acquiescement

III. Utilisation des balises « Bruit »

Les bruits sont à indiquer par les balises proposées par le logiciel. Cependant ils ne seront pas tous systématiquement notés.

1. Rires

La **balise [rire]** sera maintenue, au cours d'un tour de parole ou en tant que tour de parole.

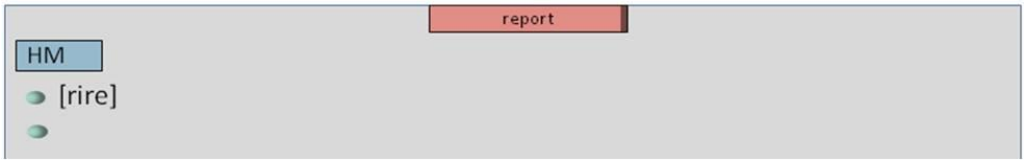

Figure 12: Exemple balise [rire]

2. Micro

Il faudra adapter la notation des bruits aux segments. Si certains bruits ont une durée trop longue ou une incidence directe sur les réalisations des locuteurs.

Par exemple, en (096/46:16), le bruit de micro est compris dans une séquence qui va de 46:12 à 46:20, alors que les quatre premières secondes sont une pause sans bruit de micro. Dans ce cas, il faudra isoler le bruit de micro dans un segment, en l'indiquant par une **balise [mic]**.

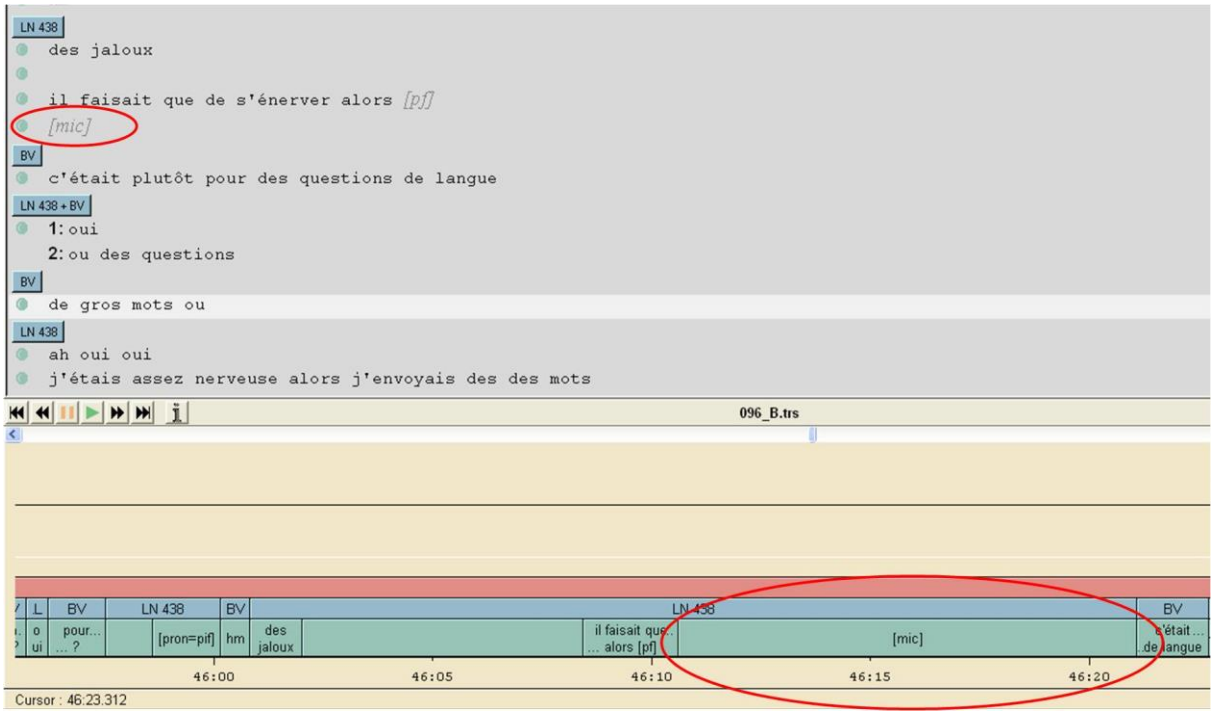

Figure 13: Exemple balise micro [mic]

3. Passages non transcrits

Dans le cadre des entretiens ESLO2, certains chercheurs commencent les enregistrements « à la sortie de la voiture ». Ces passages peuvent être plus ou moins longs et peuvent comporter de la parole. Ces passages ne seront pas transcrits.

On choisira, dans les catégories de sections, la **section [nontrans]**. On laissera sans locuteur.

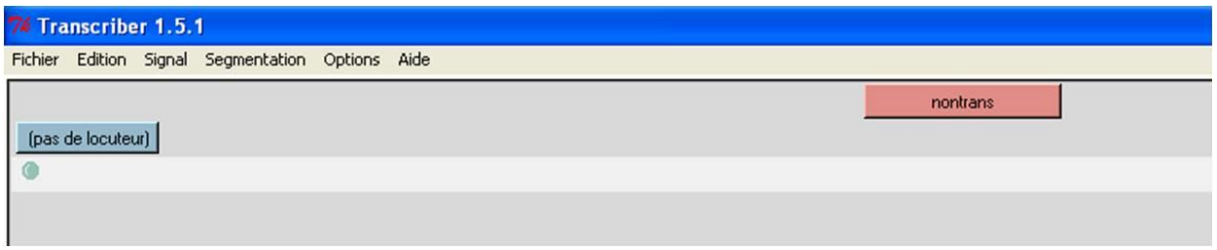

Figure 14: Exemple section [nontrans]

4. Soufflerie

♦ *Bruits de respiration*

Les bruits de respiration ne sont notés que s'ils donnent une indication sur l'expression d'un jugement :

- Inspiration marquant un embarras dans la réponse
- Expiration marquant l'agacement ou le désaccord

Les balises prévues dans Transcriber sont les suivantes :

- **Balise [i]** : inspiration
- **Balise [s]** : soupir
- **Balise [r]** : respiration

N'utiliser ces balises qu'exceptionnellement.

♦ *Clics*

Les clics, quand ils sont bien audibles, peuvent être notés par une **balise [bb]** (bruit de bouche). Quelques exemples pris dans ESLO1 :

- (109/09:31, 39:04), (110/02:07), (133/46:42), (215/07:16, 09:06)

IV. Conventions de transcription et orthographe

Rappelons que la transcription des corpus ESLO se fait en respectant deux principes :

- L'orthographe
- Ce qui a été dit

En bref, sauf s'il devait en résulter une contradiction entre ce qui a été dit et ce qu'on écrit, il faut respecter les conventions orthographiques. Voici quelques exemples :

- des nouvelles que j'ai écrits.

Attention aux difficultés de l'orthographe :

- ~~un mot croisé~~ > un mots-croisés (132/44:59)

N'hésitez pas à vous reporter au dictionnaire TLFi, aux grammaires pour résoudre les questions d'orthographe, ou à nous/vous poser la question.

1. Signes graphiques

♦ *Point d'interrogation*

Le point d'interrogation est le seul signe de ponctuation utilisé. Il sert à marquer une question, que celle-ci se réalise syntaxiquement ou uniquement par une intonation montante.

Attention : toujours faire précéder le point d'interrogation d'un espace (insertion non automatique dans *Transcriber*).

♦ *Guillemets*

Ne jamais mettre de guillemets, même dans les cas de discours rapporté.

Exemple pris dans ESLO1 (107, 19'45) :

- ~~veut dire par « on » ?~~ > veut dire par on ?

♦ *Apostrophe*

Principe d'usage de l'apostrophe

Afin de préserver la reconnaissance automatique des unités, l'apostrophe ne doit être utilisée que lorsqu'elle peut correspondre à un usage orthographique.

En conséquence, étant donné que la suite « y' » n'existe pas en français, on transcrira :

- ~~y'a~~ > y a

En revanche, on admet la transcription suivante, la suite « qu' » existant dans l'orthographe :

- qu'y a eu un changement (107/8:59)

Non usage de l'apostrophe : absence d'élision

L'élision apparait dans le cas de la chute d'un schwa devant voyelle. Si le schwa garde son contenu mélodique devant voyelle sans qu'il puisse pour autant être assimilé à « euh », il sera conservé dans la graphie. En ce sens, on ne marquera pas l'élision (par l'apostrophe).

Exemples pris dans ESLO1 :

- parce que on (107/7:49)
- parce que il (109/4:51)
- parce que ici (118/01:57)

2. Trait d'union (segmentation lexicale)

♦ *Usage normé du trait d'union*

Il conviendra d'appliquer les conventions du français en termes de trait d'union, notamment pour la graphie des nombres. Si vous avez un doute, posez la question, la réponse sera alors intégrée au Lexique-ESLO.

Rappel : Quand chacun des éléments d'un nombre est inférieur à cent, il prend un trait d'union, sauf s'il est joint par « et » : ex, dix-sept, soixante et un.

Exemples pris dans ESLO1 :

- mille neuf cent vingt-six (129/8:36)
- demi-heure (118/02:37)
- trois-quarts d'heure (118/02:38)

D'autre part, on distinguera les deux formes ci-dessous :

- cette place-là
- cette place là

♦ *Mots incomplets*

Le trait d'union est la notation arrêtée pour les mots incomplets, cas où le locuteur commence un mot et ne le termine pas. Vous ajouterez dans ce cas un tiret accolé à la partie tronquée. Veillez à laisser un espace entre le tiret et le mot qui suit.

Exemples :

- il faut les remp- remplacer
- de bien l- de bien l'écrire

♦ *Segmentation*

En revanche, la segmentation au milieu d'un mot n'est pas admise.

Exemples :

- des Port- -ugais > des Portugais (129/39:44)
- sa- -voir-livre > savoir-livre (129/1:20:19)

3. Majuscules

♦ *Emploi lexical*

Les majuscules ne sont pas utilisées pour marquer le début de phrase.

En revanche, elles sont conservées pour les noms propres, à savoir les noms de personnes et de lieux, les noms d'institutions et de marques.

Exemples pris dans ESLO1 :

- Caisse Nationale d'Epargne (078/10:49)
- Waterman (106/47:30)
- Tuperware (107/31:09)
- Martini (118/32:59)

- sécurité sociale (106/53:28)

Mais

- Académie Française (118/39:12)
- Bureau (pour le BRGM) (118/07:49)
- Ponts et Chaussées (129/1:07:03)
- l'Ecole Centrale (131/10:31)
- bons du Trésor (133/1:01:06)

Problème des institutions qui sont spécifiques

- une Université Nouvelle (542/39:54)

♦ *Assimilation à des noms propres*

Sont aussi notés avec des majuscules :

- **les nationalités :** s'il ne s'agit pas de l'adjectif, les nationalités prennent la majuscule : « les Anglais mangent des chips », en revanche « les bateaux anglais »,
- **les groupes de musique :** « Les Haricots Rouges » (096/55:33),
- **les titres de films :** « Le Cerveau » (096/56:52),
- **les titres de périodiques :** « les Sélection », i.e. *Sélection du Reader's Digest* (096/1:02:02), « la Nouvelle République » (109/48:03), « Tout l'Univers » – collection encyclopédique par fascicules – (118/34:32),
- **les modèles de véhicule ou d'avions :** Concorde (110/46:19),
- **les événements historiques :** la Guerre de Quatorze (110/51:20) (129/1:14:40), la Première Guerre Mondiale (129/47:49), la Libération (133/52:53),
- **les sites et monuments :** Gare du Nord, Palais Royal (110/52:09 > 52:14).

En revanche pas de majuscule à « dieu » dans « mon dieu » (118/00:16).

!!! ATTENTION !!!

Pour tous les doutes sur la présence ou absence de majuscule ou de trait d'union, il faut poser la question à C. Dugua et L. Kanaan-Caillol.

Nos réponses seront recensées dans le Lexique-ESLO et permettront de le compléter progressivement. Il convient alors de vous référez au Lexique-ESLO avant de poser une question.

4. Épellation et sigles

Pour les épellations et les sigles, les lettres seront inscrites en capitale.

Afin de distinguer épellations et sigles :

- les épellations sont notées avec des espaces entre les lettres, ex :
  - H A R I C O T
- les sigles sont notés avec les lettres accolées et sans point, ex :
  - TVA (542/1:04:46)
  - CGT (542/1:07:08)
- les acronymes sont notés avec la première lettre en majuscule et les suivantes en minusculet, ex :
  - Capes

5. Chiffres

Les chiffres doivent être transcrits en toute lettre :

- une quatre cent quatre Peugeot (118/13:12, 14:19, 14:30)

Cependant lorsque les chiffres dépendent d'une suite de lettres, ils seront notés en chiffres :

- ~~appartement F quatre~~ > appartement F4
- ~~CM un~~ > CM1

6. Répétitions

En cas de répétitions de termes, tous doivent être transcrits :

- tout tout tout tout tout
- oui oui oui oui oui

7. Graphies incertaines

Lorsque vous êtes confrontés à un terme dont vous ne connaissez pas la graphie (nom propre, nom commun, onomatopées…) et que vos recherches ne vous permettent pas de trouver la graphie exacte, vous devrez poser la question à C. Dugua et L. Kanaan-Caillol. A partir de là, une décision sur l'orthographe à adopter sera prise. Comme il s'agira généralement de mots non lexicalisés, de verlan, de mots inventés, en d'autres termes, de mots que l'on ne trouve pas dans le Petit Robert, pour distinguer ces mots dans la transcription, vous devrez accoler au mot le **caractère &**.

Par exemple :

- &chelou
- &bravitude

8. Marques d'affirmation et de négation

♦ *Marques d'affirmation*

Les formes d'approbation sont actuellement partagées entre « oui », « ouais » et « hm » (ou plutôt « hm hm »).

(106/49:21, 1:00:33), (107/12:19), (110/36:17), (132/48:01)

♦ *Marques de négation*

« Non » sera toujours noté « non », et non « nan » (078/06:59, *passim*). Le « nan » sera réservé aux formes enfantines du « non ».

9. Prononciation des mots étrangers

Les mots étrangers doivent être écrits de la même manière que dans leur langue d'emprunt.

Certaines réalisations spéciales peuvent avoir lieu, dans ce cas vous utiliserez la **balise « Prononciation » [pron]** pour signaler la prononciation réalisée (en utilisant l'orthographe).

- ~~katchoup~~ > ketchup [pron= katchoup]

V. Difficultés liées à l'écoute

1. Passages peu compréhensibles

Les passages peu compréhensibles sont à noter en utilisant la **balise « Prononciation inintelligible » [pron=pi]** proposée par Transcriber.

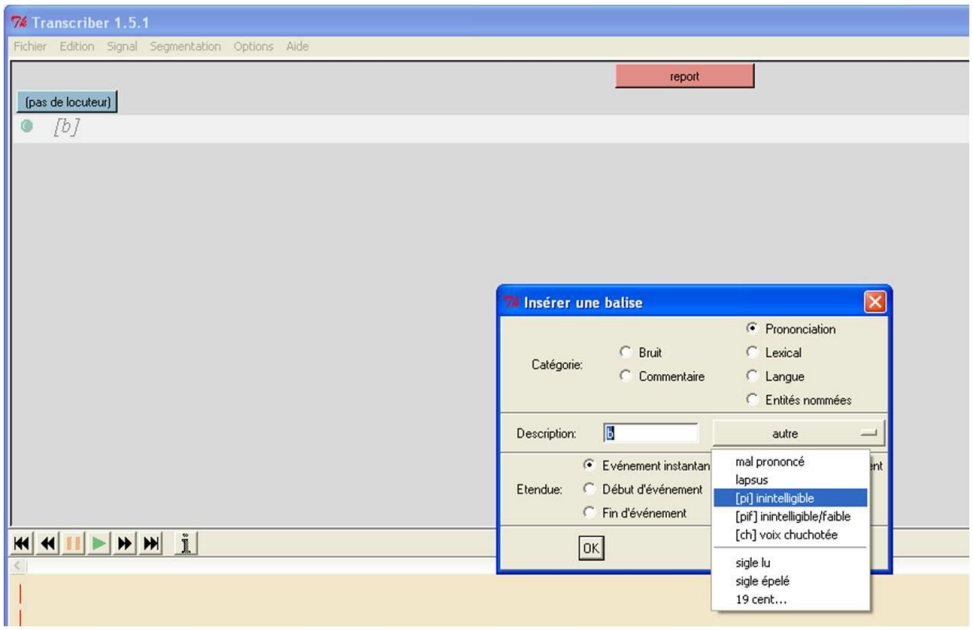

Figure 15: Créer une balise [pron=pi]

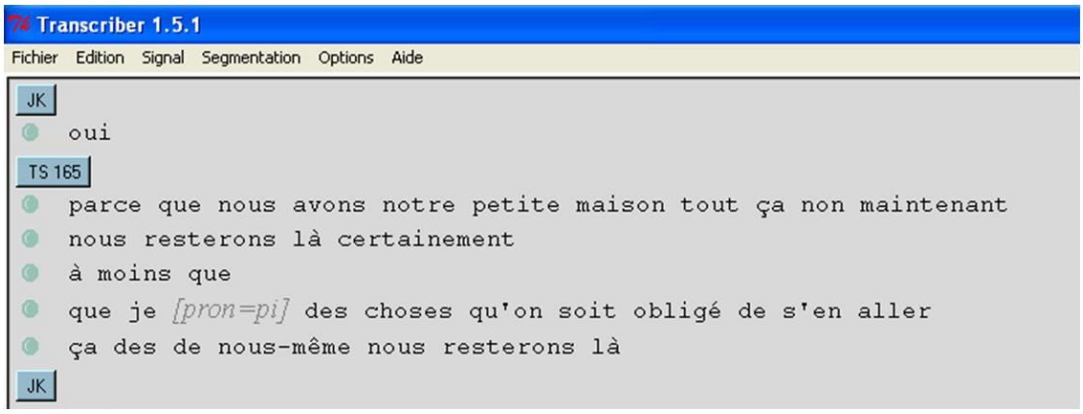

Figure 16: Exemple de balise [pron=pi]

2. Passages peu compréhensibles en raison de l'acoustique

Si des passages ne sont pas compréhensibles en raison de problèmes acoustiques, dans ce cas, il faut utiliser la balise **« Prononciation inintelligible faible » [pron=pif]**.

3. Ambiguïtés

Lorsque vous hésitez entre deux formes, le recours au contexte devrait vous permettre de faire certains choix (non exclusifs). Si cela ne suffit pas, choisissez la forme qui selon vous convient le mieux.

Exemples pris dans ESLO1 :

- ~~ça changeait~~ > ça a changé (106/32:41)
- ~~ça la fatiguait~~ > ça l'a fatiguée (109/16:00)

VI. Rétablissement des mots et des constructions

1. Mots rétablis

♦ *Elisions erronées*

Dans les cas d'élisions erronées, vous devez rétablir la seule possibilité admise par l'orthographe.

Exemples pris dans ESLO1 :

Qu' / qui

- ~~le petit qu'était malade~~ > le petit qui était malade (106/14:01)
- ~~qu'a son CAP~~ > qui a son CAP (110/11:37)
- ~~qu'était~~ > qui était (118/01:49)

Qui / qu'il

- ~~y a tout ce qui faut~~ > y a tout ce qu'il faut (109/28:51)
- ~~tout ce qui fallait~~ > tout ce qu'il fallait (139/14:23)

Qu' / qu'il

- ~~ce qu'y reste~~ > ce qu'il reste (109/52:15)

T' / tu

- ~~t'as appris~~ > tu as appris (110/37:38)
- ~~t'écris~~ > tu écris (110/44:59)

♦ *Suppressions indécidables*

Si l'enregistrement ne permet pas de décider à l'écoute de l'absence effective (suppression effective) d'une unité, vous devez préserver la notation de l'unité en question.

Exemples pris dans ESLO1 :

- ~~qui y a déjà~~ > qu'il y a déjà (096/1:07:27)
- ~~tout ce qui y a~~ > tout ce qu'il y a (139/07:15)
- ~~guide savoir vivre~~ > guide de savoir-vivre (129/1:19:52)

♦ *Ne de négation*

Si la présence/absence du premier terme de la négation est indécidable du fait d'une liaison, on doit le rétablir. Sinon, il ne doit être indiqué que s'il figure explicitement dans l'enregistrement.

- ~~on a pas~~ > on n'a pas
- on part pas

♦ *Aphérèses et apocopes*

Les aphérèses sont rétablies, quitte à faire l'objet d'une balise prononciation lorsqu'elles tendent vers une prononciation inhabituelle.

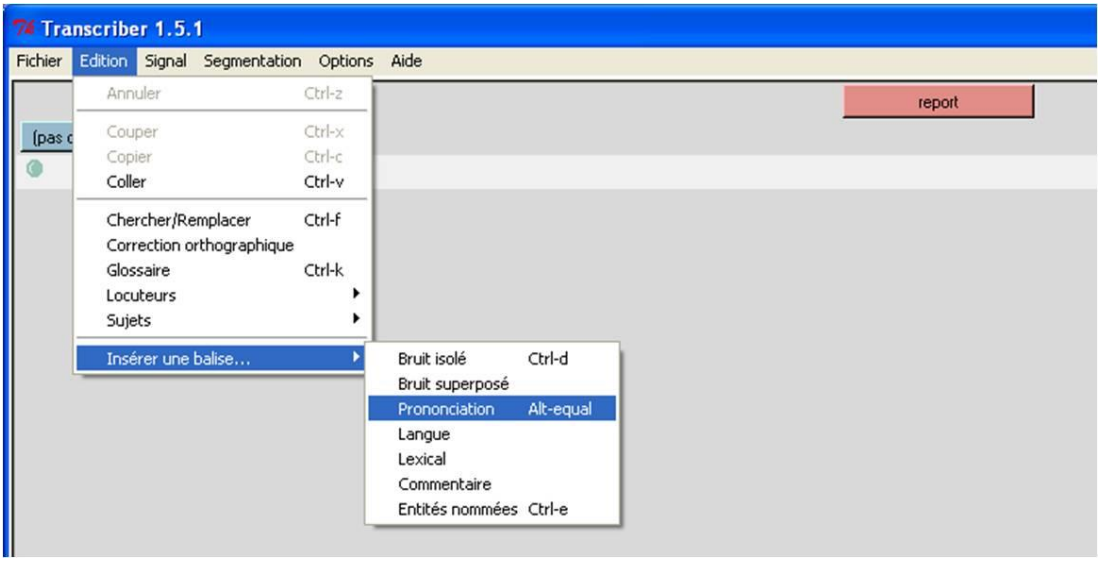

Figure 17: Insérer une prononciation particulière

Exemples pris dans ESLO1 :

- ~~tendez~~ > attendez (129/3 :08)
- ~~fin~~ > enfin (passim)
- ~~endez~~ > attendez [pron = endez]

Les apocopes ne sont pas rétablies.

- perfo pour : *perforatrice* (109/09:03) reste sous cette forme
- perfo-vérif pour : *perforatrice-vérificatrice* (109/09:26) reste sous cette forme

On accorde en nombre les apocopes lexicalisées (dictionnaire) :

- des vélos, mais des perfo

♦ *Mots rétablis liés à la prononciation*

Si pour un locuteur les types de prononciation ci-dessous (ou autres) sont généralisées, l'indiquer dans le fichier « Remarques ».

- ~~pu~~ > plus (132/43:41)
- ~~v'là~~ > voilà

2. Mots non rétablis

♦ *Lapsus*

Lorsque la forme existe dans le lexique, le lapsus ne doit pas être corrigé. Par exemple, le locuteur voulait dire « à l'attention de » et dit la forme ci-dessous, on la transcrit telle quelle :

- à l'intention de

En revanche quand la forme n'existe pas dans le lexique, vous devez la corriger et indiquer dans la **balise « Prononciation » ([pron=])**, la forme prononcée par le témoin.

- ~~oblette~~ > omelette *[pron=oblette]* (109/00:37)
- ~~faire la dastylo~~ > faire la dactylo *[pron=dastylo]* (106/05:25)
- ~~rénuméré~~ > rémunéré *[pron=rénuméré]* (110/3:42) (129/24:22)
- ~~menthélisées~~ > mentholisées [pron=menthélisées] (129/30:13)

♦ *Déformations volontaires (métalinguistique)*

Lorsque vous avez affaire à des formes délibérément déformées à des fins métalinguistiques, le procédé utilisé et le même que pour les lapsus.

- ~~il fait bieau temps~~ > il fait beau [*pron = bieau*] temps (110/23:53)

♦ *Il y a*

On ne rétablit pas le « il » de « il y a » s'il n'est pas prononcé.

Rappel : pas d'apostrophe entre « y » et « a »

- ~~il y a quelqu'un dehors~~ > y a quelqu'un dehors (110/25:20)

VII. Onomatopées et interjections

Vous trouverez ci-dessous la liste des principales onomatopées et interjections que vous pouvez être amené à rencontrer dans les transcriptions. L'orthographe des onomatopées et interjections de cette liste a été vérifiée dans Le Petit Robert. Pour tout doute graphique, il convient de poser la question à C. Dugua ou L. Kanaan-Caillol. Cette liste sera peu à peu complétée par les occurrences que vous trouverez dans les transcriptions.

ah	boum	hou	Ouille
aïe	clac	miam	pff
bah	euh	mouais	zut
ben	hé	oh	
bof	hein	ouais	
bouh	hop	ouf	

Figure 18: Liste des onomatopées et interjections

Annexes

I. Fichier « Remarques »

La version word de ce fichier vous est distribuée.

Problèmes et remarques

Nom de la transcription :

Date :

Nom du/de la transcripteur/trice :

Version de la transcription : A – B – C

Remarques (préciser le temps (minutage) s'il ne s'agit pas de remarques générales) :

Problèmes (préciser le temps (minutage) s'il ne s'agit pas de problèmes généraux) :