Outils pour l'analyse des troubles du langage et de la pensée

Projet Tuteuré de Master 1 Sciences de la Cognition et Applications

Marine RUEZ et Valentin STEYER
Année 2014/2015

Encadrants : Maxime AMBLARD, Manuel REBUSCHI

Outils pour l'analyse des troubles du langage et de la pensée

Projet Tuteuré de Master 1 Sciences de la Cognition et Applications

Marine RUEZ et Valentin STEYER

Encadrants : Maxime AMBLARD, Manuel REBUSCHI

2014-2015

Remerciements

Nous tenons à remercier particulièrement Mr. Maxime Amblard ainsi que Mr. Manuel Rebuschi pour nous avoir guidés tout au long de ce projet.

Nous voulons également remercier Mr. Michel Musiol de nous avoir mis en contact avec une classe de Master en psychologie pour faire passer nos tests.

Enfin, nous remercions toutes les personnes volontaires qui ont bien voulu réaliser l'annotation d'un extrait de texte.

Sommaire

Introduction

Dans le cadre de notre Master 1 en Sciences de la Cognition et Applications, nous avons réalisé un projet tuteuré d'une durée de 5 mois, au sein de l'équipe du projet SLAM (Schizophrénie et Langage : Analyse et Modélisation) à la MSH Lorraine.

Plusieurs personnes travaillent sur ce projet, comme Maxime Amblard, Michel Musiol et Manuel Rebuschi qui ont été nos encadrants.

Leur objectif est d'analyser le discours chez les patients atteints de schizophrénie. En effet, "le délire des schizophrènes est repéré en psychiatrie comme l'une des formes de pensée les plus radicalement déviantes. Il est d'ailleurs appréhendé le plus souvent sur la base de l'analyse des productions verbales et du discours des patients."[1]. On peut alors repérer des "ruptures" conversationnelles, qui sont typiques de cette pathologie.

Notre mission est de faire repérer ces ruptures à d'autres individus "lambda" en menant une campagne d'annotations.

[1] S*chizophrénie, logicité et perspective en première personne.* Manuel Rebuschi, Maxime Amblard, Michel Musiol. L'évolution psychiatrique n°78, pages 127-141 ; 2013.

I) Présentation du sujet et objectifs de notre travail

A) Projet SLAM

Le projet SLAM (Schizophrénie et Langage : Analyse et Modélisation) vise à rendre automatique l'étude des conversations pathologiques dans le cadre d'une approche interdisciplinaire alliant psychologie, linguistique informatique et philosophie. Les analyses sont réalisées à partir d'un corpus de textes. Il s'agit de retranscriptions de conversations réelles entre des patients schizophrènes et des psychologues venant du CHU Vinatier (Lyon) et du CHU Saint Antoine (Paris).

L'objectif de ces travaux est donc de repérer les "ruptures" conversationnelles dans ces échanges. La représentation graphique, sous forme d'un arbre, est issue de la S-DRT (Segmented Discourse Representation Theory) et permet de rendre compte de ces discontinuités.

Afin de bien comprendre le projet, la première phase de notre travail a été celle de la lecture. Il s'agissait de lire différents documents issus des travaux de l'équipe pour comprendre dans quel projet on intervenait, ainsi que des exemples de textes et d'arbres de type S-DRT. Même si nous n'allions pas réaliser des arbres, il fallait comprendre le principe afin d'être capable d'expliquer le plus clairement possible la façon d'annoter un texte, et générer un arbre ressemblant à un schéma S-DRT.

B) S-DRT

La S-DRT est une théorie née dans les années 80, qui s'est développée depuis une quinzaine d'années. Il s'agit d'une extension de la DRT (Discourse Representation Theory).
"L'avantage de cette théorie est qu'elle se base sur le calcul de la sémantique compositionnelle et qu'elle tente de rendre compte d'une pragmatique de manière simplifiée"[2]. Une conversation est interprétée par une double construction : celle d'un arbre hiérarchique reliant les actes, les relations pragmatiques, et celle représentant le contenu sémantique des actes.

[2] Fiche de présentation du projet tuteuré, Annexe 1.

C'est à dire, plus simplement, que cet arbre représente les différentes sections de phrases (des unités de sens) reliées entre elles par différents types de relations.

1) Le thème

Le contenu sémantique est en fait le thème. Le thème est l'idée générale de l'idée, la thématique de la conversation en cours. Il est modifié au fur et à mesure de la conversation. Celui-ci joue un rôle très important dans les discontinuités puisqu'il est le lieu où se joue la cohérence sémantique du locuteur schizophrène. "Si le patient vient à changer de thème de manière intempestive"[3], il y a alors une incohérence, une discontinuité.

Exemple d'une conversation entre deux individus, Pe1 et Pe2 :

Pe1 : Hier j'ai mangé un délicieux repas.
Pe2 : Ah oui ? Quoi donc ?
Pe1: Je suis heureuse qu'il y ait du soleil.

Il s'agit d'une conversation qui porte sur le repas de la veille. La conversation a comme thème "le repas". Mais Pe1 ne répond pas à la question de Pe2 dans sa deuxième intervention. Il évoque une histoire, une idée sans lien avec la discussion actuelle. Ainsi, il y a changement de thème et une discontinuité.

2) Relations rhétoriques

L'autre dimension de la S-DRT est la structuration pragmatique de la conversation, c'est à dire la forme. On compte trois types de relations :

- subordonnantes, représentées schématiquement de manière verticale;
- coordonnantes, représentées horizontalement;
- requérants, représentées obliquement.

Les relations requérantes sont générées, par exemple, par les questions qui requièrent une réponse. Les relations horizontales représentent une juxtaposition linéaire, des étapes au fur et à mesure, c'est le cas d'une narration ou d'une suite par exemple. Le type "relations

[3] *L'interaction conversationnelle à l'épreuve du handicap schizophrénique* - Maxime Amblard, Michel Musiol, Manuel Rebuschi. Recherches sur la philosophie et le langage, 2014, 31, pp.1-21.

coordonnantes" contient par exemple les relations d'élaboration, et représentent le contenu de ce qu'elles élaborent.

Voici un tableau exposant tous les types de relations :

Relations horizontales	**Relations verticales**	**Relations obliques**
Narration	Elaboration	Question
Réponse	Elaboration : explication	Question : conduite
Réponse phatique	Elaboration : prescription	Méta-question
Suite	Evaluation	Demande d'élaboration
	Phatique	Conduite
		Contre-élaboration

L'explication de chaque relation est expliquée dans le guide d'annotations dans l'annexe 2.

3) Unités de sens

Les relations joignent entre elles des unités de sens. Il s'agit d'interventions, ou d'une partie de phrase au sein d'une même intervention. Graphiquement, chaque intervention est représentée par une boite.

4) Exemple

Cet exemple est repris de "L'interaction conversationnelle à l'épreuve du handicap schizophrénique"[4] par Maxime Amblard, Michel Musiol et Manuel Rebuschi à la page 5.

[A_1] Alice : Hier soir je suis sortie.
[B_2] Benoit : Tu es allée diner?
[A_3] Alice : Oui.
[B_4] Benoit : Et alors ?

[4] *L'interaction conversationnelle à l'épreuve du handicap schizophrénique* - Maxime Amblard, Michel Musiol, Manuel Rebuschi. Recherches sur la philosophie et le langage, 2014, 31, p 5.

[A_5] Alice : [D'abord j'ai pris un poisson,][1] [puis une viande][2] [et enfin plein de fromage.][3] [C'était fantastique.][4]

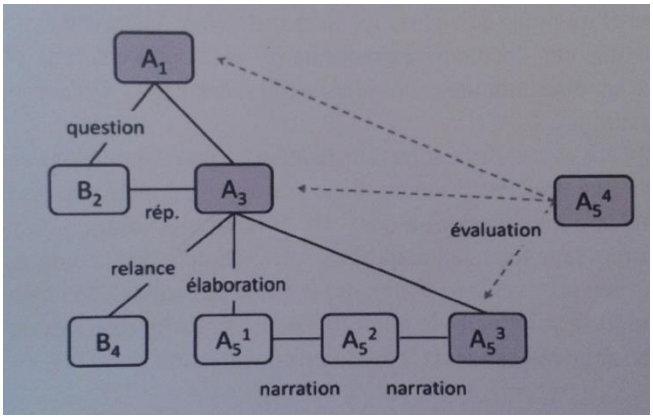


Figure 1 : Arbre de type S-DRT pour le dialogue précédemment cité

Dans ce dialogue, chaque intervention introduit des unités de sens, reliées entre elles par des relations rhétoriques.

II) La partie "technique"

A) Prise en main du logiciel Glozz

Glozz est le logiciel que les annotateurs ont dû utiliser pour analyser le texte. Il était donc indispensable de bien comprendre son fonctionnement pour pouvoir le présenter à d'autres personnes. Bien sûr, toutes les fonctionnalités du logiciel n'ont pas été utilisées, nous nous sommes donc consacrés à que ce qui allait être utilisé.

De plus, il a fallu créer des améliorations et adapter la plate-forme d'annotation. En effet, une fois le texte annoté, le logiciel Glozz ne permet pas d'avoir une visualisation d'un arbre sous forme S-DRT (On a un simple graphe avec les différentes relations et unités). Ainsi, un programme informatique réalisant cette tâche permettait de faciliter les réflexions de l'annotateur. Pour nous, il permettait de vérifier si les textes annotés des personnes sont correctement faits et analyser et comparer plus facilement tous les arbres produits par les annotateurs.

1) Qu'est-ce que Glozz ?

Glozz est un logiciel réalisé en Java. C'est une plate-forme d'annotation manuelle et d'exploration de corpus textuels. Ce logiciel a été conçu dans le cadre du projet ANR Annodis[5], et se perpétue au sein du laboratoire GREYC[6].
Glozz est utilisé en France dans le cadre de plusieurs projets d'annotation dont celui du projet SLAM. Il présente plusieurs fonctionnalités dont, par exemple, annoter manuellement des textes, importer et exporter des fichiers ...

2) Exploitation de Glozz

Le travail étant trop lourd à faire à la main, il était demandé de le réaliser sur le logiciel Glozz. Ce logiciel correspondait à notre travail, car le but de l'analyse et de l'arbre final est d'obtenir des "bouts de phrases" reliés par différentes relations. Ces bouts représentent les unités pour le logiciel, et les liaisons entre celles-ci sont des relations. Il fallait rester simple pour ne pas compliquer le travail à l'annotateur, qui était déjà assez dur à la base. De plus, il aurait été très dur d'analyser les résultats s'il fallait comparer les arbres un à un en format papier.

a) Unité

Il nous suffisait d'un seul type d'unité car nous avions qu'une seule sorte de "portion de phrase". Nous l'avons appelé "Segment". D'autres types d'unités ont été utilisés pour le pré traitement des fichiers.

b) Thème

Il est possible d'ajouter des caractéristiques pour chaque type d'unité. Ainsi, pour "Segment", il est possible d'ajouter une caractéristique, c'est à dire pour nous un thème, qui correspond à la thématique de la phrase, de l'idée. Un thème peut être utilisé plusieurs fois, mais peut aussi changer.

---

[5] ANNODIS est un projet multidisciplinaire (linguistique, logique, TAL) né de la collaboration de trois laboratoires français, CLLE-ERSS, IRIT et GREYC . Ce projet financé par L'Agence National pour la Recherche (ANR) a démarré en Décembre 2007.(site internet ERSS : http://w3.erss.univ-tlse2.fr/)

[6] Institut de recherche dans l'informatique situé en Basse-Normandie.

c) Relation

Pour les relations, nous avons inséré tous les types de relations comme vu précédemment dans le tableau. Nous avons décidé de réunir tous les types d'élaboration (Elaboration, Elaboration : explication, Elaboration : prescription, et Evaluation) car il s'agit toutes d'élaborations, mais avec une caractéristique différente. Nous avons donc laissé un type "Elaboration", mais en faisant, comme pour l'unité et son thème, une caractéristique pour choisir le type d'élaboration.

3) L'interface de Glozz

Voici ci-dessous l'interface du logiciel Glozz :

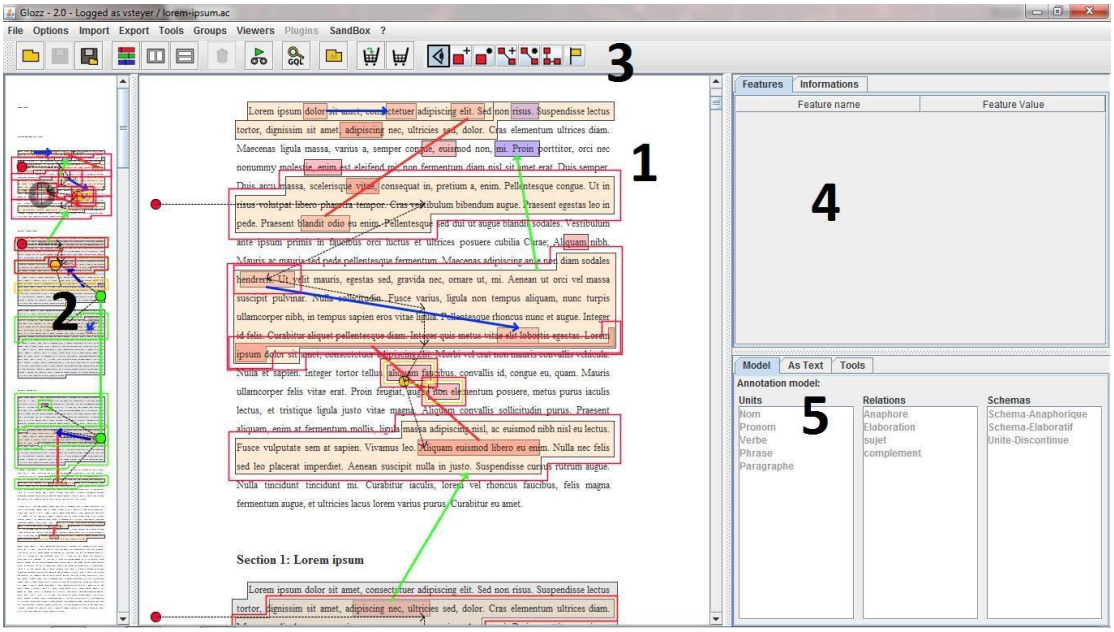


Figure 2 : Interface du logiciel Glozz

1 : C'est la vue principale du texte, c'est sur cette vue qu'on pourra ajouter/modifier toutes les annotations.

2 : La vue en Macro, c'est le même texte annoté que dans la vue n°1, mais le texte est en mode "macro", c'est à dire qu'il y a une vue globale du texte et de ses annotations, elle est utilisée pour naviguer plus facilement.

3 : C'est la barre des menus. Le groupe de boutons le plus à droite sera à utiliser pour ajouter ou modifier des annotations. Ce sera les seuls boutons à utiliser.

4 : C'est un tableau contenant les différentes valeurs pour une annotation sélectionnée.

5 : C'est le modèle d'annotation, c'est ici qu'on voit la liste de tous les types d'annotations disponibles (unités, relations, ainsi que les schémas)

Dans l'annexe n°2, qui est le guide d'annotation, vous retrouverez plus d'informations sur le logiciel.

4) Les différents types de fichier de Glozz

Le logiciel Glozz possède plusieurs types de fichiers (et certains propres à lui-même) que nous avons dû utiliser, soit pour les importer dans le logiciel, soit pour les traiter à la fin de l'annotation.

- Le format .txt, qui est un simple document avec le texte que l'on souhaite annoter.

- Le format .aa, qui respecte une architecture XML, est utilisé par Glozz pour enregistrer toutes les relations ainsi que les différentes unités. On retrouve des informations comme la position de début et de fin des unités ainsi que leurs caractéristiques (s'il y en a), ainsi que le type de l'unité ou de la relation. C'est ce fichier que nous avons dû utiliser pour présélectionner le texte, ainsi que pour traiter les résultats à la fin.
  Voici à quoi ressemble une partie d'un fichier au format .aa :

```
<unit id="vsteyer_1427327664610">
    <metadata>
        <author>vsteyer</author>
        <creation-date>1427327664610</creation-date>
        <lastModifier>vsteyer</lastModifier>
        <lastModificationDate>1427327905280</lastModificationDate>
    </metadata>
    <characterisation>
        <type>Segment</type>
        <featureSet>
            <feature name="Theme">mort symbolique</feature>
        </featureSet>
    </characterisation>
    <positioning>
        <start>
            <singlePosition index="676"/>
        </start>
        <end>
            <singlePosition index="726"/>
        </end>
    </positioning>
</unit>
```

Figure 3 : Exemple de description d'une unité dans le fichier .aa

- Le format .ac, qui est aussi un fichier texte qui contient le texte que l'on souhaite annoter, mais sous une forme plus compressée, c'est à dire sans sauts de ligne. C'est ce fichier qui est utilisé pour indiquer la position de début et de fin d'une unité. 1 correspondant au premier caractère du fichier, et ainsi de suite.

- Le format .aam, qui respecte aussi une architecture XML, est un modèle d'annotations nécessaire pour Glozz. C'est dans ce fichier qu'on va marquer tous les types qu'une unité peut avoir (Ici nous n'avons que "Segment"), mais aussi tous les modèles des relations (Narration, Question, Réponse…).

```
<?xml version="1.0" encoding="UTF-8" ?>

<annotationModel>

    <units>
        <type name="Segment">
            <featureSet>
                <feature name="Theme">
                    <value type="free" default="" />
                </feature>
            </featureSet>
        </type>

    </units>

    <relations>

        <type name="Narration" oriented="true">
        </type>

        <type name="Réponse" oriented="true">
        </type>
```

Figure 4 : Exemple d'une partie de description d'une unité et de deux relations dans le fichier .aam

- Le format .as, qui est aussi sous une structure XML, est juste un modèle de style pour Glozz, c'est à dire un fichier qui permet de sauvegarder les couleurs données aux différents types d'unités et des relations (Au format RGB).

```
<unit-style>
    <type name="Racine"/>
    <background-color b="102" g="102" r="0"/>
    <invisibility value="false"/>
</unit-style>
<unit-style>
    <type name="Segment"/>
    <background-color b="204" g="204" r="0"/>
    <invisibility value="false"/>
</unit-style>
<relation-style>
    <type name="Conduite"/>
    <line-color b="255" g="51" r="51"/>
    <background-color b="0" g="0" r="255"/>
    <invisibility value="false"/>
</relation-style>
```

Figure 5 : Exemple d'une partie de description de deux unités et une relation dans le fichier .as

Il y a un autre type de fichier qui est utilisé pour Glozz (.gql), mais nous n'avons pas eu besoin de l'utiliser.

B) Le prétraitement des fichiers XML

Après avoir utilisé le logiciel Glozz, nous avons remarqué qu'il fallait améliorer quelques petites choses pour faciliter la tâche de l'annotateur, mais aussi pour nous. En effet, en utilisant seulement le logiciel, nous aurions eu du mal à analyser et comparer les différents résultats.

Pour le pré traitement du fichier texte, c'est à dire de l'extrait que nous allions faire passer, il a fallu faire deux choses : Un fichier pour nous même pour avoir certaines informations sur le texte, et un deuxième, qui est un fichier pour Glozz, pour avoir quelque chose de plus structuré.

1) Le fichier de prétraitement XML

a) Explications

Le premier fichier à faire était un fichier XML apportant des informations sur un texte. Voici un exemple de la forme de départ d'un extrait de texte :

Ps- Bonjour à vous.
Pa- Bonjour.
Ps- Comment allez-vous ?
Pa- Ca va très bien.

Ici, "Ps" représente le psychologue, et "Pa" le patient. L'extrait est toujours une partie d'une conversation retranscrit entre un psychologue, et une autre personne qui peut être une personne témoin, ou une personne schizophrène.

On parle de tour de parole pour exprimer une phrase, une intervention de l'un des deux locuteurs. Par exemple, "Pa- Bonjour." est un tour de parole. Dans l'exemple ci-dessus, on retrouve donc quatre tours de parole, chaque locuteur en a 2. "Pa-" désigne le préfixe d'un tour de parole. Dans les textes que nous avons hérités, il y avait seulement "Pa" et "Ps" comme préfixes de tour de parole, mais nous avons vu des exemples avec par exemple "Pa24h", pour dire que c'est un patient avec le numéro 24, et que c'est un homme.

b) Programmation Java

Pour récupérer différentes informations sur le texte, nous avons réalisé un programme en Java. Le programme est générique, donc il marche pour tous les textes (tant qu'ils respectent la forme de l'exemple précédent), il faut juste changer le chemin vers le texte dans le programme.

Il aurait été trop compliqué de faire tous ces traitements à la main, car il faut penser grand, et donc imaginer traiter plus d'une centaine d'extraits. Il fallait donc développer un programme informatique pour faire ces traitements à notre place. Pour le langage, il était inutile de s'orienter vers un langage Web (JavaScript, PHP, .NET…) car nous ne voulions pas faire ces traitements à partir d'un navigateur. Nous avions donc eu le choix entre les langages Java et Python, car ce sont les deux plus utiles pour ce genre de tâche. Pour faire de l'analyse de textes, Python est plus performant et peut être plus rapide, mais nous avions plus l'habitude d'utiliser Java, et nous sommes plus performants dans ce langage. Nous avons donc choisi celui-ci, et il nous a permis de faire tout ce que l'on souhaitait.

c) Génération du fichier XML

Avec le texte en entrée du programme, nous avons pu l'analyser, et nous avons généré un fichier XML comprenant ces caractéristiques :

- La taille du texte, c'est à dire le nombre de tours de parole ;

- Le nom du patient/sujet. Dans notre exemple, nous aurons donc juste "Pa". Pour trouver le préfixe, on utilise une expression régulière de la forme : "([A-Z]\S*-)". Ce qui veut dire qu'on veut une majuscule, suivi de tout sauf un espace, 0 ou plusieurs fois, puis un tiret. "Pa4-" ou "PA-" fonctionnent, mais "pa-" ou "Pa -" ne fonctionnent pas. Pour que l'expression régulière marche, il faut que le texte suive le même modèle que l'exemple cité au-dessus. Cela nous permet donc de retrouver le nom du patient. On sait que dans le texte, il y aura toujours "Ps-" pour le psychologue, et un autre nom pour le patient, on recherche donc la forme respectant cette expression, en faisant attention de ne pas prendre le "Ps-". Donc on cherche dans le texte un préfixe respectant l'expression régulière, et étant différent de "Ps-".

- La liste de tous les tours de parole avec plusieurs informations :
    - Qui est la personne qui parle (On obtient le préfixe de la même façon que précédemment, avec l'expression régulière) ;

- Le contenu du tour de parole, donc ce qu'il y a après le préfixe ;
- La position du début du tour de parole, par rapport au texte au format .ac qui est généré par Glozz. Pour cela, on recherche la phrase dans le texte grâce à la fonction "indexOf" pour retrouver la première occurrence de la phrase. Mais parfois il y a plusieurs fois la même phrase ("Ouais" par exemple), donc il a fallu trouver un moyen pour trouver la position de toutes ces phrases ;
- La position de fin du tour de parole ;
- La position du début du contenu du tour de parole. C'est donc juste la position de début du tour de parole, plus la longueur du préfixe ;
- Et enfin, le nombre de mots qu'il y a dans ce tour de parole (exclu le préfixe).

- Quelques statistiques supplémentaires pour le fichier :
  - Le nombre de mots au total dans l'extrait, ainsi que le nombre de mots pour les deux personnes ;
  - Le nombre de tours de parole pour chaque personne ;
  - La moyenne du nombre de mots par tour de parole pour chaque personne.

Ce fichier est réservé uniquement pour nous, les annotateurs n'en n'ont pas besoin. Nous devions l'utiliser pour arriver à générer le second fichier, celui pour le logiciel Glozz.

Voici un exemple, montrant une partie de résultat généré par le programme :

```
<?xml version="1.0" encoding="UTF-8" ?>
<entretien>
    <taille>20</taille>
    <nom_sujet>Pa</nom_sujet>

    <tour_parole>
        <qui>Pa</qui>
        <contenu>Voilà mais bon il fait que je m'y remette quand même... enfin que je m'y remette je / je / je je m'y mette...
        <position_debut_tp>7</position_debut_tp>
        <position_debut>11</position_debut>
        <position_fin>132</position_fin>
        <nb_mots>26</nb_mots>
    </tour_parole>

    <tour_parole>
        <qui>Ps</qui>
        <contenu>Pourquoi ça vous plait pas... c'est tous les transports ?</contenu>
        <position_debut_tp>133</position_debut_tp>
        <position_debut>137</position_debut>
        <position_fin>194</position_fin>
        <nb_mots>10</nb_mots>
    </tour_parole>
```

Figure 6 : Partie du résultat généré par le programme dans un fichier XML

2) Le fichier pour Glozz

Comme dit dans la partie précédente, nous avions besoin du fichier XML précédent pour générer le fichier Glozz. En effet, il contient des informations utiles, comme la position de début et de fin du tour de parole de chaque personne, avec le nom aussi.

Le fichier que nous avons donc généré ensuite est un fichier XML mais avec l'extension ".aa", un fichier de Glozz. Ce que l'on voulait faire avec ce fichier, c'était présélectionner les différents tours de parole, pour arriver à bien les différencier sur le logiciel. On va donc différencier les deux personnes, une "Psychologue", et une "Autre", soit un témoin ou un schizophrène.

a) Identité cachée

L'annotateur sur Glozz n'aura pas exactement le même texte que dans l'exemple, car il ne doit pas savoir qui est le psychologue, et qui est l'autre personne (un schizophrène ou un témoin). Le but est que l'annotateur cherche des discontinuités ou non en ne sachant pas qui est l'autre personne. Par conséquence, nous avons remplacé les préfixes par des A et des B (quel que soit la personne qui commence), ainsi que le numéro de phrase. Il correspond donc au numéro du tour de parole.

En reprenant l'exemple précédent, on obtiendrait cela :

A1 : Bonjour à vous.
B2 : Bonjour.
A3 : Comment allez-vous ?
B4 : Ça va très bien.

Avec cette méthode, il est impossible de connaître les identités des personnes dans le corpus, c'est à dire qui est le psychologue, et qui est le schizophrène (ou la personne témoin) et donc, par la même occasion, d'influencer les annotateurs.

b) Générer le fichier pour Glozz

Ce que nous avons généré, c'est donc un document qui est de la même structure qu'un fichier .aa dans Glozz. Afin d'encadrer les tours de parole, il suffisait de créer une unité de la même la taille du tour du parole. Lorsqu'on sélectionne une unité avec le logiciel, on a la

position de début et de fin, ainsi que le type de l'unité. Ici, grâce au fichier généré juste avant, nous avons ces trois indications. Le type correspond au "Psychologue" ou à l' "Autre". Ce document est de la même forme que la figure n°3.

Pour récupérer les informations, nous avons donc dû faire un deuxième programme en Java pour générer le fichier pour Glozz. Java ne traite pas naturellement les fichiers XML, donc nous avons dû ajouter une librairie externe Java trouvée sur Internet qui peut traiter le XML facilement (Qui s'appelle JDOM). En récupérant les informations du fichier précédent, nous avons donc généré un deuxième fichier XML (Mais avec pour extension .aa), que nous avons importé dans Glozz pour qu'il sélectionne tous les tours de parole, en mettant une couleur différente pour chaque interlocuteur. Il y a, bien sûr, un fichier par texte.

A partir du texte original, il a fallu donc refaire un deuxième texte "formaté" en changeant les préfixes pour l'utiliser pour l'annotateur, générer un document XML pour obtenir des informations sur le texte, et générer un deuxième document .aa pour pouvoir l'importer dans Glozz pour sélectionner les différents tours de parole.

Voici ci-dessous, un exemple de texte, avec les tours de parole présélectionnés :

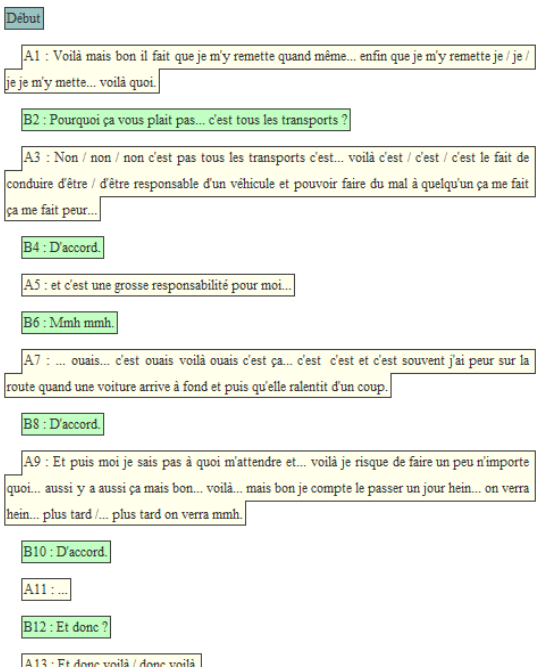


Figure 7 : Exemple d'extrait du texte formaté avec les tours de parole présélectionnés dans le logiciel Glozz

Après avoir fait le traitement du texte, tout le prétraitement était effectué, et Glozz pouvait fonctionner comme on le voulait. Mais, pour aider la personne dans son travail, nous

avons fait un arbre par rapport à ce que l'utilisateur à annoter. C'est à dire qu'à tout moment, il peut exporter son fichier.

C) La génération d'arbres : une aide pour l'annotateur

Nos tuteurs nous ont demandé de faire une génération automatique d'un arbre pour aider l'annotateur dans son travail. Avec Glozz, il est difficile de voir comment l'arbre peut être généré, nous avons donc dû faire un programme pour construire l'arbre, en fonction de ce que l'utilisateur avait annoté sur le texte.

1) Idées générales

Comme montré dans la figure n°1, il fallait générer un arbre semblable à ce schéma, donc le faire sous forme de S-DRT, même si l'utilisateur n'est pas censé connaître ce qu'est cette forme d'arbre. Il fallait donc respecter la syntaxe ainsi que l'orientation des relations.

La syntaxe à respecter était donc la suivante :

- Pour les unités, dans l'exemple, elles sont représentées par des "boites" avec écrit dans celle-ci $A_1{}^1$ par exemple. Le "A_1" correspond au numéro de la phrase, et le "^{1}" correspond au premier segment du tour de parole. Il est affiché lorsqu'un tour de parole est décomposé en plusieurs unités ;
- Sur l'exemple, les flèches en pointillés n'étaient pas à reproduire, mais les autres si (à part la relation oblique vers la droite). Nous devons donc respecter "l'inclinaison" de la relation pour faire les relations obliques, verticales et horizontales, ainsi que mettre le nom des relations sur chaque branche.

2) Simplification du modèle

Il aurait été difficile de reproduire exactement la même syntaxe. Nous avons donc modifié un peu le modèle pour obtenir quelque chose de clair pour l'annotateur, l'annotateur n'ayant aucune connaissance de la S-DRT.

Plusieurs éléments ont été mis en place. Tout d'abord, afin de simplifier la syntaxe, nous avons mis, par exemple, "A1-1", correspondant toujours au numéro du tour de parole, ainsi qu'au numéro de l'unité dans le tour de parole. Nous avons mis le "-1" même si le tour de

parole n'est composé que d'une seule unité pour avoir un équilibrage et avoir toujours la même forme.

De plus, dans chaque "boîte", nous avons rajouté une partie du texte sélectionné pour ne pas que l'utilisateur recherche de nouveau dans le texte à quoi correspondent A1 ou B4 par exemple. Pour cela, nous avons mis le texte de l'unité sélectionnée. Si le texte était assez court (moins de 20 caractères environ), nous le mettons en entier dans la "boîte". Mais s'il était plus long, alors nous avons découpé la phrase, pour avoir une syntaxe de ce genre : "Les 2 premiers mots … Les 2 derniers mots". Par exemple la phrase : "Bonjour à tous, comment allez-vous?" devient "Bonjour à … allez-vous?".

Aussi, nous avons ajouté une couleur pour le texte des unités. La couleur représente son thème. Par conséquence, si plusieurs phrases ont la même couleur, elles ont le même thème.

Enfin, nous avons également mis les relations en couleur dans le schéma. Les relations horizontales sont rouges, les relations verticales sont en jaune, et les relations obliques en bleu. Comme cela, c'est visuellement plus attractif pour l'annotateur qui peut s'y retrouver facilement, et voit la différence entre les différentes sortes de relations. Nous avons ajouté le nom de la relation sur chaque branche (de la même couleur que la catégorie).

Voici un exemple :

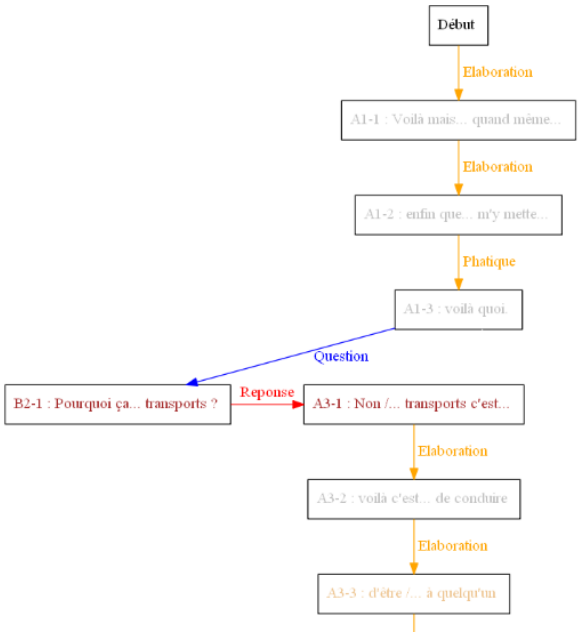


Figure 8 : Exemple d'arbre de type S-DRT généré par notre programme

3) Comment est généré l'arbre ?

Lorsque l'utilisateur annote son texte, et qu'il veut voir l'arbre, il doit exporter son travail. Grâce à un bouton du logiciel, il peut exporter toutes ses annotations (les différentes unités et relations) dans un fichier qui sera au format .aa. C'est donc ce fichier que nous traitons et qui permet la génération l'arbre. Nous avons de nouveau fait un programme en Java pour créer cet arbre. Pour se faire, il faut que l'utilisateur clique sur un bouton d'une autre petite fenêtre (faite en Java aussi). L'action sur le bouton lancera un fichier que nous avons créé aussi, au format .bat. Il s'agit d'un script qui contient les lignes de commande nécessaires à compiler les fichiers Java pour générer l'arbre.

La fenêtre ressemble à cela :

Figure 9 : Fenêtre pour générer l'arbre de type S-DRT

4) Choix des librairies

Le plus dur a été de savoir quel plugin, ou de quelle façon allait être généré l'arbre. Il existe de nombreuses librairies externes à Java pour faire des arbres, mais aussi des librairies propres à LaTeX.

a) Librairie Tikz

Nous avons donc choisi, au départ, d'utiliser la librairie "Tikz" de LaTeX pour faire un arbre. Elle était intéressante car nous pouvions donner des positions exactes pour chaque unité en donnant une abscisse et une ordonnée comme pour un repère orthonormé. En fonction d'où étaient placées les unités, on pouvait facilement les relier pour avoir des relations verticales, horizontales ou obliques. Il suffisait juste de changer l'abscisse ou l'ordonnée pour avoir deux unités une à côté de l'autre, ou une au-dessus de l'autre par exemple.
Cela était bien et marchait correctement si nous n'avions à chaque fois qu'un seul fils pour chaque unité, mais ce n'était pas tout le temps le cas. Et donc, quand il y avait deux relations verticales sur la même unité, il y avait une superposition des deux unités. Ainsi, nous avons donc choisi de faire le graphe d'une autre manière.

b) Format .dot

Nous avons décidé de faire un fichier au format .dot, qui est un fichier texte et qui permet de générer des arbres simples avec une syntaxe non complexe. En Java, nous devions écrire la syntaxe dans un fichier au format .dot, et ensuite, avec des lignes de commande dans le fichier script au format .bat (cité précédemment), nous pouvions facilement générer un graphe. Pour pouvoir générer l'arbre à partir de ce type de fichier, il a fallu télécharger un logiciel du nom de "Graphviz" pour pouvoir utiliser les lignes de commande nécessaires à la construction.

Le fichier produit par Glozz (qui a pour extension .aa, mais qui a une structure XML) a donc été traité en Java, et nous avions récupéré les différentes unités et relations pour les mettre dans des classes. Vu que le fichier nous donne juste la position de début et de fin des différentes unités, il a fallu de nouveau rechercher dans le texte à quel morceau de texte se référaient ces deux positions. Grâce aux tours de parole sélectionnés, nous pouvions déterminer quelle unité était associée à quel tour de parole, et de ce fait, mettre le préfixe correspondant dans le texte de l'unité pour qu'il soit visible aussi dans le graphe.

D) Améliorations possibles

Voici une liste de plusieurs éléments qui pourraient être améliorés :

- Faire une assistance au dessin plus ergonomique, plus beau. Le format .dot ne génère pas tout le temps des arbres très clairs par exemple ;
- Faire un système de versionning, c'est à dire voir l'évolution de l'annotateur dans son arbre. Donc il faudrait conserver toutes les étapes de son arbre dans un dossier au lieu d'avoir que l'arbre final. On aurait, bien sûr, plusieurs étapes seulement si l'annotateur avait voulu générer plusieurs fois l'arbre au cours de son travail ;
- Pouvoir modifier le chemin d'enregistrement des fichiers directement dans la fenêtre de la figure n° 9 ;
- Faire un programme avec des boutons (par exemple) pour importer les différents fichiers sur Glozz, et pour changer le nom du texte, et le nom du dossier au lieu de changer à la main dans deux fichiers. Dès qu'une nouvelle personne passe le test, il faut changer le nom du dossier dans lequel vont se stocker les fichiers. Si on change d'extrait, il faut aussi changer le nom dans les fichiers ;
- Dans le meilleur des cas, il faudrait faire un système portatif de ce que nous avons fait, c'est à dire capable de marcher sur n'importe quel ordinateur. Car pour faire faire le

travail, nous avons dû utiliser nos ordinateurs, il était donc dur de faire passer beaucoup de personnes en même temps (2 toutes les 1h30 environ).

Une fois les différents travaux réalisés, nous avons fait tester notre expérience avant la phase d'entretien finale. L'objectif était de détecter s'il y avait des erreurs au niveau de la programmation ou dans le guide. Ainsi, les testeurs, deux adultes, nous ont permis d'améliorer le code pour la génération de l'arbre, de nous donner une estimation du temps (environ un peu plus d'une heure), et de nous rendre compte qu'il fallait rester proche des annotateurs pour toutes questions éventuelles. Nous nous sommes aperçus également que le thème à choisir lors de l'annotation n'était pas clair, ce qui nous a amené à le modifier dans la vidéo.

III) Mise en place de la campagne d'annotations

A) Les différents documents

Les annotateurs devaient, à partir de supports d'explications, comprendre comment annoter un texte. L'enjeu est donc d'expliquer correctement un maximum d'informations sans donner trop d'indications, ce qui pourrait biaiser les résultats. Nous avons créé trois supports donnant un maximum d'informations sur le déroulement de l'expérience.

1) Le guide d'annotation

Il s'agit d'un document texte de 8 pages réalisé avec le logiciel Word. Nous avons essayé de réduire au maximum le nombre de pages afin de donner envie à l'annotateur de le lire. Le but est de ne pas le décourager dès les premières minutes. Le guide expose plusieurs parties. Nous présentons les objectifs et intérêts du travail de l'annotateur pour l'équipe du projet SLAM. Cette partie est utile pour donner de l'intérêt à l'annotateur à participer au projet. Nous sommes restés vague afin de ne pas biaiser leurs résultats en leur donnant trop d'indices. De plus, nous présentons les différentes manipulations possibles du logiciel Glozz. Nous avons mis beaucoup d'images afin de rendre les explications plus attractives.

Enfin, nous expliquons ce que sont les unités de sens et les relations rhétoriques et comment les réaliser sous le logiciel Glozz. Nous avons détaillé chaque relation accompagnée d'un exemple pour que l'annotateur puisse bien comprendre.

La version finale du guide d'annotation est disponible en annexe n°2.

2) La fiche de synthèse

C'est un document texte très court, d'une page, effectué avec le logiciel Word également. Il reprend les idées principales du guide d'annotation. Le but étant pour l'annotateur d'avoir un support sous la main lui évitant de chercher dans le guide et de lui faire perdre du temps.

La fiche de synthèse est présente dans l'annexe n°3.

3) Vidéo d'explication

Sachant qu'une démonstration vaut mieux qu'un long discours, nous avons été conviés à faire une vidéo d'explication pour résumer et expliquer le guide d'annotation à l'aide d'un exemple visuel, c'est à dire un exemple du logiciel à utiliser. Cela permet d'être plus concret, et donc de mieux faire comprendre la tâche à effectuer aux utilisateurs. C'est une vidéo courte d'environ 3/4 minutes. Elle présente la démonstration du logiciel et le travail à effectuer.

La vidéo a été enregistrée avec le logiciel CamStudio, qui permet de faire des vidéos de son écran d'ordinateur. On y voit donc l'interface de Glozz et par l'enregistrement de nos voix, donne l'explication des différentes étapes pour annoter un texte. Afin de correspondre au mieux à leur travail, nous avons pris pour exemple une retranscription d'un texte entre un patient atteint de schizophrénie et un psychologue (Le texte de base cité dans la partie précédente, "Quelle mort").

La voix a été enregistrée avec le magnétophone d'un ordinateur portable, et le logiciel AudaCity a permis de découper le son (Quand il y avait des phrases mal dites, ou de trop longs silences) et le logiciel Movie Maker a permis d'assembler le son et la vidéo. Les tâches pour cette vidéo ont été séparées, nous n'avons pas enregistré le son et la vidéo en même temps. Il était plus simple de séparer les deux, en cas d'erreur lors de l'élocution du texte par exemple.

B) Le passage des annotateurs

Nous avons dû organiser et trouver des volontaires pour annoter les textes. Il a fallu planifier et aller à la rencontre d'un maximum de personnes. L'idéal est d'avoir des personnes hétérogènes pour comparer au mieux les idées de chacun, c'est à dire d'avoir des personnes de sexe, d'âge, et de formation différentes.

Chaque personne a analysé à sa façon un passage d'un texte. L'objectif est donc de comparer les idées et résultats de chacun. Nous nous sommes entraînés sur un passage de texte (L'extrait de "Quelle mort"). C'est également celui que nous avons utilisé pour générer notre premier arbre et pris pour réaliser nos programmes. Ensuite, pour les passations finales, nous avons eu quatre extraits. Le but étant de, si c'était possible, faire passer tous les extraits aux annotateurs.

1) Le texte

Nos encadrants nous ont transmis 4 textes. Nous en avons choisi un pour le faire annoter. L'idéal aurait été d'en faire annoter au moins deux par personne, mais vu le peu de réjouissance des annotateurs, nous nous sommes limités à un par personne. Il s'agissait d'un extrait d'une retranscription entre un psychologue et un schizophrène.
Il y a une discontinuité dans le texte avec ce tour de parole "Pa- Ben dès que je suis souriante ça va". En effet, la personne Pa change de thème dans la conversation, qui ne devrait pas y être. On aimerait observer comment les personnes vont pouvoir l'interpréter.

On retrouve l'extrait que nous avons fait passer en entier dans l'annexe n°4.

2) Les différents individus

Le plus dur dans ce projet était de trouver des volontaires capable d'annoter. C'est un travail long et "laborieux" pour les testeurs car il demande beaucoup de concentration, et il y a beaucoup de notions à découvrir (Le logiciel et les relations par exemple). De plus, la période du mois d'avril et mai correspond aux périodes d'examens pour les étudiants, ce qui ne nous a pas facilité la tâche.

Nous sommes, dans un premier temps, intervenus dans une classe de Master 1 Psychologie durant le cours de Monsieur Musiol où un étudiant s'est porté volontaire.

Puis, nous avons planifié deux jours entiers au Pôle Lorrain de Gestion, pour le passage des étudiants venant de notre classe, et quelques personnes extérieures. Dix personnes sont venues.

Enfin, nous avons consacré un après-midi au LORIA (Laboratoire Lorrain de Recherche en Informatique et ses Applications) où deux thésards et un ingénieur sont venus annoter.
Puis, afin d'être encore plus efficaces, nous avons chacun de notre côté fait passer des individus.
Et pour terminer, grâce au contact entre Michel Musiol et les étudiants de psychologie, nous avons organisé un après-midi à la faculté de lettres à Nancy pour faire annoter des élèves de psychologie.
Au total, 22 personnes ont participé à notre expérience.

Globalement, les personnes ont réalisé ce travail sérieusement. Seulement deux ou trois personnes sont passées à côté du travail attendu, en y mettant beaucoup trop de thèmes par exemple, ou trop de relations inutiles.
Les différents retours des personnes ont été que la tâche à réaliser était fastidieuse, et qu'ils auraient aimé avoir plus d'exemples pour bien comprendre toutes les relations. Ils ont apprécié la vidéo, qui, selon eux, les ont beaucoup aidés.
De notre côté, nous avons observé que la plupart avait un peu de mal à utiliser le logiciel Glozz.

3) Traitement des résultats

Voici les différents documents qui ont dû être générés pour pouvoir faire l'analyse des différents résultats :

a) Fichier XML

Nous avons un fichier pour chaque extrait, mais vu que nous avons des résultats seulement sur l'Extrait3, nous avons fait un fichier seulement pour celui-ci. Nous avons décidé de mettre toutes les personnes dans un seul fichier, et de ne pas les séparer individuellement.

Le fichier est un fichier XML, comprenant, pour chaque personne :

- Les détails de la personne (Sexe, âge, formation) ;
- Le nombre d'unités ;
- Le nombre de thèmes ;

- Pour chaque thème, son nom, le nombre de segments portant ce thème, ainsi que toutes les portions de phrases correspondant aux thèmes ;
- Le nombre de relations ;
- Pour chaque catégorie de relations (verticales, horizontales, obliques) :
    - Le nombre de relations utilisées dans chaque catégorie ;
    - Le nombre de relations dans cette catégorie (Par exemple 5 Suite, 2 Narration…).
- le temps de réalisation entre la première unité, et la dernière unité (ou relation, cela dépend ce qu'était sa dernière création) ;
-

Voici un extrait de ce fichier :

```
<?xml version="1.0" encoding="UTF-8" ?>
<resultats>
    <personne_1>
        <sexe>Homme</sexe>
        <age>23 ans</age>
        <formation>M1 Psycho</formation>

        <unites>
            <nb_unites>26</nb_unites>
            <themes>
                <nb_themes>7</nb_themes>

                <theme_1>
                    <nom>transport</nom>
                    <nb_phrases>7</nb_phrases>
                    <phrases>
                        <phrase_1>Voilà mais bon il fait que je m'y remette quand même... enfin que je m'y remette je / je / je je
                        <phrase_2>Pourquoi ça vous plait pas</phrase_2>
                        <phrase_3>c'est tous les transports ?</phrase_3>
                        <phrase_4>Non / non / non c'est pas tous les transports</phrase_4>
                        <phrase_5>Et donc vous vous déplacez majoritairement à pied ? ou euh</phrase_5>
                        <phrase_6>A pied en bus.</phrase_6>
                        <phrase_7>Oui en bus métro enfin on est très bien desservi ici</phrase_7>
                    </phrases>
                </theme_1>

                <theme_2>
                    <nom>transport et responsabilité</nom>
                    <nb_phrases>3</nb_phrases>
                    <phrases>
                        <phrase_1>c'est... voilà c'est / c'est / c'est le fait de conduire d'être / d'être responsable d'un véhicul
```

Figure 10 : Exemple du fichier XML des résultats, généré par le programme de traitement du fichier Glozz

b) Fichier CSV

A partir du même programme Java que pour générer le document de résultats XML, nous avons aussi généré un document sous format CSV. Il regroupe donc de manière synthétique toutes les données, réponses et caractéristiques de chaque individu. Ainsi, on pourra facilement analyser les résultats.

En voici un extrait :

Sexe	Age	Formation	Nombre unit	Nombre themes	Theme + recurrei	Nombre relations	Nombre relation	Nombre relations hc	Nombre rela	Nombre Narr	Nombre repc	Nombre repc
Homme	23	M1 Psycho	26	7	Thème non défir	23	8	11	4	3	2	1
Homme	23	En préparation de t	37	10	transport	36	19	7	10	1	3	3
Homme	21	L3 Sciences Cognitiv	30	7	transport	29	9	12	8	6	3	0
Homme	53	Administrateur des	29	16	ok	28	1	19	8	5	3	10
Femme	50	Technicien territori	27	4	moyen de locom	26	5	13	8	2	3	8
Femme	23	M1 SCA	37	32	confirmation	26	10	1	15	0	0	0
Femme	22	M1 SCA	35	19	raison deplaisanc	35	12	7	16	0	0	0
Femme	22	M1 SCA	34	5	permis	33	21	3	9	0	0	0
Homme	21	M1 SCA	36	9	doute	35	6	5	24	4	0	0
Femme	24	M1 Psycho	43	11	Thème non défir	42	27	9	6	0	7	2
Homme	22	M1 SCA	41	6	Peur	40	13	21	6	8	5	6
Femme	22	M1 SCA	26	4	conduite	25	8	8	9	1	3	2
Homme	26	M1 SCA	35	12	réponse neutre	34	11	16	7	0	5	6
Femme	20	L3 Allemand	33	11	la conduite	32	13	14	5	2	2	10
Femme	24	M1 Psycho	37	7	la conduite	36	11	19	6	14	3	0
Homme	22	M1 SCA	30	11	transports	29	6	15	8	5	6	0
Homme	21	M1 SCA	38	12	transport	36	16	15	5	2	3	2
Homme	21	L2 Sciences Cognitiv	33	11	moyen de transp	33	5	18	10	0	9	1
Homme	25	M1 SCA	27	10	peur	26	8	14	4	7	5	0
Homme	24	Ingénieur informati	30	12	mode de déplace	29	12	11	6	3	3	1
Femme	21	M1 Miage	55	15	hésitation	62	11	39	12	22	1	2
Homme	25	Préparation de thè	38	7	risques liés à la c	37	26	6	5	2	3	1

Figure 11 : Fichier CSV contenant les résultats de chaque individu

Il regroupe pour chaque individu : le sexe, l'âge, la formation, le nombre d'unités, le nombre de thèmes, le thème le plus récurrent, le nombre de relations, le nombre de relations horizontales, le nombre de relations verticales, le nombre de relations obliques, ainsi que le nombre pour chaque type de relations (narration, suite, etc.…). Il comprend aussi le temps que la personne a mis pour réaliser la tâche. Le fichier CSV entier se trouve à l'annexe n° 5.

C) Les résultats obtenus

1) Résultats communs

On compte 13 hommes et 9 femmes. L'âge en moyenne est de 25,22 ans. L'âge maximal est 53 ans et minimal 20 ans. Vu que nous avons un écart assez conséquent entre l'âge minimal et maximal, et que nous n'avions aucune personne d'un âge compris entre 26 et 50 ans, la moyenne n'est pas "fiable". Il est donc préférable de calculer la médiane. Elle est de 22,5 ans.

Dans les participants, on compte : 10 étudiants de Master 1 Sciences Cognitives, 2 thésards, 1 ingénieur en informatique, 1 technicien territorial, 1 administrateur des ventes, 3 étudiants de Master 1 Psychologie, 1 étudiante de Master 1 MIAGE, 1 étudiant de L2 Sciences Cognitives, 1 étudiante de L3 langue parcours Allemand.

Pour chaque champ, on obtient :

Items	Moyenne	Min	Max	Médiane
Nombre d'unités	34,4	26	55	34,5
Nombre de thèmes	10,8	4	32	10,5

Nombre de relations	33,3	23	62	33
Nombre de relations verticales	11,7	1	27	11
Nombre de relations horizontales	12,9	1	39	12,5
Nombre de relations obliques	8,7	4	24	8
Nombre de narration	3,9	0	22	2
Nombre de réponse	3,1	0	9	3
Nombre de réponse phatique	2,5	0	10	1
Nombre de suite	3,2	0	14	2
Nombre d'élaboration	5,9	1	15	4,5
Nombre d'élaboration explication	0	0	0	0
Nombre d'élaboration prescription	0	0	0	0
Nombre d'évaluation	0	0	0	0
Nombre de phatique	5,8	0	14	5,5
Nombre de question	1,9	0	4	2
Nombre de question conduite	0,77	0	2	0,5
Nombre de méta-question	0,3	0	2	0
Nombre de demande d'élaboration	1,6	0	5	2
Nombre de conduite	1,9	0	6	1
Nombre de contre élaboration	0,09	0	1	0
Temps (en minutes)	49,9	29	104	42,5

Nous avons calculé la moyenne pour avoir une idée du nombre moyen pour chaque champ. Mais nous avons aussi pris la médiane, car c'est le point milieu de l'ensemble. C'est à dire qu'il y a la moitié des valeurs qui sont inférieures à cette médiane, et l'autre moitié des valeurs supérieures à cette médiane. On a parfois, pour quelques champs, des écarts assez importants, il est donc utile de calculer cette médiane pour savoir où se situe le point de milieu.

Le nombre d'unités reste globalement le même pour tous les individus. En revanche on observe que le nombre de relations et de thèmes sont assez diverses. En effet, pour les thèmes,

une personne en a utilisé 4, et une autre 32. Mais celle qui en a utilisé 32 s'est trompée et n'a pas très bien compris ce qu'était le thème. La moyenne des thèmes reste tout de même à environ 10.8 thèmes, donc la valeur 32 est vraiment "inhabituelle" et loin de la moyenne.
Les relations obliques sont généralement moins utilisées (en moyenne 8,7 fois par personne) par rapport aux relations horizontales (12,9) et verticales (11,7).
On observe également que la relation "élaboration" est la relation la plus utilisée (en moyenne 5,9 fois par personne).
On observe des différences dans l'emploi des relations. C'est le cas par exemple de la relation "phatique", utilisée 14 fois par un annotateur et 0 fois pour un autre. Se retrouvent dans le même cas, les relations "élaboration", "suite", "réponse phatique", "narration", et "réponse".

Thème

Enfin, voici la liste des thèmes les plus récurrents pour chaque individu : "transport", "transport", "ok", "moyen de locomotion", "confirmation", "raison deplaisance transport", "permis", "doute", "thème non défini", "peur", "conduite", "réponse neutre", "la conduite", "la conduite", "transports", "transport", "moyen de transport", "peur", "mode de déplacement", "hésitation", "risques liés à la conduite".
Ainsi, sur l'ensemble des individus, le thème le plus récurrent est "transport" qui est utilisé par quatre personnes, suivi de "conduite" mis par trois personnes.

Ensuite, on peut voir le nombre d'utilisations de chaque thème pour toutes les personnes. La liste de tous les thèmes ainsi que leur occurrence est disponible à l'annexe n° 6.
On remarque que "transport" est en tête avec 62 utilisations sur 715 au total. Mais on remarque qu'il y a aussi "transports" avec 36 utilisations, donc si on l'assemble avec le singulier, on a un total de 98 utilisations (Ce qui représente 13,70%). Le transport ou moyen de transport est encore utilisé quelques autres fois, les personnes ayant donné un nom différent pour dire quasiment la même chose. En effet, si on rajoute des thèmes comme "moyen de transport", "moyen de locomotion", ou "les transports", on obtient environ 60 occurrences en plus. Ce qui nous fait environ 158 thèmes, soit 22,10%.

Malheureusement, le deuxième thème en tête est "Thème non défini" (43), qui veut dire que les personnes ont oublié de mettre un thème sur certaines unités.

Ensuite, on retrouve les thèmes de la peur (42), de la responsabilité (37), la conduite (20 pour "conduite", et 16 pour "la conduite", donc 36 au total), ou encore du permis (25) et du moyen de transport (24) pour les plus importants.

Il aurait fallu utiliser une lemmatisation pour compter “transport” et “les transports” ensemble par exemple, et aussi réunir les synonymes (“Moyen de transport” = “Moyen de locomotion” par exemple). Cette lemmatisation est trop complexe à faire en Java, surtout si l’on veut faire un programme générique, qui marche donc pour tous les textes.

2) Comparaison : sexe, âge et formation

Nous avons cherché les différences des résultats entre les individus de sexe opposé. Mais les résultats sont semblables (Voir annexe n°7).

Nous avions eu l’idée de comparer les résultats en fonction de l’âge des individus. Mais les âges sont trop proches (sur 22 personnes, 20 ont entre 20 ans et 25 ans), ce qui ne nous permet pas d’avoir de bonnes conclusions.

Nous avons ensuite essayé de comparer les individus en fonction de leur formation. Ainsi, nous avons séparé les personnes en 5 catégories :

- Master 1 Psychologie (3 personnes)
- Non étudiant (5 personnes)
- Master 1 Sciences Cognitives (10 personnes)
- Master 1 Miage (1 personne)
- Licence 2 ou 3 (3 personnes)

On exclut les résultats de la personne en M1 Miage, car il serait faux d’analyser les résultats d’une seule personne.

On observe quelques légères différences, mais le problème est que l’échantillon est très petit. Avec davantage d’individus, les résultats seraient peut-être complètement différents. Les différences se situent au niveau des types de relations utilisés.

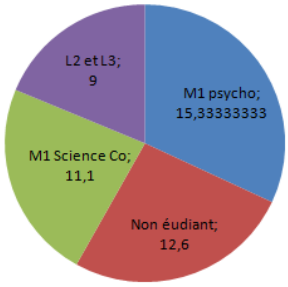
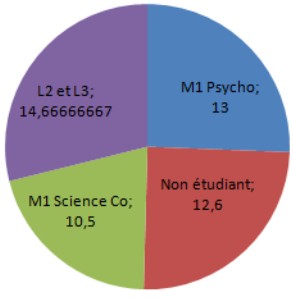
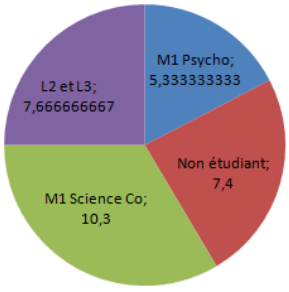
Relation verticale	Relation horizontale	Relation oblique
		

Figure 12 : Graphique des moyennes des différentes catégories de relations pour chaque catégorie d'individus

On note peu de différence pour la relation horizontale. En revanche, les étudiants en licence 2 et 3 utilisent très peu de relations verticales et obliques. Les M1 Psychologie, eux, n'utilisent pas beaucoup de relations obliques. Pour plus d'informations, voir l'annexe n°8.

Étant donné qu'il y a peu de différence entre les groupes d'individus, on peut déduire que globalement ils ont pensé la même chose, eu la même démarche.

On peut conclure que mise à part quelques annotateurs qui se démarquent, pour cause, par exemple, d'incompréhension, les résultats sont en moyenne à peu près semblables. L'utilisation des différents types de relations en revanche, semblent différentes entre les individus.

3) Résultats individuels : Discontinuité / arbres

Nous avons décidé d'analyser 17 arbres sur 22. Certains arbres ne respectent pas bien la forme d'un arbre S-DRT, car ils ont, par exemple, trop de relations, ou des sous arbres reliés à rien. Voici ci-dessous des exemples d'extraits d'arbres "incorrects".

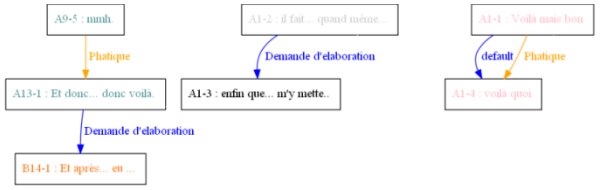


Figure 13 : Arbre S-DRT généré par un des annotateurs

Figure 14 : Arbre S-DRT généré par un des annotateurs

La figure 13 n'est pas correcte car il y a plusieurs sous-arbres, alors que normalement il ne devrait y avoir qu'un seul arbre. La figure 14 est incorrecte aussi car il y a beaucoup trop de relations liées au "Début". Mais, dans l'ensemble, on observe des arbres relativement linéaires.

Comme dit précédemment, dans l'extrait que nous avions fait passer, il y avait une discontinuité. Il s'agissait du tour de parole A15 ("Ben dès que je suis souriante ça va."), qui n'a aucun rapport avec le reste de la conversation.

La plupart des individus ont vu qu'il y avait une discontinuité dans le texte. Ils ont tous relié le tour de parole mettant en avant la discontinuité à la suite de l'arbre mais avec un thème différent et unique pour ce tour de parole.

Voici deux exemples :

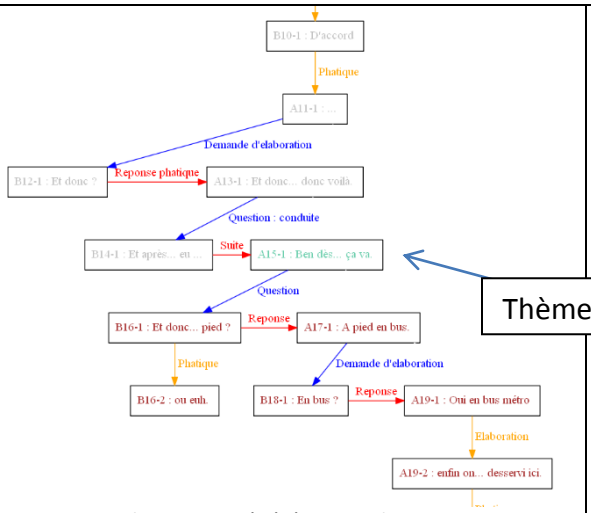


Figure 15 : Arbre S-DRT généré par un des annotateurs

Thème différent

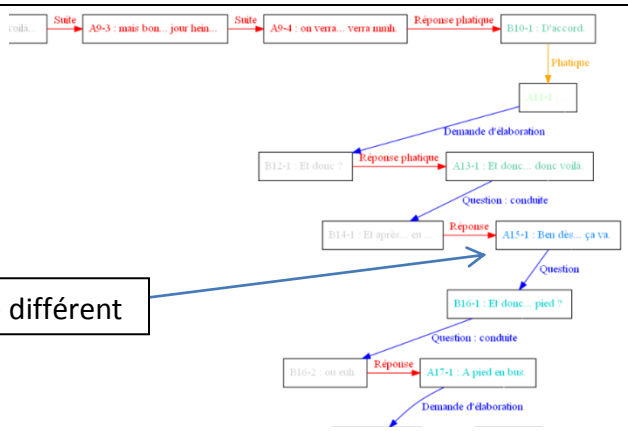


Figure 16 : Arbre S-DRT généré par un des annotateurs

Ici on voit que le thème, représenté par une couleur, est d'une couleur unique pour le tour de parole présentant une discontinuité. Les thèmes mis par ces deux annotateurs sont : "absurde" et "rapport choucroute/patate".
Ils ont donc mis en avant une discontinuité de type sémantique.

Aussi, on observe que les relations "Question" et "Réponse", sont tous plus ou moins situées au même endroit. On observe ce schéma de questions (ou conduites) et de réponses dans environ la moitié des schémas, les gens ont donc bien compris ces relations. Il peut après avoir de petites différences en fonction de comment la personne à découper le texte avec les unités.

Nous avons mis quelques exemples d'arbres dans l'annexe n°9.

IV) Conclusion

Dû à la diversité des tâches, ce projet tuteuré fut pour nous très intéressant. En effet, nous avons à la fois programmé, rédigé différents documents, organisé des passages pour les annotateurs et analysé des données.

Nous avons pu découvrir les enjeux du projet SLAM, et une manière d'analyser une retranscription d'un texte sous forme d'un arbre S-DRT. Cela nous a amené à manipuler un logiciel inconnu pour nous deux, Glozz.

Afin de pouvoir compléter et adapter ce logiciel aux annotateurs, nous avons par exemple réalisé la possibilité de générer un arbre sous forme S-DRT. Ainsi, nous avons pu améliorer nos compétences en programmation Java et découvrir de nouvelles fonctionnalités. Nous avons également découvert une librairie (Tikz de LaTeX), et un langage de description de graphe dans un format texte, DOT.

Enfin, nous avons pu mettre en avant les compétences acquises au cours de notre formation en Sciences Cognitives au travers de l'analyse de données. En effet, l'analyse linguistique, la programmation Java, ou encore l'analyse des résultats ont été vues au cours de ce cursus.

V) Bibliographie

- *Schizophrénie, logicité et perspective en première personne* - Manuel Rebuschi, Maxime Amblard, Michel Musiol. L'évolution psychiatrique n°78, pages 127-141; 2013.

- *Une analyse basée sur la S-DRT pour la modélisation de dialogues pathologiques* - Maxime Amblard, Michel Musiol, Manuel Rebuschi. TALN 2011, Montpellier; Juin et Juillet 2011.

- *L'interaction conversationnelle à l'épreuve du handicap schizophrénique* - Maxime Amblard, Michel Musiol, Manuel Rebuschi. Recherches sur la philosophie et le langage, 2014, 31, pp.1-21.

- *Ressource multi-niveaux annotée pour la pathologie mentale* - Kenny Rivalin, Nathalie Wittmann. Projet tuteuré M1 SCA; Mai 2014.

- *Interface et vérification pour la modélisation de conversations schizophréniques* - Cédric Beuzit-Laboz, Joffrey Mougel. Projet tuteuré M1 SCA; Mai 2011.

- Site internet *MSH-Lorraine, Maison des sciences de l'homme*, "http://www.msh-lorraine.fr"

- Site internet *Glozz, Glozz Annotation Platform*, "http://www.glozz.org"

- Site internet *HN, Hum-Num, la TGIR des humanités numériques*, "http://www.huma-num.fr"

Table des annexes

Annexe 1 : Fiche de présentation du projet tuteuré

Titre : Outils pour l'analyse des troubles du langage et de la pensée
Title: Tools for the analysis of language disorders and thought
Public : M1 Projet tutoré

1 Encadrement/Supervisor

Loria: Maxime Amblard
Encadrants :

Maxime Amblard Sémagramme-Loria maxime.amblard@univ-lorraine.fr
Michel Musiol Discours - Atilf michel.musiol@univ-lorraine.fr
Manuel Rebuschi LHSP-AHP manuel.rebuschi@univ-lorraine.fr

2 Description / Description

Ce projet s'inscrit dans le projet de de recherche SLAM 1 (Schizophrènes et Langage : Analyse et Modélisation) qui vise à systématiser l'étude des conversations pathologiques dans le cadre d'une approche interdisciplinaire alliant psychologie, linguistique informatique et philosophie. Il se concentrera notamment sur les conversations impliquant des personnes souffrant de troubles psychiatriques (schizophrènes).

1. Plusieurs théories du discours ont été développées pour modéliser la structure du discours. Le problème général est de définir le bon niveau de description. Dans ce projet, nous utiliserons une extension de la DRT (Discourse Representation Theory) la S-DRT qui construit un arbre de relations discursives et conversationnelles. L'avantage de cette théorie est qu'elle se base sur le calcul de la sémantique compositionnelle et qu'elle tente de rendre compte d'une pragmatique de manière simplifiée.
2. D'autre part, les psycho-linguistes étudient les troubles du langage chez les schizophrènes. On remarque que dans le cadre de l'entretien semi-dirigé (ou clinique), une partie des patients schizophrènes utilise des stratégies discursives non usuelles mais récurrentes. Il s'agit donc de les répertorier et de saisir leur nature.

Dans le cadre du projet SLAM, une étude est conduite sur l'utilisation de propriétés formelles pour modéliser les entretiens entre patients schizophrènes et psychologues. Actuellement, l'opé- ration SLAM collecte de nouvelles données afin de produire un nouveau corpus qui est constitué des transcriptions des entretiens, ainsi que plusieurs couches d'annotations :
– une annotation de type psychologique permettant de mettre en avant les discontinuités apparaissant dans l'échange
– une annotation morpho-syntaxique sur l'ensemble du corpus.
– une modélisation en sémantique formelle, au travers de la SDRT, des extraits discontinus précédemment identifiés.

Les patients de type schizophrène sont rencontrés à la fois au CH le Vinatier à Lyon et au CHU Saint Antoine à Paris. Ils passent plusieurs tests cognitifs, ainsi qu'un entretien supervisé avec pour les uns un système d'eye-tracker, pour les autres un système d'EEG. Par ailleurs, un groupe contrôle est apparié sur le celui des patients et passe le même protocole.

3 Informations diverses : matériel nécessaire, contexte de réalisation

Le but du projet tutoré est de participer à la création de la ressource, point clé de l'opération. Deux volets seront abordés.
Il s'agira de poursuivre le travail engagé lors d'un premier projet tutoré sur l'implémentation d'une interface pour l'annotation en relation discursive.

1. Prise en main et adaptation de la plateforme d'annotation GLOZZ
2. Mise en place et suivi d'une campagne d'annotation en SDRT auprès d'annotateurs :

– écriture d'un guide d'annotation
– identification d'un groupe d'annotateur
– évaluation du protocole expérimental

4 Livrable et échéancier / Deliverable and schedule

– Adaptation de la plateforme GLOZZ pour l'annotation en structures discursives.
– Rédaction d'un guide d'annotation en structures discursives des retranscriptions d'entre- tiens avec des patients schizophrènes.
– Validation du guide et de l'interface sur un groupe d'annotateurs.

Annexe 2 : Guide d'annotation

Guide d'annotation pour établir la représentation graphique de type S-DRT sur les retranscriptions d'entretiens avec des patients schizophrènes

Personnes : M. Amblard, M. Rebuschi, M. Ruez, V. Steyer

Gestionnaires de la campagne : M. Ruez, V. Steyer

Présentation du sujet

Ce travail s'inscrit dans le projet de recherche SLAM (Schizophrènes et Langage : Analyse et Modélisation) qui vise à systématiser l'étude des conversations pathologiques. Il se concentre sur les conversations impliquant des personnes souffrant de troubles psychiatriques (schizophrènes) faites avec des psychologues lors d'entretiens semi-dirigés (ou cliniques).

Afin de modéliser la structure du discours, nous utiliserons la S-DRT (extension de la DRT (Discourse Representation Theory) qui consiste à construire un arbre de relations discursives et conversationnelles.

Objectif

L'objectif est d'analyser et annoter un texte reprenant le discours entre un psychologue et une autre personne (schizophrène ou personne témoin).

Ce travail consiste à trouver les spécificités et bizarreries dans une conversation, c'est-à-dire quelque chose qui ne semble pas logique dans le déroulement de la conversation.

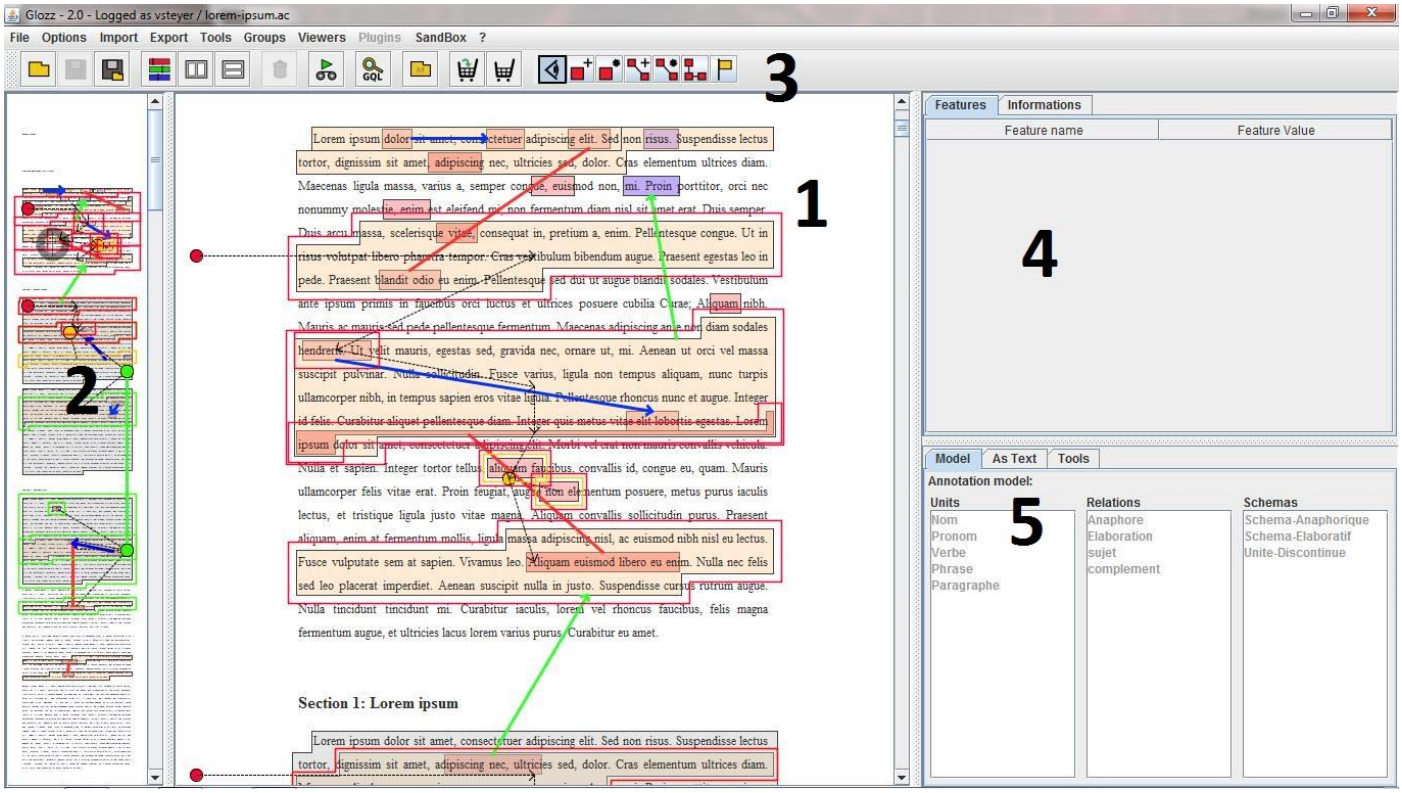
Pour cela, plusieurs annotateurs devront réaliser une analyse d'une ou plusieurs retranscriptions, permettant ainsi de comparer les résultats obtenus.

Démarche à suivre

Votre travail est le même que celui d'un annotateur. Un texte vous est mis à disposition et à partir de ce guide et suivant la démarche, vous devez l'analyser. Le résultat final est un arbre de relations discursives et conversationnelles.

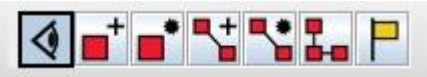
Le texte à étudier est un texte retranscrit entre deux personnes : un patient schizophrène et un psychologue ou entre un psychologue et une personne d'un groupe témoin. Les transcriptions sont données sous forme de corpus numéroté pour faciliter le repérage.
Un exemple sera détaillé afin de faciliter la compréhension des manœuvres à suivre.

Présentation de Glozz



1 : C'est la vue principale du texte, c'est sur cette vue qu'on pourra ajouter/modifier toutes les annotations.
2 : La vue en Macro, c'est le même texte annoté que dans la vue n°1, mais le texte est en mode "macro", c'est à dire qu'il y a une vue globale du texte et de ses annotations, elle est utilisée pour naviguer plus facilement.

3 : C'est la barre des menus. Le groupe de boutons le plus à droite sera à utiliser pour ajouter ou modifier des annotations. Ce sera les seuls boutons à utiliser.
4 : C'est un tableau contenant les différentes valeurs pour une annotation sélectionnée.
5 : C'est le modèle d'annotation, c'est ici qu'on voit la liste de tous les types d'annotations disponibles (unités, relations, ainsi que les schémas)
Voyons maintenant le groupe de boutons qui seront utiles à l'annotation du texte :



Voici la description des différents boutons :

 Pour retourner au mode par défaut, c'est à dire pour avoir le curseur de navigation du texte

 Pour créer une unité

 Pour modifier une unité

 Pour créer une relation

 Pour éditer une relation

 Pour accéder au sous menu de création d'un schéma

 Pour ajouter un commentaire

Sélectionner les unités de sens

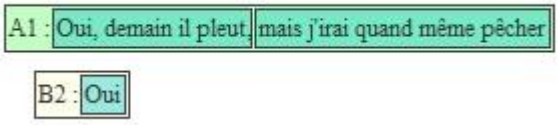
Définition

Une unité de sens est une phrase ou une partie de phrase, contenue dans un seul acte de langage. Une unité de sens peut donc couvrir au maximum une intervention, mais elle ne peut pas couvrir plusieurs interventions du même ou des deux interlocuteurs.

Si au sein d'une intervention, deux idées différentes seront exprimées, ou bien si le locuteur répond à une intervention précédente puis relance la discussion sur autre chose…, alors l'annotateur devra scinder l'intervention en autant de sous-parties que nécessaire.

Les tours de parole sont déjà pré-sélectionnés au départ, chaque couleur correspondant à un locuteur différent, sans savoir qui ils sont précisément. Les préfixes des tours de parole (A1:, B2 :, …) ne sont surtout pas à sélectionner, comme il est montré dans l'exemple.

Par exemple :



Ici, on a deux personnes différentes qui parlent. Pour la première, il semble nécessaire de séparer l'énoncé en deux sections. En effet, dans la première section, il répond à une question, alors que dans la deuxième, il apporte une information supplémentaire. Pour la deuxième personne, il n'y a pas plusieurs idées, donc on sélectionne juste le "Oui".

Créer une unité sur Glozz

Pour créer une unité, cliquez sur l'icône associée à sa création. Le sous menu de l'unité dans le menu 5 se fait en surbrillance. Cliquez donc sur le modèle d'unité souhaité (ici Section).

Une fois cela fait, vous avez deux moyens pour sélectionner l'unité :

- Un drag & drop : Mettez la souris à la position du début de la future unité, ensuite cliquez sur le bouton gauche de la souris, et gardez le appuyé jusqu'à la position de fin désirée. En relâchant le bouton, l'unité est directement créée.
- Deux clics : Cliquez au début de la future unité, ce sera donc la position de départ. Ensuite, cliquez à la fin de l'unité voulue, la position finale sera donc mise en place, et l'unité sera créée.

Ensuite, dans la vue 4, vous devez choisir le type de la section. Ici, vous devez choisir le thème de la section. Écrivez donc le thème que vous pensez le plus approprié.

Pour modifier une unité, vous pouvez cliquer sur le bouton correspondant. Ensuite, cliquez sur l'unité. En appuyant sur la touche "Suppr", vous pourrez supprimer l'unité. Sinon, vous pouvez modifier le début ou la fin de votre unité en cliquant sur la bulle correspondante au début ou à la fin de l'unité, puis de faire un drag & drop pour modifier la position.

Relier les unités par des relations théoriques

Une fois les unités sélectionnées, vous devez les relier entre elles par des relations. Le mieux est de relier ces unités qu'une fois après avoir totalement fini de créer les unités.
Vous devez choisir parmi seize relations qui vous semblent les plus appropriées.

Les différents types de relation

Relations horizontales	**Relations verticales**	**Relations obliques**
Narration	Elaboration	Question
Réponse	Elaboration : explication	Question : conduite
Réponse phatique	Elaboration : prescription	Méta-question
Suite	Evaluation	Demande d'élaboration
	Phatique	Conduite
		Contre-élaboration

Relations horizontales :

Réponse : Lorsqu'une réponse est donnée et répond à une question posée.

Exemple :
P1 : "De quelle couleur est ta chemise?
P2 : "*Ma chemise est rouge*"

Narration : Il s'agit d'une suite d'éléments dans le but de raconter une histoire.

Exemple :
P1 : "Guy a pris un repas fantastique"
P1 : "Il a mangé du saumon"
P1 : "*Il a dévoré plein de fromage*"

Réponse phatique : Lorsqu'une réponse est donnée à une question, mais sans contenu. C'est à dire qu'il répond à la question juste pour répondre. C'est une « réponse pour rien ».

Exemple :
P1 : “Est-ce que on irait se faire une promenade?”
P2 : *“Moui*, ça te dirait pas d’aller au cinéma?”

Suite: Il s’agit d’une forme d’élaboration, qui permet d’illustrer les propos.

Exemple :
P1 : “Je suis allée à la pépinière”
P2 : “Ah ah”
P3 : ”*Je suis allée à la Place Stanislas, j’ai rencontré un chat*”

Relations verticales :

Elaboration : Lorsqu’il s’agit d’une phrase qui “entre dans le détail”, apporte des informations supplémentaires, développe une idée.

Exemple :
P1 : “J’ai rencontré une femme, *elle était gentille et souriante.*”

Phatique : Recouvre le fait de parler pour parler (parler pour ne rien dire), cela n’apporte aucune information à la discussion.

Exemple :
P1 : “J’ai eu une bonne note en mathématiques.”
P2 : “*Humm*” / “*Oui*”
P1 : “J’ai eu la note de 15/20.”

Elaboration : explication : Donne une explication à la réponse.

Exemple :
P1 : “Je ne suis pas de bonne humeur, *il pleut et je suis mouillée*.”

Elaboration : prescription : Ordre formel et détaillé énumérant ce qu’il faut faire / ensemble de règle et de conseil (“selon le code de la loi...”) / Il faut / On doit.

Exemple :
P1 : “J’ai mal à la tête, *il faut que je prenne un Efferalgan*.”

Evaluation : Donne une évaluation, une « note » sur la phrase.

Exemple :
P1 : “J’étais à une soirée hier, *c’était super bien*”

Relations obliques :

Question : Lorsqu'une question est posée.

Exemple :
P1 : "*De quelle couleur est ta chemise?*"
P2 : "Elle est rouge."

Méta-question : Question "méta", autrement dit une question qui porte sur la conversation elle-même plutôt que sur le contenu.

Exemple :
P1 : "Ce soir je vais au restaurant."
P2 : "*Peux-tu répéter ?*" / "*Tu as dit que tu allais au restaurant ?*"

Question : conduite : On reste neutre. On attend une élaboration de l'autre personne.

Exemple :
P1 : "Hier je suis sortie en boite !"
P2 : *"Ah ?"*

Demande d'élaboration : On pousse l'autre à demander plus d'explication et attend une élaboration de l'autre.

Exemple :
P1 : "Hier je suis sortie en boite !"
P2 : *"Et alors ? "*

Conduite : Il s'agit d'une relance qui peut être interrogative ou non.

Exemple :
P1 : "Hier je suis sortie en boite !"
P2 : *"Ah."*

Contre-élaboration : Manifeste un désaccord, ce qui appelle à une réponse. Une riposte est attendue.

Exemple :
P1 : "Je parle cinq langues différentes"
P2 : "*C'est impossible !*"
P1 : "Si, je parle l'anglais, le français, l'espagnol, le chinois et l'italien"

Création des relations sur Glozz

Pour créer une nouvelle relation, cliquez d'abord sur le bouton correspond dans la barre des menus. Dans la vue 5, le modèle d'annotations, les différents modèles de relations se mettent en surbrillance. Cliquez sur celui voulu, et, ensuite, cliquez sur les deux unités que vous voulez relier. Attention, les relations ont un sens, en général du haut vers le bas. Votre relation est donc maintenant créée. Les unités à relier sont celles que vous avez créées, il ne faut donc pas relier les tours de parole !
La première unité sélectionnée (celle du début du texte) est à relier avec la section de phrase "Début". Si vous ne savez pas à quoi rattacher certaines unités, il faut les relier avec ce "Début".

Pour la relation "Élaboration", vous devriez ensuite choisir dans la vue 4 le type de l'élaboration. Si vous ne savez pas, laissez la valeur par défaut.

Pour modifier une relation, vous pouvez cliquer sur le bouton correspondant. Ensuite, cliquez sur la relation. En appuyant sur la touche "Suppr", vous pourrez supprimer la relation. Sinon, vous pouvez modifier l'unité de départ ou d'arrivée en cliquant sur la bulle correspondante au début ou à la fin de la relation, puis de faire un drag & drop pour modifier la position vers une autre unité.

Visualisation de l'arbre

Afin de vous permettre de visualiser votre travail sous forme d'arbre, voici les étapes à suivre :

- Dans la vue 3, vous trouverez sur gauche ce bouton :  . Cliquez dessus, puis faîtes un double-clic sur le seul fichier présent dans la fenêtre (qui devrait avoir comme nom "nom_texte.aa"). Cliquez sur "Enregistrer" si cela est nécessaire.
- Ensuite, dans la petite fenêtre ressemblant à ceci,

Cliquez sur le bouton "Rafraîchir". Attendez un peu, et une image va s'ouvrir, vous pourrez alors voir l'état d'avancement de votre arbre. Fermez l'image pour continuer à travailler.

Vous pouvez répéter ces étapes à tout moment si vous voulez voir votre arbre.

Sources

Manuel de Glozz : http://glozz.free.fr/glozzManual_1_0.pdf

Guide d'annotation pour l'identification des discontinuités décisives dans la transcription des entretiens entre schizophrène et psychologue, M. Amblard, M. Musiol, K. Fort, M. Rebuschi, N. Wittmann, K. Rivalin

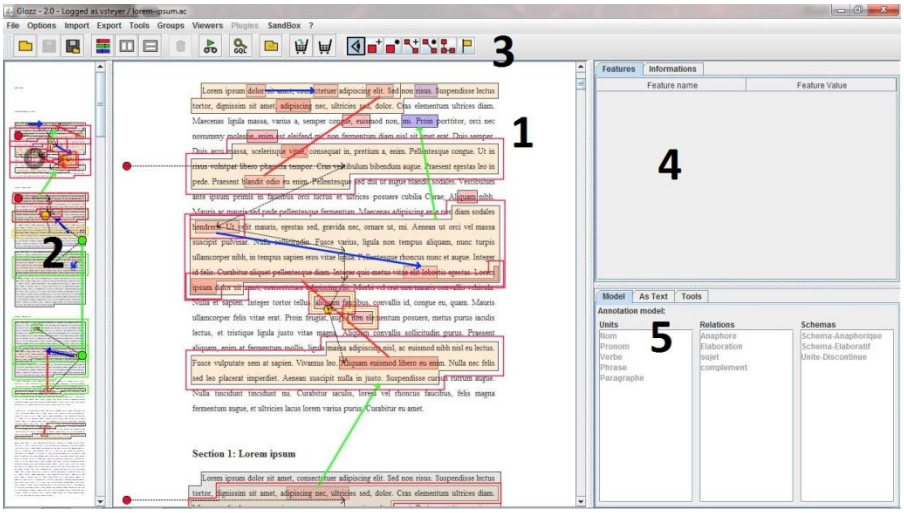
Annexe 3 : Fiche de synthèse

Fiche de synthèse

I Démarche à suivre

Vous êtes dans le rôle d'un annotateur, et vous devez annoter un texte grâce à des unités et des relations.

II Interface de Glozz



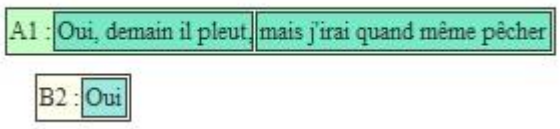
1 : Vue principale du texte
2 : Vue globale du texte et de ses annotations,
3 : Barre des menus.
4 : Tableau contenant les différentes valeurs pour une annotation sélectionnée.
5 : modèle d'annotation, c'est la liste de tous les types d'annotations disponibles

Boutons :

 pour être en mode sélection,  pour créer une unité,  pour modifier une unité,  pour créer une relation,  pour modifier une relation.

III Création unité

Définition : phrase ou une partie de phrase, contenue dans un seul acte de langage.



IV Relations théoriques

Définition : Relation entre une ou plusieurs unités.

Relations horizontales	Relations verticales	Relations obliques
Narration	Elaboration	Question
Réponse	Elaboration : explication	Question : conduite
Réponse phatique	Elaboration : prescription	Méta-question
Suite	Evaluation	Demande d'élaboration
	Phatique	Conduite
		Contre-élaboration

Annexe 4 : Extrait annoté par les différentes personnes

Début
A1 : Voilà mais bon il fait que je m'y remette quand même... enfin que je m'y remette je / je / je je m'y mette... voilà quoi.
B2 : Pourquoi ça vous plait pas... c'est tous les transports ?
A3 : Non / non / non c'est pas tous les transports c'est... voilà c'est / c'est / c'est le fait de conduire d'être / d'être responsable d'un véhicule et pouvoir faire du mal à quelqu'un ça me fait ça me fait peur...
B4 : D'accord.
A5 : et c'est une grosse responsabilité pour moi...
B6 : Mmh mmh.
A7 : ... ouais... c'est ouais voilà ouais c'est ça... c'est c'est et c'est souvent j'ai peur sur la route quand une voiture arrive à fond et puis qu'elle ralentit d'un coup.
B8 : D'accord.
A9 : Et puis moi je sais pas à quoi m'attendre et... voilà je risque de faire un peu n'importe quoi... aussi y a aussi ça mais bon... voilà... mais bon je compte le passer un jour hein... on verra hein... plus tard /... plus tard on verra mmh.
B10 : D'accord.
A11 : ...
B12 : Et donc ?
A13 : Et donc voilà / donc voilà.
B14 : Et après vous avez eu ...
A15 : Ben dès que je suis souriante ça va.
B16 : Et donc vous vous déplacez majoritairement à pied ? ou euh.
A17 : A pied en bus.
B18 : En bus ?
A19 : Oui en bus métro enfin on est très bien desservi ici.
B20 : Ouais / ouais.

Annexe 5 : Fichier CSV des statistiques

Personne	Sexe	Age	Formation	Nombre unites	Nombre themes	Theme + recurrent	Nombre relations	Nombre relations verticales	relations horizontales	Nombre relations obliques	Nombre Narration	Nombre reponse	Nombre reponse phatique	Nombre suite	Nombre elaboration
1	Homme	23 ans	M1 Psycho	26		7 Thème non c		23	8	11	4	3	2	1	5	4
2	Homme	23 ans	En préparati	37		10 transport		36	19	7	10	1	3	3	0	13
3	Homme	21 ans	L3 Sciences C	30		7 transport		29	9	12	8	6	3	0	3	3
4	Homme	53 ans	Administrate	29		16 ok		28	1	19	8	5	3	10	1	1
5	Femme	50 ans	Technicien t	27		4 moyen de lo		26	5	13	8	2	3	8	0	5
6	Femme	23 ans	M1 SCA	37		32 confirmatior		26	10	1	15	0	0	0	1	5
7	Femme	22 ans	M1 SCA	35		19 raison depla		35	12	7	16	0	0	0	7	2
8	Femme	22 ans	M1 SCA	34		5 permis		33	21	3	9	0	0	0	3	12
9	Homme	21 ans	M1 SCA	36		9 doute		35	6	5	24	4	0	0	1	3
10	Femme	24 ans	M1 Psycho	43		11 Thème non c		42	27	9	6	0	7	2	0	15
11	Homme	22 ans	M1 SCA	41		6 Peur		40	13	21	6	8	5	6	2	4
12	Femme	22 ans	M1 SCA	26		4 conduite		25	8	8	9	1	3	2	2	7
13	Homme	26 ans	M1 SCA	35		12 réponse neu		34	11	16	7	0	5	6	5	9
14	Femme	20 ans	L3 Allemand	33		11 la conduite		32	13	14	5	2	2	10	0	8
15	Femme	24 ans	M1 Psycho	37		7 la conduite		36	11	19	6	14	3	0	2	3
16	Homme	22 ans	M1 SCA	30		11 transports		29	6	15	8	5	6	0	4	1
17	Homme	21 ans	M1 SCA	38		12 transport		36	16	15	5	2	3	2	8	8
18	Homme	21 ans	L2 Sciences C	33		11 moyen de tr		33	5	18	10	0	9	1	8	2
19	Homme	25 ans	M1 SCA	27		10 peur		26	8	14	4	7	5	0	2	1
20	Homme	24 ans	Ingénieur in	30		12 mode de dé		29	12	11	6	3	3	1	4	4
21	Femme	21 ans	M1 Miage	55		15 hésitation		62	11	39	12	22	1	2	14	9
22	Homme	25 ans	Préparation	38		7 risques liés à		37	26	6	5	2	3	1	0	12

Nombre elaboration explication	Nombre elaboration prescription	Nombre evaluation	Nombre phatique	Nombre question	Nombre question conduite	Nombre metaquestion	Nombre demande elaboration	Nombre conduite	Nombre contre elaboration	Temps
0	0	0	4	1	0	1	2	0	0	42
0	0	0	6	2	2	0	1	4	1	39
0	0	0	6	0	0	2	2	4	0	38
0	0	0	0	2	2	1	2	1	0	79
0	0	0	0	2	2	0	0	4	0	52
0	0	0	5	2	0	0	0	1	0	48
0	0	0	10	1	0	0	0	0	0	40
0	0	0	9	2	1	0	0	0	0	43
0	0	0	3	2	1	0	0	6	0	42
0	0	0	12	2	1	0	2	1	0	36
0	0	0	9	2	0	0	2	2	0	67
0	0	0	1	0	0	1	3	5	0	24
0	0	0	2	3	2	0	2	0	0	46
0	0	0	5	2	2	0	1	0	0	29
0	0	0	8	3	0	0	2	1	0	104
0	0	0	5	4	1	1	0	2	0	54
0	0	0	8	3	0	0	1	1	0	58
0	0	0	3	2	0	0	4	4	0	55
0	0	0	7	1	0	0	3	0	0	40
0	0	0	8	1	2	0	1	1	1	31
0	0	0	2	2	1	0	5	4	0	92
0	0	0	14	2	0	0	3	0	0	39

Annexe 6 : Liste de tous les thèmes utilisés par les annotateurs

transport : 62	la peur liée à la conduite : 2
thème non défini : 43	apréhension : 2
peur : 42	attente : 2
responsabilité : 37	oui : 2
transports : 36	le moyen de transport : 2
permis : 25	faire du mal : 2
moyen de transport : 24	phatique : 2
conduite : 20	accident : 2
la conduite : 16	projet : 2
mode de déplacement : 14	faire mal : 2
risques liés à la conduite : 14	pied : 2
permis de conduire : 12	relance : 2
moyen de locomotion : 11	véhicules : 2
passage du permis : 10	accord : 2
hésitation : 9	le moyen de déplacement : 2
peur de l'inconnu : 8	état d'esprit : 1
les transports : 8	poids sur les épaules : 1
doute : 8	réflexion interne : 1
appréhension : 8	joie : 1
tique de langage : 7	l'incertitude : 1
répétition : 7	relance question : 1
réponse cohérente : 7	silence : 1
volonté : 7	responsabilité / conduite : 1
réponse neutre : 7	rapport choucroute/patate : 1
le passage du permis : 6	constatation : 1
danger : 6	réponse négative : 1
mode de transports : 6	étonnement : 1
situation : 6	ne pas apprécier quelque chose : 1
état émotionnel : 6	reflexion interne : 1
motivation : 6	compréhension : 1
raison deplaisance transport : 6	référence implicite au permis : 1
motivation (délayée) : 6	volonté de conduire : 1
expressif : 5	explication de la peur : 1
avoir peur : 5	quand passer le permis : 1
futur : 5	absurde : 1
souhaite plus d'informations : 5	problèmes personnels : 1
affects négatifs : 5	vehicule : 1
peur de soit : 5	reponsabilité : 1
conduite de voiture : 5	ce qu'il doit faire : 1
reprendre une activité : 4	demande de précisions : 1
décisions futures : 4	contradiction : 1
? : 4	soi-même : 1
complément : 4	constat personnel : 1
mode de deplacement : 4	cause - conséquence : 1
craintes : 4	responsabilité / danger : 1

travail : 4 encouragement à poursuivre : 4 question : 4 bus : 4 conduire : 4 trouble : 4 ok : 4 responsable : 4 question neutre : 4 réponse : 4 transport et responsabilité : 3 confirmation : 3 représentation mentale d'une action : 3 transports en commun : 3 transport en commun : 3 peur de conduire : 3 le permis : 3 rejet : 3 réponse phatique : 2 etat : 2 la responsabilité de la conduite : 2 crainte : 2	siutation : 1 relance sure réponse : 1 cause problème conduire : 1 desserte : 1 avenir : 1 type deplacement : 1 bégayement / stress : 1 absence de dissociation / confusion : 1 guide la réponse / incitation à continuer la phrase : 1 raisons déplaisance transport : 1 reporter à plus tard / éloignement peur : 1 déplacement : 1 problèmes personnels/traumatisme : 1 phrase neutre : 1 l'humeur : 1 passer le permis : 1
Total : 735	

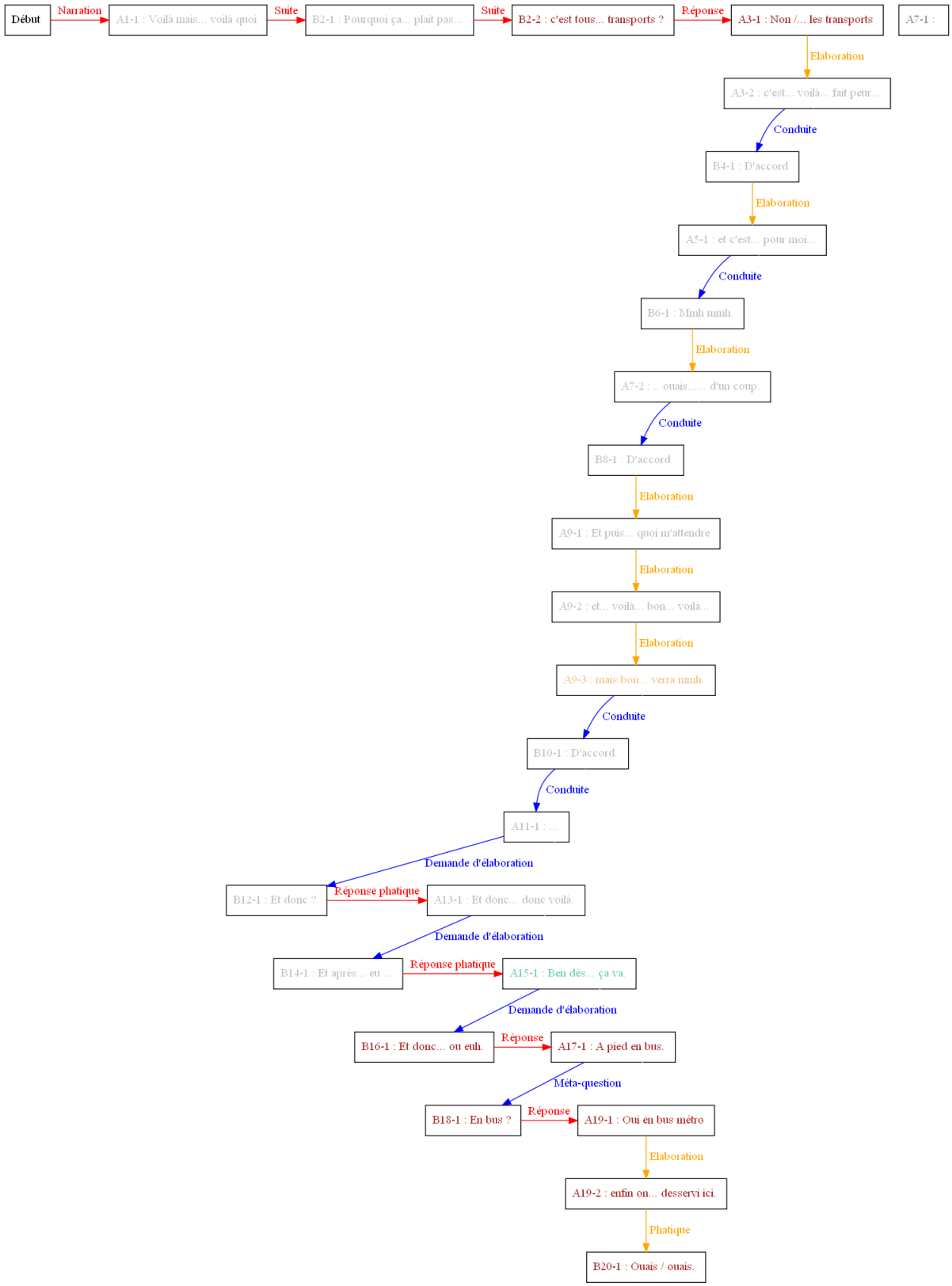
Annexe 7 : Statistiques en fonction du sexe de l'annotateur

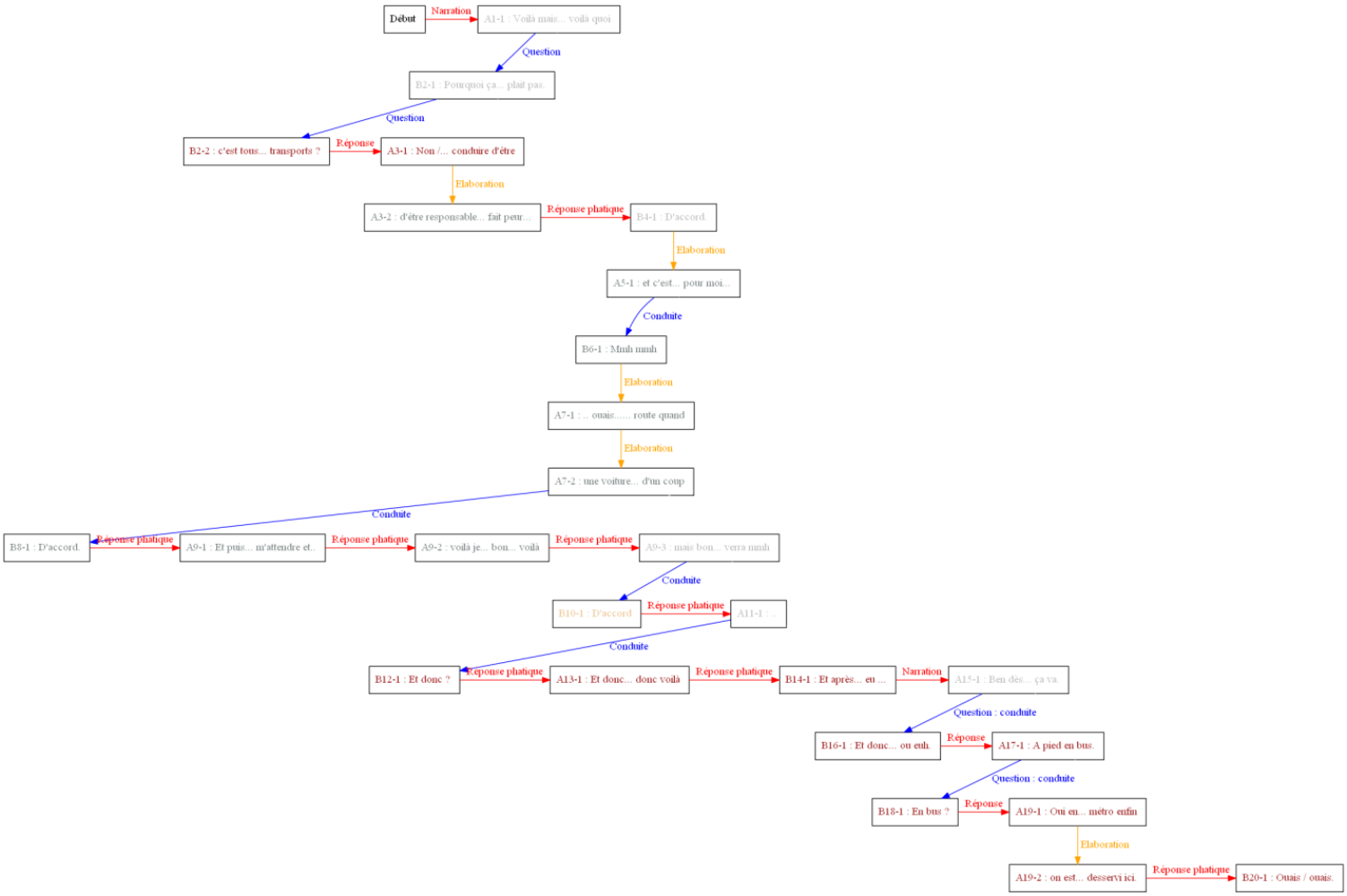
	Nombre unités	Nombre thèmes	Nombre relations	Nombre relations verticales	Nombre relations horizontales	Nombre relations obliques
Femme	34	11,81818182	35,22222222	13,11111111	13,18181818	9,25
Homme	33,07692308	10	31,92307692	10,76923077	13,07692308	8,2

Annexe 8 : Statistiques en fonction de la catégorie de l'annotateur

	Nombre unités	Nombre thèmes	Nombre relations	Nombre relations verticales	Nombre relations horizontales	Nombre relations obliques
M1 Psycho	35,33333333	8,333333333	33,66666667	15,33333333	13	5,333333333
Non étudiant	32,54545455	9,8	31,2	12,6	12,6	7,4
M1 SCA	33,9	12	31,9	11,1	10,5	10,3
L2 / L3	32	9,666666667	31,33333333	9	14,66666667	7,666666667

Annexe 9 : Quelques exemples d'arbres par les annotateurs




Début
Narration
A1-1 : Voilà mais... voilà quoi
Question
B2-1 : Pourquoi ça... plaît pas.
Question
B2-2 : c'est tous... transports ?
Réponse
A3-1 : Non /... conduire d'être
Elaboration
A3-2 : d'être responsable... fait peur...
Réponse phatique
B4-1 : D'accord.
Elaboration
A5-1 : et c'est... pour moi...
Conduite
B6-1 : Mmh mmh
Elaboration
A7-1 : .. ouais..... route quand
Elaboration
A7-2 : une voiture... d'un coup
Conduite
B8-1 : D'accord.
Réponse phatique
A9-1 : Et puis... m'attendre et..
Réponse phatique
A9-2 : voilà je... bon... voilà
Réponse phatique
A9-3 : mais bon... verra mmh
Conduite
B10-1 : D'accord
Réponse phatique
A11-1 : ...
Conduite
B12-1 : Et donc ?
Réponse phatique
A13-1 : Et donc... donc voilà
Réponse phatique
B14-1 : Et après... en ...
Narration
A15-1 : Ben dès... ça va.
Question : conduite
B16-1 : Et donc... ou euh
Réponse
A17-1 : A pied en bus.
Question : conduite
B18-1 : En bus ?
Réponse
A19-1 : Oui en... métro enfin
Elaboration
A19-2 : on est... desservi ici.
Réponse phatique
B20-1 : Ouais / ouais.

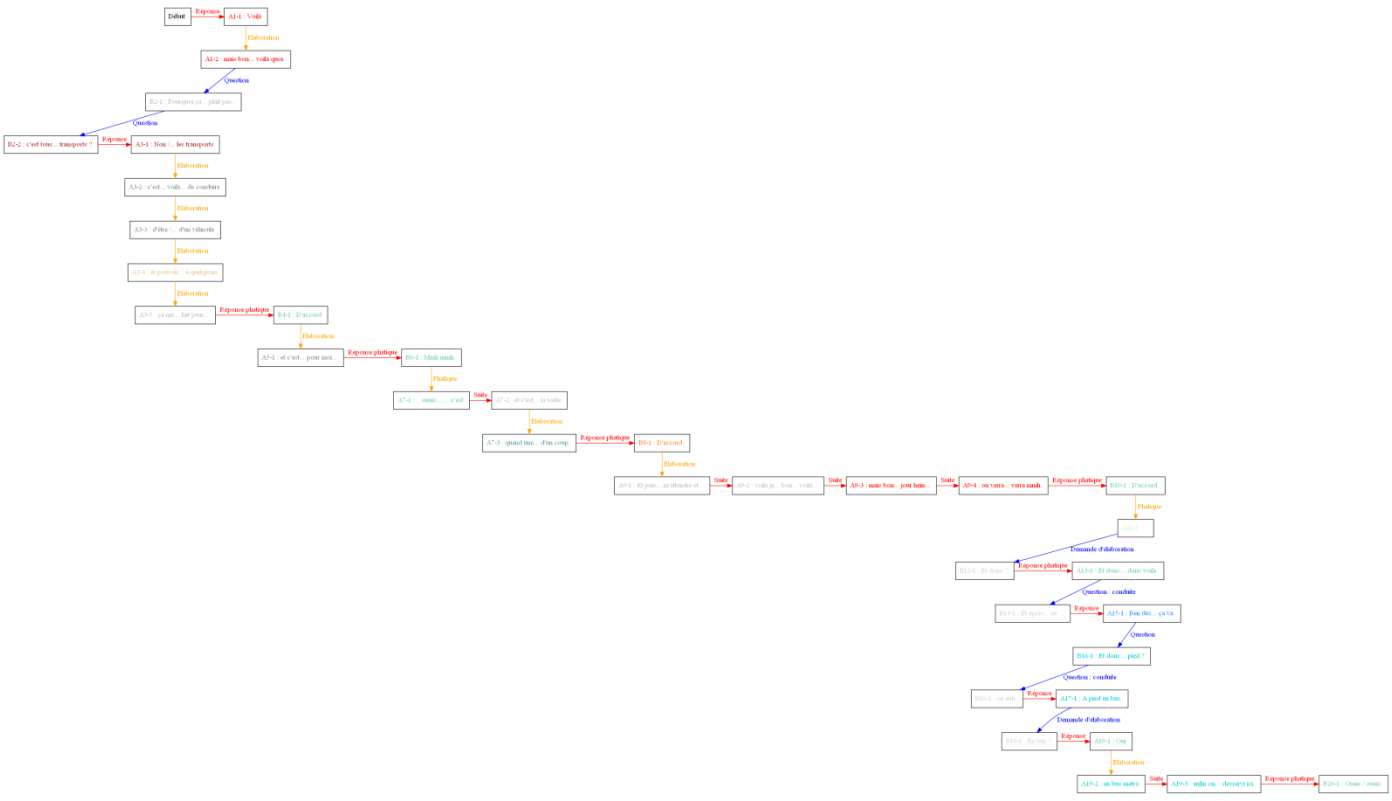


Table des matières