

GUIDE DE TRANSCRIPTION pour les RECITS de VIE - MATRICIEL

adapté du guide ESLO

Version 3 du 5 octobre 2017

1. Nommage des documents disponibles et à créer	3
2. Encodage des personnages des entretiens et des chercheurs	3
3. Tour de parole	4
Chevauchements	4
4. Fidélité à l'écoute et respect de l'orthographe	5
Orthographe	5
Contradiction entre fidélité à l'écoute et respect de l'orthographe	5
5. Règles de base	5
Segments vides, pauses et silences	6
Emploi de la majuscule	6
Emploi de l'apostrophe	6
Trait d'union (segmentation lexicale)	7
Épellation et sigles	7
Chiffres	7
Répétitions	7
Mots incomplets	7
Marques d'affirmation et de négation	8
Graphies incertaines et mots étrangers	8
Ambiguïtés	9
Onomatopées et interjections	9
6. Rétablissement des mots et des constructions	9
Elisions erronées	9
Suppressions indécidables	9
Ne de négation	10
Aphérèses et apocopes	10
Mots rétablis liés à la prononciation	10
7. Mots non rétablis	10
Lapsus	10
Déformations volontaires (métalinguistique)	10
Il y a 10	
8. Utilisation des balises « Bruit »	11
Rires	11
Conversation	11

Micro	11
Soufflerie	11
Clics	11
9. <i>Dysfluences : répétitions et révisions</i>	11
10. <i>Exemples de synthèse :</i>	12

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.

Dans ce guide, vous trouverez les conventions de transcription utilisées ainsi que la description des étapes à suivre pour faire une transcription et une relecture des enregistrements des récits de vie. 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-Matriciel qui recense certaines graphies particulières, 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.

Vous trouverez ci-dessous les consignes de transcription qui nous semblent les plus utiles. Merci de signaler les informations manquantes.

1. Nommage des documents disponibles et à créer

L'ensemble des enregistrements et des transcriptions initiales du RAHMI, ainsi que le lexique-MATRICIEL, le fichier des notes de travail pour la transcription et le classeur contenant les identifiants des entretiens sont disponibles sur le site ftp de l'IGN :

ANONYME

Les transcriptions réalisées, les fichiers de remarques, le lexique mis à jour, etc. seront déposés sur le même site ftp.

Les identifiants des entretiens sont donnés dans le fichier *descriptionEtIdentifiantsCorpusDesEspagnols.xlsx*

Deux transcriptions sont prévues :

- une transcription pure, notée A, où seuls les silences et pauses sont transcrits
- une transcription enrichie, notée B, où des balises prosodiques sont ajoutées

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

- le fichier de transcription (fichier texte, encodage UTF-8), désigné par : *<identifiant entretien>-<codeTranscription>* où le code de transcription est A ou B, par exemple : AV16-A
- le fichier « Remarques et problèmes », désigné par : *<identifiant entretien>-<codeTranscription>-Remarques*, par exemple : AV16-A-Remarques

Ces documents doivent être postés sur le ftp dans le répertoire 3-transcriptionsAnais

2. Encodage des personnages des entretiens et des chercheurs

Pour les entretiens, on a généralement un témoin (interviewé) et un intervieweur. Ces deux locuteurs sont codés. Le code du témoin correspond à deux lettres en majuscule suivies d'un chiffre (ex : AV16). Chaque chercheur a un code qu'il conserve pour toutes les transcriptions et qui prend la forme suivante : *ch_Initiales du chercheur + chiffre* (ex : *ch_JC1*). Ces codes figurent dans le tableau posté sur le ftp, *descriptionEtIdentifiantsCorpusDesEspagnols.xlsx*.

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 si le témoin a pour code AV16) :
 - * Pour le mari du témoin AV16 > AV16MAR
 - * Pour la femme du témoin AV16 > AV16FEM
 - * Pour son père > AV16PER
 - * Pour sa mère > AV16MER
 - * Pour son grand-père > AV16GRP
 - * Pour sa grand-mère > AV16GRM
 - * Pour sa fille > AV16FIE
 - * Pour son fils > AV16FIL
 - * Pour sa petite-fille > AV16PFIE
 - * Pour son petit-fils > AV16PFIL
 - * Pour sa belle-fille > AV16BFIE
 - * Pour son beau-fils > AV16BFIL
 - * Pour son ami > AV16AMI
 - * Pour son amie > AV16AMIE
 - * pour la famille > AV16FAM

- 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 :
identifiant de l'entretien + LOC, ex. AV16LOC
 Et si apparaissent plusieurs intervenants, il conviendra de les différencier sous la forme : AV16LOC1, AV16LOC2, AV16LOC3, etc.

- 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 son 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 :
identifiant de l'entretien + INC (pour inconnu), ex. AV16INC
 Et si apparaissent plusieurs intervenants, il conviendra de les différencier sous la forme : AV16INC1, AV16INC2, AV16INC3, ...

Ces codes sont aussi à utiliser dans la transcription lorsqu'un prénom et/ou un nom sont cités par le locuteur et qu'il est possible d'établir le lien de parenté. Sinon, le nom propre de personne sera transcrit (après vérification sur les dictionnaires s'il s'agit d'une personne publique). Les noms propres de personne tout comme les noms de locuteurs ou personnages créés devront être notés dans la fiche *Remarques* de l'entretien.

Si vous avez 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. merci de noter ces éléments sur la fiche *Remarques* de l'entretien

3. Tour de parole

Un tour de parole doit être attribué à un locuteur. Tout locuteur (interviewer ou témoin) est codé, les codes sont indiqués dans le fichier *descriptionEtIdentiantsCorpusDesEspagnols.xlsx*

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 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 ».

Dans le cas de chevauchement avec des marques d'acquiescement, on notera les marques d'acquiescement 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.

4. Fidélité à l'écoute et respect de l'orthographe

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

- le respect de l'orthographe ; la transcription suit également des conventions (voir plus loin), et respecte les graphies notées dans le Lexique Matriciel
- le respect de ce qui a été dit, notamment en termes de structure grammaticale.

Orthographe

Pour élimer les erreurs d'orthographe et certaines fautes d'accord, il est demandé d'exporter le fichier sous word et de passer le correcteur orthographique. Dès qu'il y a des éléments soulignés en rouge ou vert, il faut rechercher la forme adéquate et faire la transformation correspondante.

Outre les dictionnaires, les grammaires, internet, manuels de conjugaison, le Lexique-Matriciel doit être utilisé pour la transcription. Si le doute persiste, il faut indiquer le ou les mots concernés dans la fiche *Remarques* de l'enregistrement.

Prononciation inintelligible :

Les mots ou passages peu compréhensibles sont à noter en utilisant la balise « Prononciation inintelligible » : *[pron=pi]*

Contradiction entre fidélité à l'écoute et respect de l'orthographe

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. 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 TLFi4, disponible en ligne : <http://atilf.atilf.fr/tlf.htm>

5. Règles de base

Sont listés à la suite des résumés concernant les points importants. Ces points sont abordés plus en détail au-dessous.

- la seule ponctuation transcrite est **le point d'interrogation**. Il sert à marquer une question, que celle-ci se réalise syntaxiquement ou uniquement par une intonation montante. **(Attention : laisser une espace devant le point d'interrogation) ; aucun autre signe de ponctuation ne doit figurer dans les transcriptions ;**
- **ne pas utiliser de** majuscule en début de phrase (i.e. après le point d'exclamation) ni en début de tour de parole ;
- les initiales majuscules sont réservées aux noms propres de personne, de lieux, les noms d'institutions et de marques, etc. Pour Matriciel, les majuscules sont significatives (voir ci-dessous)
- **les pauses doivent être annotées (voir ci-dessous)**
- dans une même transcription, l'écriture des mots doit être homogène, aussi bien pour les noms propres (attention aux noms propres de lieux) que pour les noms communs. (ex : *et cætera* ou *et cetera*)
- **les guillemets ne sont pas utilisés**, même dans les cas de discours rapporté, et même si celui-ci est explicitement exprimé

Segments vides, pauses et silences

La pause est à noter par un segment vide. Une pause est attribuée à un locuteur. Exceptionnellement seulement, lorsqu'une pause longue n'est pas attribuable à un locuteur, on met *N*SPEAKER* comme locuteur.

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

Il faut mettre en place des conventions permettant de distinguer des pauses (courtes) ou des silences (plus longs). Les pauses et silences seront balisés entre crochets

- balise *[silence]* : silence, décrit un silence **long** dans le discours, entre deux tours de paroles ou entre des paroles d'un même locuteur. Ce silence doit être remarquable, une période sans bruit est normale entre deux tours de parole, un silence prolongé est quant à lui remarquable.
- balise *[pause]* : pause, décrit une pause **courte** dans le discours du locuteur, lorsqu'il cherche ses mots par exemple.

Ne pas abuser des balises *[silence]* et *[pause]*.

NB : les balises *[silence]* seront ensuite transformées en un point dans la transcription enrichie.

Emploi de la majuscule

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 : *Guerre de Quatorze* (110/51:20) (129/1:14:40), *Première Guerre Mondiale* (129/47:49), *Libération* (133/52:53),
- les sites et monuments : *Gare de Lyon* (celle de Paris) mais *gare de Lyon* (celle qui est à Lyon) et *Palais-Royal* (monument parisien).

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

En cas de doute sur l'emploi ou non de la majuscule ou du trait d'union, il faut poser la question à Catherine Domingues. Les réponses seront recensées dans le Lexique-Matriciel et permettront de le compléter progressivement.

Emploi 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 transcrit :

- *y'a > y a*

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

- *qu'y a eu un changement*

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 ESP_1 :

- *parce que on*
- *parce que il*
- *parce que ici*

Trait d'union (segmentation lexicale)

Il convient 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-Matriciel.

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 : *mille neuf cent vingt-six*, *demi-heure*, *trois-quarts d'heure*

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

- cette place-là
- cette place là

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
- les sigles sont notés avec les lettres accolées et sans point, ex : *ONT*
- les acronymes sont notés avec la première lettre en majuscule et les suivantes en minuscules, ex : *Capex*

Chiffres

Les chiffres doivent être transcrits en toute lettre : *une quatre cent quatre Peugeot*

Cependant lorsque les chiffres dépendent d'une suite de lettres, ils seront notés en chiffres : *appartement F4, CM1*

Les dates et les numéros dans une adresse sont à écrire en chiffres :

- 9 février 1923
- 9 février 23
- 23 cours de la Marne (pas de majuscule au nom générique : *rue, cours, boulevard, etc.*)

Répétitions

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

- *tout tout tout tout tout*
- *oui oui oui oui oui*

Mots incomplets

Le trait d'union est la notation arrêtée pour les mots incomplets, i.e. le cas où le locuteur commence un mot et ne le termine pas. Ce tiret doit être accolé à la partie tronquée et une espace sera placée entre le tiret et le mot qui suit.

Exemples :

- *il faut les remp-* remplacer
- *de bien l-* de bien l'écrire
- si le mot incomplet est un chiffre ou une date : *je suis revenu en 1950- 58 mais clandestinement*

En revanche, la segmentation d'un mot pour des commodités d'écriture n'est pas admise.

Marques d'affirmation et de négation

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

« *Non* » sera toujours noté « *non* », et non « *nan* ». Le « *nan* » sera réservé aux formes enfantines du « *non* ».

Graphies incertaines et mots étrangers

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. Dominguès. A partir de là, une décision sur l'orthographe à adopter sera prise.

Certaines réalisations spéciales peuvent avoir lieu, dans ce cas il faut utiliser la balise « Prononciation » : *[pron=xxx]* pour signaler la prononciation réalisée (en utilisant l'orthographe), par exemple : *katchoup* > *ketchup* *[pron= katchoup]*

S'il s'agit 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, il faut précéder le mot du caractère &, par exemple : *&chelou, &bravitude*

Les noms de lieux prononcés en langue étrangère, en l'occurrence en espagnol et portugais, ne sont pas annotés comme étant prononcés en langue étrangère (ils seront reconnus directement par les gazetiers).

S'il s'agit de toute évidence (indices syntaxiques et sémantiques) d'un nom de lieu et que vous ne trouvez pas de trace dans les gazetiers, dictionnaires, sur internet, il faut seulement marquer l'emplacement par une balise de prononciation : *[pron=pi]*.

Les mots étrangers et qui ne sont pas des noms de lieux sont nombreux dans les récits de vie, ils peuvent être isolés ou former des phrases complètes. Ils seront transcrits en plusieurs phases.

Dans la première phase de transcription (transcription "pure"), la place de ces mots doit être marquée dans la transcription par une double balise *[lg=espagnol-]* et *[-lg=espagnol]*, exemple : *[lg=espagnol-]descaradamente [-lg=espagnol]*

Dans la deuxième phase (après la construction de la transcription enrichie), ces mots sont transcrits par un hispanophone : la transcription est placée entre la double balise positionnée précédemment *[lg=espagnol-]* et *[-lg=espagnol]* et la traduction est ajoutée tout de suite après la balise *[-lg=espagnol]*, entre deux balises : *[trad-]* et *[-trad]*

[lg=espagnol-] pero la tierra [-lg=espagnol] [trad-] mais la terre [-trad] c'est pas de nous la terre c'est de [lg=espagnol-] el terrateniente [-lg=espagnol] [trad-] le propriétaire terrien [-trad] comme on l'appelle c'est le grand [lg=espagnol-] señor [-lg=espagnol] [trad-] maître [-trad] de l'Andalousie tu comprends ?

Pour tenir compte des mots français dont la prononciation ou la flexion ont été hispanisées ou des mots espagnols qui ont été francisés, des balises spécifiques ont été créées (elles seront utilisées principalement au moment de la transcription/traduction des séquences en langue espagnole) :

- **prononciation "à la française" d'un mot espagnol**, balise *[pronFran=xxx]* comme dans :

(VP12) : avant la guerre avant l'arrivée de la République et ça vous coûtait une amende [pronFran=multe] de trois pesetas cinq pesetas

- **prononciation "à l'espagnol" d'un mot français**, balise *[pronHisp=xxx]* comme dans :
une épicerie [pronHisp=epiceria]

Ambiguïtés

En cas d'hésitation entre deux formes, le recours au contexte, et en particulier la concordance des temps, doit permettre de faire certains choix (non exclusifs). Si cela ne suffit pas, choisissez la forme qui selon vous convient le mieux. Par exemple : ça était > ça a été

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. Dominguès. Cette liste sera peu à peu complétée par les occurrences que vous trouverez dans les transcriptions. Ne pas mettre de majuscule sur ces termes quelle que soit leur place dans le discours.

<i>[ah]</i>	<i>[aïe]</i>	<i>[bah]</i>	<i>[beh]</i>	<i>[ben]</i>
<i>[bof]</i>	<i>[bouh]</i>	<i>[boum]</i>	<i>[clac]</i>	<i>[eh]</i>
<i>[euh]</i>	<i>[hé]</i>	<i>[hein]</i>	<i>[hop]</i>	<i>[hou]</i>
<i>[miam]</i>	<i>[mouais]</i>	<i>[oh]</i>	<i>[ouais]</i>	<i>[ouf]</i>
<i>[ouille]</i>	<i>[pff]</i>	<i>[zut]</i>		

6. Rétablissement des mots et des constructions

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*
- *tout ce qui y a > tout ce qu'il y a*
- *guide savoir-vivre > guide de savoir-vivre*

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.

Exemples pris dans ESLO1 :

- *tendez > attendez*
- *fin > enfin (passim)*
- *endez > attendez [pron = endez]*

Les apocopes ne sont pas rétablies.

- *perf* 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 certains types de prononciation sont généralisées, l'indiquer dans le fichier *Remarques* : par exemple, dans VP12, prononciation quasi systématique *república* (quelques emplois cependant de *république*)

7. 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, elle doit être corrigée et la forme prononcée par le témoin indiqué dans la balise « Prononciation » ([pron=xxx]), par ex. :

- oblette > omelette [pron=oblette]
- rénuméré > rémunéré [pron=rénuméré]
- menthélisées > mentholées [pron=menthélisées]

Déformations volontaires (métalinguistique)

Lorsque des formes sont 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

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

8. Utilisation des balises « Bruit »

Les bruits sont à indiquer par les balises suivantes qui ne couvrent pas tous les cas, donc certaines marques repérables dans le discours (insistance, agacement...) ne pourront être annotées.

Rires

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

Conversation

Au cours d'un tour de parole, il peut arriver qu'une (ou plusieurs) personne(s) parle(nt) en même temps que le locuteur mais sans que cela perturbe la compréhension : il s'agit d'une conversation de fond qui sera marquée en insérant une balise [conv] au début de cette conversation, la fin de la conversation n'étant pas balisée.

Micro

Elle est balisée [mic].

Soufflerie

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 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.

9. Dysfluences : répétitions et révisions

Les dysfluences (répétitions et révisions) sont annotées dans une deuxième phase pour la production de la transcription enrichie. Elles sont encadrées par les balises correspondantes : *[revis-]xxx[-revis]* et *[repetition-]xxx[-repetition]*.

Pour annoter une répétition, il est plus commode de mettre la balise sur 1ère séquence (celle qui va être répétée) plutôt que sur la 2ème, par ex. :

(EB06) *c'est-à-dire qu'il y avait une fois [repetition-]par[-repetition] par semaine*

Pour annoter une révision, puisque c'est la dernière réalisation qui est conservée, la balise de révision est donc placée sur la ou les premières réalisations.

[revis-]on a[-revis] on est allé dans la banlieue de Barcelone

Quand il y a deux révisions à la suite, on n'en balise qu'une seule ; par ex. :

(EB06) *[revis-]il y a parce que[-revis] c'est la route de Barcelone*

Lorsque la révision correspond à un ajout de précision, la révision n'est pas notée, par ex. :

(EB06) *la plupart du village la majorité des gens du village*

Lorsqu'une phrase interrogative du chercheur est terminée par l'interviewé, pour qu'il n'y ait pas de rupture de construction, la fin de la question du chercheur peut être notée comme une révision :

(EB06) *Et c'était un Espagnol qui les commandait ces Maures-là [revis-] ou [-revis] ?*

10. Exemples de synthèse :

[revis-]on a[-revis] on est allé [repetition-]dans[-repetition] dans la [euh] banlieue de Barcelone [revis-]diriger[-revis] il allait [mic] diriger [revis-]un atelier un[-revis] une usine de textile qui travaillait pour l'armée [mdd-]quoi[-mdd] [revis-]c'est-à-dire que[-revis] à l'époque il faut dire que [revis-]les[-revis] la coupe du tissu était faite par des hommes ils étaient tous à la guerre ou presque tous

sur [repetition-]le[-repetition] le chemin il me dit tu sais maintenant ça c'est terminé tu lèves plus le poing [mdd-]là[-mdd] [lg=espagnol-]ni se te ocurra[-lg=espagnol] [trad-]n'y pense même plus[-trad] [rire] tu lèves plus le poing [mdd-]là[-mdd] [hein] c'est terminé [mdd-]là[-mdd] [hein] [rire]

