

Formation ELAN

Fonctions avancées - 1

Organisation de la session

Ce que je ne vais pas reprendre dans cette formation (sauf si besoin) :

- Ouverture et présentation générale du logiciel ;
- Création d'annotations dans un template existant ;
- Les modes de visualisation des annotations.

Ce que nous allons voir ce matin :

- Questions à se poser avant la création de template ;
- Création de template ;
- Visualisations des statistiques et réflexion sur le type d'information obtenues ;
- Les différents modes de segmentation.

Création du template

Créer un template sous ELAN, c'est

- Définir les informations attendues au niveau des analyses
- Mettre au clair les liens entre ces informations (temporel, hiérarchique, etc.)
- Formaliser ces relations dans le logiciel (acteur, lien entre eux, type linguistique, vocabulaire contrôlé)

C'est un travail long, par exemple du projet ANR « Multimodalité » de Jean-Marc Colletta où nous avons mis en place un template et un manuel de codage (présentation cet après-midi)

Il y a eu au moins 7 ou 8 versions successives pour mettre au point la grille définitive d'annotation, avec réunion des différents utilisateurs pour la faire évoluer en fonction des problématiques rencontrées lors de l'annotation

	00:00:07.000	00:00:08.000	00:00:09.000	00:00:10.000	00:00:11.000	00:00:12.000
Enfant [30]	et après heu:: // e/ y a/ elle regarde l'heure //					et:: donc elle s'en va //
TDP [1]	:: // à bouger et i(l) tombe du nid:: i(l) va dans une toile d'araignée dans une fleur après il va chez heu:: jerry:: et heu:: // jerry elle // elle le savait pas donc elle regarde elle voit l'heure					
Synt.Prop. [31]	et après elle regarde l'heure					et donc elle s'en va
Synt.Type Pr [31]	Independante					Independante
Synt.Mots [177]	et	après	elle	regarde	l'heure	et donc elle s'en va
Synt.Cat.gr [2]						
Disc.Cat.gr [75]	Conn	Conn	Pronom			
						Conn Conn Prono
Narration [31]	et après elle regarde l'heure					et donc elle s'en va
Episode [31]	A					A
ElementDuScr [31]	A4					A8
ModalitésPrag [31]	Raconte					Raconte
PlanNarratif [31]	Premier plan					Premier plan
1 - Geste(Phase [26]			Stroke			Stroke
1 - Forme du g [22]			Tête >Av + élevat			Tête inclinée >D
1 - Valeur du [22]			Cadrage			Discursive
1 - Relation G [22]			Supplément			Redondance
1 - Relation sy [22]			Synchrone			Synchrone
2 - Geste(Phase [23]			Stroke			Stroke
2 - Valeur du [20]			Cadrage			Discursive
2 - Relation G [20]			Supplément			Redondance
2 - Relation sy [20]			Synchrone			Synchrone

Les questions à se poser

Pour construire un template utilisable et pertinent, il faut se poser un certain nombre de questions :

- Quelles sont les données dont je veux disposer :
 - Parole, gestes, 1 ou plusieurs intervenants, etc
- Quelles informations voulons-nous sur ces données :
 - Liste des annotations, nombre d'annotations par type, chevauchement possible entre différentes annotations, etc

=> L'ensemble des réponses à ces questions va nous permettre de définir les acteurs (tiers) nécessaires, les liens existants entre eux (types linguistiques) et les listes d'annotation (vocabulaire contrôle)

Nous allons reprendre rapidement la création des différents éléments sous ELAN, en identifiant à chaque étape certains des questions importantes à se poser

Les acteurs

Les acteurs sont les lignes d'annotation présentes sous ELAN.

Les lignes peuvent être indépendantes les unes des autres, ou entretenir une relation de dépendance.

Par exemple, dans la grille du projet « Multimodalité » de Grenoble, on peut voir :

- Des acteurs indépendants
- Des acteurs avec un lien hiérarchique à 1 ou plusieurs niveaux (acteurs parents)

Les relations entre les acteurs peuvent être de 4 types et sont définis au moyen de types linguistiques.



Questions sur les acteurs -1

De quels acteurs vais-je avoir besoin :

- Combien de participant dont je vais analyser les productions ;
- Quelles sont les informations dont je vais avoir besoin pour chaque participants (parole, gestes en général, gestes du visage, gestes des ou d'une main, etc) ;
- La granularité des informations (temporalité ou non, lien avec les autres informations, etc).

Pour mieux savoir ce qu'on peut obtenir, il faut regarder dans l'onglet 'Affichage' et dans la rubrique 'Statistiques des annotations' (Les tableaux sont exportables sous Excel).

Statistiques des annotations

Statistiques

Annotations Annotations II Acteur Linguistic Type Participant Annotator

Acteur

Sélection de l'acteur: Synt.Prop.

☐ Show only root tiers

☒ Count contiguous annotations with the same value as 1

☐ Use media duration as observation period

Variables de Statistiques

Annotation	Occurrences	Fréquence	Durée Moyenne	Taux de durÉe	Latence
alors le petit chape...	1	0.0079874118389...	4.78	0.0381798285901...	29.78
c'est bon je sais pa...	1	0.0079874118389...	2.31	0.0184509213479...	21.519
c'est ce qu'on dit	1	0.0079874118389...	1.64	0.0130993554158...	108.99
c'est pour mieux t'e...	1	0.0079874118389...	2.26	0.0180515507560...	83.551
c'est pour mieux te ...	1	0.0079874118389...	1.555	0.0124204254095...	96.845
c'est pour mieux te ...	1	0.0079874118389...	2.899	0.0231555069210...	90.581
c'est une fille toute ...	1	0.0079874118389...	5.12	0.0408955486153...	9.469
de aller emmener ...	1	0.0079874118389...	4.37	0.0349049897361...	16.99
donc c'est la maman	1	0.0079874118389...	3.27	0.0261188367133...	6.119
elle manque de se ...	1	0.0079874118389...	5.04	0.0402565556682...	35.287
en mettant sa robe ...	1	0.0079874118389...	2.07	0.0165339425066...	59.265

Enregistrer Fermer

Nous pouvons donc obtenir des informations sur le nombre d'occurrences, leur fréquence, leur durée moyenne, etc.

Les différentes informations n'auront pas toutes la même pertinence selon nos analyses futures. On voit dans l'exemple que le nombre d'occurrences est différent pour l'acteur 'Synt.Prop' et pour 'Synt.Type Proposition'.

Statistiques des annotations

Statistiques

Annotations Annotations II Acteur Linguistic Type Participant Annotator

Acteur

Sélection de l'acteur: Synt.Type Proposition

☐ Show only root tiers

☒ Count contiguous annotations with the same value as 1

☐ Use media duration as observation period

Variables de Statistiques

Annotation	Occurrences	Fréquence	Durée Moyenne	Taux de durÉe	Latence
Ct d'adjectif	0	0.0	0.0	0.0	-
Ct de Nom	3	0.02396223551682...	1.64533333333333...	0.03942586483701...	25.32
Ct de Nom foc.	0	0.0	0.0	0.0	-
Ct de Phrase	3	0.02396223551682...	2.29133333333333...	0.05490546898088...	59.265
Ct de Verbe	9	0.07188670655047...	2.63044444444444...	0.18909398787510...	16.99
Factive	1	0.00798741183894...	2.34	0.01869054370312...	117.48
Indépendante	10	0.07987411838941...	2.5562	0.204174221427031	21.519
Indépendante Prés	2	0.01597482367788...	4.195	0.06701438532872...	6.119
Infinitive	0	0.0	0.0	0.0	-
Principale	8	0.06389929471153...	1.730375	0.11056974208647...	14.942
Principale Prés	0	0.0	0.0	0.0	-

Enregistrer Fermer

Questions sur les acteurs -2

A partir de l'exemple précédent, nous avons vu que différents types d'informations sont disponibles en fonction des acteurs

- Nombre d'occurrence, fréquence, durée moyenne, durée minimale, durée maximale, durée médiane, durée totale, pourcentage en durée sur la totalité des annotations ;

Selon les informations rentrées, ces résultats sont plus ou moins pertinent :

- Pour le cas de l'acteur « Synt.Prop », le nombre d'occurrences, 1 pour chaque, n'apporte pas grand chose. Par contre, la durée moyenne est intéressante.
- Pour le cas de l'acteur « Synt.Typ.Prop », la durée n'est pas pertinente, mais le nombre d'occurrences oui.

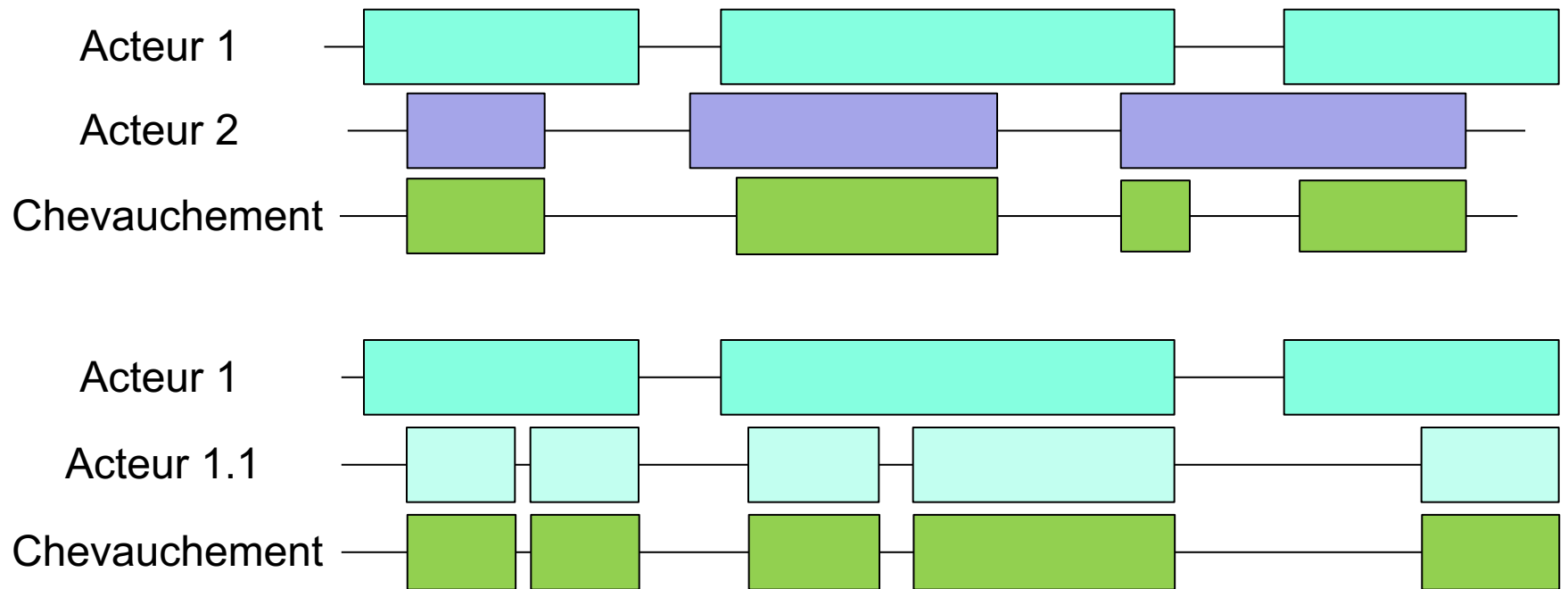
Questions sur les acteurs -3

Ces différences correspondent à des types d'acteurs différents :

- L'acteur « Enfant » est un acteur en haut de la hiérarchie (parent), sans type linguistique et sans vocabulaire contrôlé.
 - Nous n'avons pas de limite temporelle pour placer les annotations ;
 - Il n'y a pas de vocabulaire contraint pour rentrer les informations.
- L'acteur « Synt.Typ.Prop » est un acteur en bas de la hiérarchie (enfant), avec un type linguistique et un vocabulaire contrôlé.
 - Les annotations de cet acteur sont liées à celles de son acteur parent ;
 - Il existe un vocabulaire contraint pour rentrer les annotations, ce qui permet d'avoir des statistiques sur le nombre d'occurrences par catégories définies.

Questions sur les acteurs -4

Différentes fonction du logiciel vont nous permettre d'obtenir des informations de manière automatique



Les types linguistiques

Ils permettent de définir :

- Les liens existant entre 2 acteurs
- L'utilisation ou non d'un vocabulaire contrôlé

Ajouter Modifier Supprimer Importer

GRS None

Nom du type

Stéréotype

Utiliser le vocabulaire contrôlé

Lexicon connection

ISO Data Category

Alignable dans le temps

Références aux graphiques autorisés

None

None

Time Subdivision

None

Voice2

TypeProposition

Voice

Relationsynchrone

Disc.Cat.Gram

Elementduscript

Accord gestes

Les stéréotypes

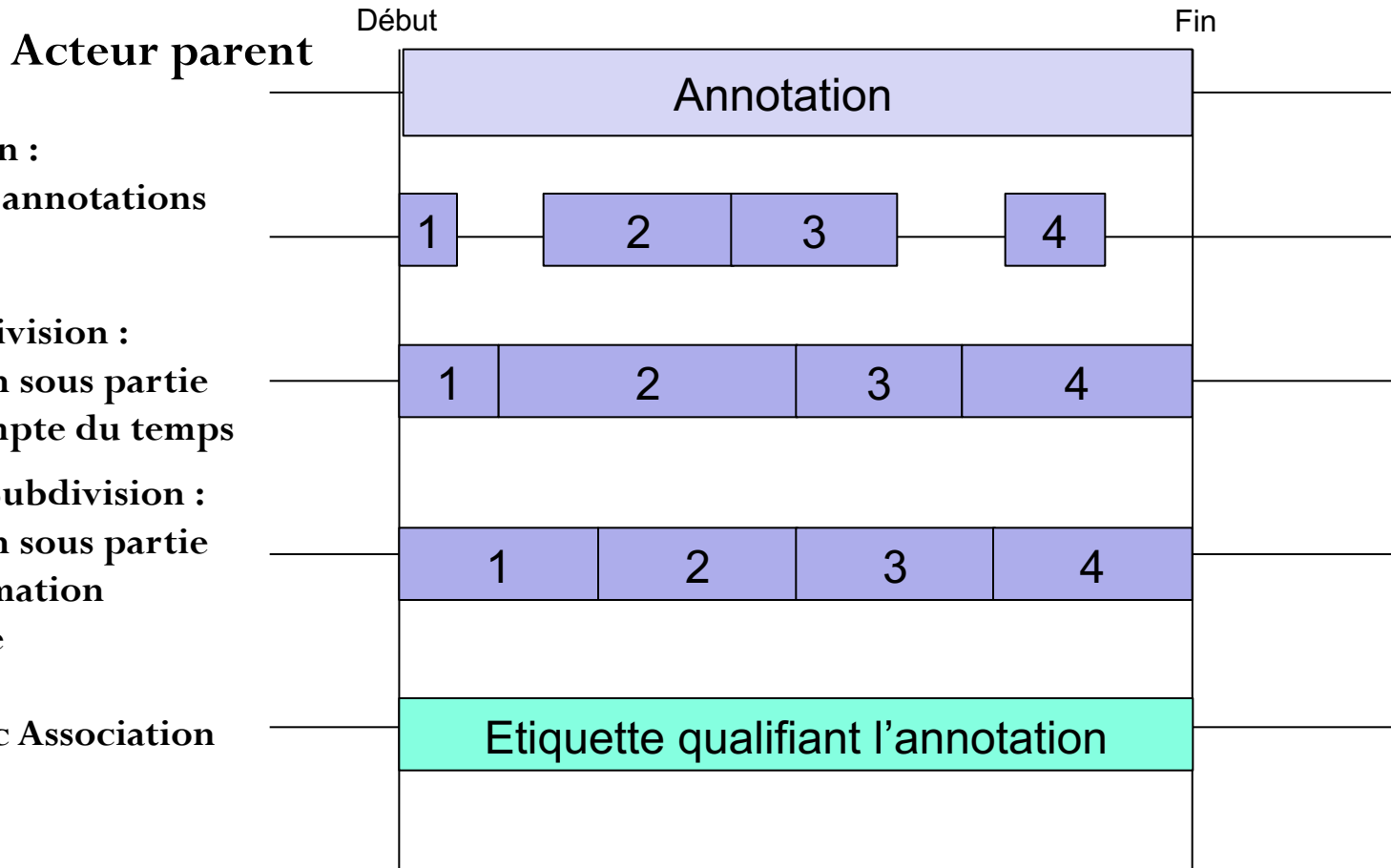
Included In Les annotations de ce type doivent être incluses dans celles de l'acteur parent, mais sans nécessairement correspondre à l'ensemble de l'annotation précédente.

Time Subdivision Les annotations de ce type doivent être incluses dans celle de l'acteur parent, mais, contrairement à l'Included In, il ne peut y avoir d'espace vide entre deux annotations. Par contre, la valeur temporelle de chaque annotation peut être adaptée au besoin de l'annotateur.

Symbolic Subdivision Les annotations de ce type sont une division temporellement arbitraire de celle de l'acteur parent.

Symbolic Association Les annotations de ce type sont une correspondance exacte de celle de l'acteur parent.

Les stéréotypes (représentation)



Vocabulaire contrôlé

L'utilisation de vocabulaire contrôlé est souhaitable dans le cas d'annotations qui se répètent régulièrement afin d'éviter les erreurs de saisie et d'uniformiser les codages.

Ex : dans le projet « Multimodalité », les différents types de propositions syntaxiques ont été entrés.

Editer le vocabulaire contrôlé

Vocabulaire contrôlé

CV courant: TypeProposition

Nom du CV: TypeProposition

Description: Les différents types de propositions syntaxiques définis pour notre analyse

Ajouter

Modifier

Supprimer

Importer

External CV

Entrées

Independante

Principale

Ct de Nom

Ct de Verbe

Ct de Phrase

Ct d'adjectif

Factitive

Infinitive

Ct d'adverbe

Ct de Nom foc.

Phrase nominale

Independante Prés

Principale Prés

Valeur de l'entrées

Independante

Description de l'entrée

ISO Data Category

Rechercher...

Ajouter

Modifier

Supprimer

More Options...

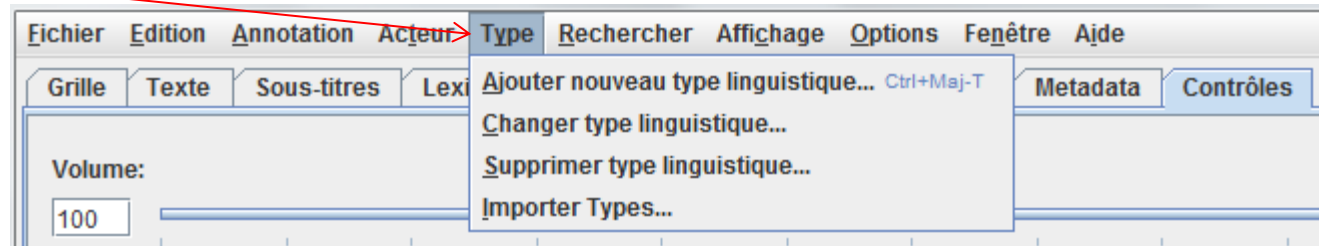
Sort A-Z

Sort Z-A

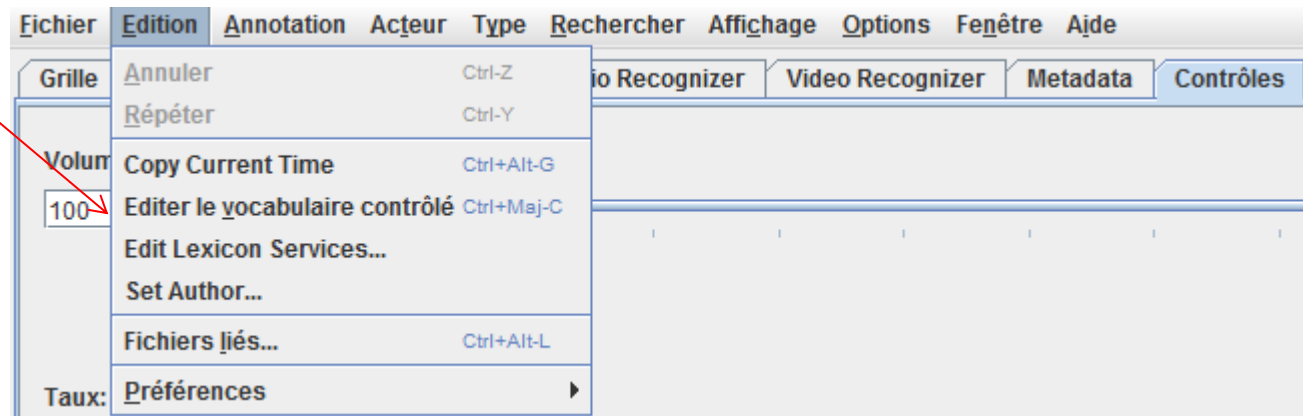
Fermer

Dans ELAN

Créer, modifier, supprimer un type linguistique



Créer, modifier, supprimer un vocabulaire contrôlé



Création d'une grille d'annotation

Pour construire une grille d'annotation, il faut donc avoir défini:

- Les acteurs à annoter
- La hiérarchie des acteurs
- Les types linguistiques nécessaires
- Éventuellement les vocabulaires contrôlés adéquats

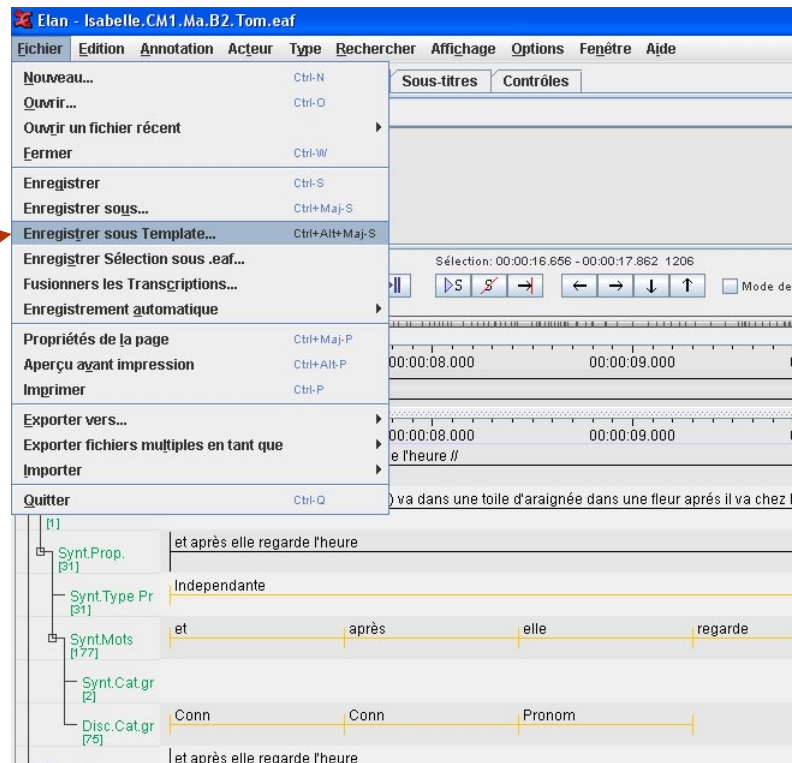
Attention, si la réflexion sur la grille se fait à partir des acteurs, dans le logiciel il faut partir dans l'ordre inverse.

Vocabulaire contrôlé => Type linguistique => Acteur

Les templates (création)

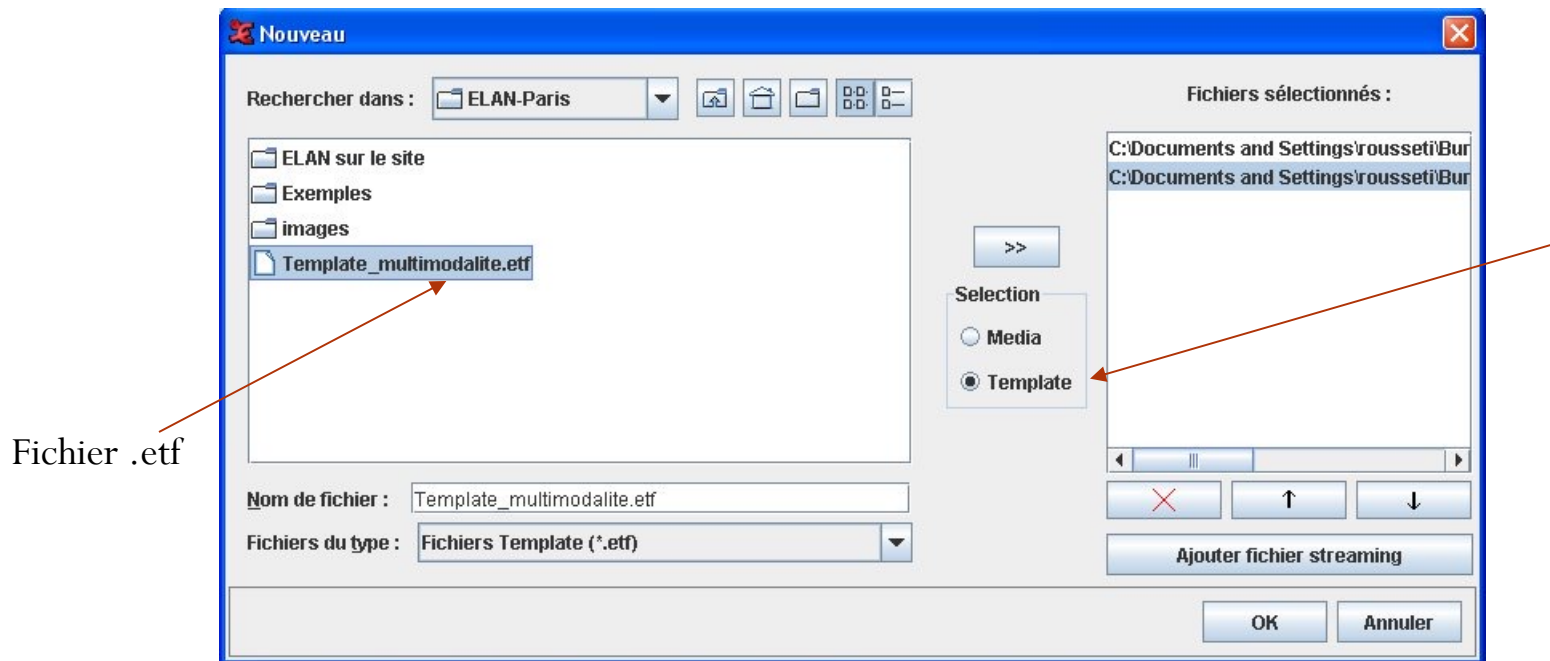
Le « template » est la sauvegarde de la grille d'annotation avec toutes les informations définies précédemment (acteurs, hiérarchisation, type linguistique, vocabulaire contrôlé)

Création d'un fichier .etf



Les templates (utilisation)

Une fois le fichier .etf sauvegardé, il faut l'intégrer au moment de la création d'un nouveau fichier, après avoir sélectionné le ou les fichiers media.



ELAN - Nom de fichier indéfini

Fichier Edition Annotation Acteur Type Rechercher Affichage Options Fenêtre Aide

Grille Texte Sous-titres Lexicon Audio Recognizer Video Recognizer Metadata Contrôles

Volume: 100 0 50 100

Taux: 100 0 100 200

00:00:00.000 Sélection: 00:00:00.000 - 00:00:00.000 0

Mode de sélection Mode de boucle

00:00:00.000 00:00:01.000 00:00:02.000 00:00:03.000 00:00:04.000 00:00:05.000 00:00:06.000 00:00:07.000 00:00:08.000 00:00:09.000 00:00:10.000 00:00:11.000

Adulte [0]

Enfant [0]

1 - Geste(Phases) [0]

1 - Valeur du Geste [0]

1 - Relation Geste/ [0]

1 - Relation synchron [0]

1 - Forme du geste [0]

2 - Geste(Phases) [0]

2 - Valeur du Geste [0]

2 - Relation Geste/ [0]

2 - Relation synchron [0]

2 - Forme du geste [0]

1 - Geste(Phases)- [0]

1 - Valeur du Geste [0]

1 - Relation Geste/

On obtient un fichier comprenant la vidéo, le signal sonore, et l'ensemble de la grille à annoter

TP

Avec la vidéo de votre choix, proposer un template permettant de récupérer les informations suivantes :

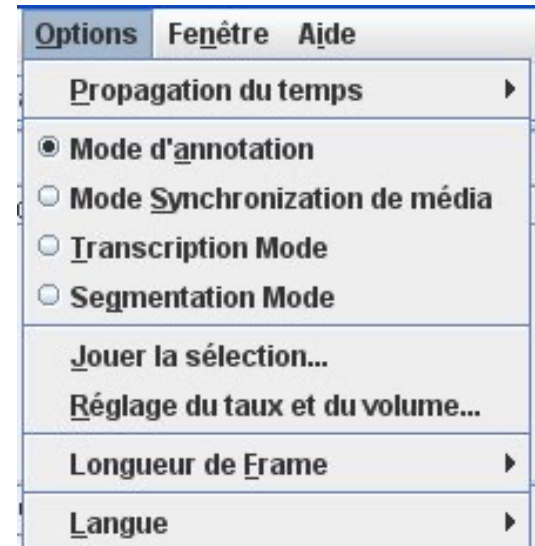
- Les paroles d'un premier intervenant ;
- Les gestes manuels de cet intervenant ;
- Les regards de cet intervenant ;
- La durée de la parole, des gestes manuels et des regards ;
- Les chevauchements entre parole, gestes manuels et regards ;
- 1 catégorisation de la parole (type de proposition), des gestes manuels (phase du geste), des regards (direction) avec pour chacun, le nombre d'occurrences et leur durée ;
- Les moments sans parole, ni geste manuel, ni regard (fonction gaps, qui fonctionne comme chevauchement au niveau de hiérarchies).

Annotation

Il ne reste plus qu'à annoter !

Depuis la version 4.1, ELAN propose plusieurs modes permettant de transcrire et d'annoter plus facilement :

- Mode de segmentation : découpage temporel ;
- Mode de transcription : remplissage facilité des annotations ;
- Mode d'annotation = mode traditionnel : permettant l'analyse et la visualisation des informations.



Mode de segmentation

Il y a 3 modes de segmentation (attention d'être dans le bon acteur – fenêtre du bas)

- Deux frappes par annotation : chaque frappe correspond à une borne (première et dernière d'une seule annotation) ;
- Une frappe par annotation : la première correspond au début des annotations, à partir de la deuxième, cela correspond à la borne finale de la première annotation qui est aussi la borne de début de l'annotation suivante ;
- Une annotation par touche pressée : chaque clic crée une annotation d'une durée définie ;

Mode de transcription

Ce mode vous permet d'avoir accès aux annotations vides que vous avez créées avec le mode segmentation. Vous devez sélectionner les annotations à remplir par type linguistique.

Si vous cliquez sur une annotation, la lecture de la vidéo et du son est lancée automatiquement (sauf si vous décliquez 'Automatic blayback') et peut être mise en boucle.

Vous pouvez taper directement vos textes dans la fenêtre de droite plus simplement que dans le mode annotation.

=> Ce mode est prévu pour les transcriptions de premier niveau. Tout ce qui concerne les annotations avec analyse nécessitant une vision plus globale se fait dans le mode annotation.

Mode d'annotation

Rappel sur la création d'annotations :

- On sélectionne la séquence temporelle souhaitée ;
- On double-clique sur la ligne à annoter ;
- On entre son texte ;
- On clique sur « Entrée » pour valider.

Il existe ensuite plusieurs fonctionnalités permettant d'aider à l'annotation.



Insertion de médias

Fichiers liés

Fichiers média liés | **Fichiers secondaires liés**

Nom de fichier	URL du fichier	Type MIME	Extrait de...	Offset	Media princi...	Statut
----------------	----------------	-----------	---------------	--------	-----------------	--------

Description du lien média

Nom de fichier
URL du fichier
Type MIME
Extrait de...
Offset
Media principal
Statut

Ajouter fichier streaming | **Ajouter...** | **Supprimer** | **Mettre à jour...**
Placer le média principal | **Placer l'extrait vers...** | **↑** | **↓**

Appliquer | **Fermer**

ions
nent

Mise en pratique

Dans la deuxième partie de cette matinée, vous allez travailler sur vos données.

- Formalisation des informations à produire grâce à votre corpus ;
- Réalisation d'un template sur papier ;
- Passage à ELAN ;
- Présentation au groupe et discussions, avis, conseils.