

ConversationAlign: An R package for Computing Linguistic Alignment and Corpus Analytics in Dyadic Conversation Transcripts

Jamie Reilly^{1, 2}, Benjamin Sacks², Virginia Ulichney², Gus Cooney³, and Chelsea Helion^{1,2}

¹ Department of Communication Sciences and Disorders, Temple University, United States ² Department of Psychology and Neuroscience, Temple University, United States ³ Wharton School, University of Pennsylvania, United States

DOI:

Software

- [Review](#) ↗
- [Repository](#) ↗
- [Archive](#) ↗

Submitted:

Published:

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC-BY](#)).

Summary

ConversationAlign executes a series of operations upon one or more conversation transcripts (i.e., two-person dialogues). **ConversationAlign** imports raw language transcripts into R, appends unique document identifiers, and concatenates all conversations into a single dataframe. **ConversationAlign** generates corpus analytics and executes text cleaning operations such as stopwords removal and lemmatization. The package vectorizes text and yokes published norms to each content word spanning more than 40 lexical, affective, and semantic dimensions. **ConversationAlign** outputs summary statistics each conversation including main effects and indices of local and global alignment for each specified dimension of interest.

Software Design

ConversationAlign was designed as a user-friendly R package with no proprietary components or input from AI (large language models). We were intentional about making the software accessible to users who do not have extensive backgrounds in computational linguistics or Natural Language Processing.

Statment of Need

Although many excellent text analysis applications exist (e.g., **Quanteda** (Benoit et al. 2018) and **Korpus** (Michalke et al. 2018)), we know of no R packages that are tailored to the unique demands of conversation analysis (but for Python see **ALIGN** (Duran, Paxton, and Fusaroli 2019)). **ConversationAlign** offers a comprehensive text processing pipeline and novel algorithms for computing linguistic alignment in 2-person dialogues. This software offers standardization and automation advantages that are in great need in a field that has historically relied heavily upon manual coding systems and subjective human judgment.

Research Impact Statement

`ConversationAlign` has supported two peer-reviewed publications to date in the cognitive neuroscience and psychological methods journals *Cortex* (Reilly et al. 2025) and *Behavior Research Methods* (article in press). The software is relatively new, and there are limited indices of uptake within the broader language research community.

Background

Conversation is among the most complex behaviors that humans routinely undertake. In a dyadic interaction, conversation partners modify the form and content of their own production to align with each other (Pickering and Garrod 2021). This process, known as linguistic alignment, occurs across many dimensions. `ConversationAlign` offers an automated approach to computing linguistic alignment across >40 distinct psycholinguistic dimensions (e.g., word length, valence, concreteness), leveraging recent advances in natural language processing to examine dynamics of human interaction at an unprecedented scale.

AI Usage

`ConversationAlign` is **NOT** a large language model (LLM). It instead indexes a static lexical lookup database populated with published norms for >100,000 English words across more than 40 unique dimensions. We used DeepSeek to troubleshoot elements of code and to generate regular expressions (regex) for complex pattern matching. We did not use AI to write this paper or generate segments of code.

References

- Benoit, Kenneth, Kohei Watanabe, Haiyan Wang, Paul Nulty, Adam Obeng, Stefan Müller, and Akitaka Matsuo. 2018. “Quanteda: An r Package for the Quantitative Analysis of Textual Data.” *Journal of Open Source Software* 3 (30): 774. <https://doi.org/10.21105/joss.00774>.
- Duran, Nicholas D., Alexandra Paxton, and Riccardo Fusaroli. 2019. “ALIGN: Analyzing Linguistic Interactions with Generalizable techNiques—a Python Library.” *Psychological Methods* 24 (4): 419–38. <https://doi.org/10.1037/met0000206>.
- Michalke, Meik, Earl Brown, Alberto Mirisola, Alexandre Brulet, and Laura Hauser. 2018. “koRpus: An r Package for Text Analysis.” <https://CRAN.R-project.org/package=koRpus>.
- Pickering, Martin J., and Simon Garrod. 2021. *Understanding Dialogue: Language Use and Social Interaction*. Cambridge University Press. https://books.google.com/books?hl=en&lr=&id=3RgXEAAQBAJ&oi=fnd&pg=PR7&dq=pickering+garrod+understanding+dialogue&ots=0qe68OV8Xs&sig=ulM_ibE3lJewbmVg-UvQvFEHhM.
- Reilly, Jamie, Virginia Ulichney, Benjamin Sacks, Anna Duncan, Sarah M. Weinstein, Tania Giovannetti, Chelsea Helion, and Gus Cooney. 2025. “Abstract Word Dropout and Cross-Speaker Misalignment of Word Concreteness Are Features of Conversation in Aging.” *Cortex* 190 (September): 286–303. <https://doi.org/10.1016/j.cortex.2025.07.003>.