

## CFPR – Conventions de transcription

### 1. Introduction

La transcription d'un enregistrement sonore tente d'être la plus fidèle possible aux paroles prononcées par les locuteurs. Néanmoins, elle ne peut pas refléter l'enregistrement avec une fidélité parfaite : Blanche-Benveniste & Jeanjean (1996), dont notre transcription orthographique reprend les principes, soulignent l'illusion auditive à laquelle même un transcripateur expert est sujet : « on écoute ce qu'on s'attend à écouter » (*idem* : 106). Les conventions de transcription adoptées ici cherchent à répondre à cette difficulté, ainsi qu'à offrir une transcription aussi lisible que possible afin de permettre différents types d'annotations et analyses. Nous appelons cette transcription « orthographique » car elle respecte l'orthographe standard sans aménagement.

Nous présenterons la transcriptions en trois points : les principes orthographiques (2), les symboles spéciaux (3) et l'anonymisation (4).

### 2. Transcription orthographique

La transcription adopte les principes généraux suivants :

- 1) L'orthographe est standard.
- 2) La transcription comprend tout ce qui est dit par le locuteur, et seulement ce qui est dit par le locuteur.

L'orthographe standard exclut tout truquage orthographique pour approximer la prononciation du locuteur, par exemple « l'gouvern'ment ». Chaque élément verbalisé est standardisé. Tout élément verbalisé est également transcrit : cela inclut les hésitations, les répétitions, etc. Inversement, si un élément n'est pas verbalisé, il n'est pas transcrit : par exemple un « ne » de négation non prononcé ne sera pas ajouté.

Dans le détail :

- Pas de ponctuation. Les majuscules ne sont conservées que pour les noms propres.
- Les majuscules ne sont conservées que pour les noms propres (David), les sigles (S.N.C.F.) et les acronymes (CROUS). Le point « . » n'est conservé que pour les sigles.
- Les nombres sont transcrits en toutes lettres (quatre-vingt-dix, nonante, etc.) à l'exception des années (1963, etc.).
- La réalisation phonique prime. Les accords non-standards sont transcrits fidèlement : « tu as vu des chevaux » ; « ils croient que c'est vrai ». Les liaisons non réalisées ne sont pas transcrites, « parce que il est pas là » ; les liaisons particulières sont indiquées par un trait d'union : « donne-moi-z-en ».

- Les phatiques et les onomatopées sont codifiés et transcrits selon l'orthographe fournie :  
*ah, aïe, areu, atchoum, badaboum, baf, bah, bam, bang, bé, bêêê, beurk, ben, bing, bon, boum, broum, cataclop, clap clap, coa coa, cocorico, coin coin, crac, croa croa, cuicui, ding, ding deng dong, ding dong, dring, hé, hé ben, eh bien, euh, flic flac, flip flop, frou frou, glouglou, glou glou, groin groin, grr, hé, hep, hi han, hip hip hip hourra, houla, hourra, hum, mêêê, meuh, miam, miam miam, miaou, oh, O.K., ouah, ouah ouah, ouais, ouf, ouh, paf, pan, patatras, pchhh, pchit, pff, pif-paf, pin pon, pioupiou, plouf, pof, pouet, pouet pouet, pouf, psst, ron ron, schlaf, snif, splaf, splatch, sss, tacatac, tagada, tchac, teuf teuf, tic tac, toc, tut tut, vlan, vroum, vrrr, wouah, zip.*
- Les mots en langue étrangère sont transcrits selon la norme de la langue d'origine.
- Quelques truquages sont adoptés et systématisés :
  - Pronom personnel élidé « tu as / t'as »
  - Expression élidée « il y a / y a »
  - Adverbe élidé « puis / pis »Inversement « enfin », « ils » et « parce que » ne sont jamais élidés. Ces truquages, comme les accords et liaisons, sont aménagés pour faciliter l'analyse.

Nos conventions n'adoptent pas le choix multiple : quand plusieurs transcriptions d'un même passage sont possibles, nous adoptons la plus plausible ou nous considérons le passage inanalysable.

### 3. Symboles spéciaux

Outre le point « . » réservé pour les sigles, la transcription réserve une série de symboles pour des cas particuliers :

- Tiret (-) :  
En fin de mot pour signaler une troncation : « une mais- maison ».
- Tiret bas ( \_ ) :  
Isolément pour signaler un passage non verbalisé (pause) : « oui \_ hier »
- Croisillon (#) :  
Isolément pour signaler un passage anonymisé : « j'ai vu # hier »
- Pourcent (%) :  
Isolément pour signaler un passage inanalysable : « il a % et ensuite »

Note : Techniquement, la transcription est découpée en « segments ». Un segment verbalisé peut contenir des lettres [a-Z], des chiffres [0-9] et les symboles [.-] ; les symboles [\_#%] doivent occuper seuls un segment à part.

#### **4. Pseudonymisation**

Suivant les principes du RGPD (Règlement Général sur la Protection des Données), nous procédons à la pseudonymisation des transcriptions.

En pratique, nous éliminons de la transcription tous les noms propres de personnes et de lieux permettant d'identifier directement des individus, ainsi que tout passage de même nature. Nous ne pouvons exclure que l'identité du locuteur puisse être reconstruite, ne serait-ce que par la voix, et ne pouvons que rendre cette ré-identification aussi difficile que possible.

Les passages anonymisés le sont dans la transcription via le symbole du croisillon (#, voir point (3)) et dans l'enregistrement par une déformation du signal (Hirst 2013).