



GUIDE DU TRANSCRIPTEUR ET DU RELECTEUR D'ESLO2¹

Version : février 2010

A partir du guide fait par Linda

¹ Il s'agit du guide pour le corpus ESLO2, même si la plupart des exemples utilisés sont tirés du corpus ESLO1.

Table des matières

Présentation	5
I. Objectifs du guide.....	5
II. Quelques mots sur les ESLOs.....	5
Phase 1 : Transcription	6
I. Outils d'aide à la transcription et à l'organisation	6
1. Base de données.....	6
2. Google Groupe	6
II. Logiciel de transcription : Transcriber	7
1. Quelques principes sur le logiciel	7
2. Installation du logiciel.....	7
◆ Téléchargement.....	7
◆ Paramétrages.....	7
◆ Installation du dictionnaire	7
3. Manuel d'utilisation simplifié	8
◆ Créer et nommer une section (ou Report)	8
◆ Créer et nommer un tour de parole	9
◆ Créer des segments dans un tour de parole	9
◆ Chevauchements	10
◆ Raccourcis clavier	11
III. Etapes à suivre pour faire une transcription	11
1. Paramètres de l'épisode (= transcription).....	11
2. Nom du fichier de transcription	11
3. Codes témoin et chercheur	11
4. Première écoute de familiarisation	12
5. Transcription.....	12
6. Après la transcription : première relecture.....	12
◆ Découpage en sections.....	12
◆ Anonymisation.....	12
◆ Problèmes et remarques	12
7. Documents à rendre : récapitulatif	13

IV. Conventions	14
1. Principes et règles de base des conventions de transcription	14
2. Principes de segmentation	15
◆ Trois niveaux de segmentation	15
◆ Règles de segmentation	15
Segmentation des sections :	15
Identification des locuteurs	16
Segmentation des segments	18
Chevauchements	19
3. Utilisation des balises « Bruit »	20
◆ Rires	20
◆ Micro	20
◆ Passages non transcrits	21
◆ Soufflerie	21
Bruits de respiration	21
Clics	21
4. Conventions de transcription	22
◆ Signes graphiques	22
Guillemets	22
Apostrophe	22
◆ Trait d'union (segmentation lexicale)	23
Usage normé du trait d'union	23
Mots incomplets	23
Segmentation	24
◆ Majuscules	24
Emploi lexical	24
Assimilation à des noms propres	25
◆ Epellation et sigles	25
◆ Chiffres	26
◆ Répétitions	26
◆ Graphies incertaines	26
◆ Marques d'affirmation et de négation	26

Marques d'affirmation	26
Marques de négation	26
◆ Prononciation des mots étrangers	27
5. Difficultés liées à l'écoute.....	27
◆ Passages peu compréhensibles	27
◆ Passages peu compréhensibles en raison de l'acoustique.....	28
◆ Ambiguïtés.....	28
6. Rétablissement des mots et des constructions.....	28
◆ Mots rétablis.....	28
Elisions erronées.....	28
Suppressions indécidables	29
Ne de négation	29
Aphérèses et apocopes	29
Mots rétablis liés à la prononciation	30
◆ Mots non rétablis	30
Lapsus	30
Déformations volontaires (métalinguistique)	30
Il y a	30
7. Onomatopées et interjections.....	31
Phase 2 : Relecture	32
I. Relecture en trois étapes	32
1. Etape 1.....	32
2. Etape 2.....	32
3. Etape 3.....	32
II. Documents à rendre.....	33

Table des figures

Figure 1: Créer une section (Report)	8
Figure 2: Créer un tour de parole	9
Figure 3: Créer un tour de parole avec chevauchement	10
Figure 4: Principes de base des conventions de transcription	14
Figure 5: Trois niveaux de segmentation: section, locuteur, segment	15
Figure 6: Exemple de section	16
Figure 7: Exemple si hésitation entre deux locuteurs	17
Figure 8: Exemple si locuteur non identifié (inconnu)	17
Figure 9: Exemple de précision sur un locuteur	17
Figure 10: Exemple de tour de parole trop long	18
Figure 11: Exemple de tour de parole découpé en segments	18
Figure 12: Exemple de chevauchement de paroles	19
Figure 13: Exemple de chevauchements avec des marques d'acquiescement	19
Figure 14: Exemple balise [rire]	20
Figure 15: Exemple balise micro [mic]	20
Figure 16: Exemple section [nontrans]	21
Figure 17: Créer une balise [pron=pi]	27
Figure 18: Exemple de balise [pron=pi]	27
Figure 19: Insérer une prononciation particulière	29
Figure 20: Liste des onomatopées et interjections	31

Présentation

I. Objectifs du guide

Dans ce guide vous trouverez l'ensemble des étapes et des conventions qui vous permettront de transcrire les enregistrements de ESLO2. Afin de vous guider au mieux dans cette tâche, une deuxième partie sera consacrée aux procédures de relectures et de vérifications de vos transcriptions.

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.

Phase 1 : Transcription

I. Outils d'aide à la transcription et à l'organisation

Deux outils sont proposés pour aider les transcrip-teurs dans leur tâche de transcription et pour faciliter l'organisation générale du groupe : informations à transmettre, questions à poser.

1. Base de données

Une base de données permet de réunir l'ensemble des informations sur les enregistrements, les locuteurs et les transcriptions ESLO2. Vous pourrez vous y reporter pour obtenir notamment les codes des locuteurs.

Pour accéder à la base, cliquer sur le lien suivant :

<http://www.univ-orleans.fr/eslo/corpus/>

Login : orleans

Mot de passe : coraleslo

2. Google Groupe

Un Google Groupe « Transcription ESLO » a été créé dans le but de faciliter la communication entre les transcriptrices et les organisatrices, et entre les transcriptrices entre elles. On peut l'intégrer uniquement sur invitation, aucune information ne sera « publique ».

<http://groups.google.fr/group/transcription-eslo?hl=fr>

Il sera possible par exemple de poser des questions sur sa transcription et de permettre ainsi à tout le monde d'y avoir accès et d'avoir accès aux réponses proposées.

Sur ce groupe sont également déposés les documents dont les transcriptrices peuvent avoir besoin, notamment le guide du transcrip-teur et du relecteur, les conventions de transcription, la trame de l'entretien (avec les numéros des questions) mais aussi une page avec les sites web à consulter pour des questions d'orthographe, et d'autres documents à venir.

On peut également imaginer créer une base de données (un lexique) de noms de ville/quartier/rue, une base de données sur les onomatopées, etc.

II. Logiciel de transcription : Transcriber

1. 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.

2. Installation du logiciel

◆ *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>

◆ *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.

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

- « **Nom du transcripateur** » : 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** ».

◆ *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.

3. Manuel d'utilisation simplifié²

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

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

◆ *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

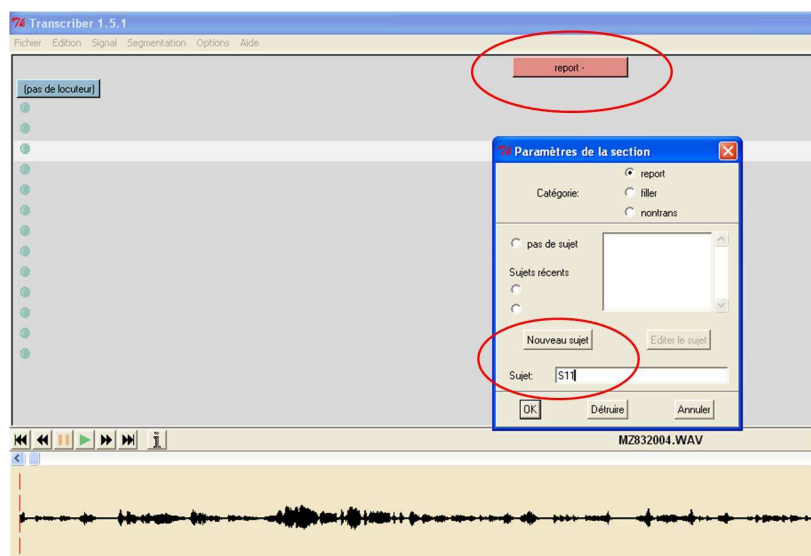


Figure 1: Créer une section (Report)

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

◆ *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 : se référer à la base de données et à la section « Identification des locuteurs » (p.16). Pour ce faire, suivre la démarche ci-dessous :

Segmentation > Créer un tour (ou CTRL t)
Créer un locuteur (si le locuteur n'a pas encore été créé)
Ou choisir le locuteur dans la fenêtre

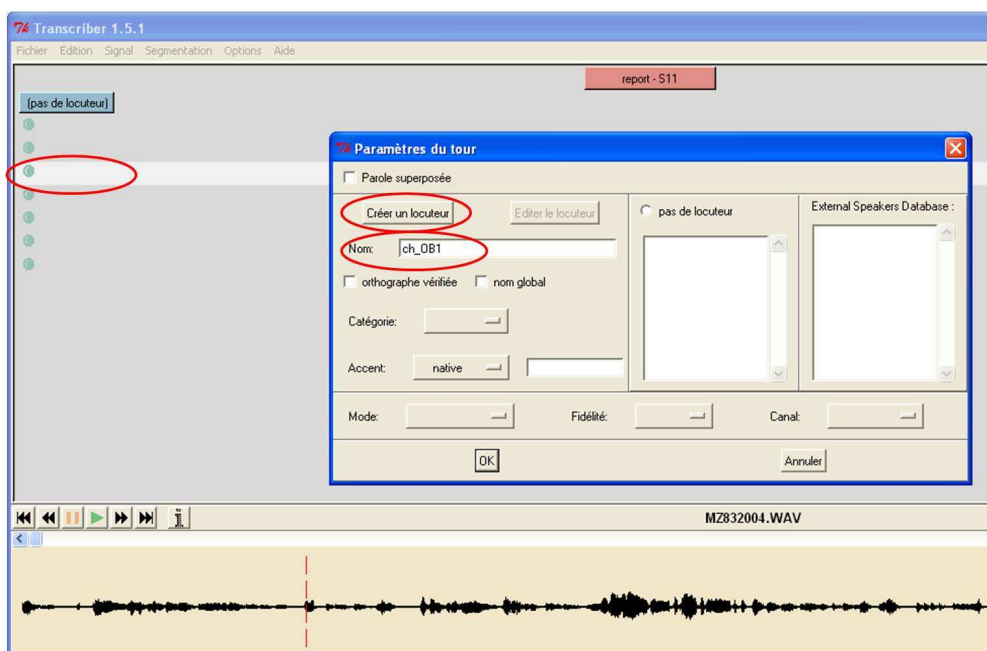


Figure 2: Créer un tour de parole

◆ *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.

◆ *Chevauchements*

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

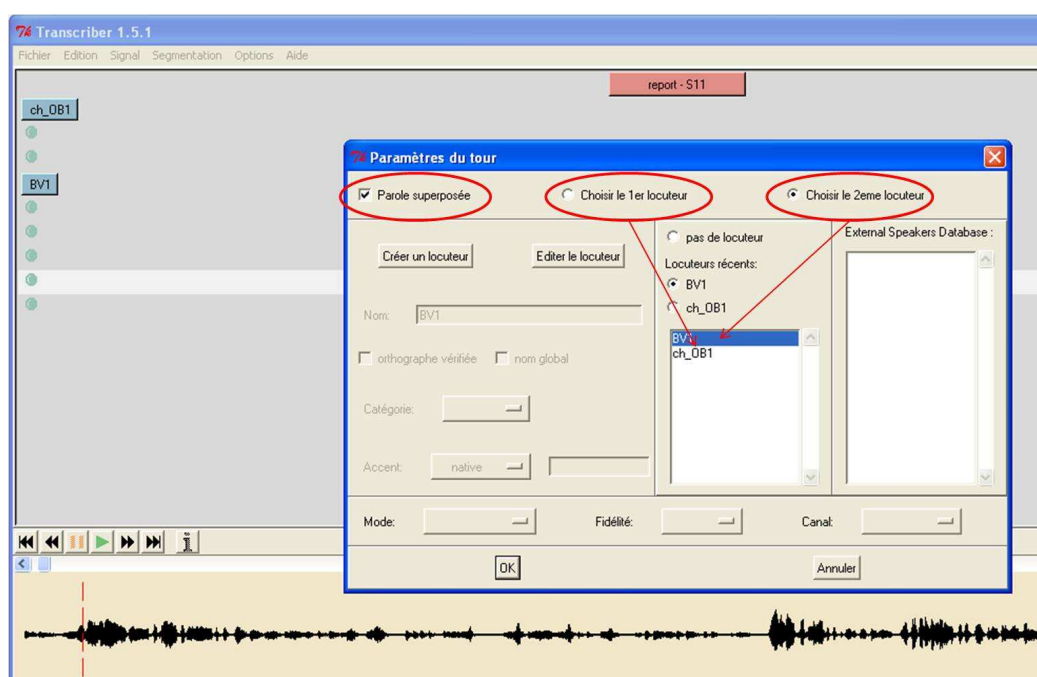


Figure 3: Créer un tour de parole avec chevauchement

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 ».

◆ *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**
Enregistrer : **CTRL s**

III. Etapes à suivre pour faire une transcription

Dans cette section, vous trouverez les 6 étapes à suivre pour toute transcription. Nous terminons par une partie récapitulant les documents à rendre pour chaque transcription.

1. 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

2. 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_1], le fichier transcription devra être appelé [ESLO2_ENT_1_A].

3. Codes témoin et chercheur

Pour chaque entretien, on a 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).

Dans Transcriber, il faut donc nommer les locuteurs à partir de ces informations. Ces codages vous seront donnés en même temps que le fichier son. Vous avez toujours la possibilité de les retrouver dans la base de données.

4. Première écoute de familiarisation

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.

5. Transcription

Les quatre étapes précédentes faites, il est alors possible de passer à la transcription du fichier son. Cette étape n'est pas détaillée ici, se référer à la section « Conventions » (p.14).

6. Après la transcription : première relecture

Avant de rendre sa transcription, il est important d'en faire une première relecture, dans le but de corriger les fautes d'orthographe et de frappe les plus visibles et de préciser la transcription sur les trois aspects suivants.

◆ *Découpage en sections*

La transcription étant faite, elle doit être découpée en sections (ou Report) qui correspondent aux questions de la trame du questionnaire. Cette étape peut éventuellement être faite en même temps que la transcription et dans ce cas, intégrer l'étape 5.

◆ *Anonymisation*

Les transcriptions et les fichiers sons qui seront mis en ligne doivent être anonymisés. Plusieurs niveaux d'anonymisation sont envisageables. *A minima*, il s'agira de masquer les endroits où le nom de famille d'une personne est prononcé. Pour ce faire, dès que vous rencontrerez un nom de famille dans une transcription, au lieu de transcrire ledit nom, vous utiliserez le codage suivant : **NPERS**

◆ *Problèmes et remarques*

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. 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_1_A_Remarques]**.

7. 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_1_A]**
- le fichier « Problèmes et remarques » Ex. : **[ESLO2_ENT_1_A_Remarques]**

³ Document disponible sur le Google Groupe.

IV. Conventions

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

1. 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 (/leletʁkəzɛkʁi/). 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 Robert, ainsi que 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.26)

Figure 4: Principes de base des conventions de transcription

2. Principes de segmentation

◆ *Trois niveaux de segmentation*

Trois niveaux de segmentation sont possibles avec Transcriber :

- En sections
- En locuteurs
- En segments

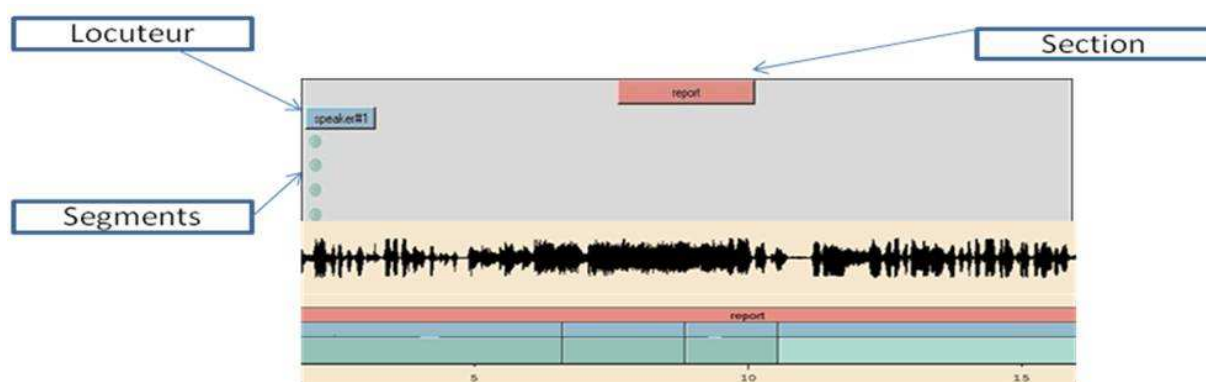


Figure 5: Trois niveaux de segmentation: section, locuteur, segment

◆ *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 6: 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 : MD 461
- Le chercheur : BV

ESLO2

Pour 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_1, les locuteurs sont :

- Témoin : BV1
- Chercheur : ch_OB1

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. De ce fait, cette hésitation sera notée par un point d'interrogation entre les deux locuteurs pour lesquels vous hésitez.

Par exemple :

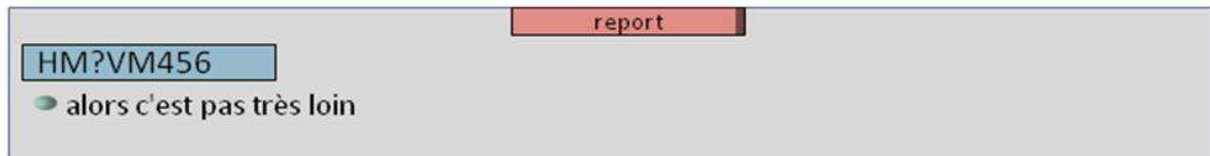


Figure 7: Exemple si hésitation entre deux locuteurs

Locuteur non identifié

Lorsque le locuteur n'est pas identifié, il peut s'agir d'une personne extérieure à l'interview, le transcripteur ne dispose alors d'aucune information le concernant. Dans ce cas, il sera noté par **INC**.



Figure 8: Exemple si locuteur non identifié (inconnu)

Si le transcripteur connaît le rapport entre le locuteur et le témoin, il l'indique dans l'intitulé du code : « Mère de *code témoin* », « Frère de *code témoin* », etc.

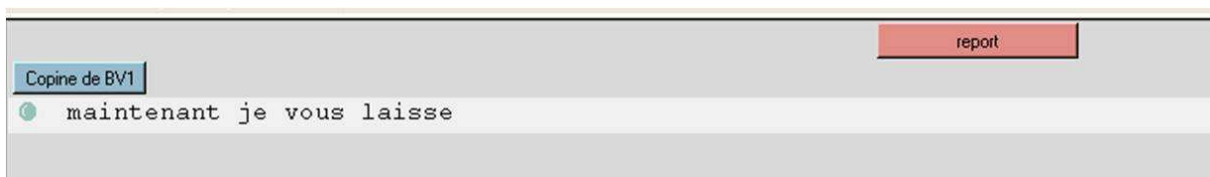


Figure 9: Exemple de précision sur un locuteur

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

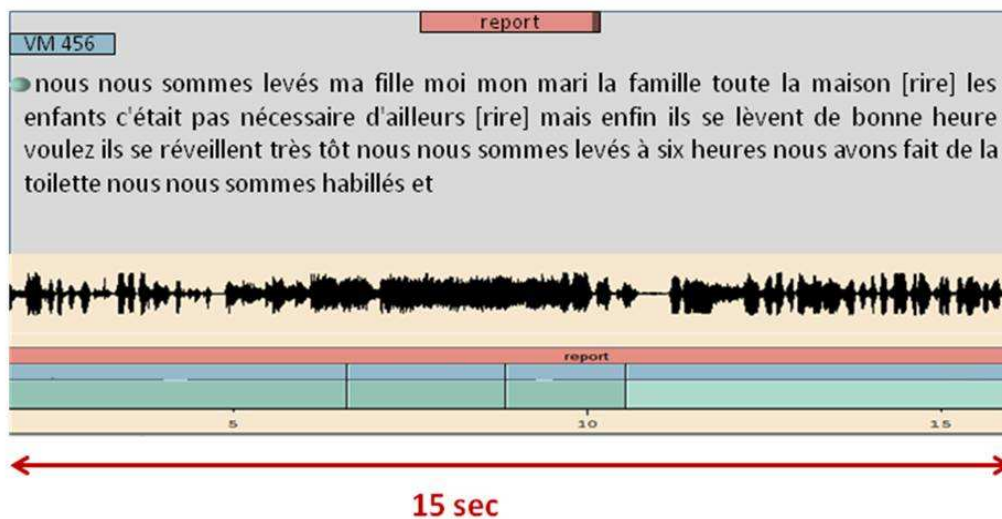


Figure 10: Exemple de tour de parole trop long

Nous lui préférons cette segmentation :



Figure 11: 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 lorsque celles-ci sont clairement attribuables. Exceptionnellement seulement, lorsque la pause n'est pas attribuable à un locuteur (pause longue), on met « **NO SPEAKER** » comme locuteur. Si la pause apparaît entre une question et la réponse, on met la pause plutôt au locuteur qui va répondre.

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.



report

JSM+RF 211

- 1: j'aime le beau papier
- 2: et les lignes

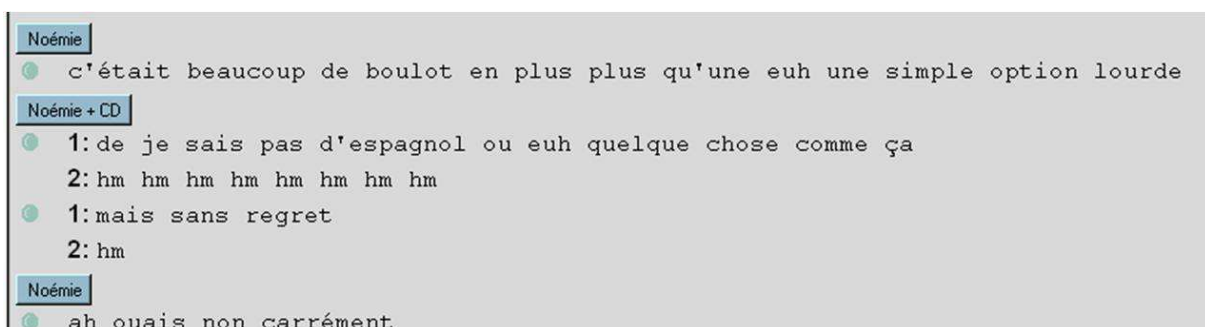
RF 211

- derrière

Figure 12: Exemple de chevauchement de paroles

Cas particulier : Chevauchement avec des marques d'acquiescement

Dans le cas de chevauchement avec des marques d'acquiescement (back channel), 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.



Noémie

- c'était beaucoup de boulot en plus plus qu'une euh une simple option lourde

Noémie + CD

- 1: de je sais pas d'espagnol ou euh quelque chose comme ça
- 2: hm hm hm hm hm hm hm hm

Noémie

- 1: mais sans regret
- 2: hm

Noémie

- ah ouais non carrément

Figure 13: Exemple de chevauchements avec des marques d'acquiescement

3. Utilisation des balises « Bruit »

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

◆ Rires

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



Figure 14: Exemple balise [rire]

◆ 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].

The screenshot displays a transcription software interface. The top section shows a list of transcription items with markers like LN 438, BV, and [mic]. The bottom section shows a timeline with a red oval highlighting a segment from 46:12 to 46:20 containing the label [mic].

L	BV	LN 438	BV	LN 438	BV
o	pour...	[pron=pi]	hm	des jaloux	
?	... ?			il faisait que ... alors [p]	
				[mic]	
					... était ...
					... de langue

Timeline markers: 46:00, 46:05, 46:10, 46:15, 46:20

Cursor: 46:23.312

Figure 15: Exemple balise micro [mic]

◆ *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.

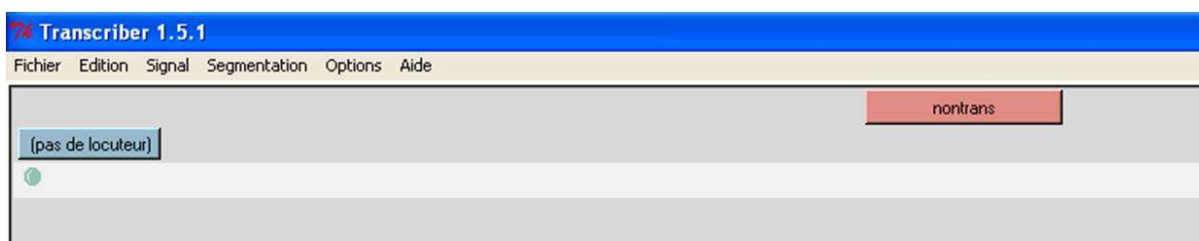


Figure 16: Exemple section [nontrans]

◆ *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

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)

4. Conventions de transcription

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 convictions que j'ai eues (à propos de 106).
- ~~et bien~~ > eh bien (*passim*)
- ~~eh la la~~ > oh là là (096/40:09)
- ~~où que c'est beau~~ > ouh que c'est beau (109/12:55)

Attention aux difficultés de l'orthographe :

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

N'hésitez pas à vous reporter au dictionnaire ou aux grammaires pour résoudre les questions d'orthographe.

◆ *Signes graphiques*

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 apparaît 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)

◆ *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.

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’écriture

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)

◆ ***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'Épargne (078/10:49)
- Waterman (106/47:30)
- Tupperware (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'École 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).

Pour tous les doutes sur la présence ou absence de majuscule ou de trait d'union, il faut poser la question sur le Google Groupe. Par ce biais, se mettra progressivement en place un lexique qui reprendra l'ensemble des termes qui ont posé questions. Lexique auquel il conviendra de se référer avant de poser une question.

◆ ***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)

◆ *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

◆ *Répétitions*

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

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

◆ *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 sur le Google Groupe. 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, pour distinguer ces mots dans la transcription, vous devrez accoler au mot le caractère **&**. Par exemple :

- &chelou
- &bravitude

◆ *Marques d'affirmation et de négation*

Marques d'affirmation

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

(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 ».

◆ *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]

5. Difficultés liées à l'écoute

◆ *Passages peu compréhensibles*

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

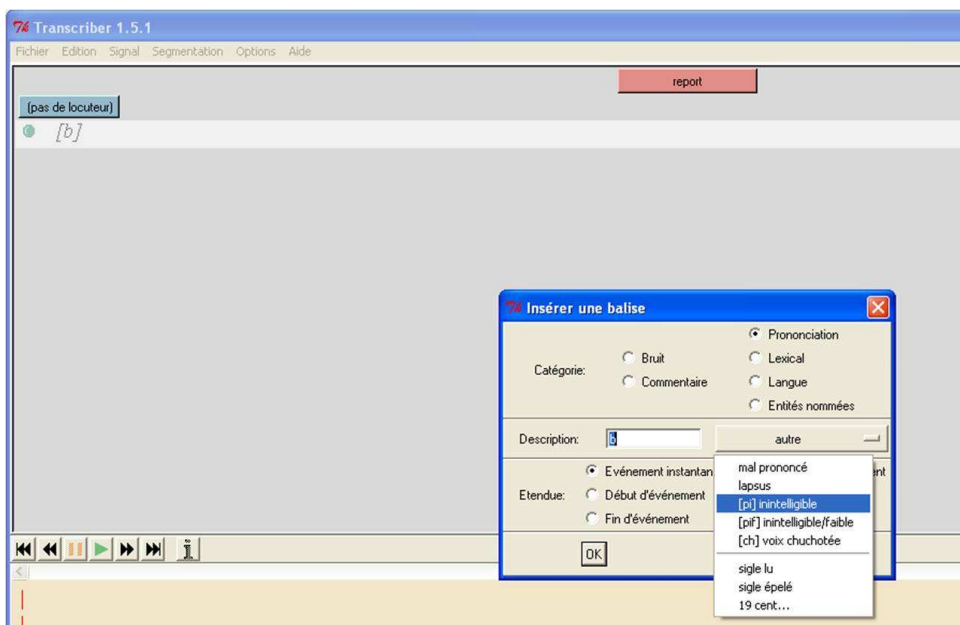


Figure 17: Créer une balise [pron=pi]

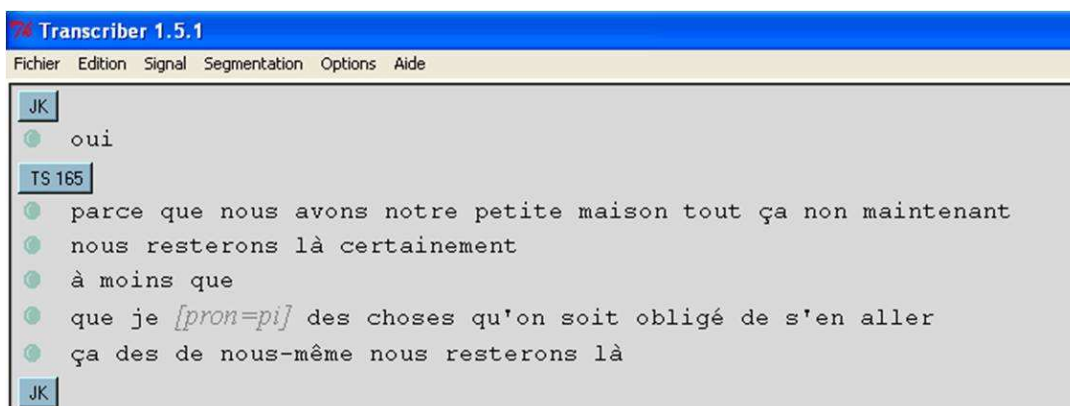


Figure 18: Exemple de balise [pron=pi]

◆ *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].

◆ *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)

6. Rétablissement des mots et des constructions

◆ *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.

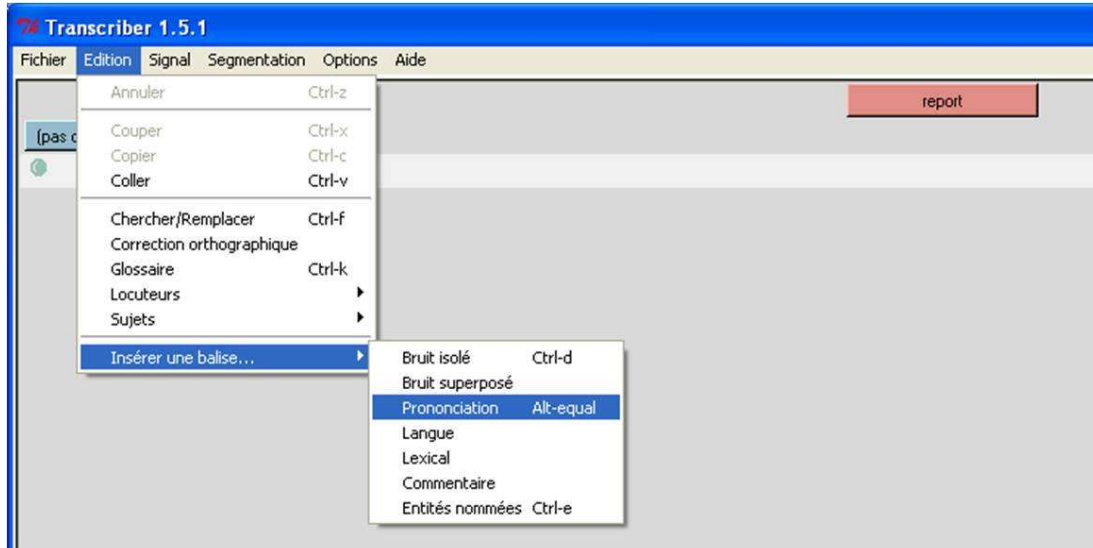


Figure 19: Insérer une prononciation particulière

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 ».

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

◆ **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 dastyle~~ > 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 est le même que pour les lapsus.

- ~~il fait bieu temps~~ > il fait beau [*pron = bieu*] 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)

7. 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 dans le Google Groupe. Cette liste sera peu à peu complétée par les occurrences que vous trouverez dans les transcriptions.

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

Figure 20: Liste des onomatopées et interjections

Phase 2 : Relecture

Une fois votre transcription terminée, plusieurs relectures sont nécessaires. Dans cette partie, vous trouverez un ensemble d'étapes qui vous permettront d'appréhender au mieux cette phase.

Dans ce projet, relire renvoie à différentes tâches : (1) corriger l'orthographe en lien avec les conventions de transcription, (2) réécouter l'enregistrement pour d'une part, diminuer le nombre de balises [pron=pi] et d'autre part, vérifier que la première transcription correspond au mieux à ce qui a été dit, (3) affiner les informations sur les métadonnées.

I. Relecture en trois étapes

Trois étapes sont nécessaires, chacune permet en effet de se concentrer sur un élément précis de la transcription.

1. Etape 1

La première étape consiste à exporter votre transcription en format texte à partir de Transcriber. Vous récupérez ensuite le fichier texte que vous copierez sur Word. Cette phase permet dans un premier temps de repérer les fautes de frappe et les fautes d'accord, mais attention ne vous fier pas uniquement à son correcteur, tous les éléments ne seront pas repérés.

Dès que vous repérez un élément à corriger, il faut faire la correction sous Transcriber, en utilisant la fonction « Rechercher ».

2. Etape 2

La deuxième étape consiste en une relecture sans le fichier sonore. Durant cette phase, il s'agira de vérifier les fautes d'orthographe, de grammaire, de conjugaison, d'accord, l'écriture des noms de ville, de pays, etc., mais aussi de vérifier toutes les particularités des conventions (espace devant point d'interrogation, transcription des onomatopées, etc.).

Vous devrez également veiller à l'homogénéité de vos transcriptions, l'écriture des mots, des noms de ville, des néologismes doit être la même dans toute la transcription.

Pour vous aider, consulter les dictionnaires, les grammaires, les sites mis à votre disposition sur le Google Groupe.

3. Etape 3

La troisième étape consiste en une vérification avec le fichier sonore. Cette phase vous permettra alors de porter votre attention sur la segmentation, et le respect des conventions liées à l'écoute (amorces, répétitions...).

II. Documents à rendre

Les deux documents à rendre sont semblables à ceux concernant la première transcription :

- le fichier Transcrire de la relecture, qui porte le même nom que le premier fichier à la seule différence de la dernière lettre : remplacer A par B.
 - o Exemple : si le fichier transcription se nomme [ESLO2_ENT_1_A], le fichier relecture correspondant, se nommera **[ESLO2_ENT_1_B]**
- le fichier « Problèmes et remarques », qui prendra la forme suivante : **[ESLO2_ENT_1_B_Remarques]**