

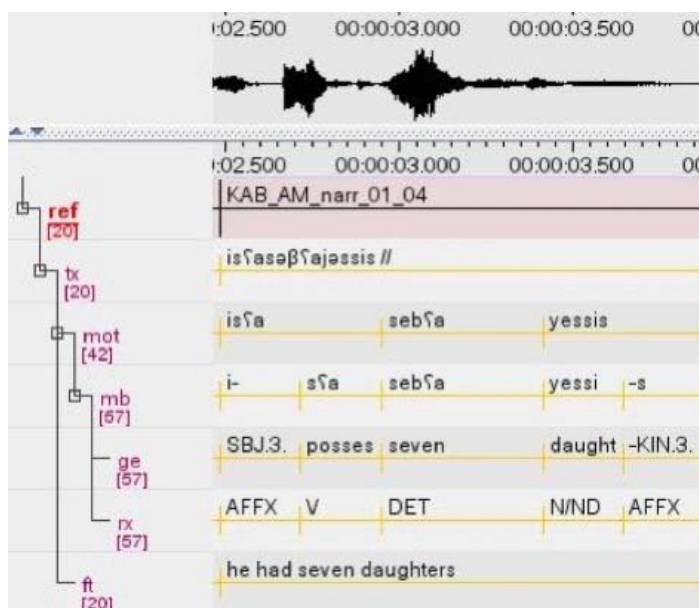
Application du logiciel ELAN à l'annotation linguistique

Elan est un logiciel développé par le Max Plank Institute for Psycholinguistics de Nijmegen aux Pays-Bas. Il permet l'annotation d'enregistrements audio ou vidéo (pour la langue des signes par exemple).

Des lignes d'annotation (tiers) synchronisées avec le son et/ou l'image sont créées puis découpées en segments temporels par l'utilisateur. Ces lignes pourront comporter par exemple la transcription de la parole de chaque interlocuteur, chacun sur sa propre ligne, ou bien dans le cas d'une vidéo, la documentation de la gestuelle du signeur.

L'utilisation d'ELAN pour l'annotation morpho-syntaxique de textes issus d'enregistrement de la parole, en lien avec le son, a été défini dans le cadre du projet de corpus afro-asiatique *CorpAfroAs* et comporte 7/8 lignes d'annotation :

- une ligne **ref** (type 'none', directement liée au temps, pas de parent) de segmentation de l'enregistrement audio en unités prosodiques (terminales et non-terminales). Cette ligne porte un identifiant d'unité généré automatiquement par la fonction *label* (du menu *tier*) une fois l'annotation achevée.
- une ligne **tx** (type 'symbolic association', enfant de **ref**) pour transcrire le contenu oral de chaque segment, en phonétique *large*.
- une ligne **mot** (type 'symbolic subdivision', enfant de **tx**) pour isoler chaque mot de la phrase (rem: dans cette configuration, les mots ne sont pas alignés avec le son)
- une ligne **mb** (type 'symbolic subdivision', enfant de **mot**) pour la segmentation des mots en morphèmes
- une ligne **ge** (type 'symbolic association', enfant de **mb**) pour gloser chaque morphème
- une ligne **rx** (type 'symbolic association', enfant de **mb**) pour donner une catégorie à chaque morphème
- une ligne **ft** (type 'symbolic association', enfant de **tx**) pour la traduction libre du contenu oral du segment.
- une dernière ligne **Mft** (type 'none', pas de parent) est prévue qui permet de définir des unités d'une taille temporelle regroupant plusieurs unités prosodiques dont les traductions individuelles seraient dépourvues de sens.



Une fois les enregistrements oraux annotés, un module permet de faire des recherches plus ou moins complexes d'occurrences de morphèmes dans des contextes particuliers. De la liste des morphèmes répondant aux critères de recherche, on peut ensuite sauter vers le segment correspondant pour l'écouter ou le retrouver dans son contexte.

La formation s'appliquera à permettre à l'utilisateur de comprendre comment paramétrer l'outil pour reproduire ce schéma tout en le rendant capable de l'adapter à ses propres besoins particuliers.

Gestion des documents

ELAN gère les annotations d'un document audio ou/et vidéo. Il génère un fichier texte au format XML (*.eaf) contenant les annotations saisies par l'utilisateur, ainsi que les références temporelles aux différents fichiers audio/vidéo (chemin d'accès et noms) auxquels ces annotations sont associées. Chaque fichier ELAN est associé à un fichier de préférences .pfsx qui conserve le paramétrage d'affichage de la dernière ouverture du fichier.

1. Création d'un fichier d'annotation (Ouverture d'un fichier audio/vidéo)

→ FILE, NEW

La création d'un nouveau fichier d'annotation ELAN commence par l'ouverture du/des fichier(s) *média* (audio/vidéo) que le logiciel va permettre d'annoter. Différents formats de fichier média sont reconnus par ELAN :

- Windows (par ordre de préférence DirectX étant la meilleure solution):
 - mpg, mpeg, wav mpg4, mp4, wma* (Windows Media Player)
 - mov†, qt (QuickTime)
- MacOS (par ordre de préférence):
 - QuickTime avec l'environnement Cocoa (QTKit)
 - QuickTime avec QuickTime for Java
- Linux: JMF

Le fichier *média* à ouvrir dans le projet ELAN est recherché dans la fenêtre de gauche et versé dans la fenêtre de droite grâce au bouton >>

4 fichiers vidéo et/ou 1 fichier audio peuvent être ouverts dans un projet ELAN. Afin que les vidéos démarrent au même point, elles peuvent être synchronisées par

→ **OPTIONS, MEDIA SYNCHRONISATION MODE.**

Pour visualiser la bande son de la vidéo, il faudra au préalable l'avoir extraite du fichier vidéo pour en constituer un fichier audio qui sera ouvert dans le projet ELAN (cf. Annexe 3)

Une fois le fichier crée, l'enregistrer en lui donnant si possible le même nom que le fichier média.

→ FILE, SAVE AS

2. Ouverture ou Importation d'un fichier d'annotation

Pour ouvrir un fichier ELAN d'extension .eaf

→ FILE, OPEN

Rechercher le fichier d'annotation précédemment créé par ELAN (*.eaf)‡

L'ouverture d'un fichier ELAN entraînera l'ouverture des fichiers média qui lui sont liés. Si les fichiers média associés au fichier d'annotation ELAN ont été renommés ou déplacés, ELAN proposera de les rechercher ou bien de passer outre.

* Contrairement aux autres fichiers 'média', la vitesse d'écoute des fichiers *.wma (Windows Media Audio) ne peut pas être modifiée.

† Pour les fichiers *.mov (c.à.d Cinepak-Quicktime-Movies) il est important qu'ils soient '*selfcontained*', c.à.d que les informations vidéo doivent être intégrées dans le fichier *.mov lui-même. Si ce n'est pas le cas, ELAN ne pourra pas jouer le fichier.

‡ EUDICO Annotation Format

Suivant le nombre de fichiers choisi dans EDIT, PREFERENCES, USER INTERFACE, ELAN propose à l'ouverture un certain nombre de fichiers récemment ouverts

→ FILE, OPEN RECENT FILE

ELAN peut importer différents formats de fichier d'annotation

→ FILE IMPORT

- Shoebox/Toolbox (*.txt)
- FLEEx (SIL Fieldworks Language Explorer)
- CHAT (*.cha)
- Transcriber (*.trs)
- CSV / Tab délimité (en provenance de Excel par exemple)
 - . CSV : Comma-separated values est un format informatique ouvert représentant des données sous forme de valeurs séparées par des virgules
 - . Tab delimited : données sous forme de valeurs séparées par des tabulations
- Praat textGrid

Une fois le fichier importé, l'enregistrer en lui donnant si possible le même nom que le fichier média.

3. Affichage/Ajout des fichiers média associés à un fichier d'annotation

Un fichier d'annotation ELAN fait référence à un/des fichier(s) média dont il représente l'annotation. Cette référence peut être perdue à cause d'un renommage ou d'un déplacement du fichier média dans un autre répertoire. Pour vérifier le nom des fichiers média associés ou associer des fichiers média au fichier d'annotation ouvert :

→ EDIT, LINKED FILE

Lorsqu'un fichier d'annotation a été importé, il est ainsi possible de lui associer (lier) ensuite un/des fichier(s) média avec le bouton

ADD

rechercher le fichier puis faire APPLY

Lorsqu'une ligne de média est sélectionnée, on peut préciser :

SET MASTER MEDIA : lequel est le maître

SET EXTRACTED FROM : le fichier vidéo dont est issu un fichier audio

4. Décalage entre annotations et image/son

Lorsqu'un fichier audio/vidéo est lié à un fichier d'annotation, il est possible qu'un décalage temporel survienne entre les deux (suppression a posteriori du message d'introduction dans le fichier audio par exemple). Pour remédier à cela:

→ ANNOTATION, SHIFT ALL ANNOTATIONS

Il est également possible de décaler seulement une annotation dans certaines conditions

→ ANNOTATION, SHIFT

5. Enregistrement d'un fichier ou d'un segment d'annotation

→ FILE, SAVE

Le fichier d'annotation (.eaf) est enregistré indépendamment des fichiers média qui ont été annotés. Ces fichiers audio/vidéo doivent être conservés tel quel pour être retrouvés lors d'une prochaine ouverture du fichier ELAN.

→ FILE, SAVE SELECTION AS

Permet d'enregistrer dans un nouveau fichier ELAN toutes les annotations du segment temporel préalablement sélectionné. La durée des segments sélectionnés est conservée ainsi que le lien avec le fichier média d'origine (intégral), mais l'annotation se retrouvera à l'origine de l'échelle du temps d'où un décalage prévisible entre annotation et audio/vidéo. (solution: ANNOTATION, SHIFT ALL ANNOTATIONS ou bien Editer le fichier audio/vidéo pour qu'il corresponde au fichier d'annotation. (cf. Annexe 3)

→ FILE, AUTOMATIC BACKUP

Paramètre ELAN pour une sauvegarde automatique toutes les 1, 5, 10, 20 ou 30 minutes

Annotations tiers et types linguistiques

Le processus d'annotation se fait en trois étapes: définition des types linguistiques et des lignes d'annotation (tiers) (voir sections 4.2.1 et 4.3), partitionnement des segments temporels (voir section 29), saisie des annotations (voir section 4.6).

Il existe 2 types de tiers :

- **indépendant**, qui contient des informations directement reliées à un intervalle temporel, c.à.d. qu'elle est synchronisée temporellement (*time-alignable*).
- **référentiel**, qui contiennent des informations qui sont liées aux annotations d'une autre tier (appelée "tier parent"). Ils ne sont généralement pas liés directement à l'axe du temps. (Certains d'entre eux peuvent l'être – mais seulement à l'intérieur de l'intervalle de temps déterminé par le tier parent, voir plus bas.)

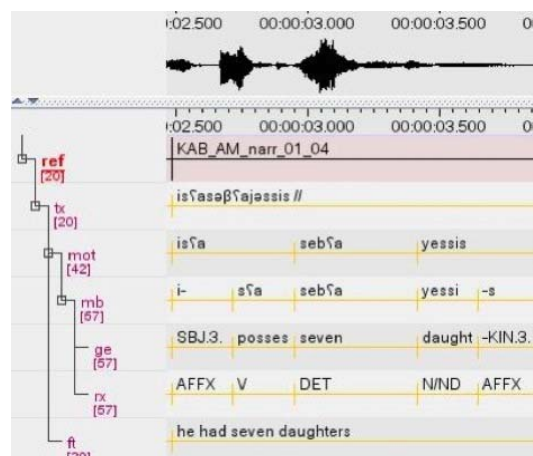
Un exemple: une tier de transcription peut être indépendante et synchronisée temporellement, en étant directement liée aux intervalles de temps de paroles du locuteur. Une tier de traduction, par contre, sera référentielle et non time-alignable: elle fera référence à la tier de transcription, et pas directement à l'axe du temps. Par définition, elle héritera de l'alignement temporel de la tier de transcription, c.à.d. de sa tier parent.

Dans la zone d'affichage des annotations, l'étiquette d'une tier référentielle aura la même couleur que celle de la tier indépendante dont elle est l'enfant.

Il est possible de construire des hiérarchies imbriquées, par exemple, la tier A peut être la tier parent de la tier B, la tier B parent de la tier C, etc.

L'illustration suivante présente le modèle d'annotation CorpA

Tier	Type / Stéréotype*	Hierarchie
ref	Indépendant : None	Sans parent
tx	Réf : symbolic association	Enfant de ref
mot	Réf : symbolic subdivision	Enfant de tx
mb	Réf : symbolic subdivision	Enfant de mot
ge	Réf : symbolic association	Enfant de mb
rx	Réf : symbolic association	Enfant de mb
ft	Réf : symbolic association	Enfant de ref



Les tiers *parent* et *enfant* sont liés de telle manière que des modifications faites sur le parent affecteront aussi les enfants (mais pas inversement):

* cf 11. Stéréotypes

- Si vous effacez une tier parent, tous ses enfants seront supprimés également. De même si vous supprimez une annotation dans une tier parent, les annotations des enfants disparaissent aussi.
- Si vous modifiez l'intervalle temporel d'une annotation d'une tier parent, l'intervalle temporel des tous ses enfants changera de la même manière. L'intervalle temporel d'un enfant ne peut pas être modifié indépendamment.

6. Afficher la dépendance entre les tiers

→ VIEW, TIER DEPENDENCIES

Pour afficher la hiérarchie des tiers dans la zone des noms des tiers :

→ CLIC-DROIT dans la zone des noms des tiers, SORT TIERS, SORT BY HIERARCHY

Stéréotypes

Chaque tier se voit assigner un type linguistique. **Chaque type linguistique spécifie un nombre de contraintes** qui s'appliquent à toutes les tiers de ce type. Ces jeux de contraintes sont résumées implicitement dans ce qu'on appelle des '**stéréotypes**'. Les cinq stéréotypes suivants sont actuellement disponibles :

- **None** : une tier de ce type sera indépendante, c.à.d que ses annotations sont directement liés à l'axe du temps et ne seront pas contraintes par d'autres tiers
- **Time Subdivision** : Une annotation de la tier parent peut être subdivisée en sous-unités dans une tier enfant de ce type. Ces sous-unités seront elles-même liés au temps. Il est à noter qu'il ne peut y avoir de trous dans les annotations, c.à.d. que les sous-unités doivent se suivre les unes après les autres. Remarque: la tier parent doit être elle-même de type alignée au temps.

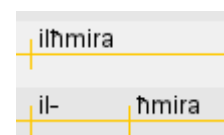
Ex., une parole transcrite sur une tier correspondant à une phrase peut être subdivisée en mots dans une tier enfant – chacun des mots étant lui-même lié à l'intervalle de temps correspondant dans le signal audio.

na me be di last born for my family									
na	me	be	di	last	born	for	my	family	

Les annotations de telles tiers sont synchronisées temporellement (time-alignable). Néanmoins, elles diffèrent des annotations de tiers indépendantes par le fait qu'elles sont contraintes dans l'intervalle temporel de l'annotation de leur parent.

- **Symbolic Subdivision** : Semblable à *Time Subdivision*, sauf que les sous-unités ne sont pas synchronisées temporellement.

Ex., un mot de la tier parent peut être segmenté dans la tier enfant en morphèmes, qui ne seront pas synchronisés temporellement. L'intervalle temporel du mot est subdivisé à parts égales pour chaque morphème composant le mot. Il serait en effet trop fastidieux de vouloir synchroniser chaque morphème avec son image sonore et cela n'aurait aucun intérêt.



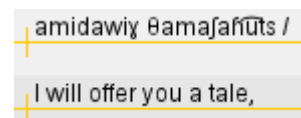
- **Included In** : Les annotations de la tier enfant sont synchronisées au son et tombent à l'intérieur des limites des annotations de la tier parent, mais il peut y avoir des trous entre les annotations enfant.

Ex., une phrase contenant des pauses peut être découpée en mots dans la tier enfant, les pauses constituant des trous dans les annotations.

so i get + three elder ones									
so	i	get			three	elder	ones		

- **Symbolic Association:** L'annotation de la tier (enfant) ne peut pas diviser l'annotation de la tier parent, c.à.d. qu'il y a une correspondance terme à terme entre l'annotation parent et celle qui s'y réfère dans la tier enfant.

Ex., une phrase dans une tier parent a exactement une traduction libre. On encore, un mot a exactement une définition.



Vous pouvez définir autant de tiers que vous voulez et pour chacune d'elles vous aurez à définir son type et son parent s'il doit en avoir un. Il est conseillé de prendre des décisions sur le type d'information que vous souhaitez entrer (et par conséquent sur le type de tiers dont vous aurez besoin) suffisamment tôt dans le processus d'annotation. Néanmoins, il est toujours possible à un stade ultérieur de changer le parent d'une tier dépendante ou de copier une tier puis de modifier la copie.

7. Définir un type linguistique

→ TYPE, ADD NEW LINGUISTIC TYPE...

La fenêtre de dialogue Add type apparaît.

Type name : Donner un nom au type que vous voulez créer

Stereotype : Choisissez son stéréotype

Use control vocabulary : permet de choisir une liste fermée de vocabulaire (prédéfinie) pour limiter le contenu de l'annotation à une membre de cette liste (cf. 26)

ADD

Plusieurs tiers peuvent être d'un même type. Par exemple (cf plus haut), les tiers **ge** (glose) et **rx** (catégorie) qui servent à documenter chaque morphème de la tier **mb** (morphem break) peuvent être du même type, lui-même de stéréotype *symbolic association* puisqu'il s'agit d'une correspondance terme à terme entre un morphème et sa glose ou sa catégorie grammaticale.

Néanmoins si une liste de vocabulaire devait être utilisée pour contrôler l'une des tiers (catégorie par exemple), il faudrait alors créer 2 types (de nom différents) mais de même stéréotype.

8. Ajouter une nouvelle tier

→ TIER, ADD NEW TIER.

La boîte de dialogue Add tier apparaît

Tier name : donnez un nom (étiquette) à la tier

Participant : si c'est pertinent, donnez le nom ou identifiant du locuteur

Annotator : si nécessaire, donnez votre nom ou identifiant

Parent tier : déroulez la fenêtre et choisissez le parent

Si la tier n'a pas de parent, sélectionnez **none** (tier indépendante)

Linguistic type : déroulez la fenêtre et choisissez le type

ADD

9. Fichier modèle (template)

Lorsqu'un projet ELAN est correctement configuré, il est possible d'enregistrer sa configuration (c.à.d. de conserver le paramétrage des *types* et des *tiers*) afin de pouvoir le réutiliser dans un nouveau projet.

→ FILE, SAVE AS TEMPLATE

Sélectionnez l'emplacement où conserver ce modèle et donnez-lui un nom.

Lors de la création d'un nouveau fichier ELAN (une fois sélectionnés les fichiers média à annoter), il est possible d'importer en même temps les types et les tiers d'un modèle.

10. Importer des types

→ TYPE, IMPORTER DES TYPES, BROWSE

Rechercher le fichier modèle (.etf) ou à défaut un fichier ELAN (.eaf) devant servir de modèle.
Si des types avaient été créés avant l'importation, ceux du même nom ne seront pas importés.

IMPORT, CLOSE

Les types sont importés (mais ça ne se voit pas), on peut le vérifier par

→ TYPE, CHANGE LINGUISTIC TYPE, CLOSE

11. Importer des tiers

→ TIER, IMPORTER DES TIERS, BROWSE

Rechercher le fichier modèle (.etf) ou à défaut un fichier ELAN (.eaf) devant servir de modèle.

IMPORT, CLOSE

Les nouvelles tiers s'affichent. Si une tier importée existait déjà avant l'importation, la nouvelle tier sera ajoutée avec le même nom suivi de -cp (pour copie)

12. Modifier les attributs d'une tier

Il est possible d'accéder aux attributs d'une tier pour en modifier son étiquette (nom), le Participant, ou le type (mais pas le parent, pour cela cf. 22)

→ CLIC-DROIT SUR L'ÉTIQUETTE DE LA TIER

CHANGE ATTRIBUTES OF

Tier name : Renommer la tier

Participant : changer le nom du locuteur

Annotator : changer le nom de celui qui fait l'annotation de cette ligne

Linguistic Type: ici, on ne peut choisir qu'un type de même *stéréotype*

CHANGE, CLOSE

Il n'est pas toujours possible de changer le type d'une tier : par exemple une tier d'un type 'symbolic' ne pourra pas être d'un type 'alignable' puisqu'il n'a pas de vrais index temporels.

13. Supprimer une tier

→ CLIC-DROIT SUR L'ÉTIQUETTE DE LA TIER

DELETE

Evidemment toutes les annotations de la tier seront perdues.

Attention, la suppression d'une tier supprimera également toutes les tiers dont elle est le parent. (un message signalant les tiers qui seront supprimées permet d'annuler la décision de suppression)

14. Copier une tier

Il peut être intéressant de copier la segmentation et le contenu d'une tier (et éventuellement ses enfants) dans une nouvelle tier. Il est alors possible de choisir pour la nouvelle tier un autre type que celui de la tier de départ. La nouvelle tier portera le nom de celle à copier, suivi de '-cp'.

→ TIER, COPY TIER

CLIQUEZ sur l'étiquette de la tier à copier

COCHEZ la case *COPY DEPENDING TIERS AS WELL* si vous souhaitez copier aussi les tiers enfants (les tiers créées porteront le nom original des enfants suivi de '-cp')

NEXT

CLIQUEZ sur l'étiquette de la **tier parent** de la nouvelle tier. (**Pour synchroniser cette nouvelle tier directement au temps, choisir *Transcription (none)***).

NEXT

CHOISISSEZ le **type linguistique** de la nouvelle tier

FINISH

RENOMMER la/les nouvelle(s) tier(s) (cf. 17)

15. Changer le parent d'une tier

Cette opération peut être délicate à cause des liens de dépendances (et des stéréotypes) entre les tiers. Pour cette raison ce processus créera une copie (-cp) de la tier concernée, sous son nouveau parent.

→ TIER, CHANGE PARENT OF TIER

CLIQUEZ sur l'étiquette de la tier concernée par le changement de parent, NEXT

CLIQUEZ sur l'étiquette de la **nouvelle tier parent**, NEXT

CHOISISSEZ le **type linguistique** de la tier, FINISH

Si le processus s'est bien passé, supprimez l'ancienne tier et renommez la nouvelle.

16. Sélectionner un segment temporel sur une tier indépendante

POSITIONNER LE CURSEUR DE LA SOURIS dans la zone d'affichage du son (ligne des temps)

CLIQUEZ au début de la zone à sélectionner et MAINTENEZ APPUYÉ LE BOUTON de la souris

DÉPLACEZ LA SOURIS jusqu'à la fin de la zone à sélectionner. (Vous pouvez faire dérouler la base des temps en maintenant le curseur dans la partie droite de l'écran, à gauche de l'ascenseur vertical.)

RELÂCHER LE BOUTON DE LA SOURIS

17. Créer un nouveau segment et son annotation

SÉLECTIONNEZ LE SEGMENT **voulu avec la souris**

CLIC-DROIT au niveau de la TIER INDÉPENDANTE

NEW ANNOTATION HERE

TAPEZ LE CONTENU DE L'ANNOTATION

CTRL/ENTRÉE pour valider (ou ENTRÉE, cf. 22)

18. Subdiviser une annotation

Lorsqu'une tier est de type *None* (indépendante), on peut subdiviser un segment existant à un emplacement choisi,

SÉLECTIONNER LE SEGMENT en cliquant dessus

POSITIONNER LE CURSEUR sur la tier à l'endroit où l'on veut créer un nouveau segment

CLIC-DROIT, **SPLIT ANNOTATION**

Le contenu du nouveau segment (et des éventuels enfants en association) sera le même que celui du segment initial (de même pour les enfants en *association*). Par contre les contenus des éventuels enfants en *subdivision* ne seront conservés que sur le segment de gauche.

Lorsqu'une tier est d'un type *Subdivision*, il est possible d'ajouter un enfant à l'annotation parent. Le nouvel enfant peut être placé avant ou après l'enfant sélectionné.

SÉLECTIONNER L'ANNOTATION à côté de laquelle on veut créer un 'sibling'

CLIC-DROIT, **NEW ANNOTATION BEFORE/AFTER**

19. Déplacer une annotation

CLIQUEZ SUR LE SEGMENT À MODIFIER (la ligne support devient bleue)

MAINTENEZ LE BOUTON *ALT* APPUYÉ

CLIQUEZ SUR L'ANNOTATION SANS LÂCHER, ET DÉPLACEZ-LA (le segment actif devient vert pendant le déplacement)

20. Modifier les frontières d'un segment

CLIQUEZ SUR LE SEGMENT À MODIFIER (la ligne support devient bleue)

MAINTENEZ APPUYÉ LA TOUCHE *MAJ*

CLIQUEZ À LA POSITION VOULUE, au début et/ou à la fin du segment

RELÂCHER LA TOUCHE *MAJ*

CTRL/ENTRÉE

Suivant le mode de propagation du temps (cf. 25) les frontières des annotations suivantes et précédentes seront modifiées ou non.

Ou bien

SÉLECTIONNEZ LE SEGMENT à modifier en cliquant dessus

MAINTENEZ APPUYÉ LA TOUCHE *ALT*

AMENER LE CURSEUR AU NIVEAU DE LA FRONTIÈRE À MODIFIER

CLIQUEZ ET RELÂCHER LA TOUCHE *ALT*, le curseur se transforme en une flèche double

DÉPLACEZ LA SOURIS dans le sens voulu

RELÂCHER À L'EMPLACEMENT VOULU

Ou encore

CLIQUEZ SUR LE SEGMENT À MODIFIER (la ligne support devient bleue)

SÉLECTIONNER AVEC LA SOURIS LA NOUVELLE ZONE que vous voulez affecter à cette annotation

CTRL/ENTRÉE

21. Saisir/modifier une annotation sur un segment existant ou dépendant

DOUBLE-CLIC DANS L'ESPACE CORRESPONDANT AU SEGMENT DE LA TIER VOULUE

Une cellule de saisie s'ouvre.

TAPER VOTRE ANNOTATION (activer le clavier phonétique si nécessaire) ou bien coller le texte du presse-papier (Ctrl/V)

CTRL/ENTRÉE (ou ENTRÉE, cf. 23) pour enregistrer et fermez la cellule de saisie

Rem : ESCAPE pour sortir de la cellule de saisie sans la modifier

Un segment *dépendant* (de type association ou subdivision) qui n'a pas encore été créé peut l'être directement en double-cliquant dans l'espace où le segment doit être créé.

22. Modifier les paramètres d'édition de ELAN

Pour simplifier la saisie dans ELAN, on pourra préférer terminer une saisie d'annotation par la touche **Entrée** plutôt que **Ctrl/Entrée**. Pour cela :

→ EDIT, PREFERENCES, EDIT PREFERENCES, EDITING

VALIDER L'OPTION : *Enter key commits changes...*

Pour éviter que ELAN recentre l'affichage d'une annotation lorsqu'on clique dessus, ce qui est déstabilisant en cas de double clic (pour édition)

DÉCOCHER L'OPTION : *Always center the active annotation*

23. Création d'un vocabulaire contrôlé

Un *vocabulaire contrôlé* est une liste de mots (ou étiquettes) devant servir à contraindre l'annotation de certains tiers (par ex: la catégorie grammaticale). Un vocabulaire contrôlé peut être affecté à un *type* de tiers. Les tiers du type concerné ne pourront contenir qu'un élément de la liste qui se déroulera lors de l'édition de l'annotation. Pour créer un vocabulaire contrôlé:

→ EDIT, EDIT CONTROLLED VOCABULARIES

CV NAME : donner un nom à la nouvelle liste, ADD

ENTRY VALUE : saisir un à un les mots de la liste, ADD

Vous pouvez associer un raccourci clavier à chaque élément de la liste et lui attribuer une couleur différente.

SÉLECTIONNER UN ÉLÉMENT DE LA LISTE

MORE OPTIONS

CLOSE

24. Affecter un vocabulaire contrôlé à un type

Pour affecter un vocabulaire contrôlé à un type :

→ TYPE, CHANGER TYPE

UTILISER LE VOCABULAIRE CONTRÔLE : choisir dans la liste déroulante

MODIFIER

Dorénavant lorsque vous double-cliquerez sur une annotation, une liste déroulante apparaîtra.

25. Mode de déplacement des annotations

Lors du déplacement d'une annotation (ou de ses frontières), il peut y avoir écrasement ou déplacement des annotations suivantes/précédentes suivant le mode de propagation défini dans

→ OPTIONS, PROPAGATE TIME CHANGES

Normal : c.à.d. **écrasement**: si vous étendez une sélection dans un intervalle temporel déjà occupé par une annotation, celle-ci sera réduite d'autant (jusqu'à être détruite). Les annotations existantes ne sont pas déplacées temporellement.

Bulldozer : si vous étendez une sélection dans un intervalle temporel déjà occupé par une annotation, celle-ci sera déplacé vers la droite/gauche. Pensez à un bulldozer qui pousserait toutes les annotations en supprimant les espaces vides entre elles. (les annotations sont préservées, mais les ancrs temporelles des annotations sont déplacées, sans conservation des intervalles entre elles)

Shift : comme pour le mode bulldozer, mais les espaces vides sont conservés (décalage temporel entre les annotations suivantes préservées)

26. Créer des annotations les unes à la suite des autres (sans trous)

CLIQUEZ, au niveau de la zone d'affichage du son, AU DÉBUT DU PREMIER SEGMENT À CRÉER (une curseur vertical rouge se place à cet endroit)


COCHEZ LA CASE *SELECTION MODE*


* DÉCLENCHER LA LECTURE AVEC LE BOUTON *PLAY* de gauche (bouton noir qui se transforme en *Pause*)

CLIQUEZ SUR LE BOUTON *PAUSE* pour arrêter la lecture (et la sélection) à la fin du segment voulu

CLIC-DROIT DANS LE SEGMENT SÉLECTIONNÉ au niveau de la tier d'annotation voulue, choisissez NEW ANNOTATION HERE

CTRL/ENTRÉE (ou ENTRÉE) pour fermer la case de saisie

CLIQUEZ SUR LE BOUTON DÉSÉLECTION  pour désélectionner le segment en cours

CLIQUEZ SUR LE BOUTON FIN DE SÉLECTION  pour amener le fil rouge en fin de sélection

RECOMMENCEZ LE PROCESSUS *

27. Segmenter au kilomètre

En passant en mode **SEGMENTATION**, il est possible de créer des segments à la volée, en jouant la bande son ou la vidéo. Les segments ainsi créés pourront ensuite être réajustés plus finement au fur et à mesure de la transcription.

→ OPTIONS, SEGMENTATION MODE

TWO KEYSTROKES : pour créer des segments non-contigus

ONE KEYSTROKE : pour créer des segments contigus

SELECTIONNEZ LA TIER sur laquelle vous voulez faire la segmentation en double-cliquant dessus. (Elle est dupliquée en rouge en dessous du signal sonore)

CLIQUEZ SUR LE BOUTON *PLAY*

TAPEZ LA TOUCHE <ENTRÉE> pour Démarrer/Terminer un segment

CLIQUEZ SUR LE BOUTON *PAUSE* pour arrêter

Les segments sont créés sur la tier sélectionnée. Un segment peut être écouté en cliquant dedans pour placer le cuseur rouge puis en cliquant sur le bouton de lecture d'une sélection. En approchant la souris d'une frontière d'un segment, celle-ci peut facilement être déplacée dans un sens ou dans l'autre, en cliquant dessus et en déplaçant la souris sans relâcher le bouton.

Un clic-droit sur un segment actif permet de le fusionner (MERGE) avec le précédent (BEFORE) ou le suivant (NEXT), ou bien de le scinder (SPLIT) en deux au niveau de la position du cuseur

Pour revenir au mode d'affichage normal :

OPTIONS, ANNOTATION MODE

28. Segmenter automatiquement à partir des silences (pauses)

Un outil de segmentation automatique basé sur les silences est disponible. Il permet de créer une tier comportant des segments de signal et des segments de silence qui pourra servir de base pour créer la tier de segmentation du son en unités prosodiques (tier **REF**)

→ ONGLET *AUDIO RECOGNIZER*

SÉLECTIONNER : **SILENCE RECOGNIZER MPI-PL** (agrandissez la zone *Parameters* grâce à la poignée de séparation (sous *Selection Mode*) entre cette zone et celle du son)

Sous *Silence level*, sélectionner SELECT BY EXAMPLE

dans la zone *Selection Panel*, sélectionner SELECTION

SÉLECTIONNEZ quelques millisecondes (~50/90 ms) de silence sur le signal audio, puis

CLIQUEZ SUR LE BOUTON PLUS (+) à droite de la fenêtre *Selections*, pour ajouter cet extrait et RECOMMENCER l'opération 3 ou 4 fois en sélectionnant des échantillons de silence plus loin dans la zone audio

Dans la zone *Settings* : Choisissez la durée minimale d'un silence (200~250 ms par exemple) et la durée minimale d'un non-silence (par exemple 80~100 ms)

Dans la zone *Progress*, cliquez sur le bouton **START**

Des barres verticales apparaissent dans la zone du signal audio, délimitant les zones de signal marquées d'un 'x' et les zones de silences marquées d'un 's'

Si vous n'êtes pas satisfait de la segmentation, Ajoutez (+) ou enlevez (-) des échantillons de silence dans la zone *Selections* ou changez les valeurs minimales de silences et non-silence dans la zone *Settings*, puis relancez **Start**.

Lorsque vous êtes satisfait de la proposition de segmentation

CLIQUEZ sur **CREATE TIERS**

Il est possible de créer soit 2 tiers indépendantes, une pour les silences (s), une pour la parole (x) ou bien une seule tier pour les 2, alternant les annotations x et s. Si l'on ne s'intéresse pas aux silences, on cochera uniquement la en face de x. Pour avoir une seule tier alternant les x et les s :

SÉLECTIONNEZ L'ONGLET **ALL SEGMENTATIONS**

CLIQUEZ SUR **CREATE**, puis **CLOSE**

Une tier *Channell* contenant les segments de silence 's' et de signal 'x' a été créée. Cette tier est de type *segmentation* (stereotype *None*) c'est-à-dire indépendante.

Pour vous servir de cette ligne comme ligne de base dans ELAN-CorpA, vous devrez changer son *type* en **REF**, ainsi que son nom en **REF** (cf. 17).

Remarque : La durée d'un segment peut être lue en le sélectionnant et en lisant cette valeur au dessus du bouton de sélection représentant une flèche pointant vers le haut. (Cette valeur pourra être reportée comme annotation d'une Pause par exemple)

29. Saisir les annotations - bis

L'option d'affichage *TRANSCRIPTION MODE* permet de saisir facilement les transcriptions et traductions, comme avec le logiciel *Transcriber*. Vous pouvez utiliser la tier **REF** pour la transcription et la dupliquer par la suite dans une tier que vous renommerez **TX**. Ou bien vous pouvez commencer par dupliquer (cf. 21) la tier **REF** puis la renommer **TX**, et faire de même pour créer la tier de traduction **FT**.

→ Menu **OPTIONS, TRANSCRIPTION MODE**

FONT SIZE : Choisissez la taille de la police d'affichage (14 ou 16 par exemple)

NUMBER OF COLUMNS : 2 si vous voulez faire la transcription et la traduction en même temps

CLIQUEZ SUR **<SELECT A TYPE>** dans la ligne 1 de la table, pour choisir le type **TX** de la tier de transcription (puis sur la ligne 2 pour choisir le type de la tier de traduction **FT** si vous voulez faire les 2 en même temps).

Si vous n'avez pas un type spécifique pour chaque tier,

CLIQUEZ SUR LE BOUTON **SELECT TIERS** décocher les lignes qui ne concernent pas les tiers à afficher, et sélectionner les tiers voulues dans les cellules Type1, Type2.

CLIQUEZ SUR LE BOUTON **APPLY**

Dans la table qui apparaît, chaque ligne correspond à un segment d'annotation. En plaçant le curseur dans une cellule de la table, le son correspondant est joué. Il suffit de transcrire dans la colonne choisie. Pour réécouter, cliquez sur le chiffre de la ligne voulue

Pour revenir au mode d'affichage normal :

OPTIONS, ANNOTATION MODE

30. Naviguer dans le texte

L'onglet **Text** permet de sélectionner une tier dont les annotations vont être affichées en continu, séparées par un point. Si on a choisi la tier contenant la transcription par exemple, on peut alors voir le texte en continu et en cliquant dans une phrase, l'annotation correspondante sera sélectionnée et affichée dans la zone annotation. De même pour la traduction.

→ **Onglet TEXT**

31. Naviguer dans le texte et corriger les annotations

L'onglet Grid permet d'afficher les annotations d'une tier ou de plusieurs tiers dépendantes **dans une table**. En cliquant sur une ligne de la table, l'annotation concernée est recherchée et présentée sur l'écran. Cela permet de naviguer facilement dans le texte. De plus les annotations sont directement modifiables dans les cellules de la table

→ Onglet GRID

SÉLECTIONNEZ LA TIER À AFFICHER, à la place de *Empty* (par exemple *REF*)

COCHEZ dans la case à gauche : **MULTIPLE TIERS WITH SYMBOLIC ASSOCIATION** pour voir ses enfants.

La table affichera la tier *REF* et ses enfants *TX* (transcription) et *FT* (traduction)

DOUBLE-CLIQUEZ dans une cellule pour l'éditer

CTRL/ENTRÉE (ou ENTRÉE) pour fermer une cellule saisie

32. Importer un fichier d'annotations tabulés

Un fichier texte tabulé comportant des lignes d'annotations séparées par des tabulations peut être importé dans ELAN, chaque ligne représentant un segment et chaque colonne une ligne d'annotation (tier). La première ligne peut contenir le nom de la tier de chaque colonne.

Par exemple

tx	ft
tak	A man,
e:girimwa	an old man,
sa:su:ba:bta	a very old man.

→ FILE, IMPORT, CSV/TAB DELIMITED

DÉCOCHEZ LES COLONNES À NE PAS IMPORTER

CHOISISSEZ *ANNOTATION* comme type linguistique pour les colonnes à importer

SPECIFY FIRST ROW OF DATA : 2 (numéro de la première ligne de données)

OK

Les tiers importées comporteront des annotations de durée toutes égales (par défaut 1s). Elles seront de type *None* (indépendantes)

Remarque: Si le texte tabulé contient des virgules, l'importation ne marchera pas. Commencez par remplacer les virgules par un symbole (par ex \$) avant de l'importer dans ELAN. Dans ELAN vous remplacerez en retour ce symbole par une virgule.

33. Importer un fichier d'annotation Toolbox

Un fichier annoté dans Toolbox comporte généralement les lignes *ref*, *tx* (text), *mb* (morphem breaks), *ge* (gloss), *ps* (part of speech) et *ft* (free translation). Il est possible d'importer directement ce type de fichier dans ELAN. Dans le modèle d'annotation *ELAN-CorpA*, il y a une ligne de plus, la tier *mot* qui contient les mots, isolés chacun dans une cellule.

Pour se conformer à ce schéma, il faudra donc dupliquer la ligne *tx* en une ligne *mot*. Pour cela, dans **Toolbox**, on peut exporter en *Standard Format* la base de données Toolbox en passant par une table de remplacement (cf. ci dessous) qui va se charger de dupliquer la ligne *tx* dans une ligne *mot*. DANS TOOLBOX :

FICHIER, EXPORTER, STANDARD FORMAT, MODIFIER

TABLE DE REMPLACEMENT : Rechercher le fichier *tx2mot.cct*

OK, donner un nom au nouveau fichier

Le fichier *tx2mot.cct* est un fichier texte qui contiendra les simples lignes suivantes :

```
"tx ' > set(tx) store(tx)
' > if(tx) clear(tx) endstore 'tx ' out(tx) 'mot ' out(tx) store(tx) endstore endif dup
```

Plus simplement dans un éditeur de texte comme **Notepad++**, on pourra dupliquer la tier *tx* en *mot* en faisant:

```
rechercher : \\tx (.*)$
remplacer : \\tx \\1\\r\\n\\mot \\1
en mode expressions régulières
```

Pour importer le fichier Toolbox dans ELAN:

FICHIER, IMPORTER

TOOLBOX FILE : rechercher le fichier Toolbox

Sélectionner **SET MARKER FIELDS**

FIELD MARKER : **REF**

STEREOTYPE : **NONE**

ADD

FIELD MARKER : **TX**

PARENT MARKER : **REF**

STEREOTYPE : **SYMBOLIC ASSOCIATION**

Pour le schéma ELAN-CorpA

AJOUTER

FIELD MARKER : **MOT**

PARENT MARKER : **TX**

STÉRÉOTYPE : **SYMBOLIC SUBDIVISION**

ADD

FIELD MARKER : **MB**

PARENT MARKER : **MOT**

STEREOTYPE : **SYMBOLIC SUBDIVISION**

AJOUTER

FIELD MARKER : **GE**

PARENT MARKER : **MB**

STEREOTYPE : **SYMBOLIC ASSOCIATION**

ADD

FIELD MARKER : **PS (OU RX)**

PARENT MARKER : **MM**

STEREOTYPE : **SYMBOLIC ASSOCIATION**

ADD

FIELD MARKER : **FT**

PARENT MARKER : **REF**

STEREOTYPE : **SYMBOLIC ASSOCIATION**

Vous pouvez enregistrer ce paramétrage pour pouvoir le réutiliser une prochaine fois (avec Load Markers)

STORE MARKERS

CLOSE

DEFAULT BLOCK DURATION (MS) : 2000 (soit 2 secondes)

OK

34. Créer des annotations à partir des chevauchements

Lorsqu'on a plusieurs tiers d'annotation (par exemple plusieurs locuteurs), on peut vouloir créer une tier indépendante comportant des segments d'annotations correspondant aux chevauchements temporels entre les tiers choisies (suivant qu'ils ont la même valeur ou pas, ou en rapport avec leur contenu). Les segments créés sur cette nouvelle tier pourront contenir la durée de chevauchements, la concaténation des annotations qui se chevauchent ou bien une valeur choisie.

→ TIER, CREATE ANNOTATION FROM OVERLAPS

Sélectionner les tiers concernées par le chevauchement (par ex: tx@SP1 et tx@SP2), NEXT

Plusieurs critères sur le chevauchement sont possibles:

- *regardless of their annotation values*: quel que soit le contenu des segments qui se chevauchent
- *and their annotation are equal* : seulement si leurs contenus sont identiques
- *and their annotation values are different* : seulement si leurs contenus diffèrent
- *according to specified constraints* : suivant le contenu de chacune des tiers qui se chevauchent

Destination tier : donner un nom de tier nouveau

Root or Child Tier : soit une tier indépendante, soit dépendante, mais il faut qu'elle soit de type alignable

Type : choisir (ou ajouter) un type alignable (stereotype None, *Time Aligned* ou *Included In*)

Value specification : soit la durée de chevauchement, soit une valeur donnée, soit la valeur d'une des tiers donnée, soit la concaténation des contenus des segments en chevauchement

FINISH

35. Supprimer les annotations à partir d'un index

On peut vouloir supprimer les annotations d'une tier à partir d'un index temporel, ou dans un intervalle de temps.

On commence par sélectionner la tier concernée en double-cliquant sur son nom. Celui-ci s'affiche alors en rouge. Puis on positionne le curseur temporel à l'emplacement concerné, ou bien on sélectionne le segment temporel voulu pour cette action.

→ ANNOTATIONS, DELETE, IN THE SELECTED TIME INTERVAL

ou bien LEFT / RIGHT OF CROSSHAIR

36. Créer une nouvelle série de tiers pour un autre locuteur

L'annotation de dialogue nécessite une série indépendante de tiers ref, tx, mot... pour chaque locuteurs, de façon à pouvoir gérer les chevauchements de parole. Lorsqu'une série a été créée ou bien importée avec un participant SP par exemple, on peut créer simplement une autre série pour SP2.

→ TIER, ADD NEW PARTICIPANT,

SELECT PARTICIPANT : SP,

SPECIFY NEW PARTICIPANT : SP2, SUFFIX

VALUE TO BE REPLACED : SP, NEW VALUE : SP2

37. Concaténer le contenu de plusieurs tiers dans une autre tier

On pourrait vouloir gagner de la place en exportant le contenu de 2 tiers en une seule. Par exemple, en concaténant les contenus des tiers *glose* et *catégorie*

→ TIER, MERGE TIERS : GE, RX, NEXT
REGARDLESS OF THEIR ANNOTATION VALUES, NEXT
DESTINATION TIER : **gerx**, NEXT
CONCATENATE THE VALUES, COMPUTE IN THE ORDER: rx, ge, FINISH

38. Changer la casse des annotations d'une tier

Après tokenisation de la tier **tx** dans **mot**, il convient de remplacer les majuscules des mots (début de phrase) par des minuscules. Ou bien on peut vouloir changer la casse des annotations de la tier catégorie.

→ TIER, CHANGE CASE OF ANNOTATION

39. Copier un groupe d'annotations

On peut copier un groupe constitué d'une annotation et de ses enfants (par exemple toute une phrase avec ses mots, morphèmes et gloses) et coller cet ensemble à un emplacement où il n'y a pas encore d'annotation, sur la même tier parent.

→ SÉLECTIONNER L'ANNOTATION PARENT à copier
CLIC-DROIT, COPY ANNOTATION GROUP
VÉRIFIER LA TAILLE DU SEGMENT SÉLECTIONNÉ
SÉLECTIONNER SUR LA MÊME TIER UN ESPACE DE MÊME TAILLE (attention de sélectionner le segment vide à partir de la fin du segment précédent)
CLIC-DROIT, PASTE ANNOTATION GROUP

Procédure d'interalignement dans ELAN-CorpA

Le processus d'interalignement tel que nous l'entendons, est celui par lequel les mots d'une phrase contenu dans une ligne (*mot* par exemple) sont segmentés en morphèmes dans une ligne en dessous (disons *mb*), chaque morphème étant ensuite glosé dans une 3ème ligne (*ge*) et étiqueté grammaticalement dans une 4ème ligne (*rx*). Durant ce processus, un alignement vertical est maintenu entre d'une part chaque mot et le premier morphème qui le constitue et d'autre part entre chaque morphème, sa glose et sa catégorie. Jusqu'à présent, ELAN ne permettait pas de s'appuyer sur un lexique pour segmenter et annoter un texte directement. Les seules solutions possibles étaient :

- d'opérer manuellement au découpage des unités et de remplir chacune d'elle
- de transférer le texte à annoter d'ELAN vers Toolbox qui lui, permet d'effectuer un interalignement à partir d'un lexique, et enfin de réimporter les données annotées dans ELAN.

L'idée était donc de simplifier la démarche en permettant à l'utilisateur d'effectuer un processus d'interalignement directement dans ELAN c.à.d que le découpage des unités en morphèmes et leur annotation s'effectuent directement dans les lignes appropriées (*mb*, *ge* et *rx*) du fichier ELAN.

Pour ce faire, un nouvel onglet « *Interlinearize* » a été créé [§].

Après avoir ouvert un fichier à annoter

- Cliquer sur l'onglet « *Interlinearize* ».

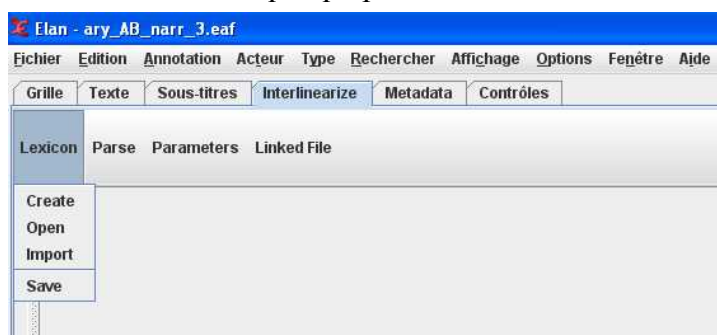
Création, ouverture, importation d'un lexique

La procédure d'interalignement repose sur l'existence d'un lexique qui peut être

- un tout nouveau lexique ELAN
- un lexique ELAN existant
- un dictionnaire Toolbox importé dans ELAN

Un sous-menu « *Lexicon* » permet de choisir

- Cliquer sur l'onglet « *Lexicon* »



Créer un lexique ELAN-CorpA

En choisissant *Create*, une fenêtre de sélection de fichiers s'ouvre. Choisissez le dossier dans lequel vous voulez enregistrer votre lexique et donnez-lui un nom. L'extension **.eaf** lui sera automatiquement ajoutée.

A l'écran s'affiche alors sur la gauche, un tableau contenant les différentes colonnes du dictionnaire, et sur la droite des éléments servant à l'interalignement et à l'édition du lexique (« Insert Record », « Insert Variant »... cf. fig. 3)

Ouvrir un lexique ELAN-CorpA

En choisissant *Open*, une fenêtre de sélection de fichiers s'ouvre. Choisissez le dossier dans lequel se trouve votre lexique (d'extension **.eaf**) sélectionnez-le, puis ouvrez-le.

[§] Depuis la version ELAN 5.0, une fonctionnalité d'interlignement a été ajoutée par le MPI à ELAN

Importer un lexique Toolbox

Une fenêtre s'ouvre permettant de choisir comme source de données un fichier dictionnaire Shoebox/Toolbox (.txt) ou un fichier dictionnaire ELAN-CorpA (.eaf).

- Choisir Fichiers du type « text files (*.txt) »
- Sélectionner un fichier dictionnaire issu de Shoebox/Toolbox
- Cliquer sur « Select »

Parmi les champs contenus dans chaque fiche du dictionnaire Toolbox, certains n'entrent pas en jeu dans le processus d'interalignement (c'est le cas des exemples, définitions etc.). ELAN-CorpA ne traite que les concepts ou champs suivants (colonne de droite):

- **Lexeme** (toute entrée dans le lexique: lexème, affixe, mot-forme),
- **Variant** (forme alternative du lexème en contexte),
- **Underlying form** (décomposition de l'entrée en ses constituants, lorsqu'il y a assimilation par ex.),
- **Glose** (sens du lexème),
- **Part of speech** (catégorie grammaticale. Dans Toolbox, la catégorie (\ps) d'un mot recouvre les différents sens du mot. Ici, chaque sens (Gloss) d'un même mot du lexique aura sa propre catégorie (Tier X),
- **Tier X** (catégorie associée à l'entrée),
- **date** (de dernière modification de l'entrée).

Ces champs doivent être liés aux champs présents dans le fichier Toolbox (colonne de gauche) afin d'être correctement traités.

Il est impératif que Lexeme, Glose et Tier X (ou Part of Speech) soient associés à un champ du fichier Toolbox. Toutefois, si vous n'utilisiez pas de champ \rx dans Toolbox, associez le champ de partie du discours de Toolbox (\ps) au champ Part of Speech de ELAN-CorpA. L'importation copiera le contenu du champ \ps dans le champ ELAN Tier X."

L'utilisateur choisit l'association voulue en appariant un à un les champs de droite aux champs de gauche.

- Sélectionner un champ dans la liste de droite
- Sélectionner un champ issu de votre fichier Toolbox dans la liste de gauche
- Cliquer sur la flèche située entre les deux listes

Le champ de droite sélectionné comporte maintenant une flèche «->» suivi du champ Toolbox approprié

Pour supprimer une correspondance, sélectionner dans la colonne de droite l'élément voulu, puis

- Cliquer sur la croix rouge

Pour déplacer la correspondance sur l'élément situé au-dessus, puis

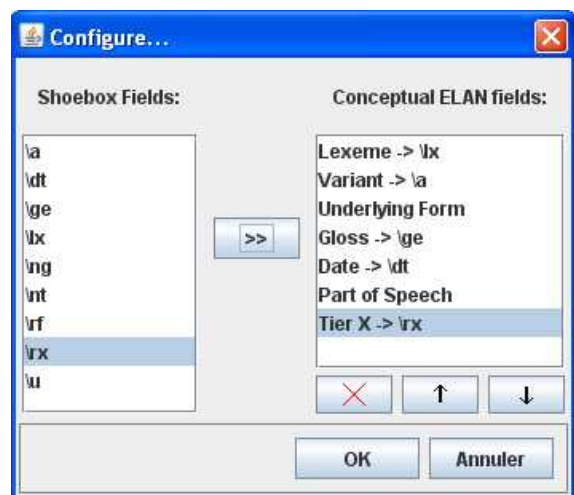
- Cliquer sur la flèche vers le haut

Pour déplacer la correspondance sur l'élément en dessous, puis

- Cliquer sur la flèche vers le bas

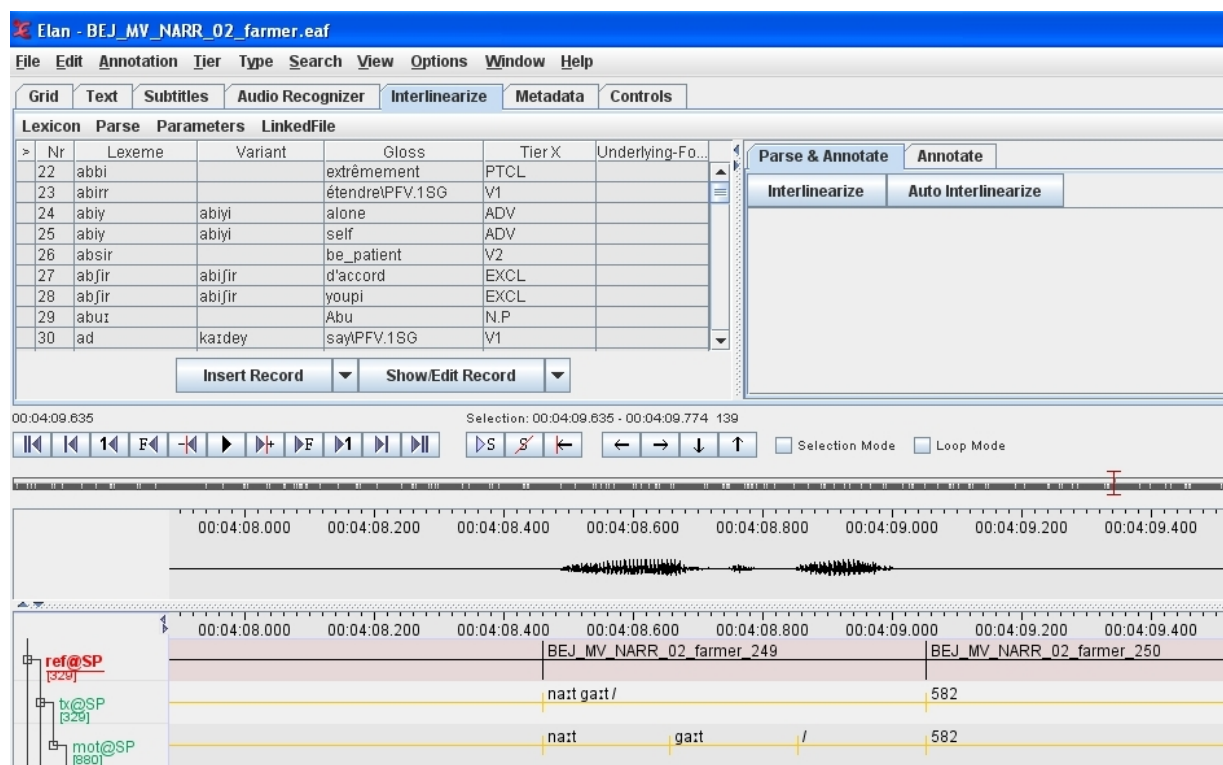
Lorsque toutes les correspondances nécessaires sont faites, cliquer sur « OK ».

Le programme récupère les éléments pertinents pour l'interalignement qu'il copie dans un nouveau fichier lexique ELAN-CorpA (auquel il faudra donner un nom); c'est ce fichier XML



nouvellement créé (d'extension **.eaf**) que le programme utilisera comme lexique pour l'interalignement (pour insérer de nouveaux morphèmes, modifier des éléments etc.).

A l'écran s'affiche alors sur la gauche, un tableau contenant les données du lexique, et sur la droite des éléments servant à l'interalignement et à l'édition du lexique (« Insert Record », « Insert Variant »...).



Ouverture automatique du lexique

Sur la droite des menus de la section *Lexique*, se trouve le menu *Linked File*. Celui-ci permet de choisir de rouvrir automatiquement ou non le lexique actuel au prochain démarrage de ELAN-CorpA. Par défaut la case devant le nom du lexique ouvert est cochée, il se rechargera donc automatiquement.

Paramétrer l'interalignement

Avant de lancer l'annotation interactive des mots, il faut choisir la tier de départ et définir les tiers d'annotation. Par défaut ces lignes sont « mot » pour la ligne contenant les mots à segmenter et annoter, « mb » pour la ligne de segmentation en morphèmes, « ge » pour la glose des morphèmes et « rx » pour l'étiquetage grammatical des morphèmes. Si ces lignes existent déjà, les annotations qu'elles contiennent déjà seront écrasées au fur et à mesure que le processus d'interalignement se fera. Ce paramétrage est à faire une fois pour toute.

Si ces lignes d'annotation n'existent pas encore dans votre fichier ELAN, faites:

- *Parameters, Interlinearize tier Parameters, configure interlinear Tiers*
- Choisissez la ligne source à segmenter et annoter : *Choose interlinear tier* (par ex *mot*)
- Cliquez sur *Ok*
- Choisissez les étiquettes de la ligne des segments (*mb*), de la glose (*ge*) et de la catégorie (*rx*)
- Cliquez sur le bouton *Create tiers*

Les lignes sont créées, le processus peut commencer.

Si ces lignes existent déjà dans votre fichier ELAN, faites:

- *Parameters, Interlinearize tier Parameters, Rename Interlinear Tiers*

- Donnez le nom de vos tiers
- Cliquez sur le bouton *Rename tiers*

Principes de fonctionnement de l'annotateur

Une entrée du lexique ELAN (que nous appelons un *Lexème*), peut être un *mot plein* (présentant éventuellement des formes alternatives appelées ici *variantes*), une *racine*, un *affixe* ou un *clitique*. Les affixes se reconnaissent conventionnellement dans ELAN-CorpA par la présence d'un tiret (-) à gauche ou à droite suivant qu'il s'agit respectivement d'un suffixe ou d'un préfixe, les clitiques par la présence d'un signe *égal* (=), et les racines peuvent se distinguer d'un mot plein par la présence d'un souligné (_). D'autres symboles peuvent être déclarés (par ex ~ pour les reduplications)

Le parseur commence par rechercher le mot en cours dans le lexique, au niveau du champ *Lexeme* ou du champ *Variant*. S'il est trouvé dans le champ *Lexeme*, il est présenté tel quel dans la section *Segmentation*. S'il est trouvé dans un champ *Variante*, c'est la forme de l'entrée lexicale du champ *Lexeme* qui est présentée dans la section *Segmentation*. L'entrée (ou les entrées en cas de polysémie ou d'homonymie) trouvée dans le lexique s'affiche dans la section *Lexique*. Un double-clic sur l'entrée du lexique valide les annotations du mot qui sont reportées dans les lignes correspondantes (glose et catégorie) de la zone d'annotation, sous le mot en cours. Le mot suivant du texte est sélectionné et le processus peut recommencer.

Si le parseur ne trouve pas le mot dans le lexique, il va commencer à rechercher une correspondance entre la fin ou le début du mot et les différents affixes/clitiques du lexique. Chaque fois qu'il trouve une correspondance, il isole l'affixe trouvé et relance la recherche sur le segment restant (qui est alors traité comme un mot). Toutes les segmentations possibles s'affichent, en fin d'analyse, dans la section *Segmentations*. Le choix de l'une d'elles (par un double-clic dessus) réduit l'affichage du lexique aux seuls morphèmes impliqués dans la segmentation choisie. Le choix de l'entrée lexicale correspondant à un morphème se fait en double-cliquant sur l'entrée du dictionnaire souhaitée (en cas de polysémie ou d'homonymie), le segment suivant étant alors automatiquement sélectionné dans la section *Segmentation*. Le dernier segment validé provoque le transfert des annotations dans le texte, et le mot suivant du texte est sélectionné et le processus peut recommencer.

Lorsqu'un mot (ou le reste d'un mot, une fois les affixes connus isolés) n'est pas trouvé dans le lexique et que le parseur ne peut pas le segmenter sur la base des affixes du lexique, il apparaît dans la section *Segmentations* précédé d'un astérisque (*). A ce stade, si ce mot comporte des affixes, il faut commencer par eux. Ensuite seulement la racine (ou le mot entier s'il n'est pas segmentable) sera entré, soit tel quel, soit sous forme de variante d'un mot existant.

Lancer le processus d'interalignement

Le parseur va chercher chaque mot à annoter dans le lexique, et s'il ne le trouve pas, il va essayer de lui trouver des segmentations possibles en fonctions des affixes contenus dans le lexique.

- Placer le curseur sur le premier mot de la ligne à segmenter, il se souligne en bleu.
 - Cliquer sur le bouton *INTERLINEARIZE* dans la section ***Segmentations***, à droite du lexique.
- Les différentes segmentations possibles s'affichent dans la section *Segmentations*, et le lexique se réduit aux seuls éléments intervenant dans les segmentations trouvées. *Le dernier segment non segmentable est précédé d'un astérisque s'il n'existe pas dans le lexique.*

Dans l'exemple ci-dessus, le mot *?arjabwa* présente 3 segmentations possibles. Les suffixes *-a*, *-b* et *-wa* trouvés dans le lexique conduisent à l'isolation d'un radical possible **?arj*.

Saisie d'une entrée et/ou d'une variante dans le lexique (*Insert record*)

Pour entrer un mot nouveau dans le lexique, qu'il s'agisse d'un lemme ou d'un affixe, on peut cliquer dans section *Lexique* sur le bouton « *Insert record* », ou plus simplement dans la zone *Segmentations* :

- CLIC-DROIT sur le mot précédé d'un astérisque (ici *?arj*)

- CLIC sur *INSERT RECORD*

une fenêtre avec le mot choisi apparaît. Il peut être modifié. Par exemple, ici le mot à ajouter est *?arjab* glosé comme nom propre *Aryab*.

- CLIQUER sur le bouton *SAVE RECORD*

Si le morphème que vous êtes en train de gloser contient d'autres éléments morphologiques qui ne peuvent être segmentés, ou que vous ne voulez pas isoler comme morphèmes séparés, vous pouvez utiliser la cellule à droite du symbole « \ » pour entrer la valeur grammaticale de ces éléments. Remarquez qu'il ne faut pas entrer le délimiteur (\) devant l'étiquette grammaticale, celle-ci sera automatiquement ajoutée dans la ligne d'annotation.

Une fois l'entrée créée, le processus peut être relancé de nouveau avec le bouton « *Interlinearize* ».

Maintenant, avec le mot *ʔarjab* entré dans le lexique, 3 segmentations complètes restent possibles.

Sélection de la segmentation et de la glose

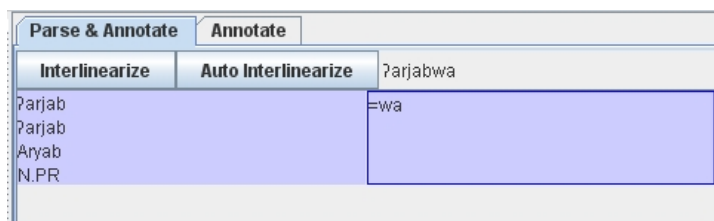
Il reste à faire le choix de la bonne segmentation

- DOUBLE-CLIC sur le premier segment de la ligne de segmentation voulue

Le lexique se réduit à l'entrée correspondante (elles pourraient être multiples en cas de polysémie ou d'homonymie).

- DOUBLE-CLIC sur l'entrée du lexique voulue.

L'annotation du premier segment s'affiche dans la section Segmentation et le segment suivant est sélectionné.



Et on recommence : un double-clic sur le

segment sélectionné permet de réduire le lexique aux entrées correspondantes, puis un double-clic sur l'entrée voulue dans le lexique valide le choix et le segment suivant est sélectionné (pour gagner du temps, on peut aussi directement double-cliquer sur l'entrée du lexique voulue, sans double-cliquer sur le segment en cours dans la section Segmentation).

	00:01:05.800	00:01:06.000	00:01:06.200	00:01:06.400	00:01:06.600	00:01:06.800	00:01:07.000
ref@SP	BEJ_MV_NARR_03_camel_078						
tx@SP	ʔarjabwa marib idari /						
mot@SP	ʔarjabwa	marib		idari			/
mb@SP	ʔarjab	=wa					
ga@SP	Aryab	=COORD					
rx@SP	N.PR	=CONJ					
ft@SP	in the direction of Aryab,						

Lorsque le dernier segment a été annoté, les annotations choisies sont automatiquement transférées dans leurs lignes correspondantes, et le mot suivant est sélectionné

Affixes

Un affixe dans le lexique ELAN-Corpa est une entrée qui comporte un tiret. Pour un préfixe on ajoutera un tiret (-) à la fin du segment et pour un suffixe au début.

Pour entrer un affixe dans le lexique, on peut aussi faire un clic-droit sur le mot contenant cet affixe et choisir « *Insert a record* ». Dans la case de saisie de l'enregistrement, on supprimera ensuite tout sauf cet affixe et on rajoutera le tiret du bon côté.

En relançant le processus d'interalignement, une segmentation sera proposée en isolant l'affixe du reste du mot. Ce reste est lui-même recherché dans le lexique. On recommencera autant de fois qu'il y a d'affixes dans le mot. Finalement le radical sera entré dans le lexique en temps que *Lexeme* ou bien comme *Variante* d'un lexème.

Fonctionnalités avancées du parseur

Morpho-phonologie (lemme et variante)

Lorsqu'une transformation morpho-phonologique apparaît à la frontière d'un radical et d'un affixe (ou de deux affixes consécutifs), il faut toujours avoir à l'esprit que le parseur recherche une correspondance entre ce qui reste à segmenter et les entrées du lexique (au niveau des champs *Lexeme* et *Variante*)

Lexicon Parse Parameters LinkedFile					
Nr	Lexeme	Variant	Gloss	Tier X	Underlying-Form
4	-a	-ya; -ai; -ait	PL		
5	-a	-á; -ai; -ei	COP.3PL	PRED.N	
6	-a	-á; -ai; -ei	COP.1PL	PRED.N	
7	-a	-á; -ai; -ei	IMP.SG.M	PRED.N	
9	-a	-ai	ORD	NUM	
599	-ai		3PL.NOM	PRO	

Interlinearize		Auto Interlinearize		rhisainheib	
rh	-is	-ai	-n	=heib	
rhi	-s	-ai	-n	=heib	
rh	-is	=ai	-n	=heib	
rhi	-s	=ai	-n	=heib	
rh	-is	-ai	-n	=hei	-b
rhi	-s	-ai	-n	=hei	-b

Dans l'exemple ci-dessus, le parseur ne peut pas trouver la segmentation correcte du mot 'rhisa:nhe:b' (rh -is -a:n =he:b) à cause de la disparition de la voyelle 'a' du suffixe '-a:na' devant le clitique '=he:b'. Lorsque '=he:b' est isolé, pour que le parseur puisse correctement isoler le suffixe '-a:na' qui se trouve déjà dans le lexique, nous pouvons entrer '-a:n' comme variante de '-a:na'.

Ajouter une variante à une entrée

Remplacer le bouton *INSERT RECORD* par *INSERT VARIANT* grâce à la flèche déroulante, puis cliquer dessus.

- Entrer la forme variante (Wordform : -a:n) et rechercher dans la liste déroulante l'entrée correspondante (-a:na).
- SAUVEGARDER l'enregistrement

Comme le parseur recherche une correspondance (avec la fin ou le début du segment restant) au niveau des *lexemes* ou des *variantes* des lexèmes, il proposera maintenant la variante '-a:n' de l'entrée '-a:na' comme satisfaisant la correspondance.

Interlinearize		Auto Interlinearize		rhisainheib	
rh	-is	-a:n	=heib		
rhi	-s	-a:n	=heib		
rh	-is	-a:n	=hei	-b	
rhi	-s	-a:n	=hei	-b	
rh	-is	-ai	-n	=heib	
rhi	-s	-ai	-n	=heib	

L'annotation peut maintenant continuer par la validation dans le lexique des différents segments du découpage; la validation du morphème '-a:n' renverra la forme de base du *lexeme* '-a:na' dans la ligne d'annotation *mb*.

Lorsque la morphophonologie est trop complexe pour permettre la segmentation correcte, même en ayant recours aux variantes des affixes ou radicaux, il est toujours possible de donner le découpage directement dans l'entrée lexicale.

Forme sous-jacente

Il peut arriver que la morphophonologie soit trop complexe et que la segmentation d'un mot soit difficile à gérer par l'ajout d'allomorphe d'une entrée lexicale. Dans ce cas, on peut donner directement le découpage du mot lors de la saisie de l'entrée lexicale. Mais attention, les différents segments composant le mot doivent alors exister dans le lexique.

- Clic sur le bouton *INSERT RECORD* ou CLIC-DROIT sur le mot
- Sélectionnez l'onglet *INSERT UNDERLYING FORM*

- Recherchez le premier segment (ici t'aáro) en déroulant la liste en face de *CHOOSE SEGMENT 1*
- idem pour le segment 2, (ici -a).

Il est possible d'ajouter d'autre segment en cliquant sur le bouton *Add Segment*

- validez par *SAVE RECORD*, puis *FERMER*

En cas d'homonymie ou de polysémie dans les entrées lexicales, il peut être difficile de choisir, parmi plusieurs, le morphème voulu pour un segment donné. Le bouton *Show* permet d'afficher le contenu de l'entrée lexicale choisie pour vérifier qu'il s'agit bien de celle voulue.

De même, il est possible dans cette fenêtre d'ajouter une entrée qui ne serait pas déjà dans le lexique et dont on aurait besoin pour la segmentation du mot en cours.

- Cliquez sur le bouton *INSERT* en face du segment en cours

une petite fenêtre *INSERT MORPHEME* s'ouvre permettant d'ajouter une entrée dans le lexique en la validant avec OK. Cette entrée s'affichera dans le segment en cours.

Il faut bien remarquer que cette méthode qui consiste à fournir au parseur la segmentation *had hoc* d'un mot ne devrait pas être utilisée systématiquement, mais au contraire dans les seuls cas où le parseur n'arriverait pas à fournir la bonne segmentation sur la base des lexèmes (lemmes, variantes et affixes) que le lexique contient. En effet ce type d'entrée spécifique dans le lexique ne permet de segmenter que ce mot-là, alors que le principe de base du parseur consistant à fournir d'une part des lemmes (avec d'éventuelles formes variantes) et d'autre part des affixes, est bien plus productif et systématique.

La fonction Auto Interlinearize

Afin de gagner du temps dans le processus d'interalignement, on peut demander le traitement automatique d'un mot avec passage au mot suivant, lorsqu'une seule segmentation existe pour le mot en cours, sans ambiguïté sur les différents segments (une seule glose). Ainsi toute une séquence peut être traitée d'un coup jusqu'au prochain mot où un choix doit être fait par l'utilisateur.

Lancement de la fonction d'auto-interalignement

On peut lancer la fonction d'auto-interalignement à partir de n'importe quel mot de la ligne de base (dans notre exemple *mot*)

- *CLIQUER* sur le premier *MOT* où doit démarrer l'interalignement automatique (sa ligne de base devient bleue)
- *CLIQUER* sur le bouton *AUTO INTERLINEARIZE*

La segmentation et l'annotation du mot se fait, si les données du lexique le permettent, jusqu'au prochain mot où une ambiguïté ou un segment inconnu arrête le processus.

Lexique des segmentations

Un lexique des mots-forme et de leurs segmentations glosées (que nous avons appelé *Parse*) peut être exporté à partir des mot-formes trouvés dans un texte interaligné. Ce lexique peut être augmenté au fur et à mesure des nouveaux textes interalignés. Il pourra servir pour l'interalignement automatique d'un nouveau texte sur la base des segmentations et annotations déjà rencontrés dans de précédents textes interalignés. (On peut envisager également un éditeur de fichier *Parse* qui aiderait à traduire les gloses d'un lexique *Parse* dans une autre langue. On se servirait ensuite de ce nouveau lexique pour relancer l'interalignement automatique sur le même texte, ce qui produira en un temps record un texte annoté dans l'autre langue.)

Création, fusion, ouverture d'un fichier de segmentations

Pour exporter le lexique des mots-formes et de leurs segmentations, aller dans la section *Lexique* et choisir le menu *Parse* :

- *PARSE, EXPORT PARSE DATA*
- Choisir le dossier destination et le nom du fichier. Une extension *.eafp* sera automatiquement ajoutée.

Pour fusionner le lexique des mots et de leurs segmentations avec un lexique *Parse* existant, aller dans la section *Lexique* et choisir le menu *Parse* :

- *PARSE, EXPORT PARSE DATA*
- Choisir le dossier destination et le nom du fichier.

Pour ouvrir un fichier *Parse* existant afin d'utiliser la fonction d'auto-interalignement

- *PARSE, OPEN PARSE DATA*

Pour utiliser le lexique *Parse* dans le processus d'interalignement automatique

- **COCHER la case** *SEARCHING PARSE DATA FILE*
- **CLIQUER** sur le bouton *AUTO INTERLINEARIZE*

Ouverture automatique d'un lexique de segmentations

Par défaut, l'ouverture d'un fichier de segmentations (*Parse*) dans ELAN est mémorisée afin qu'une prochaine ouverture le charge automatiquement. Pour supprimer le lien entre le fichier ELAN en cours et un fichier de segmentations,

- Aller dans le menu *LINKED FILE* dans la section *LEXIQUE*
- **DÉCOCHER** la case devant le nom du fichier *PARSE*

Le fichier *Parse* ne sera plus chargé à la prochaine ouverture du fichier ELAN.

Sauvegarde des fichiers liés

A la fermeture d'un fichier ELAN, une fenêtre s'ouvrira si un fichier lié n'a pas été préalablement enregistré. Par défaut, les fichiers liés seront enregistrés en quittant ELAN, mais en décochant la case devant le nom d'un fichier lié, on peut quitter ELAN sans le mettre à jour.

Annexe

1. **Synchroniser des annotations importées à une tier préalablement segmentée**

Soit un fichier ELAN contenant une tier *tx* et éventuellement une tier *ft* importée(s) de Praat, de Toolbox ou d'un fichier CSV. Une tier *ref*, segmentée au kilomètre (cf. 32) en suivant le découpage des annotations du fichier importé, peut servir de tier parent à celles importées.

Cf 4. AJOÛT D'UN FICHIER MEDIA pour ouvrir le fichier audio correspondant

Cf 13. AJOUTER UNE NOUVELLE TIER pour créer la tier indépendante *ref* (de stereotype *None*)

Cf 32. SEGMENTER AU KILOMÈTRE la tier *ref* en s'aidant du texte des annotations de *tx*

Il faut ensuite aligner les annotations de *tx* (et *ft*) avec celles de *ref*.

Si l'on a une ligne de traduction *ft*, pour éviter d'avoir à faire le même travail d'alignement pour les 2 tiers, on commencera par rendre *ft* enfant de *tx*

Cf 20. CHANGER LE PARENT D'UNE TIER. (*ft* sera en *Symbolic Association* avec *tx*)

Ainsi l'alignement des segments de *tx* avec ceux de *ref* se transmettra également à *ft*.

Cf 28. OPTIONS, PROPAGATE TIME CHANGES se mettre en mode BULLDOZER avant de déplacer les frontières

Cf 24. MODIFIER LES FRONTIÈRES D'UN SEGMENT pour faire coïncider les frontières du texte au son

Cf 20. REMPLACER LE PARENT D'UNE TIER

Une fois que les annotations de *ref*, *tx* et *ft* sont alignées, on peut rendre *tx* (et donc *ft*) enfant de *ref*.

→ TIER, CHANGE PARENT OF TIER

CLIQUEZ sur la tier indépendante *ref*, NEXT

CLIQUEZ sur la tier à modifier *tx*, NEXT

TYPE DE STEREOTYPE : *tx* : *Symbolic Association*, FINISH

Une copie de la tier initiale est créée (*tx-cp*, *ft-cp*) dont les annotations s'alignent sur celles de la tier synchronisée (*ref*)

On supprimera la/les tiers initiale(s) *tx* (*ft*) et on renommera la/les tier(s) alignée(s).

2. **Synchroniser un fichier audio avec des annotations importées (variante)**

Cf 2. CRÉER UN NOUVEAU FICHIER ELAN avec le fichier audio

Cf 32. SEGMENTER AU KILOMÈTRE en s'aidant du texte des annotations de *tx*

Assurez-vous d'avoir le même nombre de segments dans ce fichier et celui d'annotation

Cf 24. Ajustez les frontières des segments

Enregistrer le fichier ELAN et sortir

DANS NOTEPAD++ (ou équivalent)

Ouvrir le fichier ELAN contenant les annotations importées

Ouvrir le fichier ELAN contenant la segmentation

Substituer le contenu de la balise <TIME ORDER> du fichier d'annotation par celui du fichier de segmentation

Enregistrer

3. **Extraire et traiter le son d'une video**

Lancer AUDACITY

Installer au préalable le plugin **FFmpeg** cf. *Edition, Préférences, Bibliothèques*

FICHIER, OUVRIR, RECHERCHER LE FICHIER : .avi, .mov ou bien .mp4.

Le signal audio est extrait. Généralement le son en vidéo est échantillonné en 48000 Hz.

Enregistrer le fichier extrait

FICHIER, EXPORTER L'AUDIO, type: *WAV* signé 16-bit PCM

Différents traitements peuvent être apportés au signal original.

Pour standardiser en 44100 Hz (format CD) :

PISTE, RÉÉCHANTILLONNER. *Nouvelle fréquence d'échantillonnage* : 44100 Hz

Pour réduire le son en mono (taille divisée par 2) :

PISTE, PISTE STÉREO VERS MONO

Pour normaliser le son, ç.à.d pour que le signal aie une amplitude *normale* (ici -1dB). Si l'enregistrement est faible, le signal sera amplifié (et le bruit de fond également)

SÉLECTIONNER TOUT L'ENREGISTREMENT, EFFET, NORMALISER.

PREVIEW permet de se rendre compte du résultat avant traitement

OK : le signal est traité, l'exporter

Pour réduire le niveau du bruit de fond ou supprimer un bruit indésirable :

DILATER L'ÉCHELLE TEMPORELLE en sélectionnant la LOUPE puis en cliquant dans la bande son (cliquer ensuite sur le bouton au dessus pour sortir de la fonction *loupe*).

SÉLECTIONNER 500ms de signal dans une zone de silence (bruitée) ou sélectionner le bruit régulier à éliminer.

EFFET, RÉDUCTION DE BRUIT, PRENDRE LE PROFIL DU BRUIT

Audacity a mémorisé le signal à supprimer

SÉLECTIONNER TOUT L'ENREGISTREMENT (ctrl/A) ou la partie où le bruit indésirable existe

EFFET, RÉDUCTION DE BRUIT, APERÇU

Si le résultat n'est pas mauvais, faire OK : le traitement se fait sur tout l'enregistrement.

Ré-écouter avant d'exporter

Remarque : toujours exporter en changeant le nom du fichier après traitement. Par exemple en ajoutant *_mono* à la fin du nom du fichier après une réduction en mono, *_clean* après un nettoyage... et conserver les différentes versions.

Il est recommandé de comparer l'original et le fichier traité, par exemple dans PRAAT, pour s'assurer que les détériorations ne sont pas trop dommageables (cela dépend évidemment de la destination de ce fichier audio : analyse phonétique, diffusion web...)

Pour supprimer un passage :

SÉLECTIONNER LE PASSAGE à supprimer, touche SUPPR

(pour revenir en arrière: Ctrl/Z)

Pour insérer un enregistrement (pour insérer un message en début d'enregistrement par exemple)

OUVRIR le fichier à traiter

ENREGISTRER (FICHIER, NOUVEAU...) ou OUVRIR dans Audacity le second enregistrement à insérer

Dans cette 2^{ème} instance d'Audacity, SÉLECTIONNER la partie du second enregistrement à insérer et faites CTRL/C pour le copier

Revenir à l'enregistrement principal. (Dilater le temps si nécessaire, avec la LOUPE). Positionner le curseur à l'endroit où insérer le second enregistrement et faire CTRL/V pour le coller

Application du logiciel ELAN à l'annotation linguistique

Gestion des documents.....	2
2. Ouverture ou Importation d'un fichier d'annotation.....	2
3. Affichage/Ajout des fichiers média associés à un fichier d'annotation	3
4. Décalage entre annotations et image/son.....	3
5. Enregistrement d'un fichier ou d'un segment d'annotation	3
Annotations tiers et types linguistiques.....	4
6. Afficher la dépendance entre les tiers.....	5
Séréotypes	5
7. Définir un type linguistique	6
8. Ajouter une nouvelle tier	6
9. Fichier modèle (template).....	6
10. Importer des types.....	7
11. Importer des tiers	7
12. Modifier les attributs d'une tier	7
13. Supprimer une tier.....	7
14. Copier une tier.....	7
15. Changer le parent d'une tier	8
16. Sélectionner un segment temporel sur une tier indépendante.....	8
17. Créer un nouveau segment et son annotation	8
18. Subdiviser une annotation.....	8
19. Déplacer une annotation	9
20. Modifier les frontières d'un segment.....	9
21. Saisir/modifier une annotation sur un segment existant ou dépendant.....	9
22. Modifier les paramètres d'édition de ELAN	9
23. Création d'un vocabulaire contrôlé	10
24. Affecter un vocabulaire contrôlé à un <i>type</i>	10
25. Mode de déplacement des annotations	10
26. Créer des annotations les unes à la suite des autres (sans trous)	10
27. Segmenter au kilomètre	11
28. Segmenter automatiquement à partir des silences (pauses).....	11
29. Saisir les annotations - bis.....	12
30. Naviguer dans le texte.....	12
31. Naviguer dans le texte et corriger les annotations	13
32. Importer un fichier d'annotations tabulés.....	13
33. Importer un fichier d'annotation Toolbox	13
34. Créer des annotations à partir des chevauchements.....	15
35. Supprimer les annotations à partir d'un index.....	15
36. Créer une nouvelle série de tiers pour un autre locuteur	15
37. Concaténer le contenu de plusieurs tiers dans une autre tier	16
38. Changer la casse des annotations d'une tier	16
39. Copier un groupe d'annotations	16

Procédure d'interalignement dans ELAN-CorpA	17
Créer un lexique ELAN-CorpA	17
Ouvrir un lexique ELAN-CorpA.....	17
Importer un lexique Toolbox.....	18
Ouverture automatique du lexique	19
Paramétrer l'interalignement	19
Principes de fonctionnement de l'annotateur.....	20
Lancer le processus d'interalignement.....	20
Saisie d'une entrée et/ou d'une variante dans le lexique (<i>Insert record</i>)	21
Sélection de la segmentation et de la glose	22
Affixes	22
Fonctionnalités avancées du parseur	22
Morpho-phonologie (lemme et variante)	22
La fonction Auto Interlinearize	24
Lancement de la fonction d'auto-interalignement	24
Lexique des segmentations.....	24
Création, fusion, ouverture d'un fichier de segmentations	25
Ouverture automatique d'un lexique de segmentations.....	25
Sauvegarde des fichiers liés	25
Annexe	26
1. Synchroniser des annotations importées à une tier préalablement segmentée	26
2. Synchroniser un fichier audio avec des annotations importées (variante)	26
3. Extraire et traiter le son d'une video	26
Application du logiciel ELAN à l'annotation linguistique	28