

Compte rendu réunion GP2 MULTICOM - 15 juin 2017 - ILF, Paris

Participants

Antonio Balvet, Brigitte Bigi, Dominique Boutet, Stéphanie Caët, Oralie Cattan, Myriam Charpentier, Vanda Enoiu, Achille Falaise, Lydia-Mai Ho-Dac, Julien Longhi, Sara Mazziotti, Céline Poudat, Isabelle Rousset, Marie-Anne Sallandre, Ciara Wigham

Déroulement

La réunion s'est appuyée sur l'atelier Corpus annoncé à l'OJ et a été le lieu de nombreux échanges. Les présentations sont disponibles ici

https://docs.google.com/presentation/d/1CWwaLmP8_sL2ZG0OgJgl7fQ4ZiHdqddL-pTtlbeO8rl/edit#slide=id.p

Actions et livrables pour 2017-2018

Voici les actions et les livrables qui ont émergé de la réunion, ainsi que les participants qui se sont portés volontaires pour la coordination. Si vous étiez absent(e) à cette réunion, mais que vous êtes intéressés pour participer à la coordination de l'une des actions, n'hésitez surtout pas.

Livrables en ligne

1. Table des usages (coord. C. Poudat)

- Modifier le lien vers la table des usages (corli plutôt que corpus-écrits);
 - <http://explorationdecopus.corpusecrits.huma-num.fr/>
- Pointer vers la table depuis le site CORLI, et nos sites respectifs;
- Rendre cette table vivante.

2. Inventaire des guides d'annotation disponibles (coord. M.-A. Sallandre)

Toute entreprise d'annotation devrait s'accompagner d'un guide d'annotation détaillant les choix et les annotations posées sur les corpus. Néanmoins, il s'avère que le guide d'annotation est souvent ce que l'on rédige en dernier, et faute d'une version *parfaitement finalisée*, on tarde souvent à le mettre en ligne.

Lorsqu'on annote un corpus et qu'on se lance dans l'écriture d'un guide d'annotation, on récupère les guides disponibles en ligne; il serait donc pertinent de pointer vers les guides d'annotation existants depuis le site de CORLI, dans le cadre d'une rubrique dédiée. Il faudra rappeler sur le site qu'un guide d'annotation est une publication acceptée par l'HCERES et pointer donc vers les liens HAL de ces guides, de même que rajouter des mots clés (e.g. type de discours: LSF, écrits scolaires, projet, logiciel, état d'avancement + descriptif libre optionnel etc.).

3. Bonnes pratiques pour la constitution de corpus (avec bêtisier) - coord. B. Bigi / A. Falaise / C. Poudat

Exemples pour les corpus vidéo:

- CLAP du début, lumière qui s'allume toutes les minutes
- enregistrer aussi l'audio par la caméra
- téléphone portable: recherche Hertz s'entend
- choix du micro (pas cravate)

J. Lascar: J'accompagne depuis plus de 10 ans les étudiants et les chercheurs dans la constitution de leurs corpus vidéo.

Je serai ravie de vous accompagner sur ce projet, ayant beaucoup (trop :-)) d'exemples bêtisiers lors de la constitution de corpus vidéo, (oubli d'allumer le micro, micro posé à côté d'une stéréo musique d'ambiance, fermeture de l'écran qui met la camera en veille....)

Bref n'hésitez pas à me compter dans le ce groupe de travail.

De même nous organisons avec mon collègue Daniel Valero plusieurs formations de méthodologie, ingénierie de corpus et logiciels (matériels AV, transfert - synchro et montage de corpus vidéo, anonymisation des vidéos, Transana, Nvivo...), si vous avez besoin de formateurs n'hésitez pas à nous contacter.

C. Vincent: Suite à la lecture du compte-rendu envoyé par Céline, je vous informe que je suis particulièrement intéressée par le livrable 3 : "Bonnes pratiques pour la constitution de corpus (avec bêtisier)".

Si cela peut vous servir, j'ai mis à disposition sur HAL une présentation à ce sujet :

<https://hal.archives-ouvertes.fr/hal-01225952>

J'y traite de l'enregistrement des données audio, vidéo, mocap et oculométrie. J'y parle peu de la synchronisation des données multimodales, mais je suis par ailleurs en train d'écrire un chapitre qui aborde rapidement ce point et partagerai bien volontiers avec vous les informations que j'ai glanées à ce sujet.

Documents de travail et ouvrages

4. Feuille de route décrivant les besoins en termes de méthodes, calculs, visualisations (Coord. A. Balvet, B. Bigi, M. Ho-Dac, C. Poudat)

Commencer à démarrer une feuille de route décrivant les méthodes et les besoins en termes de calculs, de visualisation.

5. Ouvrage sur la visualisation des données linguistiques (coord. B. Bigi, D. Boutet)

Ce projet d'ouvrage est parti d'une remarque de B. Bigi sur le travail très spécifique des informaticiens travaillant sur des données linguistiques. Les développements informatiques réalisés (e.g. la visualisation par filtres successifs de Brigitte) sont souvent difficiles à publier, tant en informatique qu'en linguistique, malgré le travail effectué et les innovations proposées. On manque donc d'un lieu de publication de travaux de ce type et l'idée d'un ouvrage a émergé avant, peut-être d'envisager une revue en ligne par exemple.

Editeurs possibles: OpenEdition, research publishing net

6. Textes et annotations (coord. J. Longhi, C. Poudat)

Il s'agit ici de faire la synthèse des relations entre le texte et les annotations que l'on y pose en précisant les types de textes (audio, vidéo, multimodaux, écrits etc.) et les types d'annotations - les réponses, les commentaires, les corrections (e.g. écrits scolaires, commentaires sur des écrits collaboratifs de type gdoc ou sur des révisions Wikipedia), les discussions (e.g. les discussions éditoriales Wikipedia).

Cette question semble en effet transversale à nos données.

Journées et événements envisagés

7. Journée d'échanges / formation sur les bonnes pratiques (coord. C. Wigham)

Organiser une formation sur les bonnes pratiques: votre corpus est constitué, que faire?
Archivage, diffusion, stockage, valorisation.

8. Challenge Visualisation des données, en coopération avec des designers (coord. M. Ho-Dac, A. Balvet, D. Boutet)

- Challenge auprès des designers. Dominique demande une évaluation auprès de ses collègues designers.

Ch. Benzitoun et Céline Dugua ont également des contacts qui pourraient être utiles.

Le Challenge donnerait lieu à un événement qui se tiendrait dans le sud de la France (e.g. Corse, Biarritz, Fréjus, lieux dans lesquels il serait peut-être possible de réserver une villa CNRS).

9. Data session (coord. M.-A. Sallandre)

Organisation d'un défi annotation avec ELAN, dans le cadre d'un Workshop de JEP/TALN 2018.

Ce serait aussi l'occasion d'organiser une session de formation ELAN Niveau 3, plébiscitée par les membres du groupe.