Semi-Structured Interview Analysis: A French NLP Approach for Social Sciences

Robin Quillivic[1], Charles Payet[2]

[1]EPHE, CESSP (UMR 8209), ISC-PIF (UAR 3611) – robin.quillivic@ephe.sorbonne.fr

[2]ENS Paris-Saclay, ISP (UMR 7220) – charles.payet@ens-paris-saclay.fr

Abstract

Semi-structured interviews (SSIs) are widely employed in social sciences, facilitating in-depth data collection while maintaining study focus. Typically involving asymmetric roles (an interviewer and interviewee), SSIs follow predefined guides. However, current analyses predominantly scrutinize participant responses, overlooking the interviewer's language, context, and interactions. Even within guided interactions, discourse is a co-construction, yet the investigator is often rendered invisible. Drawing on conversation analysis principles, we present a French toolbox for analyzing SSIs in social sciences. Integrating multiple Natural Language Processing (NLP) approaches—semantic, thematic, and grammatical similarity—the toolbox enhances research methodology by offering precise indicators of interaction influence in interviews. The toolkit comprises three main components. The first provides a comprehensive dataset overview, including interview structure (number and length of speaking turns), and user-provided metadata. The second module analyzes the interview theme and categorizes the dialogic act using few-shot learners. The third module focuses on discussion alignment, building features using cosine similarity, and sentence embeddings. To demonstrate the toolkit's utility, we present two examples using interview transcripts produced as part of a research program on the memory of the November 2015 Paris attacks. The first explores interaction variability based on the socio-demographic profiles of the interviewer and respondent, while the second focuses on psycholinguistics, measuring how interview structure and theme are influenced by avoidance symptoms and PTSD.

Keywords: semi-structured interviews; interviewer-respondent interaction; conversation analysis; NLP; toolkit; computational social sciences