### Version définitive des rubriques des métadonnées remplies de CC07

=====================================================================

Sur les fichiers eux-mêmes (fichier audio unique .mp3 et le fichier .cha qui contient la transcription alignée avec l’audio)

L’alignement entre les fichiers texte et audio a été effectué avec un séquençage de 20 secondes environ.

Informations générales sur le mode de transcription

Par rapport aux premières versions, la plupart des passages <pron=pi> ont été compris et transcrits selon une interprétation. Exemples de nouvelles transcriptions par rapport aux premières versions :

* *c'est exact <pron=pi> l'avantage* (avant : “c'est exact mais gamin j'ai eu l'avantage”)
* *une éducation riche <pron=nonstd>* (avant : “une éducation stricte <pron=nonstd>”)
* *une bataille qui a été sanglante d'ailleurs* (avant : “<pron=pi> d'ailleurs”)
* *commissaire de police qui était dans la lutte* (avant : “commissaire de police qui était dans la rue <pron=lieu>”)

Afin d’évaluer les variations entre transcripteurs, une écoute par un 2ème transcripteur a été effectuée sur la version transcrite et annotée telle que déposée. Les variations sur les premières lignes du fichier texte (environ 300 lignes) (version déposée → remarque ou autre proposition d’un 2ème transcripteur) (les numéros de lignes sont indicatifs et susceptibles d’avoir changé en fonction de corrections ultérieures) :

* l.12, l.13 les balises <pause> indiquées → ce sont surtout des respirations, mais peuvent être annotées comme des pauses
* l. 18 “à <repetition> à” → pas sûr qu’il y ait deux “à” dans
* l.26 “<euh> le 25 août 2009” → plutôt qu’un <euh>, un “leeeee”
* l.48 “à écrire à et à beaucoup” → “à écrire <revision> à <\revision> et à beaucoup”
* l. 59 “ou avant la guerre d'” → “<repetition> ou <\repetition> ou avant la guerre d'”
* l.60 <repetition> eh non eh non </repetition> → répétition ou emphase comme pour la l.38 “<emphase> oui oui </emphase> oui”
* l.118 “<euh> c'était dans la région” → “<euh> <mic> c'était dans la région” (balise <mic> non soumise à indication exhaustive)
* l.164 “eues à ce moment-là ?” → “<euh> eues à ce moment-là ?”
* l.186 “vous coupiez” → “<revision> vo- <\revision> vous coupiez”
* l.192 “<repetition> que le le </repetition>” → “<revision> que le le </revision>” ?
* l.198 “pas mis une hache” → ”pas <revision> me- </revision> mis une hache”
* l.215 “pouvoir rejoindre” → “pouvoir <si> rejoindre” ?
* l. 227 “ça comme si” → “<rireentendu> ça comme si” (ou plutôt il parle en riant, créer une balise double <rireparle></rireparle> ?)
* l. 228 “<repetition> une fi- </repetition> une fille” → “<revision> une fi- </revision> une fille” (révision si mot coupé, comme en ligne 242)
* l.299 “moi disons que” → prononcé “queee”
* l. 325 “que elle elle savait lire” → peut-être “que <repetition> elle </repetition> elle savait lire”

Beaucoup de variations correspondent à des balises non soumises à indication exhaustive dans le guide de transcription (ex. <mic> <si> <pause>). La diction n’est parfois pas assez nette pour distinguer entre prononciations des mots ou des révisions.

Contexte de l’enregistrement

Remarques sur l’oral (en relation avec le guide de transcription)

Exemples de corrections fréquentes :

* ajout des balises mdd : *<mdd> enfin </mdd>*
* Remplacements entre balises <repetition></repetition> qui sont en fait des <revision></revision> et inversement
* Ajouts de balises <emphase></emphase> et <rireentendu>
* Rectifications de la balise <pron=[API]>

Beaucoup d’interjections et d’onomatopées et notamment successives, ex. *<mdd> oh eh bien euh voilà </mdd>*. A VERIFIER : <DL> lorsqu'il y a une succession de mdd, d'interjections/onomatopées, on groupe tout dans un même <mdd></mdd>. Si oui, est-ce qu'on laisse/ajoute les chevrons des interjections/onomatopées, ou est-ce qu'on les efface ? <mdd> <oh> <eh> bien <euh> voilà </mdd> --- <mdd> oh eh bien euh voilà </mdd> Je re-pose la question, juste pour être sûre. Car il y a beaucoup de cas groupés et non groupés, donc ça fera beaucoup de modifications dans les deux cas.

Plusieurs successions ou d’inclusion de balises <revision></revision>, <repetition></repetition> et <mdd></mdd>. Dans ce cas, selon le guide de transcription, tout est regroupé dans <revision></revision>. Ex. *<revision> comme moyen de de voilà </revision>* donc sans indiquer <repetition> de </repetition> ni <mdd> voilà </mdd>. A CONFIRMER.

Ajouter les remarques de Carmen sur l'espagnol.

Incertitudes dans la transcription mais une interprétation est choisie

Une incertitude quand au choix de transcription concernant *en 37* dans *</mdd> fidélité en 37 au drapeau <bc> en 37 <bc> <revision> il a*.

Interprétation des Npr de lieux et de personnes

Pour les noms de lieux :

Pour les noms de personnes :

A VÉRIFIER QUE PRIS EN COMPTE :

* 916 Julia/NP165 et Julieta/NP165SUR

<DL> est-on sûr que le deuxième prénom est le surnom du premier ? Ou est-ce juste une révision du locuteur ?

<CD> pour moi oui, Julieta est bien le diminutif de Julia, et donc considéré comme un surnom.

### Commentaires Déborah puis Carmen, Catherine, Laurence

=====================================================================

Commentaires Déborah/réponses, et Carmen, Catherine, Laurence

Notes Déborah

===========

CC07

Ajout du header (@Begin ... @End)

Ajout d'une \* à chaque tier

Alignement avec séquençage de 20 secondes environ

Balisage des NprLieu avec les identifiants de gazetier : ok

Vérification des codes de noms propres de personnes : ok

==================================================================================================================

Corrections systématiques :

==================================================================================================================

Tous les [] remplacés par des <> (<euh>, <silence>, <pause>, <pron=nonstd>, <pron=pi>...)

<mdd> euh/hein </mdd> --- <euh/hein>

<mdd> et bien </mdd> --- <mdd> eh bien </mdd>

<mdd> mais/oui </mdd> ? >> corriger si

Ajouts de majuscules, par exemple :nécessaire

<mdd> ah/eh/hé/oh/bah </mdd> --- <ah/eh/hé/oh/bah>

<rire/silence/pleur> --- <rire/silence/pleurentendu>

<revis> </revis> --- <revision> </revision>

XXX là --- XXX-là

connait/parait/complètement/coté --- connaît/paraît/complétement/côté

<audiomanquantsep/fus> --- effacés

les républicains --- les Républicains

les espagnols --- les Espagnols

==================================================================================================================

Corrections notables :

==================================================================================================================

Exemples de retranscription :

38-39 c'est exact mais gamin j'ai eu l'avantage

c'est exact <pron=pi> l'avantage

44 une éducation stricte <pron=nonstd>

une éducation riche <pron=nonstd>

81 <pron=pi> d'ailleurs

une bataille qui a été sanglante d'ailleurs

254 commissaire de police qui était dans la rue <pron=lieu>

commissaire de police qui était dans la lutte

Exemples de corrections fréquentes :

enfin --- <mdd> enfin </enfin>

<repetition></repetition> qui sont en fait des <revision></revision> et inversement

Ajouts de <emphase></emphase> et <rireentendu>

Rectifications de <pron=[API]>

La plupart des passages <pron=pi> compris et transcris

<CD>

==================================================================================================================

Questions/Réponses :

==================================================================================================================

<DL> lorsqu'il y a une succession de mdd, d'interjections/onomatopées, on groupe tout dans un même <mdd></mdd>

Si oui, est-ce qu'on laisse/ajoute les chevrons des interjections/onomatopées, ou est-ce qu'on les efface ?

<mdd> <oh> <eh> bien <euh> voilà </mdd> --- <mdd> oh eh bien euh voilà </mdd>

Je re-pose la question, juste pour être sûre. Car il y a beaucoup de cas groupés et non groupés,

donc ça fera beaucoup de modifications dans les deux cas.

<LJ> On regarde avec Catherine, car je me suis posée la question également...

<DL> Même question avec <revision></revision> :

lorsqu'il y a une succession de <revision> <repetition> et/ou <mdd>, on groupe tout dans <revision> ?

En laissant/ajoutant les balises et/ou chevrons ?

Exemple : 509-510 <revision> comme moyen <repetition> de </repetition> de <mdd> voilà </mdd> </revision>

--- <revision> comme moyen de de voilà </revision>

<LJ> à confirmer, mais normalement je crois que lorsque c’était en révision, on enlève toute indication de mdd, comme l’idée est de ne pas traiter (automatiquement) les tronçons entre les balises de révision.

916 Julia/NP165 et Julieta/NP165SUR

<DL> est-on sûr que le deuxième prénom est le surnom du premier ? Ou est-ce juste une révision du locuteur ?

<LJ> Voir Catherine pour les noms de personnes.

<CD> pour moi oui, Julieta est bien le diminutif de Julia, et donc considéré comme un surnom.

855 </mdd> fidélité [pron=pi] au drapeau <bc> [pron=pi] <bc> <revision> il a

</mdd> fidélité en 37 au drapeau <bc> en 37 <bc> <revision> il a

<DL> je pense entendre "en 37", qu'en pensez-vous ? Sinon je remettrai <pron=pi>

<LJ> je crois que je n’aurai pas deviné, mais en écoutant, ça me semble être cela. En tout cas, c’est plausible.

156 Argelès-sur-Mer

<DL> j'ai mis le même id que pour "Argelès", mais il faudrait peut-être compléter le gazetier en y ajoutant "Argelès-sur-Mer"

<LJ> oui merci, je l’ajoute au dictionnaire sur le ftp. C’est possible que des oublis se reproduisent comme les dictionnaires avaient été construits sur les premières versions des transcriptions, et certains noms de lieux ont pu ne pas être indiqués à ce moment-là. (Note : pour l’annotation automatique, cela créé des doublons : Argelès-sur-mer est annoté avec Argelès et Argelès-sur-mer).

<DL> Désormais je vérifie les codes de noms propres de personne (je réviserai également FR14 et FS15).

<LJ> Ok

==================================================================================================================

Remarques à voir ensemble à la fin ou pour la suite (par exemple pour donner dans les métadonnées par récit, une indication de la difficulté de la transcription et de la correspondance entre l’audio et la transcription ?) :

==================================================================================================================

<LJ> 4/10/2019 : écoute CC07.cha ligne 1 à ligne 335, version de Déborah du 4/10/2019

<pause> et <silenceentendu> à remplacer par <si>

l.12, l.13 : les <pause> sont surtout des respirations ? mais pour moi, on peut garder en pause.

l. 18 : je ne suis pas sûre qu’il y ait deux “à” dans “à <repetition> à”

l.26 : “<euh> le 25 août 2009” : plutôt qu’un <euh>, un “leeeee”, comment l’indiquer ?

l.48 : “à écrire à et à beaucoup” : “à écrire <revision> à <\revision> et à beaucoup” ?

l. 59 : “ou avant la guerre d'” : “<repetition> ou <\repetition> ou avant la guerre d'” ?

l.60 <repetition> eh non eh non </repetition> : répétition ou emphase comme pour la l.38 “<emphase> oui oui </emphase> oui” ?

l.118 “<euh> c'était dans la région” → “<euh> <mic> c'était dans la région” (mais ce n’est pas très important, car beaucoup de cas et non traité ensuite a priori)

l.159 “imposaient <silenceentendu> et là évidemment” : durée similaire aux <pauses> précédentes ? → “imposaient <si> et là évidemment”

l.164 “eues à ce moment-là ?” → “<euh> eues à ce moment-là ?”

l.186 “vous coupiez” → “<revision> vo- <\revision> vous coupiez”

l.192: “<repetition> que le le </repetition>” → “<revision> que le le </revision>” ?

l.198 : “pas mis une hache” → ”pas <revision> me- </revision> mis une hache”

l.215 : “pouvoir rejoindre” → “pouvoir <si> rejoindre” ?

l. 227 : “ça comme si” → “<rireentendu> ça comme si” (ou plutôt il parle en riant, créer une balise double <rireparle></rireparle> ?)

l. 228 : “<repetition> une fi- </repetition> une fille” → “<revision> une fi- </revision> une fille” (révision si mot coupé, comme en ligne 242)

l.299 : “moi disons que” : indiquer queee ?

l. 325 “que elle elle savait lire” → “que <repetition> elle </repetition> elle savait lire” (pas sûr)