Oralisation de l'écrit ou routines de l'interaction ?
Les phrases préfabriquées des interactions dans les discussions Wikipédia,
du contraste français-allemand aux grands modèles de langue

Céline Poudat (Université Côte d'Azur, UMR Bases, Corpus, Langage**)**

Les pages de discussion Wikipédia occupent une position singulière dans la communication médiée : purement graphiques, elles ont des caractéristiques propres, tout en mimant une pluralité de genres écrits et oraux (courriel, clavardage), produisant un écrit parfois fortement oralisé. Nous faisons ainsi l'hypothèse d'un gradient d'oralisation dans la CMC, que la comparaison des discussions wiki avec les corpus de CoMeRe permet de situer empiriquement. Nous discuterons ce gradient en lien avec la synchronicité des échanges, sans y voir de déterminisme : la présence de marqueurs dits oraux dépend tout autant de la situation de communication (registre, audience, normes du groupe) que des propriétés techniques du dispositif. Se pose alors une question plus fondamentale : ces formes partagées avec l'oral relèvent-elles d'une oralisation de l'écrit, ou des routines de l'interaction, qui ont été plus largement décrites à l'oral qu'à l'écrit ?

Les phrases préfabriquées des interactions (PPI) offrent une entrée privilégiée sur cette question : formes transversales aux interactions orales et écrites, elles permettent d'évaluer dans quelle mesure les routines de l'oral se sont greffées sur les interactions écrites, et quelles formes sont au contraire spécifiques à l'écrit médié. Les données du projet ANR PREFAB documentent cette double dynamique : certaines PPI massives à l'oral sont significativement absentes des discussions wiki, tandis qu'émergent des formes propres au genre, relevant notamment d'un registre de contestation argumentative. Parmi celles-ci, les PPI polémiques de clôture, nettement représentées dans un genre intrinsèquement argumentatif, ont fait l'objet d'une analyse contrastive français-allemand.

Nous présenterons enfin les premiers résultats d'une expérimentation consacrée aux capacités interprétatives des grands modèles de langue : identification et catégorisation des PPI par des modèles ouverts, selon plusieurs stratégies de prompting, évaluées sur nos données annotées. Ces modèles, entraînés sur ce type d'écrit mais alignés de manière à ne pas reproduire ses routines conflictuelles, peuvent-ils les interpréter sans les mobiliser ?