



HAL
open science

Guide d'annotation pour le sous corpus de test d'ALLIES (FullTest-CleanAnnotations)

Marie Tahon, Anthony Larcher

► **To cite this version:**

Marie Tahon, Anthony Larcher. Guide d'annotation pour le sous corpus de test d'ALLIES (FullTest-CleanAnnotations). 2023. hal-04555362

HAL Id: hal-04555362

<https://hal.science/hal-04555362>

Preprint submitted on 22 Apr 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Guide d'annotation – ALLIES-clean

Marie Tahon, Anthony Larcher (nom.prenom@univ-lemans.fr)

Contenu

35 fichiers, soit environ 10h d'audio

L'annotation se fera en deux étapes :

- 1) Correction de la segmentation en tour de parole et de la transcription
- 2) Correction ou ajout de la segmentation de la parole superposée
- 3) Correction ou ajout d'événements sonores de types présence de bruit ou de musique.

Préambule description du contenu des fichiers trs.

Les fichiers de données sont déjà annotés et segmentés en tour de parole.

Une balise <Turn> contient la transcription d'un unique locuteur avec éventuellement plusieurs segments <Sync> de ce même locuteur.

Exemple 0

```
<Turn speaker="spk20" startTime="0.000" endTime="3.724">
<Sync time="0.000"/>
France Inter , il est 19 heures , le Journal , Christophe Hondelatte .
</Turn>
<Turn startTime="3.724" endTime="8.623">
<Sync time="3.724"/>
</Turn>
<Turn speaker="spk1" startTime="8.623" endTime="11.403">
<Sync time="8.623"/>
[b] bonsoir à tous , nous sommes le 21 juin , et c'est l'été !
</Turn>
```

Locuteurs :

Pré-traitement Vérifier le genre de chaque locuteur et supprimer les doublons

Exemple :

```
<Speaker id="spk1" name="Christophe_HONDELATTE" check="no" type="male"
dialect="native" accent="" scope="global"/>
<Speaker id="spk2" name="Christophe_HONDELATTE" check="no" type="unknown"
dialect="native" accent="" scope="global"/>
<Speaker id="spk3" name="Christophe_HONDELATTE" check="no" type="unknown"
dialect="native" accent="" scope="global"/>
```

Phase 1 : correction de la segmentation et transcription

Segmentation et transcription des tours de parole mono-locuteur.

Ces zones respectent les règles de transcription classiques (type ESTER, REPERE). Elles ne devraient pas poser de problème particulier car ces zones ont déjà été corrigées notamment lors de la mise en place de campagnes d'évaluation. Il faudra cependant être vigilant sur les fichiers récents (après 2014).

Les tâches à réaliser sur les zones mono-locuteur sont les suivantes :

- 1) Vérification des informations locuteur en entête du fichier (genre, identité, langue)
- 2) Correction des frontières temporelles.
- 3) Correction de la transcription si nécessaire

Segmentation et transcription des zones de parole superposée.

Par contre, une attention spécifique devra être portée sur les zones de parole où au moins deux locuteurs sont présents. Le fait qu'un locuteur soit présent ou non n'est pas lié à son intensité sonore. C'est-à-dire que même la présence d'un second locuteur à une faible intensité, en fond sonore, devra être annotée comme parole superposée.

Les tâches à réaliser sur les zones de parole superposée sont les suivantes :

- 1) Correction des frontières temporelles
- 2) Ajout des zones de parole superposée manquantes
- 3) Transcription des zones de parole superposées.

1.1. Cas de deux locuteurs simultanés : S'il y a deux locuteurs identifiables, on indique leur identifiant (comme sur l'exemple), en utilisant la case « overlapping speech » de Transcriber.

- S'il n'est pas possible de transcrire car le segment est inintelligible (brouhaha) on indiquera [nontrans] (exemple 1)
- S'il est possible de transcrire le locuteur principal (1^{er} locuteur) on indiquera sa transcription.
- Si un mot est coupé par un segment de parole superposée, on utilisera les [] pour indiquer la fin ou le début du mot comme illustré sur l'exemple 3.

La transcription du 2^{ème} locuteur ne doit pas être ajoutée. Le champ de transcription doit rester vide.

1.2. Cas d'au moins trois locuteurs simultanés : S'il y a plus de 3 locuteurs, on indiquera l'identifiant locuteur 3+. Aucune transcription ne doit être ajoutée, le champ réservé à la transcription doit rester vide (exemple 2).

Exemple 1

```
<Turn speaker="spk7 spk2" startTime="275.658" endTime="276.260">
<Sync time="275.658"/>
<Who nb="1"/>
[nontrans]
<Who nb="2"/>
</Turn>
```



```

Franck_HERIOT
• eh bien le chiffre est éloquent , depuis le début du siècle seuls 5 procès d'assises ont été révisés .
• [i] le dernier en date est celui de ce tunisien de 35 ans , Rida Daalouche acquitté le 8 mai dernier par la cour d'assises de l' Hérault ,
• [i] après avoir été condamné pour un meurtre qu'il n'avait , donc pas commis .
• [i] le cas le plus connu restant évidemment celui du capitaine Dreyfus
Franck_HERIOT + Christophe_HONDELATTE
• 1: [nontrans]
• 2:

```

Exemple 2 : 00_trs_originaux/20200514.0000.LCP_CaVousRegarde.trs

```

<Speaker id="spk1" name="3+" check="no" type="" dialect="native" accent=""
scope="local"/>

<Turn speaker="spk1" startTime="439.537" endTime="440.592">
<Sync time="439.537"/>
</Turn>

```

Exemple 3 : 00_trs_originaux/19990621.1900.inter_fm_dga.trs

```

<Sync time="1066.786"/>
[i] face à Israël , à Toulouse . la rencontre Pascal Pierozzi a débuté il y
a une heure ,
</Turn>
<Turn speaker="spk13 spk3" startTime="1070.655" endTime="1071.902">
<Sync time="1070.655"/>
<Who nb="1"/>
[nontrans]
<Who nb="2"/>
</Turn>
<Turn speaker="spk14" startTime="1071.902" endTime="1073.942">
<Sync time="1071.902"/>
4 minutes à jouer , euh bonsoir Christophe ,
</Turn>
<Turn speaker="spk13 spk3" startTime="1073.942" endTime="1075.052">
<Sync time="1073.942"/>
<Who nb="1"/>
[nontrans]
<Who nb="2"/>
</Turn>
<Turn speaker="spk14" startTime="1075.052" endTime="1134.546">
<Sync time="1075.052"/>
part vers une première victoire facile , elle mène

```

```

• [i] face à Israël , à Toulouse . la rencontre Pascal Pierozzi a débuté il y a une heure ,|
Pascal_PIEROZZI + Christophe_HONDELATTE
• 1: [nontrans]
• 2:
Pascal_PIEROZZI
• 4 minutes à jouer , euh bonsoir Christophe ,
Pascal_PIEROZZI + Christophe_HONDELATTE
• 1: [nontrans]
• 2:
Pascal_PIEROZZI
• part vers une première victoire facile , elle mène
• 71 à 55 . vous le savez la présence des Yougoslaves

```

Doit être corrigé comme suit

- modification de l'identifiant locuteur spk14->spk13
- correction de la segmentation des zones d'overlap,
- ajout des transcriptions du locuteur principal (ici spk13) dans les zones de parole superposée, utilisation des crochets pour les mots coupés par le segment.

```

<Sync time="1066.786"/>
[i] face à Israël , à Toulouse . la rencontre Pascal Pierozzi a débuté il y
a une heure ,
</Turn>
<Turn speaker="spk13" startTime="1070.655" endTime="1071.060">
<Sync time=1070.655"/>
une
<Turn speaker="spk13 spk3" startTime="1071.060" endTime="1071.615">
<Sync time="1070.060"/>
<Who nb="1"/>
heure il res[te]
<Who nb="2"/>
</Turn>
<Turn speaker="spk14" startTime="1071.615" endTime="1073.859">
<Sync time="1071.615"/>
[res]te 4 minutes à jouer , euh bonsoir Christophe ,
</Turn>
<Turn speaker="spk13 spk3" startTime="1073.859" endTime="1074.276">
<Sync time="1073.859"/>
<Who nb="1"/>
et la Youg[oslavie]
<Who nb="2"/>
</Turn>
<Turn speaker="spk13" startTime="1074.276" endTime="1134.546">
<Sync time="1074.276"/>
[Youg]oslavie part vers une première victoire facile , elle mène

```

● [i] face à Israël , à Toulouse . la rencontre Pascal Pierozzi a débuté il y a une heure ,

Pascal_PIEROZZI

● une

Pascal_PIEROZZI + Christophe_HONDELATTE

● 1: heure il res[te]

2:

Pascal_PIEROZZI

● [res]te 4 minutes à jouer , euh bonsoir Christophe ,

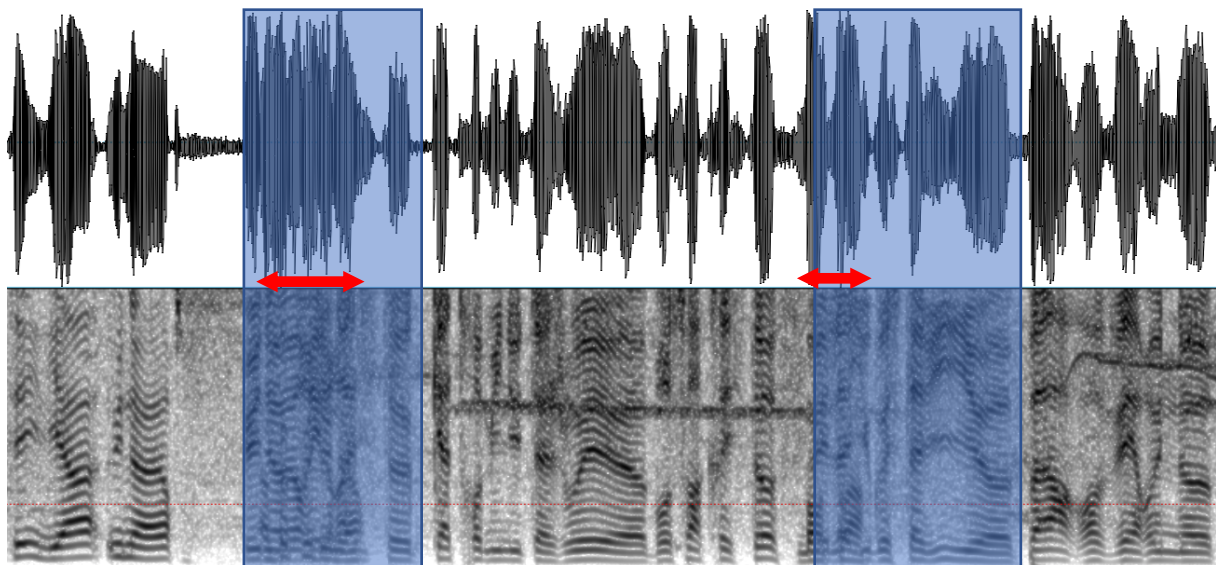
Pascal_PIEROZZI + Christophe_HONDELATTE

● 1: et la Youg[oslavie]

2:

Pascal_PIEROZZI

● [Youg]oslavie part vers une première victoire facile , elle mène



En bleu les zones annotées en parole superposée dans le fichier original. En rouge, la zone de parole superposée réelle à corriger.

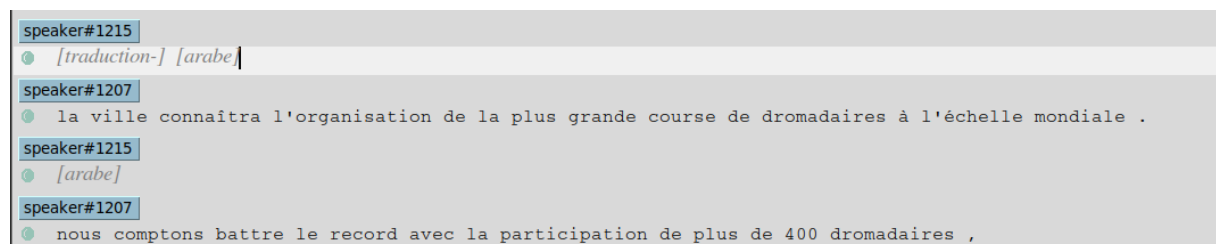
1.3. Cas des traductions : Lorsque le locuteur principal est traduit en fond par une voix off, le segment doit être annoté comme une zone de parole superposée. On doit donc voir apparaître deux locuteurs simultanés, le locuteur principal et le traducteur. Dans un second temps on ajoutera une balise de type *Event* afin de spécifier le cas traduction, et éventuellement la langue traduite.

Lorsqu'un locuteur parle dans une langue étrangère, on ne cherchera pas à transcrire le texte même lorsqu'il est l'unique locuteur ou qu'il est le locuteur principal dans une zone de parole superposée. On indiquera alors la balise [nontrans] + une balise *Event* qui indique la langue employée.

Exemple 4 : 00_trs_originaux/20070716.2045.tvme.trs

```
<Speaker id="spk9" name="speaker#1207" check="no" type="female"
dialect="nonnative" accent="marocain" scope="local"/>
<Speaker id="spk17" name="speaker#1215" check="no" type="male"
dialect="native" accent="" scope="local"/>

<Event desc="r" type="noise" extent="instantaneous"/>
  une première à l'échelle mondiale .
</Turn>
<Turn speaker="spk17" startTime="420.354" endTime="422.846">
<Sync time="420.354"/>
<Event desc="traduction-" type="noise" extent="instantaneous"/>
<Event desc="arabe" type="noise" extent="instantaneous"/>
</Turn>
<Turn speaker="spk9" startTime="422.846" endTime="427.996">
<Sync time="422.846"/>
la ville connaîtra l'organisation de la plus grande course de dromadaires à
l'échelle mondiale .
</Turn>
<Turn speaker="spk17" startTime="427.996" endTime="429.359">
<Sync time="427.996"/>
<Event desc="arabe" type="noise" extent="instantaneous"/>
</Turn>
```



Le locuteur spk17 parle en arabe (et semble être une femme, il faut donc corriger le genre de ce locuteur), la journaliste spk9 parle en français avec un accent arabe. Dans cet exemple, spk17 parle sur l'ensemble de la durée, la journaliste spk9 ne traduit que sur certains segments qui doivent être indiqués en overlap.

Ce segment doit être corrigé suivant :

```
<Speaker id="spk9" name="speaker#1207" check="no" type="female"
dialect="nonnative" accent="marocain" scope="local"/>
```



```
<Speaker id="spk17" name="speaker#1215" check="no" type="female"
dialect="native" accent="" scope="local"/>

<Event desc="r" type="noise" extent="instantaneous"/>
  une première à l'échelle mondiale .
</Turn>
<Turn speaker="spk17" startTime="420.354" endTime="422.846">
<Sync time="420.354"/>
[nontrans]
<Event desc="arabe" type="noise" extent="instantaneous"/>
</Turn>
<Turn speaker="spk9 spk17" startTime="422.846" endTime="427.996">
<Sync time="422.846"/>
la ville connaîtra l'organisation de la plus grande course de dromadaires à
l'échelle mondiale .
<Event desc="traduction-" type="noise" extent="instantaneous"/>
<Event desc="arabe" type="noise" extent="instantaneous"/>
</Turn>
<Turn speaker="spk17" startTime="427.996" endTime="429.359">
<Sync time="427.996"/>
[nontrans]
<Event desc="arabe" type="noise" extent="instantaneous"/>
</Turn>
```

```
● [r] un grand événement puisqu'il s'agit de se mettre à l'épreuve [r] du record [ent=prod.doc-]Guinness [-ent=prod.doc]
de plus de 400 dromadaires , [r] une première à l'échelle mondiale .
speaker#1215
● [nontrans][arabe]
speaker#1207 + speaker#1215
● la ville connaîtra l'organisation de la plus grande course de dromadaires à l'échelle mondiale .[traduction-] [arabe]
speaker#1215
● [nontrans][arabe]
speaker#1207
● nous comptons battre le record avec la participation de plus de 400 dromadaires ,
```

Phase 2 : ajout des événements sonores

Les événements sonores tels que le bruit et la musique, doivent être annotés indépendamment de la transcription dans des fichiers séparés. On aura donc x TRS files : transcription, bruit, musique et

2.1 Annotation de la musique

Il peut y avoir de la musique en fond sonore du discours du journaliste, ou bien de la musique seule. Cette musique peut-être un jingle, mais nous ne ferons aucune distinction. Un unique label sera utilisé pour ce type d'annotation.

Exemple : fichier TRS de transcription (19981207.0700.inter_fm_dga.trs)

```
<Section type="report" startTime="0.000" endTime="3596.600">
<Turn speaker="spk23" startTime="0.000" endTime="1.500">
<Sync time="0.000"/>
Patricia Martin , que voici , que [top] voilà !
</Turn>
<Turn speaker="spk28" startTime="1.500" endTime="2.624">
<Sync time="1.500"/>
oh , bonjour [top] Nicolas Stoufflet .
</Turn>
<Turn speaker="spk23" startTime="2.624" endTime="3.765">
<Sync time="2.624"/>
France Inter [top] , 7 heures .
</Turn>
<Turn start="3.765" end="7.947">
<Sync time="3.765"/>
</Turn>
<Turn speaker="spk28" startTime="7.947" endTime="9.632">
<Sync time="7.947"/>
le journal , Simon Tivolle .
</Turn>
<Turn speaker="spk38" startTime="9.632" endTime="62.142">
<Sync time="9.632"/>
bonjour ! lundi 7 décembre . deux incendies cette nuit en région parisienne , dans une maison de retraite de Livry-Gargan en Seine-Saint-Denis ,
```

Exemple : fichier TRS de musique (19981207.0700.inter_fm_dga.music.trs) spk1=présence de musique dans l'audio, quelque soit le reste du contenu audio.

```
<Section type="report" startTime="0.000" endTime="3596.600">  
<Turn speaker="spk1" startTime="0.000" endTime="9.270">  
</Turn>  
<Turn speaker="spk1" startTime="10.209" endTime="64.180">  
</Turn>
```

2.2 Annotation des bruits

- 1) BV : bruit vocaux (hum, raclement de gorge, toux, rires, etc)
- 2) BS : brouhaha/parole humaine en fond sonore (pas considéré comme overlap car pas possible d'identifier clairement combien de locuteurs sont présents)
- 3) B : bruits extérieurs (bruits de micro, froissement de papier, chocs)

L'annotation se fera de façon similaire à celle de la musique (19981207.0700.inter_fm_dga.noise.trs) en utilisant les 3 tags ci-dessus. Ces tags peuvent être utilisés comme 3 locuteurs différents, sauf si une autre solution semble plus pertinente.

2.2 Annotation de la qualité audio (channel) (pas prioritaire)

Il s'agit d'indiquer s'il y a un changement de qualité audio, tel que du téléphone, du studio ou de la VOIP. Par défaut on ne mettra aucun label.

Crédits

Ce travail a été financé par le projet Européen ESPERANTO Horizon 2020, dans le cadre d'un financement Marie Skłodowska-Curie No 101007666 et par le projet ANR Gender Equality Monitoring. La tâche d'annotation a été réalisée par l'entreprise Elyadata.