

Fiche de projet tutoré / Project form

TLP : campagne d'annotation

Encadrement / Supervisors

- encadrant principal / main supervisor

Manuel Rebuschi LHSP-AHP manuel.rebuschi@univ-lorraine.fr

- autres encadrants / other supervisors

Maxime Amblard Sémagramme-Loria maxime.amblard@univ-lorraine.fr

Michel Musiol Discours - Atilf michel.musiol@univ-lorraine.fr

Description / Description

Contexte

Ce projet tutoré s'inscrit dans le projet de de recherche SLAM (Schizophrénie 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. Nous utilisons une extension de la DRT (Discourse Representation Theory), la S-DRT, qui construit un arbre de relations discursives et conversationnelles. Cette théorie se base sur le calcul de la sémantique compositionnelle et tente de rendre compte d'aspects pragmatiques de manière simplifiée.
2. D'autre part, les psycholinguistes 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 utilisent 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 qui vise à modéliser les entretiens entre patients schizophrènes et psychologues. Les patients schizophrènes sont rencontrés dans un contexte médical. Ils passent plusieurs tests cognitifs, ainsi qu'un entretien supervisé. Par ailleurs, un groupe contrôle est apparié sur le celui des patients et passe le même protocole. Les données collectées permettent de produire un nouveau corpus constitué des transcriptions des entretiens, ainsi que de 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 annotation sémantico-pragmatique formelle, basée sur la SDRT, des extraits discontinus précédemment identifiés.

Le but du projet est de conduire une campagne d'annotations en SDRT (la 3^e couche) qui soit conforme aux règles de l'art. Une difficulté est en effet de produire une annotation convergente entre annotateurs naïfs qui permette de corroborer les intuitions des psycholinguistes relativement aux discontinuités conversationnelles.

Réalisation

Un projet tutoré 2016-2017 a permis de finaliser l'adaptation de la plate-forme GLOZZ et le protocole d'annotation, et de conduire une première campagne avec des résultats intéressants. Le projet 2017-2018 se focalisera sur l'analyse des résultats. Il s'agira concrètement de mettre en œuvre le protocole pour conduire une campagne élargie, et d'analyser l'intégralité des résultats extraits des deux campagnes.

Informations diverses : matériel nécessaire, contexte de réalisation / Various information: material, context of realization

La campagne d'annotation sera conduite à l'aide de la plateforme d'annotation GLOZZ dans une version adaptée lors du précédent projet tutoré (2016-2017).

Etapes du projet :

1. Biblio : lectures sur le contexte scientifique (SLAM, la SDRT, méthodologie des campagnes d'annotation)
2. Prise en main de la plateforme GLOZZ et des résultats du précédent projet tutoré
3. Mise en place et suivi d'une campagne d'annotation en SDRT auprès d'annotateurs :
 - recrutement d'un groupe d'annotateurs
 - conduite de la campagne d'annotation
4. Analyse des résultats

Livrables et échéancier / Deliverable and schedule

- novembre : biblio
- février : prise en main de Glozz
- février : recrutement des annotateurs
- mars : campagne
- avril-mai : analyse et rédaction du rapport

Bibliographie / References

- [1] M. Amblard, M. Musiol, and M. Rebuschi. Une analyse basée sur la S-DRT pour la modélisation de dialogues pathologiques. In *Traitement Automatique des Langues Naturelles - TALN 2011*, page 6, Montpellier, France, June 2011.
- [2] M. Amblard, M. Musiol, and M. Rebuschi. Schizophrénie et Langage : Analyse et modélisation. De l'utilisation des modèles formels en pragmatique pour la modélisation de discours pathologiques. In *Congrès MSH 2012*, Caen, France, Dec. 2012.
- [3] N. Asher and A. Lascarides. *Logics of Conversation*. Studies in Natural Language Processing. Cambridge University Press, 2003.
- [4] C. Barras, E. Geoffrois, Z. Wu, and M. Liberman. Transcriber : a free tool for segmenting, labeling and transcribing speech. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC)*, pages 1373–1376, Granada, Spain, May 1998.
- [5] O. Baude, C. Blanche-Benveniste, M.-F. Calas, P. Cappeau, P. Cordereix, L. Goury, M. Jacobson, I. De Lamberterie, C. Marchello-Nizia, and L. Mondada. *Corpus oraux, guide des bonnes pratiques 2006*. CNRS Editions, Presses Universitaires Orléans, 2006.
- [6] C. Benzitoun, K. Fort, and B. Sagot. TCOF-POS : un corpus libre de français parlé annoté en morphosyntaxe. In *Traitement Automatique des Langues Naturelles (TALN)*, pages 99–112, Grenoble, France, 2012.
- [7] K. Fort. *Les ressources annotées, un enjeu pour l'analyse de contenu : vers une méthodologie de l'annotation*

manuelle de corpus. PhD thesis, Université Paris XIII, LIPN, INIST-CNRS, December 2012.

[8] L. Huber, E. Laurier, *Trouble du langage et de la pensée : Campagne d'annotation*, Université de Lorraine, juin 2017

[9] H. Kamp and U. Reyle. *From Discourse to Logic*. Kluwer Academic Publishers, 1993.

[10] V. Steyer, M. Ruez, *Analyse des Troubles du Langage et la Pensée*, Université de Lorraine, juin 2015