
Towards computational segmentation of impromptu stories in biographical interviews

Fritz Kliche, Michael Corsten, Laura Maleyka & Ulrich Heid
Universität Hildesheim

[kliche, corsten, maleyka, heidul]@uni-hildesheim.de

Sociological analyses of impromptu stories produced in biographical interviews have so far mainly been carried out manually, both quantitatively (e.g. Habermas & Reese 2015) and qualitatively (e.g. Corsten & Maleyka 2024).

The analyzed life stories show interesting patterns: periods in the interviewees' lives are often narrated as a sequence of chronicle-like 'chapters' (Habermas & Reese 2015: CHR), plus the discussion of single events (SEN) in the narrated time. The manual analyses suggest that the latter follow a narrative organization as described by Labov & Waletzky (1967): (1) introduction, (2) (pre-)orientation, (3) complication/crisis, (4) resolution, (5) coda.

The objective of our work is to develop computational linguistic procedures for CHR and SEN identification. In addition to these two levels, we also segment the narrations into propositions (one inflected verb per proposition, cf. Vauth et al. 2021); and we classify them a.o. according to narrative coherence (e.g. indicated by shifts to a new 'chapter'), temporal coherence (e.g. reference to time points), report, motivational coherence (e.g. causative or adversative expressions), as well as metacommunicative statements (e.g. comments for the interviewer). This fine-grained classification is intended to guide the CHR and SEN classification.

Our data come from over 300 transcribed life stories of diverging length (median: 1437 words), of which so far 16 have been manually analyzed in detail and used as gold data for classifier training.

In a kind of inside-out procedure, the computational tasks are (i) to automatically segment the interview transcripts into propositions, (ii) to automatically classify these propositions, and later to identify (iii) SEN and CHR. We report about the first two tasks: for (i), we compare syntax-based and LLM-based segmentation into propositions; for (ii) we report about both 'traditional' ML-based and LLM-based classification experiments, and we discuss the nature and distribution of indicators for tasks (ii) and (iii).

References: • Corsten, M & L. Maleyka (2024). Die Präsentation der biographischen Visitenkarte. Erzählstimulus und Selbsteinführung am Interviewbeginn. In *BIOS* 37(1+2), 157–181. • Habermas, T. & E. Reese (2015). Getting a Life Takes Time: The Development of the Life Story in Adolescence, Its Precursors and Consequences. In *Human Development*, 58, 172–201. • Labov, W. & J. Waletzky (1967). Narrative analysis. Oral versions of personal experience. In J. Helm (ed.), *Essays on the verbal and visual arts*. University of Washington Press, 12–44. • Vauth, M., Hatzel, H. O., Gius, E. & C. Biemann (2021). *Automated Event Annotation in Literary Texts*. CHR 2021.