

ANR-06-CORP-006

Échange de corpus d'apprentissage multimodaux (MULCE)



Guide d'utilisation de Tatiana pour les corpus Mulce

Version du 09/04/2009

Référence du document

Tajariol, F. & Chanier, T. (2009). *Guide d'utilisation de Tatiana pour les corpus Mulce*, Mulce.org
[http://mulce.univ-fcomte.fr/metadata/doc/mce_guideTatiana.pdf]

Auteurs

Federico Tajariol & Thierry Chanier



1. Table des matières

1. Table des matières	2
2. Introduction.....	2
2.1. Motivations	2
2.2. Quelques mots sur ce guide et les démos associées	2
2.3. Le container d'un corpus distinguable Mulce.....	3
3. Installation du logiciel Tatiana	4
3.1. ETAPE 1 : préalable au téléchargement de Tatiana.....	4
3.2. ETAPE 2 : téléchargement et lancement de Tatiana.....	5
4. ETAPE 3 : visualisation des fichiers du corpus Copeas.....	5
5. ETAPE 4 : Règles de visualisation	9
5.1. ETAPE 5 : exportation de la transcription en format .xls.....	12
6. Bibliographie	13

2. Introduction

2.1. Motivations

Ce guide est destiné à permettre aux utilisateurs (chercheurs ou enseignants) du site Mulce (<http://mulce.org>) de pouvoir utiliser hors ligne les corpus d'apprentissage (http://mulce.univ-fcomte.fr/metadata/mce_LETECorpus-fr.pdf) dont les données ont été préparées pour le logiciel Tatiana, soit en vue d'un alignement des transcriptions des interactions en ligne avec les vidéos, soit en vue de consultation des analyses déjà réalisées avec ce même logiciel.

L'utilisateur peut, sur le site Mulce, consulter l'intégralité de nos transcriptions, ainsi que le détail du contexte (scénario pédagogique, protocole de recherche, etc.), faire des recherches dans ses données structurées et inter-reliées.

L'équipe du projet Mulce a choisi le logiciel Tatiana, développé par les collègues de l'équipe Armines, comme un exemple de logiciel d'aide à l'analyse de données multimodales présentes dans nos corpus et converties au format requis par le logiciel.

L'équipe Mulce n'est donc en rien responsable des problèmes éventuels de fonctionnement de Tatiana ni de limitations de fonctionnalités que l'utilisateur aurait pu observer. Pour toute demande en rapport avec cela, merci de prendre contact avec l'équipe Armines.

2.2. Quelques mots sur ce guide et les démos associées

Nous disposons à l'heure actuelle de deux types de corpus distinguable (composé à partir de données extraites des corpus d'apprentissage globaux) aménagés pour Tatiana :

- 1) Les corpus simples dans lesquels figurent les données de transcriptions et les vidéos. L'utilisateur peut alors aligner ses données dans Tatiana et se livrer à quelques visualisations ou analyses. Nous décrivons cela dans ce guide. Le lecteur peut consulter nos vidéos de démonstrations en rapport avec ce sujet :
 - http://mulce.univ-fcomte.fr/metadata/doc/demo_tatiana_1_flash/

- http://mulce.univ-fcomte.fr/metadata/doc/demo_tatiana_2_flash/
- 2) Des corpus disposant des mêmes données que précédemment avec, en plus des analyses réalisées par nos soins dans Tatiana. Ce cas n'est pas traité dans ce guise, voir pour cela la vidéo de démonstration :
 - http://mulce.univ-fcomte.fr/metadata/doc/demo_tatiana_3_flash/

2.3. Le container d'un corpus distinguable Mulce

Chacun de nos corpus se présente sous forme d'une archive ZIP. Ce container est au format IMS-CP.

Dans notre exemple il s'agit du corpus "copeas-T5_contexte_simple", extrait de la session T5 de la formation Copéas. Le corpus se présente dans une archive ZIP, ici "copeas-T5_contexte_simple-CP.zip". Une fois décompressée les choses se présente telles que dans la figure 1

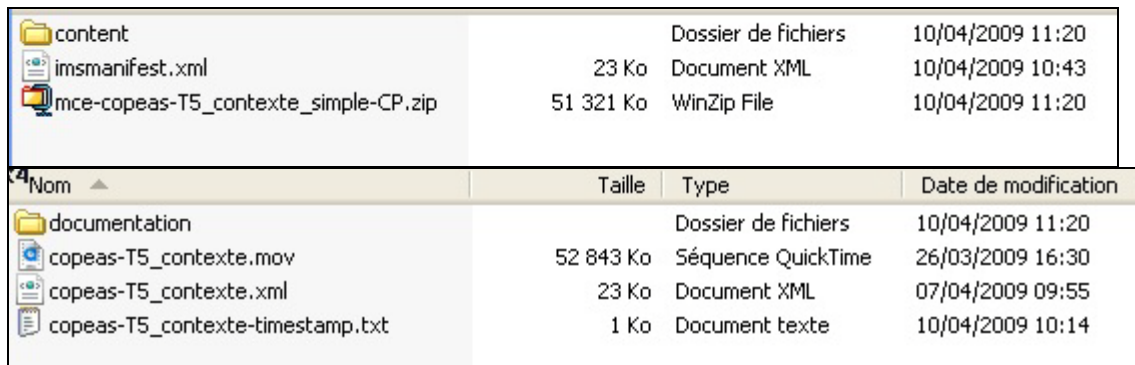


Figure 1 – En haut, le container, archive ZIP du corpus et le résultat décompressé. En bas, le contenu du dossier "content".

Dans notre exemple, le fichier vidéo est inclus dans le container. Ce n'est pas toujours le cas. Lorsque le volume de la vidéo est trop grand (au-delà d'une centaine de Mo), nous préférons que l'utilisateur la télécharge séparément à partir de notre site. Toutes les indications sont fournies dans les données du corpus distinguables (voir, par exemple, la partie "ressources" du fichier de visualisation montré en figure 2).

On notera la présence du dossier "documentation" où l'on pourra lire les informations sur le corpus en cliquant sur "index.xml" dont le contenu s'affichera dans son navigateur Internet (voir le résultat dans la figure 2). A titre d'exemple, voir cet extrait de corpus <http://mulce.univ-fcomte.fr/metadata/doc/visu-corpus/index.xml> . La démo précédente (Chanier, 2009c) le montre aussi.



Figure 2 Visualisation dans son navigateur Internet d'une sélection d'information sur le corpus.

3. Installation du logiciel Tatiana

Le logiciel Tatiana (*Trace Analysis Tool for Interaction ANALysis*), développé par l'équipe Armines (*Ecole des Mines de Saint-Etienne*), permet d'analyser des données relatives à des interactions humaines. Tatiana permet de rejouer des traces synchronisées (sessions audio-vidéo enregistrées et transcriptions), de conduire des analyses qualitatives (catégorisation et annotation) et d'exporter ces analyses pour ultérieures analyses quantitatives.

Ce document décrit les étapes à suivre pour pouvoir télécharger et lancer le logiciel Tatiana (§2.1, 2.2), visualiser les fichiers du corpus Copeas (§2.3), construire une analyse interprétative (§3) et exporter les dimensions de cette analyse (§3.1).

3.1. ETAPE 1 : préalable au téléchargement de Tatiana

Avant de télécharger TATIANA, il faut impérativement :

1. Pour le système opératif Windows, vous devez télécharger le fichier nommé `ensharpendecoder_win.exe` (disponible à l'adresse <http://www.techsmith.com/codecs/ensharpen/default.asp>) et l'installer dans le dossier nommé Program Files de votre ordinateur. Il s'agit d'un fichier permettant la visualisation des fichiers .mov du corpus Copeas.
2. Mettre à jour le Java™ Standard Edition Runtime Environment (jre).
3. Mettre à jour le lecteur QuickTime (NB : pour utiliser Tatiana, vous devez installer QuickTime à partir de la version 7.5).

Au cas où cet ordre ne serait pas respecté, le QuickTime Java Library ne pourra pas être mis à jour et donc Tatiana ne sera pas en mesure de lire les fichiers .mov. Pour résoudre ce problème **sous Windows**, il faut procéder ainsi :

- a) Vous devez repérer le dossier d'installation de Java (typiquement, C:\Program Files\Java). Dans un dossier, dont le nom commence par jre, il faut chercher un fichier nommé QTJava.zip.
- b) Copiez-le. Ensuite, repérez le dossier jre qui se termine par le numéro le plus élevé (dans notre cas, jre 1.6.0_07). Ouvrez ce dossier, et identifiez le dossier nommé lib et ouvrez-le. A l'intérieur, il y a un dossier nommé ext: vous devrez coller dans ce dossier le fichier QTJava.zip.

3.2. ETAPE 2 : téléchargement et lancement de Tatiana

Le logiciel Tatiana (version 2.10.3) doit être téléchargé à partir de l'adresse suivante <http://code.google.com/p/tatiana>

A cette adresse <http://code.google.com/p/tatiana/wiki/DownloadLinks?tm=2> vous trouverez de la documentation technique sur Tatiana.

Décompressez le fichier .zip dans un dossier, que vous placerez dans C:\Program Files de votre ordinateur. Parmi les nombreux fichiers, vous trouverez le fichier Tatiana.exe : double-cliquez et vous lancerez Tatiana.

4. ETAPE 3 : visualisation des fichiers du corpus Copeas

Dans cette section, vous trouverez la procédure pour visualiser les fichiers d'un corpus distinguable. L'archive du corpus étnat compressé vous trouverez notamment les fichiers suivants (voir figure 1 aussi) :

- a. copeas-T5_contexte.mov (fichier vidéo)
- b. copeas-T5_contexte.xml (fichier de la transcription au format Tatiana)
- c. copeas-T5_contexte-timestamp.txt (fichier contenant des valeurs à insérer dans Tatiana)

2. Ouvrez Tatiana, en double-cliquant sur l'icône.

A l'ouverture, vous trouverez un corpus déjà créé, intitulé ArminesDemo. Si vous voulez, vous pouvez le supprimer (choisissez File / Delete Corpus).

3. Vous devez maintenant créer un nouveau corpus

- a. dans le menu File choisissez New Corpus
- b. insérez un nom (par exemple, T5_contexte)

1. Importez le fichier .xml contenant la transcription de la vidéo

- a. Cliquez sur le menu File et choisissez Add file to Corpus
- b. Cliquez sur Browse, cherchez le fichier copeas-T5_contexte.xml. Sélectionnez-le et appuyez sur le bouton Next.
- c. Dans la fenêtre qui apparaît (nommée Add File To Corpus / Specify Synchronization Information), vous devez rentrer des informations contenues dans le fichier copeas-T5_contexte-timestamp.txt qui se trouve dans le dossier que vous avez nommé T5_contexte. Ouvrez ce fichier, et dans la partie **** Synchronization .xml **** vous trouverez ces informations :

First timestamp : 1108377539000

Last timestamp : 1108377881000

Delta : 0

- d. Insérez ces valeurs dans les champs correspondants de la fenêtre nommée Add File To Corpus /Specify Synchronization Information et puis appuyez sur le bouton Finish. Un dossier nommé trace vient d'être créé dans l'arborescence du corpus (figure 3). A l'intérieur, vous trouverez le fichier copeas-T5_contexte.xml que vous venez d'importer.

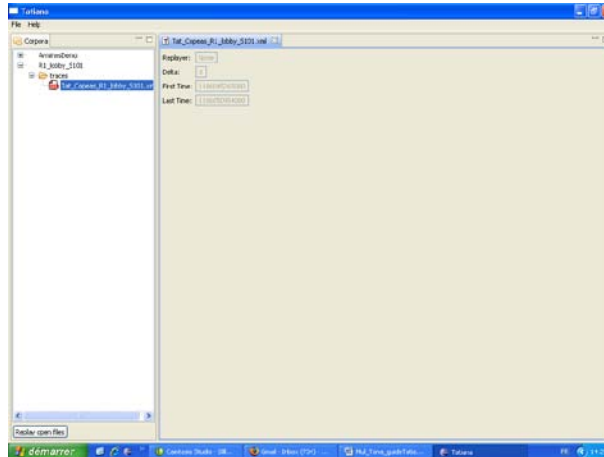


Figure 3 - le fichier .xml est ajouté au dossier "trace"

2. Visualiser le fichier importé, en créant un jeu d'analyse

- a. Dans le menu File, choisissez New Replayable et ensuite Replayable from filter. Une fenêtre s'ouvrira, vous invitant à choisir un filtre. Sélectionnez le filtre null-filter.xml et puis appuyez sur OK.

NB : selon la taille du fichier .xml, cette opération peut prendre un certain temps.

- b. Un onglet untitled s'ouvrira : à l'aide du menu déroulant, choisissez le fichier copeas-T5_contexte.xml et ensuite appuyez sur Run Filter. Vous observerez qu'un tableau s'est affiché, dont les champs contiennent le contenu de la transcription.
- c. Dans le menu File, choisissez Save et nommez le replayable (par exemple, copeas-T5_contexte). Un dossier nommé replayables vient d'être créé dans l'arborescence, et à son intérieur vous trouverez le fichier copeas-T5_contexte. Vous pouvez maintenant ajuster la taille des colonnes du tableau de manière à rendre plus lisible le contenu de leurs champs (Figure 2). Pour ce faire, double-cliquez sur les lignes verticales qui séparent les colonnes.

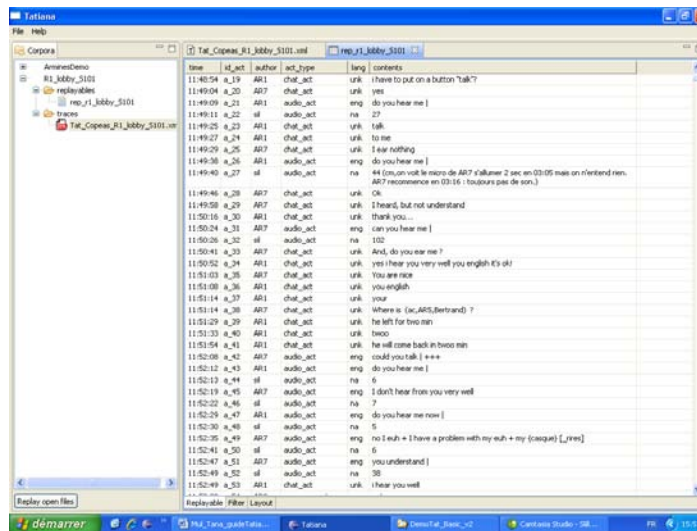


Figure 4 - visualisation des champs de la transcription

3. Importez le fichier vidéo .mov

- a. Cliquez sur le menu File et choisissez Add file to Corpus.
- b. Cliquez sur Browse, cherchez le fichier copeas-T5_contexte.mov et sélectionnez-le. Cliquez sur Next.
- c. Comme dans l'étape 3(c), une fenêtre apparaît (nommée Add File To Corpus), où vous devez rentrer des valeurs qui se trouvent dans le fichier copeas-T5_contexte-timestamp.txt, plus précisément dans la partie

**** Synchronization .mov ****:

First timestamp : 0
 Last timestamp : 342000
 Delta : 1108377539000

Collez ces valeurs dans les champs de la fenêtre Add File To Corpus.

- d. Cliquez sur Finish. Dans le dossier nommé trace, le fichier Copeas_R1_lobby_S101.mov vient d'y être ajouté.

NB : étant donné la taille importante des fichiers vidéo du corpus Copeas, l'opération d'importation du fichier .mov peut prendre un certain temps.

4. Visualisation en synchronie des fichiers .xml et .mov

- a. Si vous ne l'aviez pas encore fait, double-cliquez sur le fichier copeas-T5_contexte qui se trouve dans le dossier replayables.
- b. Double-cliquez sur le fichier copeas-T5_contexte.mov qui se trouve dans le dossier traces. Vous devrez avoir deux onglets ouverts : l'onglet copeas-T5_contexte (contenant la transcription) et l'onglet copeas-T5_contexte.mov (Figure 5).

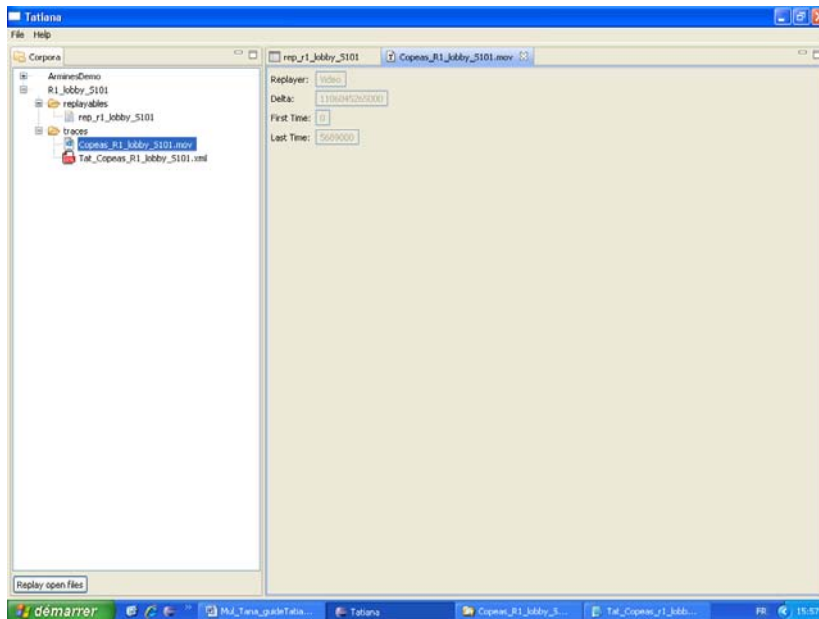


Figure 5 - le fichier .mov est ajouté

- c. Cliquez sur le bouton Replay open Files qui se trouve en bas à gauche de la fenêtre de Tatiana. Vous aurez trois fenêtres ouvertes sur votre écran (Figure 6):
 - i. la fenêtre principale contenant la transcription de la session (onglet copeas-T5_contexte) : selon les dimensions de votre écran, réduisez la taille de la vidéo et la taille de la fenêtre affichant la transcription.
 - ii. la fenêtre vidéo, affichant le copeas-T5_contexte.mov
 - iii. la fenêtre du Remote Control, qui vous permet de contrôler l'avancement de la session T5_contexte.

NB : Si vous ne voyez pas apparaître la fenêtre de la vidéo, il se peut que vous ayez oublié d'installer le codec nécessaire (cf. ETAPE 1).

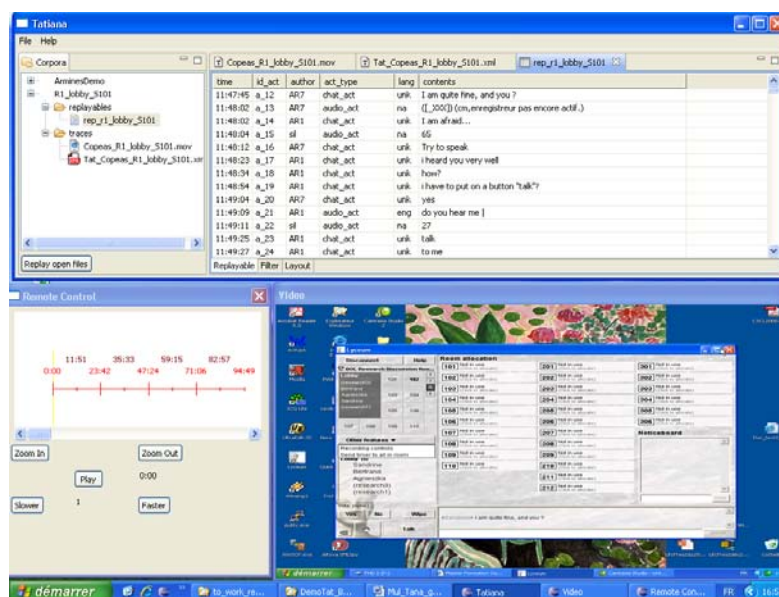


Figure 6 - le Remote Control (en bas à gauche), la vidéo (en bas) et la transcription de la session (en haut)

Vous pouvez rejouer la session (bouton Play) : vous verrez que les lignes dans le tableau de transcription sont surlignées au fur et à mesure que la vidéo avance. Appuyez sur le bouton Pause et choisissez un point de départ différent (déplacez le curseur jaune sur l'axe horizontal rouge et cliquez) : vous noterez la synchronisation entre le fichier .mov et la transcription de la session. Vous pouvez aussi rejouer un intervalle temporel précis (*drag & drop* avec le curseur jaune sur l'axe horizontal).

5. ETAPE 4 : Règles de visualisation

Chaque transcription en forme tabulaire (replayable) peut être visualisée sous une forme graphique, facilitant ainsi le travail d'interprétation du chercheur.

Voici la procédure à suivre selon notre exemple T5_contexte :

1. Sélectionnez le fichier copeas-T5_contexte dans le dossier replayables.
2. Cliquez sur File et choisissez Visualise as Score (qui se trouve en dessous de New Analysis). **NB** : Attention à bien sélectionner le fichier dans l'arborescence : même si la fenêtre du fichier copeas-T5_contexte est ouverte, la commande Visualise as Score ne sera pas disponible.
3. Un onglet nommé copeas-T5_contexte s'ouvrira (figure 7). Chaque ligne ou rectangle représente un acte produit par les acteurs pendant la session.

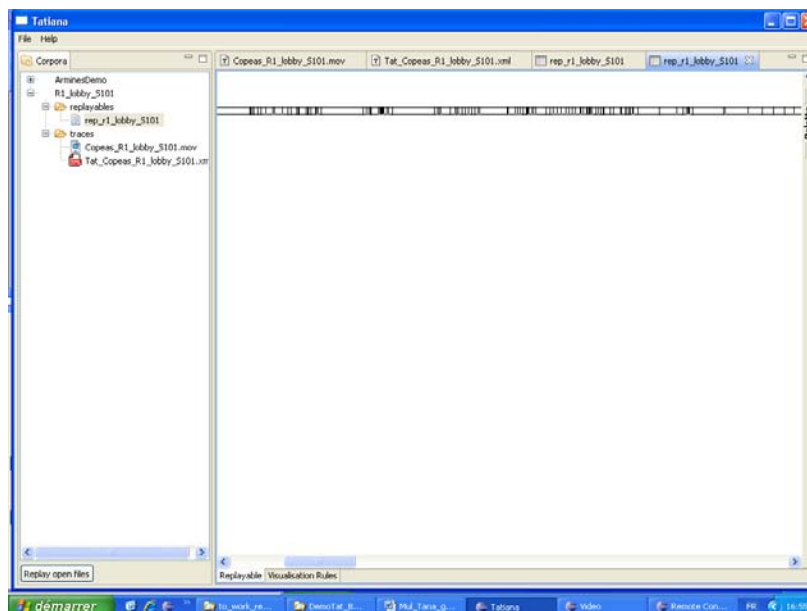


Figure 7 - view as score

4. Vous pouvez maintenant définir des règles de visualisation. Par exemple, nous allons créer des règles pour visualiser les actes produits par chaque acteur pendant la session. Cliquez sur l'onglet Visualisation Rules en bas de la fenêtre : un tableau vide apparaîtra. Pour créer une règle (dans l'ordre):

- a. vous devez choisir le facet (une des colonnes du tableau copeas-T5_contexte) : pour cet exemple, choisissez author;
- b. vous devez choisir sa valeur (menu facet value) : pour cet exemple, choisissez author qui a la valeur AR1 (nom anonymisé d'un apprenant) ;
- c. puis choisissez sa forme de visualisation à l'aide du menu Property: par exemple, on peut visualiser les actes produits par l'acteur AR1 sur une ligne horizontale différente des autres acteurs. En conséquence, nous choisissons relative position (qui modifie la position verticale des actes produits par chaque acteur). D'autres formes de visualisation sont disponibles (cf. p.21-22, TatianaDocumentation_V2.10_20081001) ;
- d. pour terminer, à l'aide du menu Property Value, définissez la position verticale de la ligne des actes produits par AR1 (en pixel, par exemple 20).
- e. cliquez sur le bouton Create rule et la règle apparaîtra dans le tableau ;
- f. cliquez maintenant sur l'onglet Replayable en dessous de Create rule et vous verrez le résultat de la règle que vous avez créée (Figure 8) : les actes produits par l'acteur AR1 sont visualisés en dessous de l'ensemble des actes.

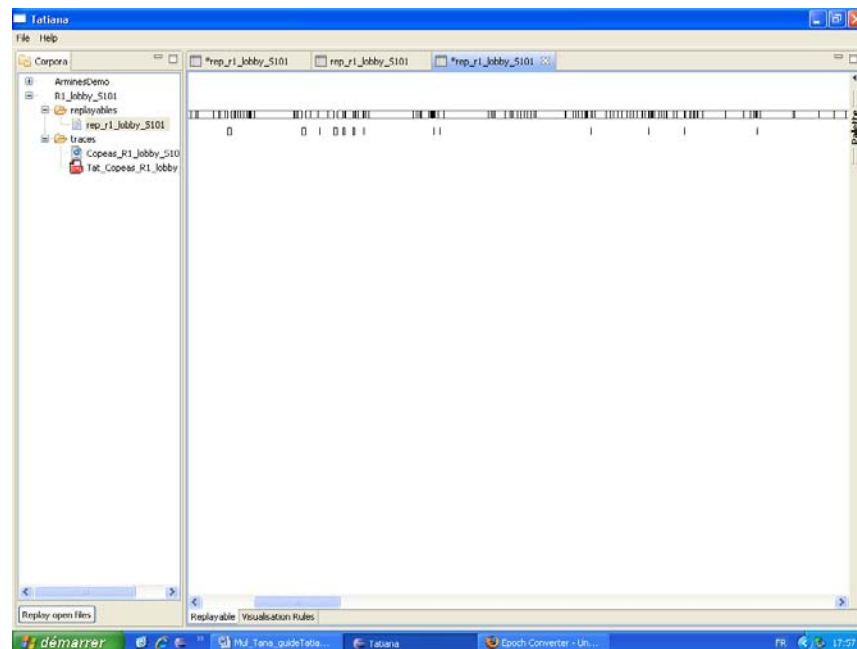


Figure 8 - view as score selon la règle « author »

5. Répétez l'étape 4 pour définir d'autres règles. Par exemple, ajoutez le facet colour pour l'acteur AR1, avec une valeur bleu. Vous pouvez ajouter d'autres règles de visualisation (Figure 9),

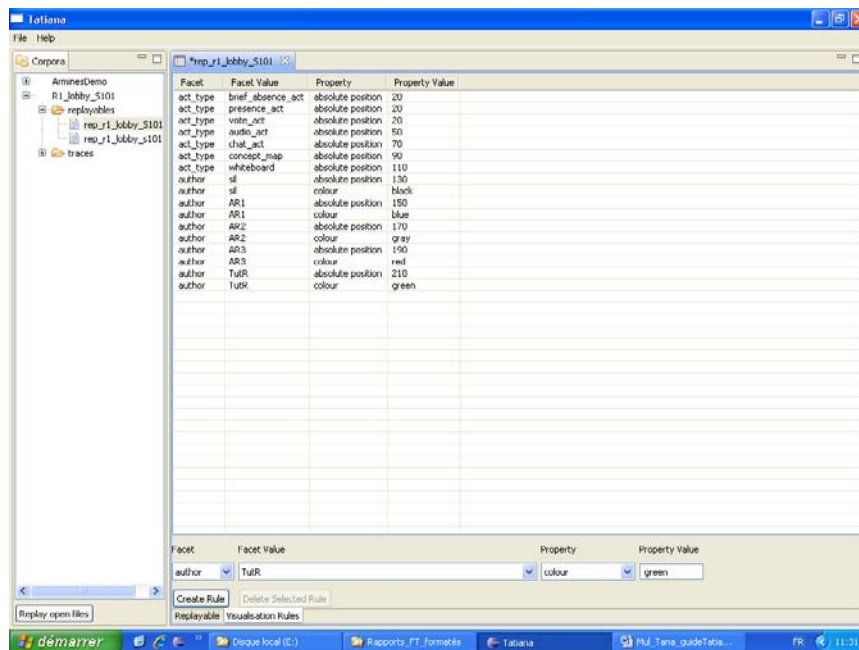


Figure 9 – d'autres règles de visualisation

Avec ces règles, nous nous focalisons sur certains types d'actes produits par quatre acteurs de la session (AR1, AR2, AR3 et TutR ; l'acteur sil est le silence).

Ces règles visualisent ainsi la session (Figure 10) : à partir du haut, la première ligne visualise les actes de type brief_presence_act, presence_act et vote_act (absolute position 10) ; la 2^{ème} ligne est occupée par les actes de type audio_act (absolute position 40) ; la 3^{ème} ligne par les actes chat_act (absolute position 60). La 4^{ème} et 5^{ème} ligne montrent respectivement les actes de production concept_map et whiteboard (absolute position 90 et 110). Ensuite, les lignes montrant les acteurs, auxquels nous avons ajouté des couleurs : l'acteur sil (6^{ème} ligne, absolute position 130), AR1 (7^{ème} ligne, absolute position 150), AR2 (8^{ème} ligne, absolute position 170), AR3 (9^{ème} ligne, absolute position 190), TutR (10^{ème} ligne, absolute position 210). Nous avons aussi attribué des couleurs aux acteurs : noir pour le silence (property value black), pour AR1 (property value blue), AR2 (property value gray), AR3 (property value red) et TutR (property value green).

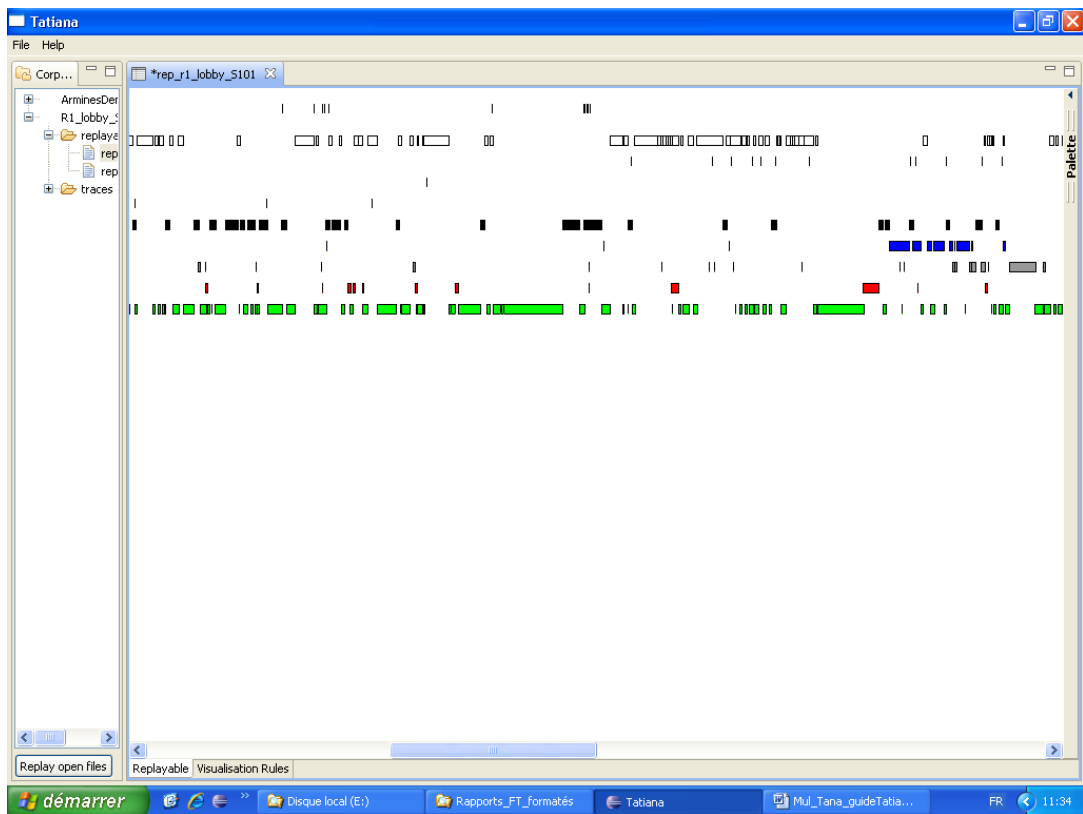


Figure 10 – visualisation obtenue avec toutes les règles de visualisation

NB : les règles que vous avez créés ne sont pas enregistrées dans un fichier à part mais sont enregistrées dans le replayable nommé copeas-T5_contexte. Donc, lors d'une prochaine séance de travail, pour voir ces règles de visualisation, vous devrez sélectionner le fichier copeas-T5_contexte, choisir la commande Visualise as Score et cliquer sur l'onglet Visualisation Rules.

6. Pour faciliter la lisibilité des actes, vous avez à votre disposition une fonctionnalité de zoom :
 - a. Cliquez sur le bouton de gauche de la souris : un petit menu apparaîtra avec les commandes Zoom In et Zoom Out.
 - b. Répétez l'étape 6a autant de fois que vous voudrez. Vous remarquerez qu'en bas de la fenêtre un curseur apparaîtra : cela vous permettra de choisir l'intervalle temporel que vous voudrez.

5.1. ETAPE 5 : exportation de la transcription en format .xls

Tatiana permet d'exporter chaque replayable en format xls, bien utile lorsqu'on veut produire une analyse statistique.

Pour ce faire, vous devez sélectionner le replayable que vous voulez exporter et choisissez dans le menu File la commande Export sous Excel.

6. Bibliographie

- Armines Team – Ecole des Mines de Saint-Etienne (2008). *Tatiana Documentation* [http://lead.emse.fr/Documents/TatianaDocumentation_V2.10_20081001.pdf]
- Armines Team – Ecole des Mines de Saint-Etienne (2008). *Tatiana First Steps User Guide*. [http://lead.emse.fr/Documents/TatianaFirstStepsUserGuide_V2.10_20081001.pdf]
- Armines Team – Ecole des Mines de Saint-Etienne (2008). *Tatiana Technical Documentation*. [http://lead.emse.fr/Documents/TatianaTechnicalDocumentation_V2.10_20081001.pdf]
- Chanier, T. (2009a). Démonstration de l'alignement transcriptions et vidéo Mulce dans Tatiana [vidéogramme], Mulce.org [http://mulce.univ-fcomte.fr/metadata/doc/demo_tatiana_1_flash/]
- Chanier, T. (2009b). Démonstration des règles de visualisation après alignement des transcriptions et vidéo Mulce dans Tatiana [vidéogramme], Mulce.org [http://mulce.univ-fcomte.fr/metadata/doc/demo_tatiana_2_flash/]
- Chanier, T. (2009c). Démonstration des analyses réalisées sur un corpus Mulce dans Tatiana [vidéogramme], Mulce.org [http://mulce.univ-fcomte.fr/metadata/doc/demo_tatiana_3_flash/]
- Mulce (2009). Notion de corpus d'apprentissage. Mulce.org [http://mulce.univ-fcomte.fr/metadata/mce_LETECorpus-fr.pdf]
- Tajariol, F. & Chanier, T. (2009). *Guide d'utilisation de Tatiana pour les corpus Mulce*, [rapport projet ANR], Mulce.org [http://mulce.univ-fcomte.fr/metadata/doc/mce_guideTatiana.pdf]