

Démarrer rapidement avec ELAN-CorpA

1. Création du fichier d'annotation ELAN

→ **FICHIER, NOUVEAU**

Sélection : Media

Rechercher le fichier *média* (extension : *wav*) à annoter, dans la fenêtre de gauche et le verser dans la fenêtre de droite grâce au bouton >>

Sélection : Template

Rechercher le fichier *modèle* à utiliser (par ex : *Corpo1.etf*) dans la fenêtre de gauche et le verser dans la fenêtre de droite grâce au bouton >>

→ **OK**

2. Enregistrement du fichier d'annotation

→ **FICHIER, ENREGISTRER SOUS**

Rechercher le dossier voulu et donner un nom au nouveau fichier ELAN

→ **ENREGISTRER**

3. Contrôle de la hiérarchie des acteurs

CLIC-DROIT dans la zone des étiquettes des acteurs

→ **TRIER PAR ACTEUR, TRIER PAR HIERARCHIE**

4. Segmentation des unités prosodiques

AMENER LE CURSEUR AU DEBUT DE LA ZONE A SEGMENTER

COCHER LA CASE 'MODE DE SELECTION'

→ **CLIQUER** à l'emplacement de la fin du segment voulu

ECOUTER LE SEGMENT en cliquant sur le bouton


CLIQUER sur un autre emplacement si nécessaire


CLIC-DROIT dans la zone sélectionnée, au niveau de la ligne **REF**

CLIQUER sur 'NOUVELLE ANNOTATION ICI'

CTRL/ENTREE ou **ENTREE** pour fermer la case de saisie

Pour sélectionner le prochain segment :

CLIQUER SUR LE BOUTON  pour amener le fil rouge à la fin de la zone précédemment sélectionné

← **CLIQUER SUR LE BOUTON**  pour désélectionner le segment précédemment sélectionné

5. Etiquetter et numéroter les segments

→ **TIER > LABEL AND NUMBER**

SELECTIONNER ref@SP

INCLUDE LABEL PART : codeLangue_InitialesChercheur_NARR(/CONV)_N°

INSERT OTHER DELIMITER : _

OK > CLOSE

6. Saisie de la transcription

DOUBLE-CLIQUER sous l'unité à transcrire, au niveau de la ligne tx

TAPER LE TEXTE (S'il s'agit d'une transcription en mots prosodiques, mettre des tirets (-) là il y aurait une frontière de mots grammaticaux, et des espaces entre les mots prosodiques. Les tirets pourront être supprimés par la suite, par rechercher/remplacer dans la ligne TX)

CTRL/ENTREE ou **ENTREE** pour fermer la case de saisie

7. Création des cellules de la ligne mot (tokenisation)

→ **ACTEUR, TOKENISER**

ACTEUR SOURCE : tx

ACTEUR DE DESTINATION : mot

DELIMITEUR DE TOKEN : - <ESPACE> (tiret, suivi d'une espace)

COCHER 'PRESERVER' si vous avez déjà des annotations en dessous de la ligne 'mot'

→ **COMMENCER, FERMER**

8. L'onglet interlinéarisation

→ **CLIQUER SUR L'ONGLET 'INTERLINEARIZE'**

9. Créer le lexique

→ **CLIQUER SUR 'LEXICON > CREATE'**

Rechercher l'emplacement voulu et donner un nom au lexique

→ **ENREGISTRER**

10. Enregistrer le lexique

L'enregistrement du lexique est indépendant de celui du fichier d'annotation (Ctrl/S). A tout moment il peut être demandé :

→ **CLIQUER SUR 'LEXICON > SAVE'**

11. Associer un lexique au fichier ELAN

Lorsqu'un nouveau fichier d'annotation est créé, on peut lui associer un lexique ELAN existant :

→ **CLIQUER SUR 'LEXICON > OPEN'**

RECHERCHE LE FICHIER d'extension .eafp

SELECT

12. Le lexique associé au fichier ELAN

Pour voir quel lexique est associé au fichier ELAN en cours, et éventuellement le dissocier :

→ **CLIQUER SUR 'LINKED FILE'**

Pour dissocier le lexique du fichier ELAN, décocher la case devant le nom du fichier

13. Paramétrer l'interalignement

→ **CLIQUER** sur 'PARAMETERS > INTERLINEARIZE TIER PARAMETERS > CHANGE INTERLINEAR TIERS' pour préciser le noms des lignes d'annotation

SAISIR LES NOMS DES LIGNES D'ANNOTATION dans les cases dont les étiquettes ne sont pas les mêmes que ceux par défaut (étiquettes de la ligne de départ : MOT ; ligne de découpage en morphèmes : MB ; ligne de glose : GE ; ligne de catégorie : RX)

→ **CLIQUER SUR 'RENAME TIERS'**

Il est possible de créer de nouvelles lignes de segmentation et annotation en choisissant

→ 'PARAMETERS > INTERLINEARIZE TIER PARAMETERS > CREATE INTERLINEARIZE TIERS'

14. Démarrer le processus d'interalignement

SELECTIONNER UN MOT dans la ligne MOT en cliquant sur sa ligne de support qui devient bleue

→ **CLIQUER SUR LE BOUTON 'INTERLINEARIZE'** dans la zone des segmentations

Le mot apparaît, précédé d'une étoile, signifiant qu'il n'a pas été trouvé dans le lexique (vide).

15. Insérer un mot simple dans le lexique

CLIC-DROIT SUR LE MOT, dans la zone des segmentations

→ **INSERT RECORD**

Dans la fenêtre de saisie d'une entrée lexicale, saisir dans

GLOSS : la glose correspondant au mot

**** : (optionnel) une étiquette correspondant à une valeur de flexion ou dérivation...

TIER X : la catégorie du mot

→ **SAVE RECORD**

16. Relancer l'interalignement et valider

→ **CLIQUER SUR LE BOUTON 'INTERLINEARIZE'** dans la zone des segmentations

Le mot apparaît sans étoile, il est bien dans le lexique

DOUBLE-CLIC sur la ligne correspondante du lexique pour valider le mot

La segmentation du mot (mb) et les annotations de ces derniers provenant du lexique (ge, rx) sont ajoutées sous le mot traité

17. Utiliser l'extension de glose \ dans la saisie d'une entrée

Lorsqu'un mot comporte un élément grammatical qui ne peut pas être isolé d'un radical, on utilisera la case à droite de l'antislash (\) dans le champ glose de la fenêtre de saisie de l'entrée.

ad?iho:b		/
a-	d?i	=ho:b /
1SG-	do\PFV	=when .
PNG-	V1	=CONJ .

Lexeme: d?i
Variant: d?i
Gloss: do \ PFV
Tier X: V1
Gloss: do \ NAR.1SG
Tier X: V1

18. Insérer un affixe dans le lexique

CLIC-DROIT SUR LE MOT à segmenter, dans la zone des segmentations

→ **INSERT RECORD**

Dans la fenêtre de saisie de l'entrée lexicale,

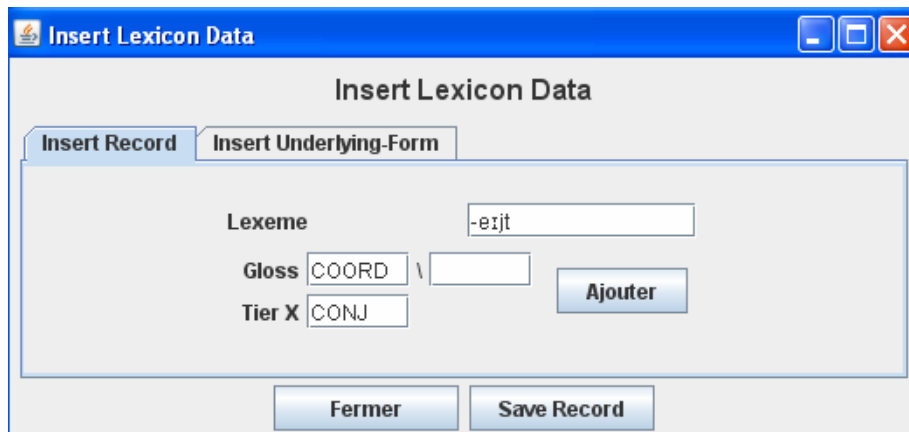
LEXEME : supprimer du mot à segmenter tout sauf l'affixe, puis lui ajouter un tiret (ou =) à gauche pour un suffixe ou un tiret à droite pour un préfixe

GLOSS : l'étiquette morpho-syntaxique correspondant à l'affixe

**** : laisser vide

TIER X : la catégorie de l'affixe

→ **SAVE RECORD**



19. Insérer une variante d'une entrée dans le lexique

CLIC-DROIT SUR LE MOT contenant la variante à entrer, dans la zone des segmentations

→ **INSERT VARIANT**

Dans la fenêtre de saisie de la variante,

WORDFORM : Editer la forme de surface de la variante si nécessaire, ajouter un tiret (ou =) à gauche s'il s'agit d'un suffixe ou un tiret à droite pour un préfixe

LEXEME : Dérouler la fenêtre et choisir l'entrée qui est l'objet de la définition d'une variante

**** : Si nécessaire, Entrer une étiquette morphologique caractérisant la variante par rapport à l'entrée

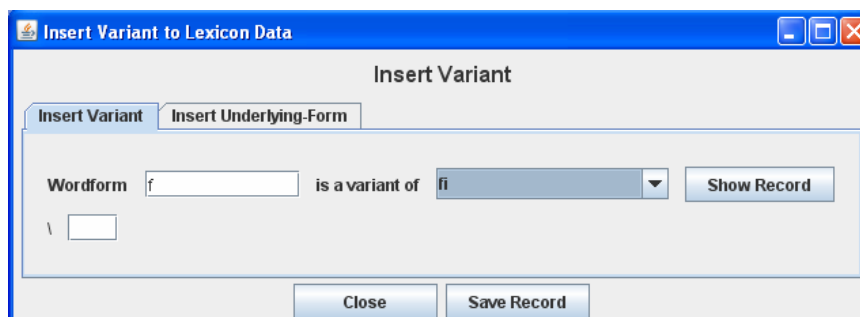
→ **SAVE VARIANT**

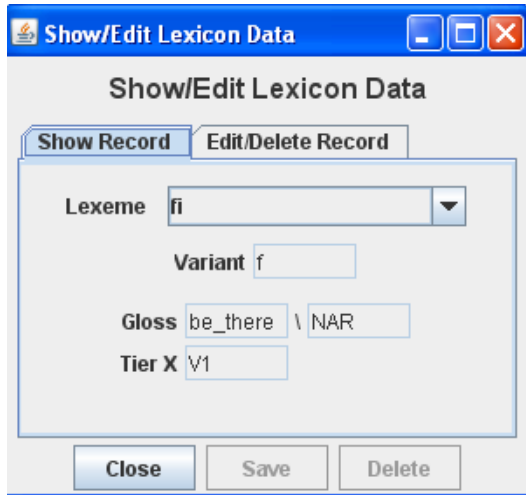
Exemple :

i:fe:jt < i:- f -e:jt < i:- fi -e:jt

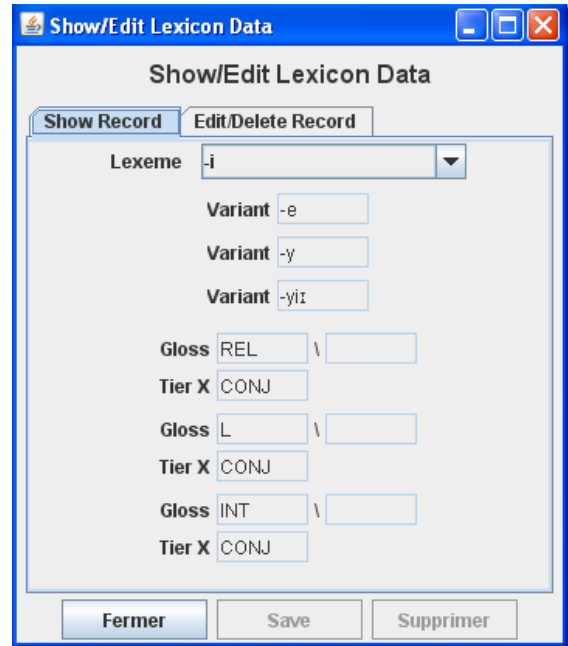
BEJ_MV_NARR_01_shelter_008					
i'fərgib i:fe:jt //					
ifərgib			i:fe:jt		//
i-	fərg	=ib	i:-	fi	-e:jt //
DEF.M-	Shərg	=LOC.SG	NAR.3SG.	be_there	-COORD .
DET-	N.PR	=POSTP	TAM.PNG-	V1.IRG	-CONJ .
"I was over there at Shərg and					

'f' variante contextuelle de 'fi'





exemple de variante d'un lexème



exemple d'un morphème (suffixe) ayant plusieurs valeurs et plusieurs variantes

La mémoire des formes segmentées et annotées

Il est possible de créer un lexique des mots-formes rencontrés dans un texte annoté, avec leur segmentation et annotations. On peut par la suite ajouter à ce lexique les mots-formes nouveaux d'un autre texte. Ce lexique pourra ensuite servir à accélérer l'annotation.

20. Créer un lexique des mots-formes (*Parse Lexicon*)

→ **PARSE > EXPORT PARSE DATA**

RECHERCHER l'emplacement où enregistrer le fichier et **LUI DONNER UN NOM**

21. La fonction d'auto-interalignement

La fonction d'auto-interalignement permet de sauter le processus de validation de la segmentation et du choix dans le lexique s'il y a une solution unique dans le lexique pour le mot en cours. Lorsqu'un lexique 'Parse' est sélectionné, le processus d'interalignement recherche d'abord si ce mot a déjà été trouvé auparavant et dans ce cas insère sa segmentation et l'annotation de ses morphèmes directement dans les lignes d'annotation, puis passe au mot suivant. Pour cela

→ **PARSE > OPEN PARSE DATA** (si aucun lexique 'Parse' n'est déjà ouvert)

COCHER LA CASE 'SEARCH IN PARSE DATA FILE' qui est apparu

CLIQUER sur **AUTO INTERLINEARIZE**

Le processus s'arrête dès qu'une ambiguïté apparaît ou qu'un mot n'est pas connu.

Simple *Annotation* de texte

Le processus d'interalignement de ELAN-CorpA consiste en la segmentations des mots en morphèmes, puis en l'annotation des morphèmes. Il est possible d'utiliser la fonction d'annotation, sans passer par celle de segmentation, en utilisant l'onglet **ANNOTATION**.

22. Paramétrer l'Annotation

→ **CLIQUER** sur 'PARAMETERS > ANNOTATION TIER PARAMETERS > CHANGE ANNOTATION TIERS' pour préciser le nom des lignes d'annotation

SAISIR LES NOMS DES LIGNES D'ANNOTATION dans les cases dont les étiquettes ne correspondent pas à celles par défaut (étiquettes de la ligne de départ : mb ; ligne de glose : ge ; ligne de catégorie : rx)

→ **CLIQUER SUR** 'RENAME TIERS'

Il est possible de créer de nouvelles lignes d'annotation (par exemple en lien avec un autre lexique) en choisissant

→ 'PARAMETERS > ANNOTATION TIER PARAMETERS > CREATE ANNOTATION TIERS'

La ligne de départ peut être la ligne 'mot' mais dans ce cas le lexique devra être un lexique des mots-formes puisque le processus de segmentation en morphèmes sera inopérant.

ELAN-CorpA à partir de Praat

23. Ouvrir un fichier Audio dans Praat

→ **READ, READ FROM FILE**

RECHERCHER le fichier WAV

OUVRIR

24. Créer un fichier d'annotation

SELECTIONNER L'OBJET SOUND dans la listes des objets de Praat

→ **ANNOTATE**

→ **TO TEXTGRID**

ALL TIERS NAMES : supprimer '*Mary John bell*' et taper '**ref@SP**'

WHICH OF THESE ARE POINT TIERS : supprimer '*bell*'

→ **APPLY, CANCEL**

25. Vérifier que le format du fichier d'annotation

→ **PRAAT, PREFERENCES, TEXT WRITING PREFERENCES**

CHOISISSEZ UTF-8

OK

26. Délimiter les segments

SELECTIONNER L'OBJET SOUND ET L'OBJET TEXTGRID dans la listes des objets de Praat (avec la touche CTRL)

→ **CLIQUER** sur EDIT

SELECTIONNER 5 A 10 SECONDES de signal audio avec la souris

CLIQUER sur le bouton SEL (en bas à gauche de la fenêtre)

CLIQUER DANS LA ZONE DU SON à l'emplacement souhaité pour la fin du segment précédent

ECOUTER LE SEGMENT en cliquant sous le segment délimité, dans la zone en dessous du texte (celle du bas, en jaune)

CLIQUER AILLEURS DANS LA ZONE DU SON si l'emplacement de la fin du segment ne vous convient pas

CLIQUER DANS LE PETIT ROND sur la frontière créée dans la zone de texte pour valider la frontière du segment (ou taper sur la touche ENTREE)

27. Saisir les transcriptions

CLIQUER DANS LA ZONE DE TEXTE (jaune) du segment à transcrire (une zone de saisie s'ouvre en dessous de la barre des menus de la fenêtre)

TAPER LE TEXTE

CLIQUER DANS LE SEGMENT SUIVANT pour continuer

28. Utiliser le système de saisie de Praat pour entrer de la phonétique

A défaut d'un clavier virtuel adapté (par ex : *AFU*), on utilisera le système de saisie des caractères spéciaux de Praat : antislash (\) suivi d'une ou deux lettres, par exemple \ff pour le symbole ϕ . Pour trouver ces codes de saisie, dans la fenêtre d'édition (sur la droite)

→ **HELP > PHONETIC SYMBOLS**

29. Convertir le code de saisie Praat en Unicode

Avant d'enregistrer le fichier pour une utilisation dans ELAN, il faut convertir le système de codage des caractères phonétiques de Praat en Unicode

→ **EDIT > CONVERT ENTIRE TEXTGRID TO UNICODE**

30. Enregistrer le fichier d'annotation textGrid

→ **FILE, WRITE TEXTGRID TO TEXT FILE**

RECHERCHER l'emplacement voulu et **DONNER UN NOM** au fichier (extension automatique .TextGrid)

OK

31. Importer le fichier d'annotation de Praat dans ELAN

→ **FICHIER, NOUVEAU**

Sélection : Media

Rechercher le fichier *média* (extension : *wav*) à annoter, dans la fenêtre de gauche et le verser dans la fenêtre de droite grâce au bouton >>

Sélection : Template

Rechercher le fichier *modèle* à utiliser (par ex : *Corpo1.etf*) dans la fenêtre de gauche et le verser dans la fenêtre de droite grâce au bouton >>

CLIC-DROIT sur l'étiquette **ref@SP**

→ **DELETE REF@SP > DELETE > OUI**

→ **FILE > IMPORT > PRAAT TEXTGRID FILE > BROWSE**

Rechercher le fichier d'annotation *TextGrid* provenant de Praat

OUVRIR > NEXT > FINISH

32. Copier la ligne ref@SP pour créer une ligne tx@SP

→ **TIER > COPY TIER**

SELECTIONNER la LIGNE DE DEPART **ref@SP > NEXT**

SELECTIONNER la LIGNE PARENT **ref@SP > NEXT**

SELECTIONNER LE TYPE **tx > NEXT**

33. Renommer la ligne ref@SP-CP en tx@SP

- CLIC-DROIT SUR L 'ETIQUETTE ref@SP-CP
- CHANGE ATTRIBUTE OF ref@SP-CP
- TIER NAME : TAPER tx@SP
- CHANGE

34. Etiquetter et numéroter les segments

- TIER > LABEL AND NUMBER
- SELECTIONNER ref@SP
- INCLUDE LABEL PART : codeLangue_InitialesChercheur_NARR(/CONV)_N°
- INSERT OTHER DELIMITER : _
- OK > CLOSE

35. Importer les autres lignes d'annotation

- TIER > IMPORT TIERS > BROWSE
- RECHERCHER le fichier MODELE (par ex : *Corp01.etf*)
- SELECT > IMPORT > CLOSE

Démarrer rapidement avec ELAN-CorpA	1
1. Création du fichier d'annotation ELAN.....	1
2. Enregistrement du fichier d'annotation.....	1
3. Contrôle de la hiérarchie des acteurs.....	1
4. Segmentation des unités prosodiques.....	1
5. Etiquetter et numéroter les segments	1
6. Saisie de la transcription	1
7. Création des cellules de la ligne mot (tokenisation)	2
8. L'onglet interlinéarisation	2
9. Créer le lexique	2
10. Enregistrer le lexique	2
11. Associer un lexique au fichier ELAN	2
12. Le lexique associé au fichier ELAN	2
13. Paramétrer l'interalignement.....	2
14. Démarrer le processus d'interalignement.....	3
15. Insérer un mot simple dans le lexique.....	3
16. Relancer l'interalignement et valider	3
17. Utiliser l'extension de glose \ dans la saisie d'une entrée.....	3
18. Insérer un affixe dans le lexique	3
19. Insérer une variante d'une entrée dans le lexique	4
La mémoire des formes segmentées et annotées.....	5
20. Créer un lexique des mots-formes (<i>Parse Lexicon</i>).....	5
21. La fonction d'auto-interalignement.....	5
Simple <i>Annotation</i> de texte	5
22. Paramétrer l'Annotation.....	6
ELAN-CorpA à partir de Praat.....	6
23. Ouvrir un fichier Audio dans Praat	6
24. Créer un fichier d'annotation	6
25. Vérifier que le format du fichier d'annotation	6
26. Délimiter les segments	6
27. Saisir les transcriptions.....	7
28. Utiliser le système de saisie de Praat pour entrer de la phonétique	7
29. Convertir le code de saisie Praat en Unicode.....	7
30. Enregistrer le fichier d'annotation textGrid	7
31. Importer le fichier d'annotation de Praat dans ELAN	7
32. Copier la ligne ref@SP pour créer une ligne tx@SP	7
33. Renommer la ligne ref@SP-CP en tx@SP	8
34. Etiquetter et numéroter les segments	8
35. Importer les autres lignes d'annotation	8