

## Synchronisation d'un fichier d'annotation Toolbox avec le fichier audio d'origine pour une utilisation dans ELAN

On suppose au départ que la base de données Toolbox est constituée de phrases annotées référencées par le champ principal (par exemple \ref) dont le contenu sera de la forme *label\_num*, c'est-à-dire une étiquette représentant l'identité du texte suivi d'un numéro représentant le numéro de la phrase dans le texte. (par exemple : \ref Sarkozy&recherche.01 pour la première phrase du texte Sarkozy&recherche)

*Dans le cas contraire, voir l'annexe en fin de document*

Le principe de la synchronisation sera :

- 1) créer un fichier ELAN à partir du fichier Audio, segmenté suivant le même découpage que le fichier d'annotation Toolbox. Les annotations de la tier *ref* auront les mêmes valeurs *intitulé\_numéro* que les \ref des fiches Toolbox
- 2) Exporter ce fichier ELAN vers Toolbox
- 3) Ouvrir dans Toolbox le fichier précédent et le fusionner (*merge*) avec le fichier Toolbox original. Enregistrer le nouveau fichier Toolbox
- 4) Importer le nouveau fichier Toolbox indexé dans ELAN

1) **Dans ELAN**, créer un fichier d'annotation pour le fichier audio

- File, New
- Sélectionner le fichier *audio* correspondant au fichier Toolbox
- Le verser dans la fenêtre de droite avec la bouton >>
- OK

Renommer la tier *default*

- Clic-droit sur l'étiquette de la tier *default*
- Choisir *Change attribute of default*
- *Tier name* : donner à la tier le même nom que celui du marqueur d'enregistrement de la base de données Toolbox (ici *ref*)
- *Participant* : Mettre *SP* pour un seul locuteur

Segmentation au kilomètre, correspondant aux fiches de la base de données Toolbox

- Options, Segmentation Mode
- Cocher *One keystroke per annotation*
- Cliquer sur le bouton *début de piste* 
- Lancer la lecture 
- Taper la touche <Entrée> pour terminer le segment précédent et démarrer le suivant  
*Attention* : taper la touche <Entrée> dès le démarrage pour créer la frontière gauche du premier segment

Régler la position des frontières des segments

- Pour écouter le contenu d'un segment, sélectionner le son correspondant en le balayant avec la souris et cliquer sur le bouton *play* 
- Pour déplacer une frontière, placer le curseur vers l'extrémité du segment à déplacer.
- Lorsque le curseur prend la forme d'une double flèche horizontale, cliquer et déplacer la frontière dans le sens voulu.  
*Attention* : afin d'éviter de créer des *trous*, pour déplacer une frontière de la gauche vers la droite, le segment de gauche doit être sélectionné (en vert), et pour déplacer une frontière de la droite vers la gauche, le segment de droite doit être sélectionné.

Ajouter, supprimer des frontières (correction de la segmentation au kilomètre)

Ajouter un segment

- Sélectionner le menu *Options, Annotation mode*
- Sélectionner avec la souris la zone correspondant au nouveau segment à ajouter
- Clic-droit au niveau de la tier d'annotation, *Create new annotation here*

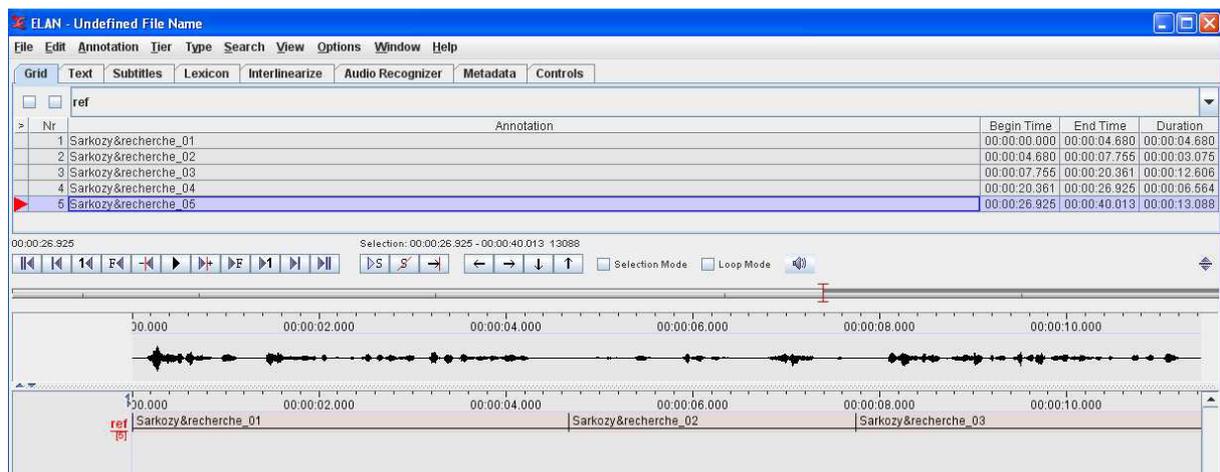
Fusionner un segment avec le suivant

- Sélectionner le segment en cliquant dessus au niveau de la tier d'annotation (segment en bleu)
- Clic droit dessus le segment, *Merge with next annotation*

Étiqueter les segments

- Sélectionner dans le menu *Tier, Label and Number*
- *Include label part* : donner le nom de l'étiquette utilisée dans le fichier Toolbox pour identifier les fiches (ex : Sarkozy&recherche)
- *Insert other delimiter* : ajouter le séparateur '.' pour la numérotation des segments
- *Minimal number of integer digits* : 2 si le nombre de segment est inférieur à 100, 3 sinon (ex : 2)

Le résultat de l'étiquetage des segments doit correspondre aux intitulés des fiches de Toolbox. (ex : Sarkozy&recherche\_05)



Cas de plusieurs locuteurs

Dans le cas de plusieurs locuteurs, il y aura une tier de segmentation par locuteur :

- Tier name : ref@SP1, Participant : SP1, labélisation : label\_SP1\_01...
- Tier name : ref@SP2, Participant : SP2, labélisation : label\_SP2\_01...

## 2) Exporter le fichier ELAN vers Toolbox (en fait ici Shoebox)

- File, Export as, Shoebox file
- Use Shoebox database type : Cliquer sur le bouton ...
- Rechercher le fichier .typ correspondant au type de la base de données Toolbox à synchroniser (généralement dans le sous-dossier *Settings* du dossier *Toolbox*)
- OK, donner un nom au fichier (extension .txt), Enregistrer

## 3) Fusionner les index temporels dans la base de données Toolbox

- Ouvrir le projet contenant la base de données à synchroniser
- Ouvrir le fichier exporté de ELAN : Fichier, Ouvrir...

Cette base de données doit contenir le même nombre de fiches, avec les mêmes intitulés que celles de la base à synchroniser. Elles contiennent les champs ELANBegin, ELANEnd avec les index temporels correspondant pour chaque fiche, ainsi que le champ ELANParticipant qui permet de distinguer différents locuteurs (dialogue par exemple...)

- Sélectionner la base à synchroniser en cliquant dans sa fenêtre
- Sélectionner dans le menu *Base de données, Fusionner*
- Sélectionner la base de données provenant de ELAN
- OK

Les fiches de la base provenant de ELAN qui ont la même clé (ici : contenu de *ref*) que celles de la base d'origine Toolbox verront leur contenu fusionner : les champs *index* et *participant* vont s'ajouter aux annotations de chaque fiche d'origine.

#### 4) Importation dans ELAN

La base de données Toolbox indexé au temps est maintenant importée dans ELAN

- File, Import, Shoebox file
- *Shoebox file* : Cliquer sur les ... pour rechercher la base à importer
- *Shoebox type file* : Cliquer sur les ... pour rechercher le type de la base de données Toolbox (dans le sous dossier *Settings...*)
- *All marker are Unicode* : à cocher
- OK

The screenshot shows two windows of the ELAN software. The top window, titled 'sarko3.txt', displays a linguistic analysis of the sentence 'Nos universités bénéficient depuis 2008 d'augmentation de leurs moyens'. The bottom window, titled 'sarkoX.txt', shows the same analysis with additional ELAN markers: \ELANBegin (0.000), \ELANEnd (4.680), and \ELANParticipant (SP). A dialog box titled 'Sélectionner une base de données' is open in the foreground, with 'sarkoX.txt' selected in the list. The dialog contains the text: 'Fusionner les enregistrements et les champs d'une autre base vers la base active. Base de laquelle on fusionne les enregistrements et les champs'. Buttons for 'OK', 'Annuler', and 'Aide' are visible, along with a checkbox for 'Chemin'.

This screenshot shows the 'sarko3.txt' window after the import operation. The linguistic analysis is identical to the previous screenshot, but now includes the ELAN markers: \ELANBegin (0.000), \ELANEnd (4.680), and \ELANParticipant (SP). The dialog box is no longer present.

## Transformation d'une base de données *textes* en base de données *phrases*

Si la base de données est constituée d'un ou plusieurs textes, chacun d'eux faisant l'objet d'une fiche unique dans laquelle les phrases annotées ne comportent pas de référence, il faudra au préalable générer un champ de référence pour chaque phrase qui sera du type *label\_num* (disons \ref), puis exporter cette base de données au format standard, et ensuite la ré-importer sous un nouveau type dont le marqueur d'enregistrement sera celui de la référence de la phrase (\ref).

Pour générer automatiquement un champ de référence pour chaque phrase d'une fiche comportant plusieurs phrases annotés, dans Toolbox :

- Commencer par créer un champ de référencement (par ex : \ref) s'il n'existe pas déjà
- Aller dans le menu 'Outils', 'Découper/numéroter du texte'
  - *Balise de référence* : choisir le marqueur pour le champ de référencement (\ref)
  - *Balise de texte* : choisir le marqueur du champ contenant le texte (\tx)
  - 'Ok'.

Dans la fenêtre qui s'ouvre :

- Sélectionner *Enregistrement en cours* pour ne traiter que la fiche en cours
- *Nom du texte*
  - *Utiliser le contenu du champ* : Choisir le champ contenant l'identifiant du texte (\id par exemple)
  - Ou bien
    - *Utiliser ce nom* : entrer un mot pour identifier le texte en question
    - *A partir du numéro* : 01 pour une numérotation sur 2 chiffres, 001 pour 3.
    - 'OK'

Chaque phrase est maintenant précédée d'une référence numéroté dans le nouveau champ de référencement (\ref)

Il peut arriver que la fonction 'Découper/numéroter du texte' n'insère pas le champ de référencement (\ref) devant chaque phrase. Dans ce cas, insérer à la main une étiquette de champ de référencement vide (\ref) devant chaque phrase. Puis faire

- 'Outils', 'Re-numéroter du texte', 'OK'

Exporter la base de données *texte* (pour ensuite la ré-importer en phrases)

- 'Fichier', 'Exporter', Sélectionner *Standard format*, 'OK'
- Donner un nom à la nouvelle base (par exemple le même nom que l'identifiant choisi pour référencer les phrases)

Importer la base nouvellement créée

(S'il n'existe pas déjà) commencer par créer un type de base de données ayant le marqueur de référencement (\ref) comme champ principal.

- Sélectionner le menu 'Projet', 'Type de base de données'
- Sélectionner le type de la base d'origine (habituellement *Text*)
- Clic sur le bouton 'Copier'
  - *Nom* : donner un nom au nouveau type (par ex *refText*)
  - *Marqueur d'enr.* : le marqueur choisi pour le champ de référencement (\ref)
- 'Fichier', 'Ouvrir', Rechercher la base de données de phrases à importer
- Sélectionner le type de base de données pour les phrases (*refText*)
- 'OK'

La base de données de phrases référencées par une étiquette de type *label\_num* est maintenant prête pour la synchronisation audio.

T134-C27-N.db	
\id	T134-C27
\tx	pínám né mǎtìè gbàsèè.
\vm	pín H -ám né mǎtìè gbàsèè.
\vma	nom_propre D -moi être Matie gbàsèè
\vam	n Fct -PP préd Npr Npr
\tr	Je m'appelle Mathieu Gbasè.
\dt	23/Mar/2007
\tx	né tò kó..
\vm	né tò kó..
\vma	être conte de
\vam	préd n fct
\tr	C'est le conte de...
\dt	23/Mar/2007
\tx	kó nàà kó kpáá-dìpà, nè fèé
\vm	kó nàà kó kpáá-dìpà, nè BHa fe F
\vma	de mère de Kpáá-dìpà qui Acc mourir D
\vam	fct n fct Npr sub mV v F
\tr	de la mère de Kpáá-dìpà qui va mourir que je va
\dt	29/Mar/2007
\tx	kpáá-dìpà, kpáá-dìpà, ʔá-nè nàà kǎà, ʔ
\vm	kpáá-dìpà, kpáá-dìpà, ʔá-nè nàà kó ʔà, ʔ
\vma	Kpáá-dìpà Kpáá-dìpà top mère de elle ʔ
\vam	Npr Npr sub n fct pr ʔ

T134-C27-NN.db	
\id	T134-C27
\ref	T134-C27.001
\tx	pínám né mǎtìè gbàsèè.
\vm	pín H -ám né mǎtìè gbàsèè.
\vma	nom_propre D -moi être Matie gbàsèè
\vam	n Fct -PP préd Npr Npr
\tr	Je m'appelle Mathieu Gbasè.
\dt	23/Mar/2007
\ref	T134-C27.002
\tx	né tò kó..
\vm	né tò kó..
\vma	être conte de
\vam	préd n fct
\tr	C'est le conte de...
\dt	23/Mar/2007
\ref	T134-C27.003
\tx	kó nàà kó kpáá-dìpà, nè fèé nè
\vm	kó nàà kó kpáá-dìpà, nè BHa fe H nè
\vma	de mère de Kpáá-dìpà qui Acc mourir D in
\vam	fct n fct Npr sub mV v Fct m
\tr	de la mère de Kpáá-dìpà qui va mourir que je vais rac
\dt	29/Mar/2007
\ref	T134-C27.004
\tx	kpáá-dìpà, kpáá-dìpà, ʔá-nè nàà kǎà, nàà
\vm	kpáá-dìpà, kpáá-dìpà, ʔá-nè nàà kó ʔà, nàà
\vma	Kpáá-dìpà Kpáá-dìpà top mère de elle mère
\vam	Npr Npr sub n fct pr n
\tr	Kpáá-dìpà, Kpáá-dìpà, voilà que sa mère est morte.
\dt	23/Mar/2007