

Manuel de transcription de données multimodales dans Second Life

Inès Saddour, Ciara R. Wigham, Thierry Chanier

► **To cite this version:**

Inès Saddour, Ciara R. Wigham, Thierry Chanier. Manuel de transcription de données multimodales dans Second Life. 2011. <edutice-00676230>

HAL Id: edutice-00676230

<https://edutice.archives-ouvertes.fr/edutice-00676230>

Submitted on 4 Mar 2012

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

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



Projet ARCHI21¹

Manuel de transcription de données multimodales dans *Second Life*

Auteurs :

Inès Saddour, Ciara R Wigham, Thierry Chanier

Version du 1^{er} septembre 2011



Licence <http://mulce.org>

¹ EU Education and Culture Lifelong Learning Programme: KA2 Languages, IIP-eacea-KA2-MultProj-ARDNM

Table des matières

1	<i>Outil de transcription</i>	3
2	<i>Types d'actes dans Second Life</i>	4
2.1	Actes de type « verbal »	5
2.1.1	Modalité audio (tpa)	5
2.1.2	Modalité clavardage	7
2.1.3	Modalité silence	7
2.2	Actes de type « non-verbal »	8
2.2.1	Entrée et sortie dans l'environnement (es)	8
2.2.2	Déplacement (mov)	8
2.2.3	Kinésique (kin)	9
2.2.4	Modalité production	10
3	<i>Métadonnées</i>	11
4	<i>Références</i>	13

1 Outil de transcription

La transcription des données recueillies dans l'environnement *Second Life* est faite sur le logiciel ELAN (EUDICO Linguistic ANnotator). C'est un logiciel développé pour créer des annotations complexes sur les ressources vidéo et audio de données multimodales.

Une fenêtre d'ELAN est composée de plusieurs zones :

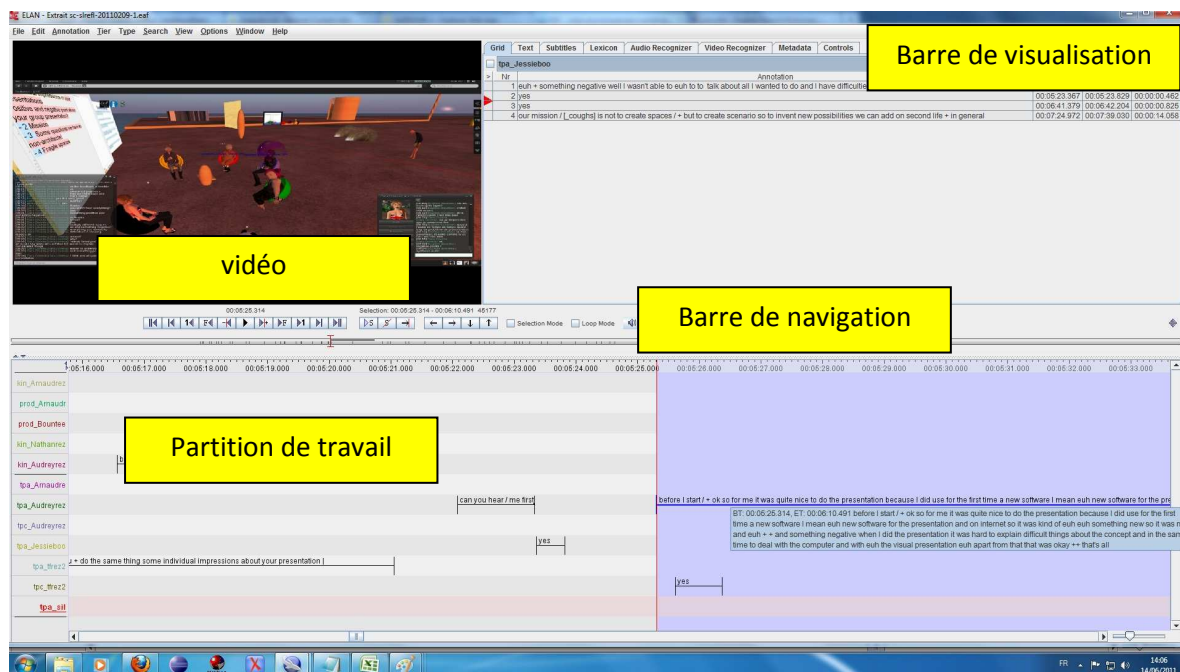


Figure 1. Interface d'ELAN

Comme l'image le montre, la vidéo s'insère en haut à gauche de la fenêtre. A droite, la barre de visualisation permet de visualiser les annotations, les acteurs, ainsi que la durée de chaque annotation. La barre de navigation permet de naviguer dans les annotations, vérifier leurs contenus et se déplacer sur l'axe temporel. Enfin, la partition de travail permet d'insérer les annotations et les gérer. Elle contient aussi les acteurs, ainsi que les différents types d'actes transcrits.

Nos annotations rendent compte de tous les actes verbaux ou non-verbaux effectués par les acteurs / avatars présents dans l'environnement et aussi de tous les actes visibles dont les acteurs ne sont pas connus. Si l'identité de l'acteur n'est pas identifiée l'acteur est '*unknown*' (par exemple, un objet peut bouger dans l'environnement sans que l'on puisse identifier l'acteur qui l'a bougé).

Une annotation indique donc une modalité liée à un acteur (tour de parole audio, clavier, déplacement, kinésique etc.) ou pas. En effet, elle peut aussi indiquer un silence ou un commentaire.

Chaque annotation est alignée temporellement à la source audio ou vidéo ou elle peut référer à d'autres annotations existantes quand il s'agit d'une annotation commentaire. Le contenu écrit des annotations est toujours en Unicode et la transcription est conservée dans un format XML.

Tout fichier de transcription est lié à un fichier de métadonnées réalisé sur le logiciel IMDI editor, disponible et téléchargeable à l'adresse suivante : <http://www.lat-mpi.eu/tools/imdi/editor/>. A chaque modification du fichier des métadonnées l'application doit être rechargée à nouveau de l'internet à l'adresse indiquée. Le fichier concerné est ensuite ouvert dans le logiciel et modifié.

2 Types d'actes dans *Second Life*

Les actes trouvés dans l'environnement *Second Life* (SL) sont classés et représentés dans le graphique Figure 2.

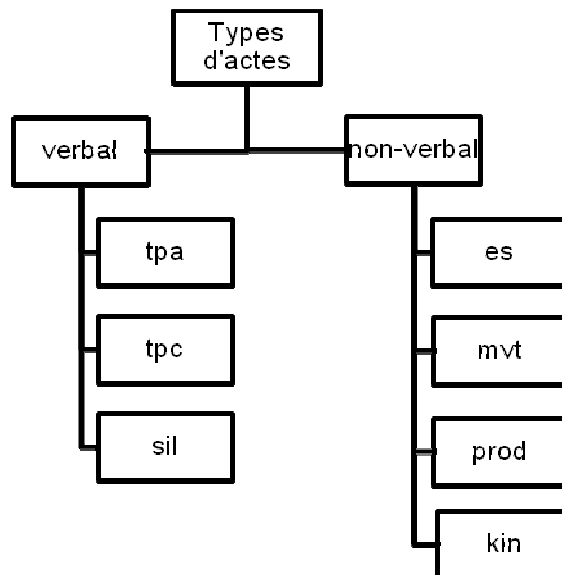


Figure 2. Types d'actes transcrits dans l'environnement *Second Life*

Chaque acte effectué par un acteur est codé dans une ligne individuelle comme suit : Type d'acte_code acteur. La liste des codes d'acteurs est disponible dans un fichier Excel nommé ARCHI21_acteurs_110304.xls.

Prenons par exemple la ligne tpa_tfrez2, elle correspond à un tour de parole audio effectué par l'acteur tfrez2.

Les deux types d'actes, verbal (v) et non-verbal (nv) sont codés comme des types linguistiques sur le logiciel ELAN. Chaque ligne d'annotation est classée dans l'une des deux catégories de la manière représentée dans la capture d'écran dans Figure 2.

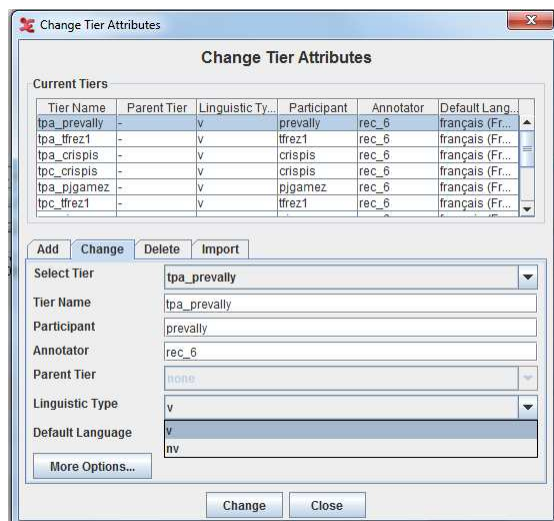


Figure 3. Classification des actes sur ELAN

2.1 Actes de type « verbal »

Les actes de type « verbal », codés v, se composent des actes figurant dans tableau 1.

<i>canal modalité</i>	<i>valeur</i>	<i>Début, Fin</i>
tpa	Tour de parole audio	Temps début = début intervention Temps fin = fin intervention
tpc	Tour de parole clavardage	Temps début = temps fin (donc action presque instantanée)
sil	silence	Intervalle temporel s'insérant entre deux tpa du même acteur d'une durée supérieure à 3 secondes et de n'importe quelle durée entre deux tpa distincts produits par deux acteurs différents

Tableau 1. Types d'actes de type « verbal »

2.1.1 Modalité audio (tpa)

Conventions	Symbole	Exemple
L'intonation montante est signalée par un slash /	/	donc alors pour demain / je + je dois encore + euh sélectionner tous les images
Quand un mot n'est pas complètement articulé à cause d'une interruption, ceci	/	second li/ (au lieu de second life)

est indiqué par un slash / sans espace.		
L'interrogation ayant une intonation montante ou descendante est signalée par un		vous l'avez dans votre inventaire
L'allongement (insistance) est signalé par une mise en majuscule suivie de deux points (ajout du 7/07/07, d'après ICOR)	MAJ:	attention on dit PRO:thèse
L'élision non standard est notée par une antiquote. Si l'élision non standard se produit en fin de mot, faire suivre l'antiquote d'un espace	`	J`sais pas
Les chevauchements en début ou en fin de tour de parole sont indiqués par les parenthèses	()	(ah ok [_chuckles] ok) merci (merci au revoir)
Les chevauchements à l'intérieur d'un tour de parole sont indiqués à l'intérieur du tour de parole entre (< >) en indiquant le code_acteur concerné avec la balise stp, séparé du texte par des guillemets.	(<stp cd="code">... </stp>)	(<stp cd="pjgmez">[_laughs]</stp>)
L'alternance codique est signalée par les accolades. Le transcripteur indique le passage d'une langue à l'autre en notant entre accolades la partie du message qui se réalise dans l'autre langue. Si une partie du message est entre accolades, alors cela signifie que l'autre langue mise entre accolades est minoritaire.	{}	like it's like a {mise en abîme} I don't know
Les pauses sont signalées par le signe + (1s) ++ (2 s) +++ (3 s)	+, ++, +++	c'est bientôt fini c'est notre dernière soirée / + qu'est-ce que vous avez pensé de cette semaine intensive
Les passages inaudibles ou trop incertains sont codés avec des XXX en majuscules mis entre crochets et précédés par un tiret bas	[_XXX]	yes [_XXX]
Les passages incompréhensibles sont codés par des xxx minuscules	xxx	we just could say that we + hm + xxx ++ [_chuckles] I don't know

Tous les tpa relevant de l'audio indiquant des sons ne relevant pas de la production langagières sont représentés par un verbe conjugué à la troisième personne mis entre crochets et précédé par un tiret bas.	[_chuckles] [_sighs] [_laughs] [_coughs] [_breathes in microphone]	gratte-ciel ok {grattacielo} en italien [_chuckles]
Les propos formulés d'une manière incompréhensible ou différente de l'emploi des natifs sont mis entre []. L'équivalent ou la réinterprétation des propos sont intégrés à la transcription sans codes particuliers.	[...]	Je vais [va]

Tableau 2. Conventions de transcription de tours de parole audio

2.1.2 Modalité clavardage

Les contenus de tpc sont tirés directement des fichiers de clavardage, quand ils existent. Si, lors des transferts, les accents ont été modifiés (accents remplacés par signes bizarres), on veillera à les rétablir. Mais il ne faut en aucun cas, réécrire les contenus (correction, changement casse, etc.).

Les temps des tours de parole figurant sur les fichiers de clavardages ne correspondent pas au temps durant lequel les tours de parole ont eu lieu. Ces temps sont alors relevés lors de la transcription de la vidéo.

Le clavardage entre les chercheurs (ce que nous appellerons, clavardage privé) ne pouvait pas être enregistré comme le clavardage entre les avatars et les tuteurs (clavardage public). Il est transcrit directement à partir de la vidéo en sachant que l'incertitude peut-être grande. Le clavardage privé est codé différemment dans les lignes d'annotation pour le différencier du clavardage public. Nous codons alors ces lignes de clavardage privé par : tpc_code acteur_pri. On ajoute donc au nom de l'annotation -pri (privé).

2.1.3 Modalité silence

Le tour de parole « sil » est une sous-catégorie de la modalité audio (tpa) tout comme dans les transcriptions multimodales adoptées dans d'autres projets de la base de données Mulce (cf. Rapport d'activité, tâcheTmult1).

Cependant, dans le projet ARCHI21 est considéré un tour de parole de type silence tout intervalle temporel séparant deux tpa. Si cet intervalle sépare deux tpa du même acteur, il faut qu'il couvre une durée supérieure à 3 secondes. Si toutefois le silence sépare deux tpa distincts par deux acteurs différents, celui-ci peut avoir une durée qui peut varier et qui peut durer au minimum une seconde.

2.2 Actes de type « non-verbal »

Il est important de noter ici que les actes non-verbaux ne sont transcrits que s'ils sont visibles par le transcrivieur. En effet, la visibilité de certains avatars est parfois réduite ou limitée pendant un moment à cause d'un acte de production ou de l'affichage d'une image.

2.2.1 Entrée et sortie dans l'environnement (es)

L'apparition et la disparation des avatars, leur entrée et sortie dans l'environnement SL et dans l'espace allouée à la session est signalée par le code es. La ligne d'annotation est ainsi nommée es_code acteur et indique dans son contenu si l'avatar apparait ou disparaît par les codes **appear** et **disappear**.

2.2.2 Déplacement (mov)

Cette modalité concerne le déplacement des avatars dans l'espace alloué lors de la session. Le déplacement peut être horizontal (marcher, courir, reculer, etc.) et aussi vertical (s'asseoir, se mettre debout, voler, planer, sauter, atterrir, etc.) Nous codons tous les déplacements possibles par la forme de base en anglais, *the base form* (ce qui correspond à l'infinitif sans to). Voici la liste complète des possibilités :

Annotation code	Description
sit_down	L'avatar s'assied
stand_up	L'avatar se lève
fly	L'avatar vole
take_off	L'avatar décolle
land	L'avatar atterrit
turn	L'avatar se retourne pour changer de direction
walk_backwards	L'avatar avance en marchant
move_backwards	L'avatar se déplace en reculant (par exemple, sur sa chaise)
walk_forward	L'avatar recule
jump	L'avatar saute
run	L'avatar court ou avance rapidement

Tableau 3. Différents types de déplacement des avatars dans SL

2.2.3 Kinésique (kin)

La kinésique désigne la gestuelle, et tous les mouvements effectués par les parties du corps des avatars. Il ne faut pas confondre ce qu'on appelle ici kinésique et le déplacement le positionnement et le changement des avatars dans l'espace, exemple, marcher, voler, courir etc., codé par **mov**.

Les gestes relevés sont codés par la forme de base en anglais, *the base form* (ce qui correspond, l'infinitif sans *to*). Tableau 4 liste ces gestes corporels ainsi que leurs codes.

Annotation code	Description	Exemples
spectator_pose	L'avatar s'incline en avant en baissant la tête. L'utilisateur de l'avatar a choisi si le système active cette kinésique à partir de 2, 5, 10 ou 30 minutes d'inactivité. Cependant un utilisateur peut ne pas activer cette option.	-----
type	L'avatar tape sur le clavier mais le clavardage n'est pas forcément destiné au groupe, il ne sort donc pas dans le fichier de clavardage. Si c'est le cas, l'acte est codé comme un tpc.	-----
drink	L'avatar fait un geste répétitif ou pas qui indique qu'il boit quelque chose pendant la session. Ces gestes sont codés ainsi que leur durée. Parfois ce geste empêche de voir l'avatar tape sur son clavier.	-----
move(X)	L'avatar fait bouger une partie de son corps. On indique donc de quelle partie il s'agit entre parenthèses séparant les mots par un _.	<p>move(arm_L) : L'avatar bouge le bras gauche</p> <p>move(arm_R_L) : L'avatar fait bouger ses deux bras, la droite et la gauche</p> <p>move(upper_body) : l'avatar fait bouger le haut de son corps tout en restant assis sur sa chaise</p> <p>move(legs_apart) : l'avatar écarte les jambes</p> <p>move(arms_apart) : l'avatar lève latéralement ses bras</p>
look_up	L'avatar lève la tête et regarde en haut	-----

lower_head	L'avatar baisse la tête	-----
stretch	L'avatar s'étire	-----
shake head	L'avatar fait bouger sa tête horizontalement	-----
nod	L'avatar fait un geste de sa tête pour acquiescer ce qui a été dit	-----
dance	L'avatar effectue debout des mouvements de danse	-----
lean_forward	L'avatar se penche en avant	-----
lean_back	L'avatar se penche en arrière	-----
clap_hands	L'avatar applaudit	-----
rotate_head	L'avatar fait pivoter sa tête	-----
eat(X)	L'avatar mange quelque chose (X)	eat(popcorn)
smoke	L'avatar fume	-----
point	L'avatar lève le bras et fait un geste déictique vers un objet dans l'environnement	-----
animate(X)	L'avatar active une animation désignée par X	animate(flashing_light)

Tableau 4. Différents types de kinésique dans les données recueillies dans Second Life

Deux actes de type kin ayant lieu simultanément et lié à un même acteur sont indiqués dans l'annotation par les deux codes correspondants séparés par un espace.

2.2.4 Modalité production

Les actes productions correspondent aux actions accomplies par les avatars afin de produire ou afficher des objets dans l'environnement *Second Life*. Ces actions sont de différents types classés en différentes catégories dans Tableau 5. Nous codons d'abord l'objet dont il est question, puis on met entre parenthèses l'action effectuée la désignant par un verbe à la forme de base suivi par le destinataire de cette action (à qui s'adresse l'action) et aussi par le nom / titre de l'objet. Le destinataire, s'il y en a un de spécifique, est désigné dans le tableau par X (et par le code acteur dans

les annotations) et le nom de l'objet par Y (ce qui correspond au nom donné au fichier attaché aux annotations).

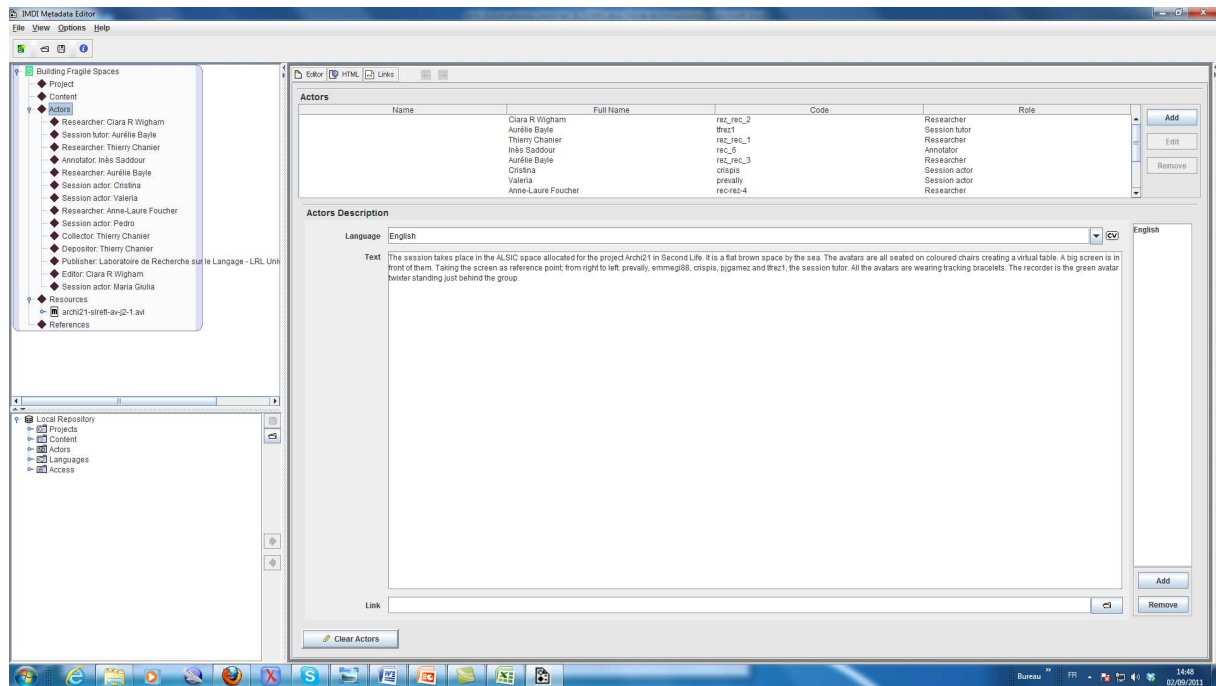
Annotation code	Description	Exemples
img(post, X)	L'avatar affiche une image sur un écran. X est le nom du fichier de l'image.	img(post, image1.jpg)
vol_up	L'avatar monte le son	-----
vol_down	L'avatar baisse le son	-----
obj(rez, X)	L'avatar rezze un objet	obj(rez, chair)
obj(move, X)	L'avatar bouge un objet désigné par X	obj(move, chair)
obj(open, X)	L'avatar ouvre un répertoire. Souvent c'est le chercheur qui filme la session qui effectue ce type de prod.	obj(open, inventory)

Tableau 5. Différents types de production des acteurs dans Second Life

3 Métadonnées

L'interface ELAN permet de visualiser les métadonnées dans la barre de visualisation dans l'option métadonnées. Comme mentionné plus haut, tout fichier de transcription est lié à un fichier de métadonnées réalisé sur le logiciel IMDI Metadata Editor, disponible et téléchargeable à l'adresse suivante : <http://www.lat-mpi.eu/tools/imdi/editor/>.

Le logiciel IMDI Metadata Editor signifie (ISLE Meta Data Initiative) et il a été conçu pour décrire des ressources et des données multimodales. Il permet de rentrer plusieurs informations sur les acteurs et aussi de fournir une description de la vidéo transcrite. Cette description peut être faite dans l'espace Actors description, comme le montre l'image suivante :



Les métadonnées qui nous intéressent sont les suivantes :

- Le nom, le titre, la date et le lieu de la session.

- Le titre et le code du projet global (archi21), ainsi que le nom, l'adresse postale, l'adresse électronique et l'affiliation de la personne responsable à contacter au sujet du corpus.

- La liste des acteurs : Plusieurs types d'acteurs sont présents. Les informations fournies dans les métadonnées dépendent du type.

- ⇒ Chercheurs : noms, prénoms et codes.

- ⇒ Collecteurs : noms, prénoms et codes.

- ⇒ Déposants : noms, prénoms et codes.

- ⇒ Diffuseurs : noms, prénoms et codes.

- ⇒ Editeurs : noms, prénoms et codes.

- ⇒ Informateurs : il s'agit des avatars des apprenants qui ont fait la formation. Pour l'anonymisation des données, il faut fournir le prénom, le code d'acteur, la date de naissance, l'âge et le sexe.

- ⇒ Tuteurs de langue : nom, prénom et code.

- La description des acteurs et du cadre global : cette description porte sur le positionnement général des avatars.

- Les différentes ressources : Il faut attacher les fichiers vidéo et texte du clavier. Pour les vidéos il faut indiquer leur durée.

Quand on sauvegarde le fichier des métadonnées on obtient un fichier imdi. Le fichier obtenu est en format xml. Ce fichier peut être attaché au fichier ELAN en cliquant sur Metadata de la barre de visualisation (voir Figure 1 plus haut).

4 Références

Ciekanski, M., Chanier, T. & A. Vetter (2008). Rapport d'activité Tâche Tmult1 : Multimodal, modèles, transcription, Annexe technique : Multimodal, conventions de transcription et outils de transcription. [<http://mulce.org>]

Max Planck Institute for Psycholinguistics (2001). *ELAN* [logiciel]. The Netherlands : Max Planck Institute for Psycholinguistics. [<http://www.lat-mpi.eu/tools/elan/>]

Rosedale, P. (2011). *Second Life v.2.7.2 (233432)* [logiciel]. San Fransisco : Linden Lab. [<http://www.secondlife.com>]