

# Le corpus BrainKT: Etudier l’instanciation du common ground par l’analyse des indices verbaux, gestuels et neurophysiologiques

Eliot Maës<sup>1</sup> Thierry Legou<sup>2</sup>

Leonor Becerra-Bonache<sup>1</sup> Philippe Blache<sup>2</sup>

(1) Aix Marseille Univ, CNRS, LIS, Marseille, France

(2) Aix Marseille Univ, CNRS, LPL, Aix-en-Provence, France

prenom.nom@lis-lab.fr, prenom.nom@univ-amu.fr

## RÉSUMÉ

---

La quantité croissante de corpus multimodaux collectés permet de développer de nouvelles méthodes d’analyse de la conversation. Dans la très grande majorité des cas, ces corpus ne comprennent cependant que les enregistrements audio et vidéo, laissant de côté d’autres modalités plus difficiles à récupérer mais apportant un point de vue complémentaire sur la conversation, telle que l’activité cérébrale des locuteurs. Nous présentons donc BrainKT, un corpus de conversation naturelle en français, rassemblant les données audio, vidéo et signaux neurophysiologiques, collecté avec l’objectif d’étudier en profondeur les transmission d’information et l’instanciation du common ground. Pour chacune des conversations des 28 dyades (56 participants), les locuteurs devaient collaborer sur un jeu conversationnel (15min), et étaient ensuite libres de discuter du sujet de leur choix (15min). Pour chaque discussion, les données audio, vidéo, l’activité cérébrale (EEG par Biosemi 64) et physiologique (montre Empatica-E4) sont enregistrées. Cet article situe le corpus dans la littérature, présente le setup expérimental utilisé ainsi les difficultés rencontrées, et les différents niveaux d’annotations proposés pour le corpus.

## ABSTRACT

---

### **Studying common ground instantiation using audio, video and brain behaviours : the BrainKT corpus**

An increasing amount of multimodal recordings has been paving the way for the development of a more automatic way to study language and conversational interactions. However this data largely comprises of audio and video recordings, leaving aside other modalities that might complement this external view of the conversation but might be more difficult to collect in naturalistic setups, such as participants brain activity. In this context, we present BrainKT, a natural conversational corpus with audio, video and neuro-physiological signals, collected with the aim of studying information exchanges and common ground instantiation in conversation in a new, more in-depth way. We recorded conversations from 28 dyads (56 participants) during 30 minutes experiments where subjects were first tasked to collaborate on a joint information game, then freely drifted to the topic of their choice. During each session, audio and video were captured, along with the participants’ neural signal (EEG with Biosemi 64) and their electrophysiological activity (with Empatica-E4). The paper situates this new type of resources in the literature, presents the experimental setup and describes the different kinds of annotations considered for the corpus.

**MOTS-CLÉS :** corpus, multimodalité, conversation naturelle, common ground, EEG.

**KEYWORDS:** corpus, multimodal, natural conversation, common ground, EEG.

---



## Références

MAËS E., LEGOU T., BECERRA L. & BLACHE P. (2023). Studying common ground instantiation using audio, video and brain behaviours : the brainkt corpus. In *Proceedings of the 14th International Conference on Recent Advances in Natural Language Processing*, p. 691–702.