### Version définitive des rubriques des métadonnées remplies de FR14

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

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

L’alignement a été effectué avec un séquençage d’environ 20 secondes.

Dans les transcriptions initiales, des parties audio manquantes étaient indiquées par une balise *<audiomanquantsep>*. Ces balises ont été effacées car a priori il n’y a pas d’audio manquants à ces endroits-là.

Certains cas où le son du micro baisse et semble s’éteindre. À la reprise du son, il semble que l’on ait raté des phrases mais pas de texte retranscrit. PAS DE BALISES pour cela.

Informations générales sur le mode de transcription

Des cas de coupure de mots, pour lesquelles un trait d'union après le mot coupé a été ajouté par rapport aux transcription initiales, ex : *et <*revision*> ap- </*revision*> après* (ici balise *<revision>* et non *<repetition>* qui est uniquement pour un (des) mot(s) complet(s))

Des cas d’utilisation de <eh> / <hé>

Rq Déborah (?)

>> conservé pour l’interpellation ?

>> même entrée dans le TLFi,

>> par souci de cohérence avec « eh bien », choisir systématiquement la forme « eh »

Contexte de l’enregistrement

Par rapport à la transcription initiale, des chevauchements entre les tours de paroles ont été indiqués avec la balise <bc> ("BackChannel"). Ces cas correspondent au moment où un des locuteurs (L1) parle et un autre locuteur (L2) se contente d'acquiescer en même temps (par exemple en utilisant *oui* ou *mmh mmh* et rien d’autre). Il n’y a pas de nouveau tour de parole créé pour L2 mais une balise <bc> est ajouté dans le tour de paroles de L1.

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

Peu de mots espagnol sont employés.

Ajouter d’autres remarques de Carmen sur l'espagnol.

Incertitudes dans la transcription mais une interprétation est choisie

Interprétation des Npr de lieux et de personnes

Pour les noms de lieux :

Laurence : “Pour certains récits, existence entre crochets de [Bordeaux], par exemple "fort du Hâ [Bordeaux]", alors que Bordeaux n'est pas cité : c'était pour l'annotation automatique, mais ces balises sont à supprimer (vous pouvez les supprimer, mais nous pouvons le faire ensuite aussi)...”

Des cas qui ont été corrigés par rapport aux transcriptions initiales :

* camp de Hennebont (et non camp de Embourg)
* Lloret de Mar (et non Lloret del Mar)

Carmen : ajout d’une balise sur le lieu *Val d’Aran* dans la transcription → voir si besoin d’ajouter l’identifiant dans le dictionnaire de référence des NPr par récit.

Pour les noms de personnes :

<DL> Code Npr : 197 “Quesada” : à anonymiser ? (FR14 se demande si c’est le vrai nom de NP137SUR)

359 “FR14FAM” ?

NP136 = NP139 + NP141 ? Supprimer NP136 ?

### Commentaires Déborah puis Carmen, Catherine, Laurence

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

Commentaires Déborah/réponses, et Carmen, Catherine, Laurence

Notes Déborah

===========

FR14

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

Ajout d'une \* à chaque tier

Alignement avec séquençage de 20 secondes environ

<CD> raison objective pour cette durée ?

<DL> inspiré d'une recommandation de Transcriber, un autre logiciel de transcription.

Balisage des NprLieu avec les identifiants de gazetier : ok

Vérification des codes de noms propres de personnes : ok

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

Corrections effectuées :

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

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

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

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

<LJ> ok, dans "<mdd> hein mais </mdd>" l. 57, est-ce que c'est prévu d'indiquer <hein> ?

<DL> <hein> est déjà dans le guide de transcription

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

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

rire / [rireentendu] --- <rireentendu>

<revis> --- <revision>

<LJ> oui, juste, je pense que l. 118, "<revision> c'étaient les vacances </revision>" n'est en fait pas une révision.

<DL> effectivement, l.118 rectifiée.

<espagnol> --- <lg=espagnol>

>> par une expression régulière ?

>> les balises uniques <espagnol> suivent immédiatement le mot auquel elles s’appliquent ; est-ce possible de les remplacer en utilisant une expression régulière ?

XXX là --- XXX-là

<LJ> oui mais pour certain cas, pas forcément, ex. "le bonhomme là" où il y a un phrasé particulier mais à voir

<audiomanquantsep> --- effacés

<LJ> pourquoi ?

<DL> parce qu'il n'y a plus d'audio manquant à ces endroits-là

Ajouts de majuscules, par exemple :

les républicains --- les Républicains

les espagnols --- les Espagnols

Exemple de retranscription :

l. 7-9 (comparer FR14\_V0.cha et FR14.cha)

Exemples de correction :

l.58 : (trait d'union après un mot coupé)

"et <repetition> ap- </repetition> après"

>> oui

l. 27 : pour l'aider <euh> <revision> pour <euh> à </revision> avec ses enfants

<euh> pour l'aider <revision> pour <euh> a- </revision> avec ses enfants

l. 38-40 : mais il n'y avait pas que des

Espagnols il y avait des Italiens il y avait des Belges il y avait

de tout

mais y avait pas que des

Espagnols y avait des Italiens y avait des Belges y avait

de tout

Exemple de remplacement de chevauchements/tours de paroles avec une balise <bc> ("BackChannel") : l.767-769

<LJ> pourquoi est-ce lié aux tours de paroles ? le <bc> est bien lié au son ? Est-ce qu'il ne faut pas une balise double <bc>...</bc> ?

<DL> le <bc> s'utilise quand un locuteur1 parles et qu'un locuteur2 se contente d'acquiescer en même temps que L1 parle

(L2 dit "oui" ou "mmh mmh" et c'est tout). Dans ce cas, on ne crée pas un tour de parole juste pour L2 qui dit "oui" ou "mmh".

On ajoute simplement <bc> (pas </bc>) dans le tour de L1.

Exemples avec <emphase> :

342. le gars me prend la main parce qu'il voulait m'amener <hein>

<repetition> il voulait m'amener </repetition> et le patron

le gars me prend la main parce qu' <emphase> il voulait m'amener </emphase> <hein>

il voulait m'amener et le patron

600. <emphase> oui </emphase> oui parce que <repetition> y a

Exemples divers :

472. <mdd> ah </mdd> d'accord

<ah> au Fort du Hâ <mdd> oui </mdd>

611-613. c'étaient pas des valises les

<lg=espagnol> baùl </lg=espagnol> <trad> </trad> <revision> elle

disait

>> modification effectuée, que je ne ferai donc plus dans les prochaines transcriptions.

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

Questions/Réponses :

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

-Remplacer <silence> par <silenceentendu> ?

<LJ> oui pour ne pas que silence soit reconnu par l'annotation automatique. Ou alors <si> qui regroupe pauses et silences ?

-Révision ou répétition :

l. 58 : et <repetition> ap- </repetition> après

l. 68-69 : <mdd> donc oui </mdd> <revision> nous a- </revision> nous avons fait

>> revision (repetition uniquement pour mot(s) complet(s))

-l. 91-96 : "oui" considéré comme mdd ?

<LJ> J'avoue que je ne trouve pas cela évident pour les mdd. Ici pour les "oui", ils indiquent une confirmation de ce qui est dit avant, une manière d'appuyer. Personnellement je ne les garderai pas, mais selon le guide de transcription et selon comment ont été pensés les mdd dans le guide de transcription, il faudrait conserver les mdd. Peut-être qu'il faut que vous définissiez une "règle" selon les entretiens ou pour le corpus. De mon côté, j'avais par exemple enlevé tous les mdd sur les oui/non/c'est ça, qui étaient des réponses à des questions.

<DL> c'est ce que je pense aussi.

-Dans quel cas <eh> / <hé> >> conservé pour l’interpellation ?

>> même entrée dans le TLFi,

>> par souci de cohérence avec « eh bien », choisir systématiquement la forme « eh »

-Gazetier : par exemple avec l'Espagne, l'identifiant est "5-id=2510769" ou "2510769" ?

>> format d’annotation : <lieu id=2510769>Espagne </lieu>

-Les lieux : des cas comme "la gare", "Saint-Jean", "la gare Saint-Jean", "un hôpital en Espagne",

"chez madame FR14", "les écoles", "la cour de l'école" ?

Exemple l. 75 : "à Saint-Jean là la gare Saint-Jean ?"

Faut-il un code ? Une balise <lieu=identifiant> la gare Saint-Jean </lieu> ?

>> la prestation concerne l’annotation des lieux Npr

<LJ> En fait, les balises de lieux ne concernent que les entités nommées de granularité à la ville ou au-dessus (département, région...). Certaines EN sont peut-être d'une plus petite granularité (par exemple les chantiers navals) mais ils sont normalement indiqués dans les dictionnaires de Npr par entretien. Si vous avez un doute, vous pouvez les noter à part et nous demander. Pour certains récits, j'ai ajouté entre crochets [Bordeaux], par exemple "fort du Hâ [Bordeaux]", alors que Bordeaux n'est pas cité : c'était pour l'annotation automatique, mais ces balises sont à supprimer (vous pouvez les supprimer, mais nous pouvons le faire ensuite aussi)...

-Tout grouper dans une même <mdd>, y compris onomatopée/interjection ? --- <mdd> hein mais </mdd>

Ou séparer ? --- <hein> <mdd> mais </mdd>

>> Oui, tout grouper.

-L'heure, en chiffres ou en lettres ? >> "4 heures du matin"

-Emphase prosodique ? --- 406-407. il a été <emphase> tué

</emphase> celui-là

>>> non, uniquement des emphases lors de répétition.

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

Nouvelles remarques/questions :

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

815, 831. <audiomanquant> : le son du micro baisse et plus rien. Quand le son revient, il semble qu'on ait est raté des phrases.

<DL> Mais ce n'est pas le même cas de figure que <audiomanquantfus>, il n'y a pas de texte transcrit correspondant à ces blancs.

<LJ> C’est vrai que cela arrive : est-ce que l’on crée une nouvelle balise <coupure> ou cela demande trop d’interprétation ?

678-679. les Allemands sont partis ça était une liesse

les Allemands sont partis ça a été une liesse

<DL> Cas de figure qui revient plusieurs fois : c'est peut-être en lien avec ce dont vous parlez dans le guide de transcription (passé simple et imparfait). Dans ce genre de cas, je mettrais du passé composé, je pense que c'est ce que l'interviewée veut exprimer.

<CD> c'est aussi mon avis, et c'est effectivement ça que je voulais dire dans le guide d'annotation

<LJ> Oui pour le passé composé de mon côté

717. la marine

la Marine

<DL> mot à mettre en majuscule ?

<CD> je pense que oui

<LJ> Oui : ajouté au guide de transcription

<DL> Il y plusieurs fois ce cas :

821-822. mais demandez ça à <repetition> NP152 </repetition>

NP152 il est encore en vie il pourra vous le dire

En fait ce n'est ni une répétition, ni une emphase. L'interviewée termine un 1er énoncé et en commence un 2e avec une reprise.

Donc, est-ce que je laisse <repetition>, je remplace par <emphase>, ou je ne mets rien...?

<CD> d'accord pour rien, en effet ce n'est pas une répétition

<LJ> c’est vrai : je dirais aucune balise

Gazetier :

Je n'ai pas trouvé d'occurrences de "Hennebont" (qui est dans le gazetier).

En revanche j'ai une occurrence de "camp de Embourg" 657 (qui n'est pas dans le gazetier).

Dans le gazetier il est écrit "Lloret", mais dans la transcription c'est "Lloret Del Mar" 717, 719.

<CD> je crois bien que c'est "LLoret de Mar", le nom de la ville (pas "del") ; il faut vérifier que l'identifiant dans le gazetier de l'entretien correspond bien à celui de "LLoret de Mar" dans Geonames.

Je laisse Laurence répondre pour "Hennebont" et "camp de Embourg"

<LJ> effectivement, ce n’était pas corrigé dans la transcription mais elle dit bien « camp de Hennebont ». Pour « Lloret Del Mar », c’est une erreur car la ville se nomme bien (Lloret de Mar, j’ai mis à jour le dictionnaire sur le FTP.

<DL> Code Npr :

197 “Quesada” : à anonymiser ? (FR14 se demande si c’est le vrai nom de NP137SUR)

359 “FR14FAM” ?

NP136 = NP139 + NP141 ? Supprimer NP136 ?

F5 option : je ne connaissais pas et ça m'a l'air pratique. Donc je vais tester aujourd'hui en alignant FS15.

En revanche, que j'utilise cette méthode ou la mienne, les bullets fonctionnent. Donc le bouton F4 (attention F4, et non pas F5), en ayant placé le curseur quelque part dans le paragraphe que l'on veut écouter, ça fonctionne.