

Analyse exploratoire d'un corpus d'apprentissage collaboratif: détection de reformulations multimodales lors des échanges enseignant-étudiant

Amélie Ngatsi-Imafouo

Ecole Nationale Supérieure des Mines de Saint-Etienne
158 Cours Fauriel - 42023 Saint-Etienne cedex2 – France

Résumé

Les traces obtenues suite à une interaction avec un environnement informatique (d'apprentissage, de travail, de collaboration) sont (doivent être) porteuses d'indications sur l'activité réalisée au sein de cet environnement. Dans le cadre de l'apprentissage collaboratif assistée par ordinateur, il s'agit pour les chercheurs d'exploiter les traces pour la compréhension du parcours cognitif des apprenants, au cours de l'activité collaborative d'apprentissage. C'est dans ce cadre que se situe notre étude. A travers l'outil Lexico3, nous nous intéressons particulièrement aux reformulations lors des échanges enseignant/étudiant, notamment la façon dont les étudiants prennent note à propos des concepts importants énoncés par l'encadrant de projet.

Abstract

The traces obtained after an interaction with an IT environment (for learning, work or collaboration) should give information about the activity realized through this environment. In Computer Supported Collaborative Learning (CSCL), researchers need to make use of these traces to analyze and understand the learning situation. The work presented here is within this frame. The Lexico3 tool helps us to detect reformulations in a corpus made of discussions between a teacher and a couple of students about tutoring computer programming projects. For example, we analyze the link between the students' note-taking and the teacher's speech.

Keywords: reformulation, interaction traces, collaborative learning, trace analysis

1. Introduction

Les interactions médiatisées par ordinateur laissent de traces qui sont/doivent être porteuses d'indices sur le déroulement de l'activité dont elles sont issues. Pour les chercheurs qui utilisent ces traces afin de comprendre le processus (cognitif) qui a eu lieu au cours de l'activité ou pour des concepteurs d'environnement qui souhaitent comprendre comment les utilisateurs se sont appropriés leur environnement, il est primordial de fournir des outils pour rendre aisé l'accès au contenu des traces d'interaction.

Une fois recueillies, ces traces sont plus ou moins mises en forme pour constituer un corpus. Ce dernier sert de base de travail aux chercheurs en sciences cognitives qui l'utilisent pour décrypter/détecter des événements intéressants dans le déroulement de l'activité. Dans le cadre de l'apprentissage collaboratif assistée par ordinateur (Computer Supported Collaborative Learning ou CSCL), il s'agit pour les chercheurs d'exploiter les traces pour la compréhension du parcours cognitif des apprenants. C'est dans ce cadre que se situe notre étude. Nous analysons un

corpus de traces recueillis lors de séances de présentation des projets informatique aux étudiants d'une école d'ingénieur à travers l'outil Tatiana (Trace Analysis Tool for Interaction Analysts). C'est un outil d'analyse de traces permettant d'analyser les interactions issues d'outils collaboratifs. Il permet entre autres la collecte, la synchronisation, la visualisation, et l'analyse des corpus d'interactions médiatisées (Dyke et al., 2009a ; 2009b). Nous nous intéressons particulièrement aux reformulations lors des échanges enseignant/étudiant, notamment la façon dont les étudiants prennent note à propos des concepts importants énoncés par l'encadrant de projet.

Tout d'abord, nous présentons quelques cas d'analyse exploratoire de corpus écrits (ou oraux transcrits) dans des domaines divers et variés. Plusieurs logiciels libres ou commerciaux offrent des fonctionnalités riches d'analyse textuelle (statistique textuelle, lexicométrie, etc.) et soutiennent ces études. Le cadre de notre étude est ensuite présenté plus en détails ainsi que les analyses que nous avons effectuées. Nous terminons par les perspectives de ce travail.

2. Analyse exploratoire de corpus dans divers domaines

Un certain nombre de travaux se sont penchés sur l'analyse exploratoire de données. Ces travaux se situent dans des domaines divers. Bon nombre s'appuient sur des corpus constitués de transcriptions d'entretiens. Ainsi, (Gelinat-Chebat et al., 2004) analysent des transcriptions d'entrevues avec des jeunes sur le thème du tabagisme, afin d'évaluer l'impact des messages dissuasifs et brochures de lutte de contre le tabagisme sur ces jeunes. Le logiciel SATO est utilisé dans une démarche itérative de construction de catégories lexicales sur ces données. D'autres travaux portant sur ce même corpus d'entrevues ont combiné divers logiciels (Alceste, DTM, Lexico3) pour valider et compléter l'analyse précédente (Daoust et al., 2006). (Peyrat-Guillard, 2006) s'appuient de façon complémentaire sur les logiciels Alceste et WordMapper pour analyser un corpus d'entretiens dont le but est d'étudier les diverses formes d'implication au travail. Ces outils permettent de générer des classes de discours, caractérisées par un certain vocabulaire spécifique. Pour ces auteurs, l'outil Alceste donne une meilleure vue d'ensemble du corpus, tandis que WordMapper met en évidence l'articulation des divers thèmes entre eux. Pour étudier l'interaction entre intervieweur et interviewés dans des entretiens semi-directifs, Bonneau (2008) utilise les logiciels Alceste et Lexico et obtient une « cartographie d'indices linguistiques, thématiques et énonciatifs » qui facilite l'analyse d'entretiens semi-directifs. Le logiciel DTM a été utilisé par Cailliau et Poudat (2008) pour l'observation des caractéristiques lexicales dans les dialogues (téléphoniques retranscrits) entre agents EDF Pro et des clients. Le corpus qui comprend plus de 1000 conversations téléphoniques retranscrites est filtré (étiquetage morpho-syntaxique et désambiguïsation, lemmatisation) et les oppositions significatives entre les discours des deux catégories d'acteurs sont alors analysées.

Certaines analyses exploratoires se sont appuyées sur des corpus purement textuels. (Diard and Trebucq) se sont intéressés à l'analyse exploratoire des normes internationales d'audit interne. Grâce au logiciels Alceste et Tropes, les auteurs identifient les principales thématiques abordées dans ces normes et déduisent que le réseau des relations entre acteurs de l'audit interne est incomplet. Ils ont ainsi pu recommander une meilleure structuration des textes des normes internationales. D'autres travaux comme ceux de Bonin et Dallo (2003) ont utilisé les logiciels Lexico3 et Hyperbase pour réaliser une étude lexicométrique sur des textes et présentent une étude comparative des fonctionnalités des deux outils pour conclure en leur complémentarité. L'article de Salem et Fleury (2008) donne un ensemble d'analyses exploratoires réalisées grâce au logiciel Lexico : ces analyses portent sur divers domaines et traitent par exemple de la comparaison de corpus (la comparaison de textes produits par des groupes islamistes de

tendance djihadite pour mettre en lumière l'évolution de leur vocabulaire qui traduit l'évolution de l'idéologie et de leurs buts affichés) ou encore de l'analyse de discours (analyse des discours du roi d'Espagne aux forces armées).

Clérot et al. (2004) proposent quant à eux le modèle « monomaniacal », un modèle statistique dans le cadre de l'analyse exploratoire de corpus textuels. Chaque document du corpus est découpé en paragraphes auxquels seront associés un thème unique. La combinaison des thèmes des paragraphes d'un document déterminera un/des thème(s) du document.

Comme nous pouvons le constater, dans le cadre de l'analyse de textes assistée par ordinateur, divers outils et logiciels ont vu le jour. Pour permettre de développer des chaînes de traitements en utilisant plusieurs de ces outils, un format XML-TEI (Text Encoding Initiative) pour échanger les corpus annotés a été proposé (Daoust and Marcoux, 2006).

Notre travail a la particularité de s'appuyer sur un corpus contenant à la fois des données issues de transcriptions d'échanges oraux mais aussi du texte.

3. Traces d'interaction dans le domaine CSCL

3.1. Définition de traces

Une trace est « une marque d'une action quelconque, d'un événement ou d'une situation passée ». Elle peut aussi être comprise comme « une quantité infime de quelque chose ». (Settouti et al., 2006) parlent de systèmes à base de traces modélisées pour les Environnements informatiques pour l'Apprentissage Humain (EIAH). Les auteurs s'intéressent à la notion de trace numérique qu'ils définissent comme la trace de l'activité d'un utilisateur qui utilise un outil informatique pour mener à bien cette activité, s'inscrivant sur un support numérique: « la trace est alors une empreinte restante relatant l'historique et la chronologie de l'interaction de l'apprenant ».

Choquet et Iksal (2007) considèrent comme trace « toute donnée fournissant de l'information sur une section d'apprentissage ».

Alain Mille ¹ parle de traces informatiques comme des éléments laissés dans l'environnement à la suite d'une activité. Ces traces peuvent être laissées volontairement ou non et peuvent être considérées comme des « indices » d'activité par des « observateurs avertis ». De point de vue informatique, une trace est une séquence d'éléments laissés dans l'environnement informatique par l'environnement lui-même ou par l'utilisateur de l'environnement (il faut un environnement logiciel pour garder trace). Cela peut-être des événements, des actions, des annotations associés à des objets informatiques localisés dans le temps et l'espace.

Tous ces sens ont en commun le fait que la trace est toujours liée à un événement situé temporellement (dans le passé), événement qu'elle décrit de façon partielle. Tout travail visant à exploiter des traces cherche donc en quelque sorte à reconstituer/observer/analyser un événement à partir d'une représentation partielle de ce dernier.

3.2. L'exploitation des traces d'interaction en CSCL

L'exploitation des traces d'activité est un des moyens privilégiés pour observer les EIAH, notamment lorsque cet apprentissage est collaboratif. La première étape consiste à collecter efficacement les traces. Elle aboutit à un ensemble de documents de nature et de format

¹ En ligne : <http://fing.tumblr.com/post/54385428/alain-mille-suivre-nos-traces>, consulté le 30/11/09.

divers issus d'outils différents: texte plat ou structuré, image, enregistrement audio ou vidéo et éventuellement des transcriptions de l'audio, etc.

Cette collecte permet ensuite la création d'un corpus de traces. A cette étape, un formatage des traces peut être réalisé et elles peuvent être enrichies par des documents annexes, des annotations, etc. Pour ce qui est du formatage, la nécessité d'aller vers une standardisation de la trace s'impose. Ce formatage devrait agir sur le stockage, sur la traçabilité des traces et sur leur accès ultérieur. Il existe des modèles théoriques qui formalisent ce qu'est une trace et les transformations qu'on peut lui apporter afin de mieux l'exploiter (Settouti et al., 2006). Des outils d'exploitation doivent ensuite être construits pour permettre d'exploiter le corpus de traces. Ces outils fournissent des fonctionnalités telles que la possibilité d'enrichir le corpus de façon manuelle et/ou automatique, effectuer des analyses automatiques ou faciliter des analyses des chercheurs.

Pour ce faire, ces outils permettent :

- de visualiser les traces d'interaction de façon intelligente (permettre aux chercheurs de rejouer une séance de travail/d'apprentissage et de naviguer au sein de la séance en synchronisant l'écrit, les enregistrements vidéo, etc.) ;
- d'interroger les traces d'interaction et d'en extraire une information particulière ;
- de regrouper des éléments de traces, les agréger, d'ajouter des annotations pour faciliter des analyses ultérieures par exemple.

Ces analyses peuvent s'appuyer sur le contenu et/ou sur la structuration du corpus de traces. Elles peuvent également faire intervenir des ressources extérieures (bases de connaissances).

En restant dans le domaine des EIAH, les différentes utilisations des traces selon le type d'utilisateur sont abordées par Settouti et al. (2006)

- Enseignant: les traces sont utilisées pour reconstituer des éléments de perception de l'activité des apprenants ;
- Apprenant: 2 grandes utilisations des traces: la visualisation de ses propres traces dans un format qui fait sens pour l'apprenant (miroir de son activité), ou l'assistance à base de traces (facilitateur de son activité) ;
- Enseignant-concepteur: la trace est un élément d'amélioration continue de la qualité d'un scénario-pédagogique (Ils rejoignent en ce sens Choquet et Iksal, re-ingénierie) ;
- Chercheur: la trace est comme une « matière première permettant d'étudier la caractéristique du processus d'apprentissage, les interactions et aidant à modéliser l'apprenant ».

Ces auteurs classifient alors l'utilisation des traces selon les quatre buts suivants: prise de conscience de l'activité par l'enseignant et l'apprenant (*awareness*), refléter l'activité (*mirroring*), faciliter l'activité (*guiding*), analyser et comprendre une situation d'apprentissage.

Le contexte de notre travail correspond au cas du chercheur (analyser et comprendre une situation d'apprentissage au travers des traces d'interaction).

4. Notre étude

4.1. Cadre de notre étude

Dans le cadre de ce travail, un corpus de traces d'interaction a été recueilli au travers de l'outil TATIANA. Le corpus a été recueilli dans le cadre de séances de présentation des projets informatique aux étudiants d'une école d'ingénieur. Il s'agit de discussions enregistrées (audio/vidéo) entre l'encadrant de projet et les étudiants, discussions au cours desquelles ces derniers peuvent utiliser un outil de prise de notes partagé pour leur prise de notes. Les enregistrements audio ont été transcrits et mis sous format XML, ainsi que les notes prises dans l'éditeur partagé.

4.1.1 Traces textuelles d'interaction

Les caractéristiques de la partie textuelle du corpus (issue de l'éditeur partagé) correspondent aux caractéristiques des textes que l'on peut retrouver sur certains blogs ou certains forums de discussions. En effet, les étudiants effectuent leur prise de note pendant que l'encadrant explique en quoi va consister le projet informatique qu'ils auront à réaliser. On retrouve donc des fautes de frappe (« celleu » pour « cellule », « fonctoin » pour « fonction »), une utilisation du langage de type « SMS » ou abréviations (« surface 2D » pour « surface à deux dimensions », info=informations, pc=ordinateur, prof=professeur, etc.). Toutes ces caractéristiques ne rendent pas aisées une analyse future qui se basera sur ce texte.

4.1.2 Traces transcrites d'interaction

L'encadrant et les élèves sont en situation d'échanges spontanés autour du projet informatique. On retrouve donc dans les traces de leurs échanges oraux (transcrits) des mots du langage parlé (« euh, euhm, hein, bah », etc.), les hésitations caractérisées par des répétitions (« on peut les les les implémenter... » « euhm: quoi d'autre quoi d'autre »). (Dister, 2008) s'est intéressée à l'étiquetage morphosyntaxique des corpus oraux et note que les textes transcrits sont soumis à divers problèmes liés à l'écoute (enregistrement, nombre de locuteurs) ainsi qu'à la transcription elle-même (tout transcrire ? – sons « inaudibles », mots amorcés, conventions de transcription). L'auteur constate également que l'un des problèmes de cet étiquetage est celui de l'unité de traitement dans les corpus oraux. Pour les corpus écrits, les concepts de phrase, de paragraphe sont souvent utilisés mais à l'oral ces concepts sont inopérants. L'auteur s'interroge sur l'unité de traitement à choisir (tour de parole ? silence? pause longue? pause brève? unité intonative). Nous avons rencontré le même type de questionnement avec le corpus que nous avons utilisé. Le problème de l'unité de traitement lors de la recherche d'information dans les traces d'interaction a également été souligné dans (Dyke et al., 2009b).

Comme nous l'avons mentionné précédemment, notre utilisation du corpus de traces correspond dans la classification de (Settouti et al., 2006) au cas du chercheur. Pour comprendre la situation d'apprentissage, ce dernier explore le corpus dans le but de détecter des « phénomènes » particuliers. Il peut s'agir de « phénomènes » d'ordre divers, selon le contenu du corpus et selon les objectifs du chercheur. Par exemple, l'on pourra s'intéresser aux argumentations/contre-argumentations dans un corpus de réunions de travail, à l'émergence de leadership dans les discussions entre apprenants, ou encore à la détection des prises de décision au cours d'une réunion. Dans le cadre de notre exploration, nous nous sommes intéressés au phénomène de reformulation multimodale des énoncés dans les traces d'interaction.

4.2. Pourquoi s'intéresser à la reformulation multimodale des énoncés au cours de l'échange enseignant/étudiant

Nous nous intéressons particulièrement au phénomène de reformulation au cours de l'interaction entre encadrant et étudiants. La reformulation telle que recherchée ici consiste à une redite d'un énoncé qui a précédemment été formulé. Ce type de redite peut intervenir de d'un mode/média d'expression vers un autre (exemple de l'oral, audio/transcription, vers l'écrit, texte/éditeur, et permettre diverses analyses).

4.2.1. De l'encadrant vers les étudiants

Ce type de reformulation peut permettre de savoir si les concepts clés du projet ont été assimilés par les étudiants. Ce type de reformulation pourra être observé entre les paroles de l'enseignant

et les prises de notes des étudiants, mais également lors des prise de paroles par les étudiants (questions à l'encadrant, par exemple). Ce pourra également être pertinent si l'on s'intéresse à de l'enseignement vicariant, par exemple l'on pourra analyser si les explications que l'enseignant donne suite à une question d'un élève X déclenche une certaine prise de note chez un autre élève Y.

4.2.2. Des étudiants vers l'encadrant

Ce type de reformulation peut mettre en exergue des explications que l'enseignant donne en plus selon ce que les étudiants écrivent dans leur prise de notes par exemple, ou alors suite à une question d'un étudiant, l'encadrant va-t-il réutiliser des termes de l'étudiant ou corriger par exemple un concept pas bien compris par l'étudiant ?

De plus, ce corpus comporte des enregistrements vidéos de l'échange encadrant/étudiants. Une fois ces reformulations localisées, en utilisant ces enregistrements vidéo, l'on pourra synchroniser les points de reformulation textuels/oraux avec des signes visuels pour étayer l'analyse. Par exemple, un hochement de tête de l'étudiant Y montrant son accord pour certaines notions à partir des explications données par l'encadrant suite à une question venant de l'étudiant X.

4.2.3. La non-reformulation

La non-reformulation peut également être intéressante à analyser. En effet, il peut s'agir par exemple de concepts importants énoncés par l'encadrant et qui ne sont jamais repris/reformulés (ni à l'écrit ni à l'oral) par les étudiants. L'on peut alors se demander si ces derniers n'en n'ont pas cerné l'importance ou si tout simplement ces concepts étaient déjà connus et maîtrisés par les étudiants.

4.3. Ce qu'on entend par reformulation (dans notre contexte)

La reformulation peut donc consister à expliquer un énoncé précédemment mentionné (oralement or par écrit), à répéter un énoncé précédemment mentionné, à utiliser d'autres mots pour dire une chose déjà dite. Sur le plan linguistique, il existe des marqueurs permettant de signaler différentes façons de reformuler : utilisation des deux points, des parenthèses, de l'apposition et de la virgule, de formules comme *c'est-à-dire...* etc. Ainsi, divers types de reformulations sont possibles : par utilisation de synonymes des mots de l'expression initiale, par utilisation des généralisations/spécialisations des mots de l'expression initiale, par utilisation de termes ayant un lien/relation particulière avec des mots de l'expression initiale, par ajout ou suppression de nouveaux termes (adverbes, adjectifs, par exemple), par l'utilisation des abréviations, par utilisation de la négation d'un terme (notamment un verbe, un adjectif, un adverbe), par exemple : « *le sac n'est pas propre* » peut être reformulé « *le sac est sale* ».

Dans le cadre de cette étude, nous nous sommes intéressés à la reformulation en termes de redite d'un énoncé déjà exprimé. Les types de reformulations ci-dessus font partie des perspectives de ce travail.

5. Expérimentations

5.1. Le corpus et son découpage, l'outil d'analyse

Chaque corpus correspond à une séance de discussion entre un encadrant et un binôme d'étudiants sur un projet en informatique. Chaque corpus a été découpé manuellement en une

sorte de « tour d'intervention » par intervenant (encadrant ou étudiant). Un tour d'intervention est soit un tour de parole par l'encadrant ou par un étudiant ou un tour d'écriture dans l'éditeur partagé de prises de notes par un des étudiants.

Un extrait du corpus préparé pour l'analyse sous Lexico3 :

<num=23><user=Professeur><tool=transcript>ok alors euhm ce projet d'une: ce projet du jeu de la vie euh ce quelque chose de très euhm: \$...
 <num=35><user=françois><tool=tchat>utilisation scientifique et artistique \$...
 <num=289><user=EtudiantD><tool=transcript>aimh: c'est vrai qu'au début 'fin on n'avait pas: on n'avait pas pensé que: par exemple la surface euh \$

Le logiciel Lexico3 a été utilisé pour faciliter notre analyse. Il permet dans un premier temps d'obtenir le dictionnaire associé à notre corpus, avec pour chaque terme sa fréquence. Ce logiciel permet également de localiser les diverses apparitions de chaque terme au sein du corpus, ce qui permet d'avoir le contexte (textuel) autour du terme. La fonctionnalité de détermination des segments répétés est intéressante : elle permet de détecter toutes les expressions qui sont répétées dans le corpus et de localiser le contexte d'apparition de ces expressions. Selon la structuration du corpus, on peut également visualiser le contexte d'apparition de ces expressions par intervenant, par « paragraphe » (un paragraphe correspond à un tour d'intervention dans notre cas). Pour chaque tour d'intervention, l'outil correspondant est associé (transcription, texte issu du tableau partagé). Ainsi l'on peut visualiser les segments répétés par intervenant (ce qui permet de détecter par exemple les reformulations des expressions de l'encadrant par les étudiants) mais également par outil (ce qui permet de détecter par exemple les reformulations des paroles, transcrites, de l'encadrant par les étudiants dans leur prise de notes).

5.2. Résultats

Nous avons choisi de générer tous les segments répétés d'une longueur au moins égale à 2 (2 termes au moins dans le segment) et avec une fréquence au moins égale à 2 (le segment est répété au moins une fois).

Ainsi, l'on constate que l'écriture de certaines expressions dans l'éditeur de prise de note par les étudiants apparaît « juste » après que l'encadrant ait prononcé ces expressions oralement. Ainsi les « expressions-clés » du projet sont détectés par les étudiants et mis sous forme de notes. L'on constate également que les étudiants réutilisent des expressions/termes de l'encadrant au moment de poser des questions à propos du projet, par exemple pour les expressions « interface graphique », « scientifique et artistique », « machine de Turing ».

5.2.1. 1^{er} groupe(B)

Pour le 1^{er} corpus, les données transcrites et les données écrites sont mises ensemble et classées suivant leur ordre d'apparition (chronologie) lors de la séance d'apprentissage.

Statistique du 1^{er} corpus (groupeB) : Principales caractéristiques lexicométriques

| Nombre des occurrences | Nombre des formes (mots différents) | Fréquence maximale | Nombre des hapax (mot apparaissant une seule fois) |
|------------------------|-------------------------------------|--------------------|--|
| 5552 | 991 | 236 | 510 |

Cas de l'expression « scientifique et artistique » : reformulation encadrant vers étudiants, oral vers écrit :

<num=32><user=Professeur><tool=transcript>et puis euhm : il y a un travail qui est à la fois je dirais scientifique et artistique
 <num=35><user=francois><tool=tchat>utilisation scientifique et artistique \$

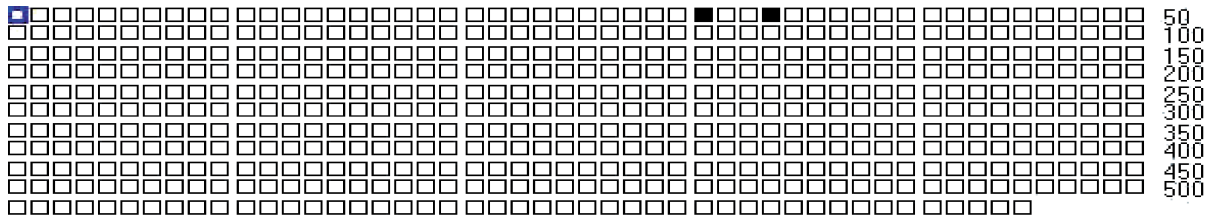


Figure 1 : Carte des paragraphes contenant l'expression « scientifique et artistique ». Un paragraphe correspond à un tour d'intervention dans notre cas

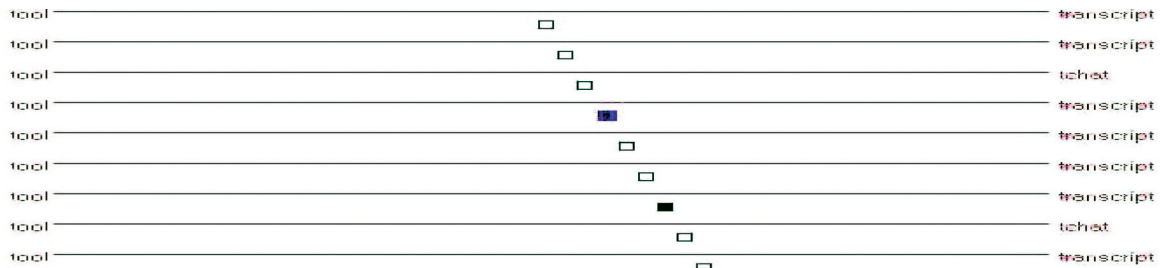


Figure 2 : Apparitions de l'expression « scientifique et artistique » selon les outils

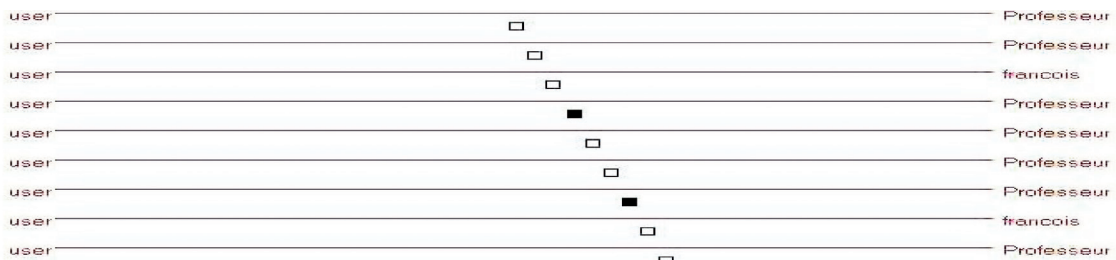


Figure 3 : Apparitions de l'expression « scientifique et artistique » selon les intervenants

En recoupant les trois graphiques, nous constatons que l'étudiant effectue une prise de note de l'expression « scientifique et artistique » juste après que l'encadrant ne l'ait utilisé oralement. Il en est de même pour d'autres expressions importantes du projet informatique dont il est question : *interface graphique, machine de Turing, etc.*

<num=38><user=Professeur><tool=transcript>démontrer que ce qu'on sait faire avec le jeu de la vie c'est euh c'est similaire à ce qu'on fait faire avec une **Machine de Turing** etc \$
 <num=41><user=Armand><tool=tchat>(**machine de turing** ?) \$

Reformulations couplées, non détectées automatiquement : Pour d'autres expressions, Lexico3 (comme tout autre logiciel d'analyse textuelle) ne permet pas de détecter la reformulation car il n'y a pas de segments répétés (les étudiants utilisent un langage non naturel ou alors changent la

forme du mot initial, ou encore couplent plusieurs déclarations de l'encadrant et reformulent en une seule expression). C'est le cas par exemple de l'expression « *surface en deux dimensions* » dite oralement par l'encadrant et noté par les étudiants « *surface 2D* »

<num=60><user=Professeur><tool=transcript>qu'il existe une **surface à deux dimensions** euh:: euh:: ortho orthonormées \$

<num=61><user=Professeur><tool=transcript>dans lesquelles avec des cases dans lesquelles on peut placer des cellules \$

<num=62><user=Professeur><tool=transcript>on appelle ça des **cellules** c'est des objets euh euh: peut importe \$

Prise de note de l'étudiant :

<num=63><user=francois><tool=tchat>**surface 2D -> cellules** \$

Ou encore les phrases suivantes de l'encadrant

<num=66><user=Professeur><tool=transcript>euh: chaque étape d'euh du jeu est une étape y a une génération \$

<num=67><user=Professeur><tool=transcript>et euh une cellule peut **survivre** au cours d'une étape ou elle peut **disparaître** ou il peut **naître** une nouvelle cellule \$

Qui ont été « reformulées » par l'étudiant dans l'éditeur de prise de note par

<num=68><user=francois><tool=tchat>**survie/disparition/naissance** \$

La carte des paragraphes fournie par Lexico3 nous montre la chronologie entre la parole de l'encadrant et la prise de note des étudiants, qui nous laisse penser que c'est bien une reformulation.

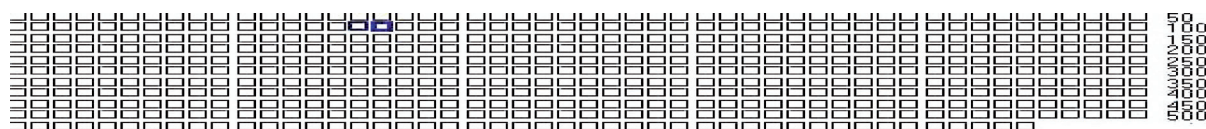


Figure 4 : Carte des paragraphes pour « survie/disparition/naissance »

Certaines reformulations n'ont également pas pu être détectées « automatiquement » à cause de faute de frappe faites par exemple par les étudiants lors de la prise de notes ou d'une refonte totale de la formulation initiale. Par exemple, l'intervention de l'enseignant :

<num=140><user=Professeur><tool=transcript>mais en fait **on va représenter que les cellules utilisées** \$

<num=141><user=Professeur><tool=transcript>que les que les cas utilisés par des organismes \$

Et la prise de note des étudiants :

<num=146><user=Armand><tool=tchat>**on peut ne s'occuper que des celleu** \$ <num=152><user=Armand><tool=tchat>**occupée et celles adjacente,** \$

La non-reformulation de certaines expressions : Il est question d'un projet sur « *le jeu de la vie* » et à aucun moment les étudiants ne reprennent cette expression qui est le concept central de leur futur projet et qui est pourtant maintes fois utilisée par l'encadrant comme le montre la figure suivante.

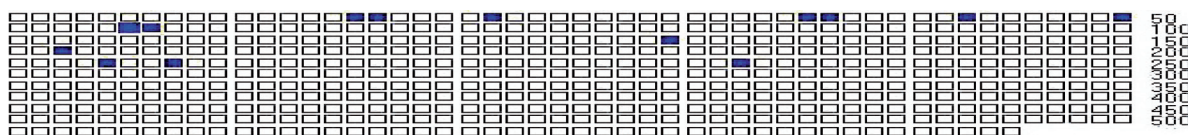


Figure 5 : Carte de « paragraphes » montrant les utilisations de l'expression « jeu de la vie » (dans certains cas jeux de la vie), essentiellement par l'encadrant

En analysant de près, l'on se rend compte qu'au tout début de l'entretien, l'encadrant a posé la question concernant le *jeu de la vie* aux étudiants et ceux-ci ont répondu avoir fait des recherches sur Internet sur le sujet avant. L'on peut donc penser qu'ils ne l'ont pas mentionné dans leur prise de notes car c'est donc un concept connu pour eux.

5.2.2. 2nd groupe (C)

Principales caractéristiques lexicométriques

| Nombre des occurrences | Nombre des formes | Fréquence maximale | Nombre des hapax |
|------------------------|-------------------|--------------------|------------------|
| 4813 | 922 | 244 | 504 |

Pour les analyses du second groupe, la partie prise de notes des étudiants n'a pas été manuellement découpée en tours d'interventions et est représentée par un seul bloc-paragraphe (le dernier paragraphe de la carte des paragraphes). Ceci ne nous permet malheureusement pas de retracer la chronologie entre la prise de notes et l'oral comme dans l'analyse précédente.

Reformulation encadrant vers étudiants, oral vers écrit :

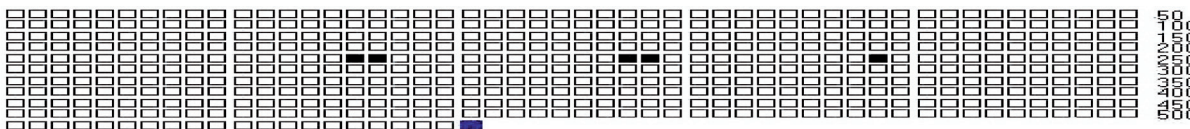


Figure 6 : Carte de paragraphes avec les apparitions de « machine de turing »

Cette expression est utilisée à plusieurs reprises par l'enseignant, les étudiants la reprennent dans leur prise de notes. Les explications de l'enseignant ne les ont visiblement pas satisfaits puisqu'ils notent aussi de regarder sur wiki (encyclopédie Wikipedia).

<num=522><user=etudiantd><tool=redaction> (machine de turing ? ? regarder sur wiki)

Reformulation et ajout, oral vers écrit :

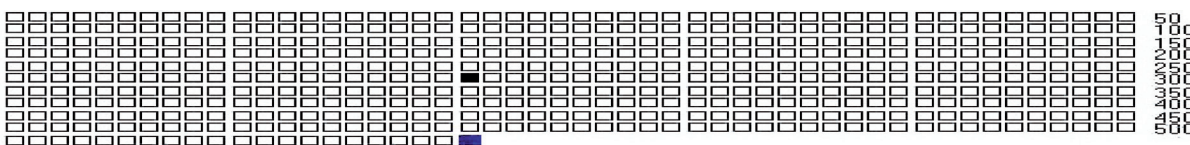


Figure 7 : Carte de paragraphes pour l'expression « passer du temps sur »

<num=272><user=professeur><tool=transcript>passer du temps sur des des representation intelligentes des des structures \$

....Suite de ce que dit le professeur

<num=273><user=professeur><tool=transcript>pour quelles occupent peu de place en memoire donc vous pour pouvoir représenter les grosses structures \$

<num=274><user=professeur><tool=transcript>passer plus du temps sur l'interface graphique puisque ça aussi est tout a fait legitime d'avoir le \$

<num=522><user=etudiantd><tool=redaction>... j ' ai aussi le cormen **passer du temps sur : structure de donnee , graphisme reflechir sur les structures** systeme d ' accompte de point

Cette expression nous a intéressés car apriori elle va nous aiguiller vers les concepts importants sur lesquels les étudiants devront passer du temps, donc sur le cœur de leur projet. Elle est prononcée par l'enseignant et reprise par les étudiants dans leur prises de notes. En analysant ce qui est dit/écrit autour de cette expression, l'on constate que les étudiants font une sorte de résumé des divers points mentionnés (même à divers moments) par l'enseignant comme étant importants.

Reformulation oral vers oral :

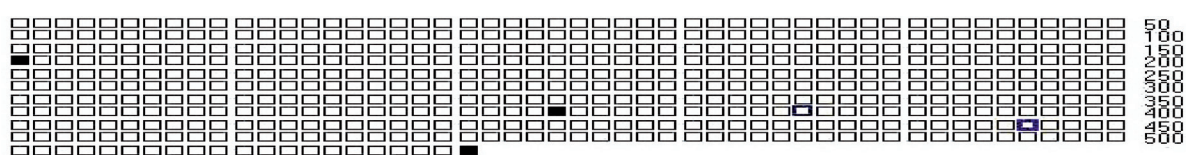


Figure 8 : Carte des paragraphes avec les apparitions de « structure de donnée » ou « structures de données »

Pour cette expression, une reformulation oral-oral a lieu puisque l'expression est tout d'abord utilisée par l'enseignant, elle est ensuite reprise par élève qui l'utilise pour poser une question et l'élève l'intègre également dans sa prise de note.

Reformulation de l'étudiant vers l'encadrant (oral vers oral et oral vers écrit) :

Pour cette expression, une reformulation de l'étudiant vers l'enseignant a lieu. L'étudiant explique au tout début ce qu'il a compris du projet et utilise cette expression trouvée sur Wikipedia. L'enseignant va la reprendre pour appuyer l'idée exprimée par l'étudiant. Et dans la prise de notes, l'étudiant va écrire cette expression aussitôt. L'enseignant la re-utilisera à d'autres reprises dans ses explications.

6. Conclusions

Le logiciel Lexico3, par sa fonctionnalité de retrait des segments répétés d'un corpus, nous a aiguillé vers bon nombre de reformulations potentielles et les différents modes de visualisation (par intervenant, par outil utilisé, par tour d'intervention pour la chronologie) nous ont permis de valider ces reformulations. Par exemple, nous avons pu (manuellement) focaliser sur les interventions étudiant succédant à une intervention de l'encadrant pour en déduire une éventuelle reformulation et pour analyser la compréhension du projet par les étudiants. De façon automatique, il est toutefois délicat de détecter toutes les reformulations. Des prétraitements comme la lemmatisation du corpus permettront de retrouver les segments répétés contenant des mots de même racine et aiguilleront automatiquement vers des reformulations. L'utilisation de ressources externes de type ontologie (du domaine du corpus des traces par exemple) ou dictionnaires propres à un corpus permettrait, grâce aux relations entre mots/concepts de détecter des reformulations plus élaborées.

De façon parallèle à cette étude, à travers l'outil Tatiana, des chercheurs se sont intéressés à la reformulation. Une fois les tours d'interventions créés manuellement, ces chercheurs les catégorisent

manuellement en scrutant le corpus. Ils lient ensuite manuellement les tours d'intervention de même catégorie (reformulation), selon la chronologie et selon le contexte d'apparition de ces tours d'intervention. Une visualisation automatique est alors générée par l'outil et ils peuvent ainsi analyser les types de reformulations. Les fonctionnalités offertes par Lexico3 (détection de segments répétés, visualisation sous divers angles du corpus) peuvent aiguiller plus aisément vers les points de reformulation, qui seront validés ou non et ensuite catégorisés.

Références

- Bonin E. and Dallo A. (2003). Hyperbase et Lexico 3, outils lexicométriques pour l'historien. *Histoire et Mesure*, vol. (XVII)N°3/4 : 389-402, mis en ligne le 03 avril 2007. URL : <http://histoiremesure.revues.org/index840.html>. Consulté le 27 juillet 2009.
- Cailliau F. and Poudat C. (2008). Caractérisation lexicale des contributions clients agents dans un corpus de conversations téléphoniques retranscrites. In *JADT2008*.
- Choquet C. and Iksal S. (2007). Modélisation et construction de traces d'utilisation d'une activité d'apprentissage : une approche langage pour la réingénierie d'un EIAH. In *Sticef*, vol.14. N° spécial *Analyse des traces d'utilisation dans les EIAH*.
- Clérot F., Collin O., Cappé O. and Moulines E. (2004). Le modèle monomaniaque : un modèle statistique simple pour l'analyse exploratoire d'un corpus de textes. In *CIFT2004*, pp 45-57.
- Daoust F. and Marcoux Y. (2006). Logiciels d'analyse textuelles : vers un format XML-TEI pour l'échange de corpus annotés. In *Actes des JADT2006*.
- Daoust F., Dobrowolski G., Dufresne M. and Gélinas-Chebat C. (2006). Analyse exploratoire d'entrevues de groupe : quand Alceste, DTM, Lexico et SATO se donnent la main. In *JADT2006*.
- Diard C. et Trebucq S. Analyse exploratoire des normes internationales d'audit interne. En ligne. http://cermat.iae.univ-tours.fr/IMG/pdf/Communication_Diard_-_Trebucq.pdf.
- Dister A. (2008). Etiquetage morphosyntaxique des corpus oraux. In *Journée Rhapsodie-Paris X*, 24 janvier.
- Dyke G., Lund C. and Girardot J.J. (2009a). Un outil flexible pour l'analyse de traces. Gestion, synchronisation, visualisation, analyse et partage de corpus d'interactions médiatisées par ordinateur avec Tatiana. In *ICT 2009*.
- Dyke G., Beigbeder M., Lund K. and Girardot J. (2009b). Les Traces d'interactions humaines, un nouveau domaine d'application pour la RI. In *Actes de la conférence CORIA09*.
- Gélinas-Chebat C., Daoust F., Dufresne M., Gallopel K. and Lebel M.E. (2004). Analyse exploratoire d'entrevues de groupe : les jeunes français et le tabac. In *JADT2004*.
- Peyrat-Guillard D. (2006). Alceste et WordMapper : l'apport complémentaire de deux logiciels pour analyser un même corpus d'entretiens. In *JADT2006*.
- Salem A. and Fleury S. (2008). Explorations textométriques. In *JADT2008*.
- Settouti L.S., Prié Y., Mille A. and Marty J.C. (2006). Système à base de traces pour l'apprentissage humain. In *LREC 2006*.
- Logiciel Sato: URL : <http://www.ling.uqam.ca/ato/visibilite.html>.
- Logiciel Alceste : URL : http://www.image-zafar.com/index_alceste.htm.
- Logiciel Lexico3 : URL : <http://www.cavi.univ-paris3.fr/Ilpga/ilpga/tal/lexicoWWW/>.
- Logiciel WordMapper : URL : <http://www.grimmersoft.com/>.
- Logiciel DTM :Data and Text Mining: URL: <http://ses.telecom-paristech.fr/lebart/>.
- Logiciel Tropes: URL: <http://www.acetic.fr/tropesfr.htm>.