

CORPUS comprises 800 television commercials, with 100 commercials sampled per decade, sourced from YouTube compilations, segmented into individual video files. The corpus is divided into a verbal and a visual subcorpus.

Stage/ Component	Description	Tool/ Methods	Output
Verbal Subcorpus	Textual transcripts of audiovisual commercials	1) Video-to-audio conversion; and 2) Automatic transcription using Whisper Large v2	800 verbal transcripts
Visual Subcorpus	Frame-based visual descriptions of commercials	1) Video decomposition into still frames; and 2) Automatic annotation using GPT-image-1.5	800 visual tagging files
Lexical Multidimensional Analysis (LMDA)	Separate analysis of verbal and visual subcorpora to identify discourse dimensions	Lexical Multidimensional Analysis 1) Content words are extracted 2) Frequency counts are compiled 3) Factor analysis is applied 4) High scores on each factor result in interpreted dimensions	Lexical dimensions of variation representing underlying discursive patterns
Multimodal Analysis	Integration of the two semiotic modes: verbal and visual dimensions	Canonical Correlation Analysis Cross-modal comparison and discourse-based interpretation	Multimodal discourse profiles
Statistical Analysis (ANOVAs)	Examining the diachronic variation in the verbal, visual, and cross-modal discourses across decades	Analysis of Variance (ANOVAs)	Statistically validated differences between dimensions and subcorpora