### Version définitive des rubriques des métadonnées remplies de AV16

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

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

Le fichier audio a été généré à partir d’une fusion des fichiers segmentés initiaux fournis par le RAHMI, l’alignement étant impossible à partir du fichier mp3 unique du RAHMI. Par ailleurs, des bouts d’audio du fichier mp3 unique ont été ajoutés aux segments car ils manquaient dans les fichiers segmentés initiaux. Le fichier mp3 déposé ici correspond à une durée de 3 h 24’ 49”.

L’alignement entre texte et audio n’a pu être effectué après 3 h 21’ 28” qui correspond au dernier repère. À certains endroits, l’alignement ne fonctionne pas directement (message : “Media marker not found at cursor position. If you are planning to link sound to text then please use Sonic mode “F5“). Il s’agit des cas où de nouveaux paragraphes ont été créés par rapport à la 1ère version de l’alignement et pour accéder à l’audio il faut placer le curseur dans le paragraphe initial correspondant ou sur la bullet elle-même directement. Cela pourrait être rectifié.

L'alignement n'est pas aussi propre qu’en moyenne pour les autres fichiers car :

* Il n’est pas évident de choisir où délimiter entre les interlocuteurs car beaucoup de chevauchements de paroles.
* Les corrections sont effectuées après l'alignement et, comme expliqué précédemment, il y a eu beaucoup de corrections notamment de retranscriptions qui font que parfois le texte coïncide moins avec l'alignement.

Informations générales sur le mode de transcription

Beaucoup de corrections ont été effectuées par rapport aux premières versions de transcriptions et d’annotations, concernant également les codes Npr, les noms de lieux, les passages en langue étrangère, des passages pas ou mal transcrits.

En conséquence, moins de temps a été consacré à annoter les répétitions de petits mots, dont certaines sont donc peut-être non annotées.

Contexte de l’enregistrement

Beaucoup de chevauchements de parole entre les interlocuteurs, CH\_EE1 se répète beaucoup et AV16 parle parfois en flux continu. Il est donc moins évident de choisir où délimiter entre les interlocuteurs.

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

Des passages contiennent plusieurs balises <mdd> propres à la mise en place de l’entretien, ex. *<mdd> là on peut couper là </mdd>*.

Utilisation de l’espagnol et de l’allemand :

* *<lg=allemand> frei Spanier </lg=allemand> <trad> Espagnols libres </trad>*
* *<lg=allemand> rote Spanier </lg=allemand> <trad> Espagnols rouges </trad>*
* *il disait <lg=allemand> Spanier Spanier Fliegeralarm Fliegeralarm </lg=allemand> <trad> l’Espagnol l’Espagnol aviation aviation alarme aux avions </trad>*

Ajouter les remarques de Carmen sur l'espagnol.

Incertitudes dans la transcription mais une interprétation est choisie

Des corrections par rapport aux premières versions :

* *à l'hôpital Saint-Louis de Perpignan* à la place de *à l'hôpital cette nuit de Perpignan*

Des choix mais incertains selon interprétations :

* *le trajet Caris* : cela pourrait aussi être d’après l’intonation “le trajet car ici”.
* *<lg=espagnol> se acabó </lg=espagnol> <trad> c'est fini </trad>* : cela pourrait être aussi “sale cabot".

Interprétation des Npr de lieux et de personnes

Pour les noms de lieux :

* Lieu correspondant à une gare entre deux communes : *San Clodio Quiroga*. San Clodio et Quiroga sont deux municipalités partageant une même gare de train nommée San Clodio-Quiroga. Annoté avec deux identifiants *<lieu id=3110764> San Clodio </lieu> <lieu id=3112395> Quiroga </lieu>* plutôt qu’un seul identifiant correspondant à la gare San Clodio-Quiroga (id=3110763). San Clodio est cité par ailleurs seul à un autre endroit.
* Lieu cité : *aux Aubrais*, prononcé [obroué], interprété comme le lieu de la gare des Aubrais à Fleury-les-Aubrais. L’identifiant associé est celui de la ville de Fleury-les-Aubrais (dans GeoNames id=3018280).
* *<lieu id=2982625> Rouffignac*. Il y a au moins deux Rouffignac en Dordogne : Rouffignac-Saint-Cernin-de-Reilhac et Rouffignac-de-Sigoulès. Ici c’est l’identifiant de Rouffignac-Saint-Cernin-de-Reilhac car plus d’habitants et certainement plus connu comme proche de Lascaux.
* Plusieurs lieux à Bordeaux ont été transcrits par rapport aux premières versions :
  + *Ravezies* : place et quartier de Bordeaux, de l'autre côté de la base sous-marine par rapport à la rue Achard
  + *Bassins à flot*, nom propre figé au pluriel et avec majuscule.
  + *arc de triomphe*, en minuscules car il s’agit de portes à Bordeaux qui ont une forme d'arc de triomphe, notamment la Porte de Bourgogne qui est sans doute le monument dont parle AV16.
  + *Grand Théâtre*, avec majuscules car monument à Bordeaux
* Des noms de fleuves et rivières qui pourraient être confondus avec des régions ou départements :
  + *<lieu id=3021044> Dordogne* (la rivière)
  + *<lieu id=2997869> Loire* (le fleuve et non un des départements contenant le nom Loire)

A VOIR : remplacer les lieux en France avec les id de la BDNyme ? (Id Geonames pour l’instant car plus facile à trouver)

Pour les noms de personnes :

A VOIR ET AJOUTER ?

* *<lg=espagnol> del Caudillo </lg=espagnol> <trad> du Chef </trad>* : <LJ> Je comprends que “el Caudillo” est Franco, est-ce qu’il ne faudrait pas indiquer cette appellation dans la liste des noms de personnes ? Avec une majuscule ?

<DL> 30-31 "Caudillo" ("Chef") : surnom de Franco, à remplacer par un code Npr (NP1SUR ?) ?

* "Hélène" --- "AUTRE"
* 906 "Voix de femme" --- "AUTRE2"
* 2613-2613 "Alvarez" --- "NP104"
* 2739-2740 "AV16LOC2" --- "NP111"
* 3098 "monsieur Vásquez" --- "monsieur NP107"
* 3592 "NP164" --- "NP166"
* 3839 "AV16AMI" --- "CH\_EE1"

DES INCERTITUDES SUR LES NPR :

* <DL> 485-486 "la mère s'appelait NP94 la fille aussi NP94"

Le premier NP94 correspond à "Matilde" (la mère), et le deuxième correspond à un prénom ressemblant, "Matildine" peut-être, (la fille). Faut-il un code NPr, autre que NP94, pour la fille ? Ajouter un code NP94FIE ?

* <DL> 643 "Corinne" : une des infirmières qui soignent AV16 ? Code Npr ?
* <DL> 1601 1608 "16LOC1" : Michel (le chauffeur), plutôt un code Npr "NP\*\*\*" ?
* <DL> 3002 3007 "l'ingénieur espion NP106" --- "l'ingénieur NP111"

Il n'y a pas d'autres occurrences de NP106, je ne crois pas qu'il y ait un "Sala", je pense que "Espinoza" (NP111) est l'ingénieur catalan (NP106).

* <DL> 3065-3068 "Gérard Catalan" : la deuxième identité d'AV16. Code Npr ? (AV16SUR ?)
* <DL> 3120 "AV16FIE" ? Ou plutôt AV16FEM ?
* <DL> 3181, 3326 : "AV16AMI"/"AV16AMI2" --- "NP108" : ajouter NP108 (Alberto Brandosa) à NP108SUR

### Commentaires Déborah puis Carmen, Catherine, Laurence

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

Commentaires Déborah/réponses, et Carmen, Catherine, Laurence

Notes Déborah

===========

AV16

Fichier audio

=========

Fusion à partir des fichiers segmentés (car alignement impossible à partir du fichier mp3 unique fourni par le Rahmi)

Déborah : le 24/10/2019 « Une question avec AV16 : il manque, j'estime, environ 2min20 d'enregistrement, lignes 191-241, entre les fichiers segmentés 3 et 4.

Soit il manque cet extrait-là dans les fichiers segmentés (il y aurait peut-être un 70e fichier mystérieux), soit deux extraits ou plus ont été intervertis (au moment de la segmentation, pas au moment de la fusion).

Si c'est fastidieux à vérifier pour Laurence (à son retour), je pourrai réviser AV16, sans l'aligner dans un premier temps, et ainsi voir ce qu'il en est.

J'ai ajouté dans Drive la transcription AV16 alignée jusqu'à la ligne 191, puis l'extrait manquant indiqué. »

Laurence :

Ok entre segments 1 et 3

Manque entre : 191-241 / segments 3 et 4. Dans le fichier brut Angel Villar « brut » du CD, l’audio existe bien entre 8 min 52 s et 11 min 22 s

Ok entre segments 4 et 25 (lignes 1142-1143 de la fin du segment 24, répétées au début du segment 25)

Manque aussi entre : 2453 et 2474 / segments 39 et 40. Dans le fichier brut Angel Villar « brut » du CD, l’audio existe bien entre 1 h 53 min 09 s et 1 h 54 min.

Ok entre segments 40 et 69

Durée des fichiers :

3h24min47 : AV16-ANGEL VILLAR brut.mp3 (fichier du CD d’origine)

3h21min28 : AV16.mp3 (fusion brute des segments du CD d’origine)

3h24min49 : AV16-v2.mp3 (fusion version 2 avec les 2 extraits manquants)

Le 31/10/2019 : Pb pour aligner AV16-v2.mp3 après 3h21min28…

Remarques Deborah

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

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 nécessaire

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

<rire/pleur> --- <rireentendu/pleurentendu>

<pause/silence> --- <si>

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

<espagnol></espagnol> --- <lg=espagnol></lg=espagnol>

XXX là --- XXX-là

<audiomanquantsep/fus> --- effacés

Ajouts de majuscules, par exemple :

les républicains --- les Républicains

les espagnols --- les Espagnols

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

Corrections notables :

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

Transcription :

481 \*CH\_EE1 : oui j'ai connu

482 \*AV16 : tu as connu ?

483 \*CH\_EE1 : <revision> j'é- j'é- j'é- </revision> j'étais là-bas

<LJ> D’accord !

Cela n’a rien à voir, mais par exemple à cet endroit, deux bullets sont ligne fin 476 et ligne fin 486. Si l’on clique F4 dans le texte entre les lignes 477 et 485, cela indique une erreur : “Media marker not found at cursor position. If you are planning to link sound to text then please use Sonic mode “F5“. Par contre F4 fonction sur les lignes 485 et 486.

<DL> c’est parce que les bullets ne fonctionnent que sur le paragraphe où elles sont insérées. Et en corrigeant je crée parfois de nouveaux paragraphes au sein d’une même bullet. Si on veut jouer la bullet, il faut placer le curseur dans le paragraphe correspondant ou directement sur la bullet elle-même. Comme je l’expliquais en bas de page, si j’en ai le temps, je rectifierai ces cas-là.

3692 vers l'Isle-de-Noé

<LJ> ligne 3697 à présent : j’ajoute le Npr de lieu au dictionnaire : L’Isle-de-Noé, id=2998135

“L’” en majuscule a priori pour ajouter l’id dans la transcription : “<lieu id=2998135> L’Isle-de-Noé </lieu>”

4017 4022 dans le .cha du 6 novembre \*CH\_EE1 : on va faire <revision> un petit un petit cro- </revision> un petit

4018 croquis <mdd> là on peut couper là </mdd>

4019 \*CH\_OD1 : de repasser la question

4020 \*CH\_EE1 : ça il <repetition> faut faut </repetition> faut

4021 \*CH\_OD1 : essayer de le

4022 \*AV16 : <pron=pi> </revision> du fait qu'il voyait arriver des FTP de

4023 partout alors ils étaient que quatre

4024 \*CH\_EE1 : on y va on y va

4025 4030 \*AV16FIE : il vaut mieux que tu quest-

<LJ> effectivement, il y a pas mal de mdd propres au format audio dans ce passage (et ensuite li 4031)

Ici aussi le lien entre audio et texte ne se fait pas F4 que sur la ligne 4030, c’est-à-dire la dernière ligne entre les deux bullets.

Correction :

30-31 <lg=espagnol> El Ferrol del Caudillo </lg=espagnol> <trad> Le Ferrol du chef </trad>

<lg=espagnol> del Caudillo </lg=espagnol> <trad> du Chef </trad>

<LJ> Je n’entends en effet pas de “Ferrol”

Je comprends que “el Caudillo” est Franco, est-ce qu’il ne faudrait pas indiquer cette appellation dans la liste des noms de personnes ? Avec une majuscule ?

L 29: “<lieu id=3110776> San Claudio </lieu> Quilova” —> je pense que c’est en fait San Clodio Quiroga, soit deux municipalités qui partagent une même gare de train nommée San Clodio-Quiroga.

<DL> Bien joué. J’avais repéré Quiroga, mais je n’avais pas pensé à la gare.

Proposition : “<lieu id=3110764> San Clodio </lieu> <lieu id=3112395> Quiroga </lieu>”

Je modifie le dictionnaire Npr lieux.

(Ou alors indiquer la gare ? Estación de San Clodio ou San Clodio-Quiroga avec id=3110763)

À modifier également San Claudio en San Clodio : lignes 93 et 131

<DL> Ok, donc je corrige et annote San Clodio et Quiroga, mais je ne touche pas à San Clodio-Quiroga (je corrige juste l’orthographe).

<LJ> oui, ça me semble le mieux

229 <lg=espagnol> gardievo </lg=espagnol>

<lg=espagnol> gallego </lg=espagnol>

<LJ> oui, effectivement “le <lg=espagnol> gallego </lg=espagnol> <trad> galicien </trad>”

- Lieux :

739-740 au <pron=pi>

aux Aubrais

(Mais pas sûr, car prononciation [obroué])

<LJ> Je pense que c’est bien les Aubrais. Cela correspond à un lieu-dit qui est aussi le lieu de la gare des Aubrais à Fleury-les-Aubrais. Je propose de prendre l’identifiant de la ville de Fleury-les-Aubrais, dans GeoNames : 3018280. Je l’ajoute au dictionnaire : Aubrais:5-id=3018280.

1173 à l'hôpital cette nuit de Perpignan

à l'hôpital Saint-Louis de Perpignan

<LJ> D’accord

2243 "<pron=pi>" --- "Ravezies" (place et quartier de Bordeaux, de l'autre côté de la base sous-marine par rapport à la rue Achard)

<LJ> ok, super

2092 2102 "Bassins à flot" : lieu à Bordeaux, volontairement mis au pluriel et avec majuscule.

<LJ> ok, “du Bassins à flot” ou “de Bassins à flot” ? Même si j’imagine que l’on dit plutôt des Bassins à flot

<DL> Tout à fait, “Bassins à flot” est un nom propre, c’est figé. Par exemple, on dit “ce soir, je vais à Bassins à flot”.

<LJ> c’est noté !

3950 3958 "Arc de Triomphe" --- "arc de triomphe" en minuscule (de même dans les lignes suivantes), car il n'y a pas de monument "Arc de Triomphe" à Bordeaux. Il y a des "Portes" qui ont une forme d'arc de triomphe, notamment la "Porte de Bourgogne" qui est sans doute le monument dont parle AV16.

<LJ> ok, “nous on était du côté de l'arc de triomphe” -- “nous on était du côté <repetition> de </repetition> de l'arc <revision> del </revision> de triomphe”

4180, 4187 "grand théâtre" --- "Grand Théâtre" (monument de Bordeaux)

<LJ> D’accord

4475 on voit les photos

du coup parce que on voit les photos <bc> à Rouffignac <bc> c'est <repetition> Rouffignac </repetition> <bc> Rouffignac

<LJ> D’accord

- Langues :

1456 1458 <lg=espagnol> rojo español </lg=espagnol> <trad> rouge espagnol </trad>

<lg=allemand> rote Spanier </lg=allemand> <trad> Espagnols rouges </trad>

<LJ> Du coup, suppression de “rojo español” dans le texte car non prononcé ?

“Rouge espagnol” -- “Rouges espagnols” (avec ou sans majuscules r/R ?) comme avant “malades espagnols rouges”

<DL> Je n’entends pas “rojo español” donc je l’ai supprimé.

<LJ> ok

1458-1459 1462-1463 <anglais> free </anglais> <trad> </trad> <lg=espagnol> español </lg=espagnol>

<lg=allemand> frei Spanier </lg=allemand> <trad> Espagnols libres </trad>

(peut-être légèrement prononcé à l'espagnol [eSpanier])

Ligne 1464 : “espagnol libre” -- “Espagnols libres”

2767-2769 2772-2774 il disait <lg=allemand> Spanisch Spanisch </lg=allemand> <trad> l’Espagnol l’Espagnol </trad> l'alarme [pron=pi] <repetition> aviation </repetition> aviation <revision> alors nous avions </revision>

il disait <lg=allemand> Spanier Spanier Fliegeralarm Fliegeralarm </lg=allemand> <trad> </trad> aviation aviation alarme aux avions

<LJ> oui c’est vrai que balise <repetition> non pertinente.

2770, 2819, 2870, 2884, 3505 : <pron=pi> ---- <lg=allemand> </lg=allemand> <trad> </trad>

<LJ> Pour le lien audio-texte : la partie entre les bullets ligne 2878-2886 : le F4 ne fonctionne pas.

Faire écouter à quelqu’un qui parle allemand (je crois qu’il y a un étudiant allemand au laboratoire).

- Npr :

"Hélène" --- "AUTRE"

906 "Voix de femme" --- "AUTRE2"

2613-2613 "Alvarez" --- "NP104"

2739-2740 "AV16LOC2" --- "NP111"

3098 "monsieur Vásquez" --- "monsieur NP107"

3592 "NP164" --- "NP166"

3839 "AV16AMI" --- "CH\_EE1"

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

Questions/Réponses :

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

- Gazetier :

<DL> Ajouter "l'Isle-de-Noé" (l. 3692)

<DL> Ajouter "Rouffignac" (l. 4475)

<DL> Dordogne (le "ruisseau" l.1447) : "5-id=2992132"

C'est le même id que "Montpon" : c'est normal ou j'ai commis une erreur de copier-coller ? Je ne trouve plus les gazetiers dans le dossier sur le FTP. Pour le moment, j'ai remplacé par l'id "PAIHABIT0000000049526714"

<DL> Je signale juste que j'ai annoté tous les "Vichy", mais que ça concerne souvent "la politique de Vichy", "le gouvernement de Vichy"... (comme pour le "mur de l'Atlantique").

<DL> Je n'ai utilisé que "Loire:5-id=2997869" (je suppose que c'est l'id du fleuve)

Et je n'ai pas utilisé "Loire:5-id=DEPARTEM0000000000000002".

<DL> Je n'ai pas eu à utiliser "Belchite:5-id=3128365"

<DL> Je n'ai pas eu à utiliser "Ferrol:5-id=3123493"

<LJ> Je remets tous les dictionnaires car en effet ils ont disparus.

Pour AV16, je mets le nouvelle version :

Ajouts :

L’Isle-de-Noé:5-id=2998135

Rouffignac:5-id=2982625 (il y a au moins deux Rouffignac en Dordogne : Rouffignac-Saint-Cernin-de-Reilhac et Rouffignac-de-Sigoulès, je mets l’id de Rouffignac-Saint-Cernin-de-Reilhac car plus d’habitants et certainement plus connu comme proche de Lascaux ?)

Quiroga:5-id=3112395

Aubrais:5-id=3018280

Modifications :

Dordogne:5-id=3021044 (l’erreur était bien dans le fichier sur le FTP)

San Clodio:5-id=3110764 (à la place de San Claudio:5-id=3110776)

Suppressions :

Loire:5-id=DEPARTEM0000000000000002

Belchite:5-id=3128365

Ferrol:5-id=3123493

Note pour plus tard : penser ensuite à remplacer pour les lieux en France avec les id de la BDNyme.

- Incertitudes :

<DL> 29 "Quilova" ?

<LJ> Quiroga:5-id=3112395 ajouté au dictionnaire Npr lieux

<DL> 2394 "Caris" ? ("car ici" peut-être ?)

<LJ> Je supprimerais “Caris”, et remplacerais par “car ici” qui semble logique avec l’intonation ou par <pron=pi> ou par <revision> </revision>.

<DL> 2885 : <lg=espagnol> se acabó </lg=espagnol> <trad> c'est fini </trad>

Je me demande s'il ne dit pas plutôt "sale cabot"

<LJ> Je trouve que l'intonation se rapproche plus de “se acabó” mais à voir.

-Codes Npr :

<DL> 30-31 "Caudillo" ("Chef") : surnom de Franco, à remplacer par un code Npr (NP1SUR ?) ?

<LJ> oui d’accord : à confirmer par Catherine

<DL> 485-486 "la mère s'appelait NP94 la fille aussi NP94"

Le premier NP94 correspond à "Matilde" (la mère), et le deuxième correspond à un prénom ressemblant, "Matildine" peut-être, (la fille). Faut-il un code NPr, autre que NP94, pour la fille ?

<LJ> oui d’accord : ajouter NP94FILLE ? (je ne sais plus le code pour désigner la fille de)

<DL> NP94FIE

<DL> 643 "Corinne" : une des infirmières qui soignent AV16 ? Code Npr ?

<DL> 1601 1608 "16LOC1" : Michel (le chauffeur), plutôt un code Npr "NP\*\*\*" ?

<DL> 3002 3007 "l'ingénieur espion NP106" --- "l'ingénieur NP111"

Il n'y a pas d'autres occurrences de NP106, je ne crois pas qu'il y ait un "Sala", je pense que "Espinoza" (NP111) est l'ingénieur catalan (NP106).

<DL> 3065-3068 "Gérard Catalan" : la deuxième identité d'AV16. Code Npr ? (AV16SUR ?)

<DL> 3120 "AV16FIE" ? Ou plutôt AV16FEM ?

<DL> 3181, 3326 : "AV16AMI"/"AV16AMI2" --- "NP108" : ajouter NP108 (Alberto Brandosa) à NP108SUR

Observation globale :

Il y a eu beaucoup de corrections à effectuer : code Npr, noms de lieux, passages en langue étrangère, passages pas ou mal transcrits... Qui ont pris beaucoup plus de temps, j'ai donc moins pris le temps d'annoter les répétitions de "petits" mots. J'ai préféré procéder ainsi, de sorte à avoir assez de temps pour les transcriptions restantes.

<LJ> Oui, c’est mieux comme cela. On l’indiquera éventuellement dans les métadonnées du fichier lors du dépôt.

L'alignement n'est pas aussi propre que d'habitude, pour deux raisons :

- Beaucoup de chevauchements de parole entre les interlocuteurs, CH\_EE1 se répète beaucoup, et AV16 parle parfois en flux continu. Il est donc moins évident de "trouver une place" où segmenter.

- Les corrections sont effectuées après l'alignement et, comme expliqué précédemment, il y a eu beaucoup de corrections (notamment de retranscriptions) qui font que parfois le texte coïncide moins avec l'alignement.

Si j'en ai le temps, je rectifierai cela à la fin de la prestation.

<LJ> Ok, les corrections ont en effet dû pas mal décaler le texte.