



© James Yang 2014 for Scientific American

Text to talk: foundations of interactive language modeling for conversational AI and talking robots

12.09.2022

Andreas Liesenfeld, Ada Lopez, Mark Dingemans

{andreas.liesenfeld, ada.lopez, mark.dingemans} @ru.nl

Center for Language Studies

Radboud University

Overview

Language in interaction allows collaboration and exchange at an unrivaled level in the natural world. It provides humans with communicative tools that helped us to thrive as a species, and as social individuals. A key to its success is its flexible use in social interaction, an aspect of language that computational linguistics and NLP struggle to get a grip on.

But progress has been made towards technology that aspires to keep up with the prowess of human language and sociality. Enabled by advances in speech recognition, the likes of Siri and Alexa have entered the lives of many. Increasing amounts of data and more machine learning architectures promise more robust voice user interfaces.

At the intersection of language theory and tech, this tutorial introduces strands of both the language sciences and technological fields that share an interest in understanding how language is used in interaction. Drawing on linguistics, cognitive science and the study of human interaction, we review the theoretical and empirical foundations necessary for progress in technological fields such as voice user interfaces (VUI), social robots, and conversational AI.

You will learn the basics of interactive language modeling, exploring elements and dynamics of conversation. Drawing on techniques from dialog modeling, NLP and signal processing, you will dive into exploring structure and variation in conversational speech data in a hands-on tutorial. Some experience working with Python and Jupyter required.

The tutorial concludes with a discussion of the implications of what we know (and don't know) about interactive language use for designing and building next-generation interactive language technology. We discuss some technological limits of current products, and touch upon societal and ethical issues that emerge alongside the rise of voice AI. This tutorial might appeal to anyone interested in understanding why talking machines struggle to hold up their end of a conversation, whether coming from an academic, engineering or design background.

Outline

Introduction (45min): Language and technology in conversation

Hands-on (120min): Tutorial on interactive models of conversational structures

Discussion (30min): Scientific, technological, and societal aspects of conversational AI

Introduction: language and technology in conversation

Social interaction is the primordial home of language. It is where we acquire language and use it to accomplish tasks day in, day out. What may seem fluid and effortless to us, continues to pose serious challenges to NLP and language technology (Dingemanse and Liesenfeld 2022). To date, no machine can “lead a half-decent coherent conversation with a human” (Kopp and Krämer 2021). Understanding why it is so hard for machines to keep up with human conversationalists requires understanding the workings of language in interaction. With implications for language technology in mind, the lecture covers key findings and current trends in the study of human interaction.

Hands-on: tutorial on interactive models of conversational structures

Beyond NLPs mainstay methods that are built on and for text, the tutorial introduces methods for processing data that is - talk not text, sequential not flat, dyadic not single-minded. In this hands-on session you will learn how to work with conversational data using NLP and signal processing techniques. We examine patterns in interaction, learn to understand the current limits of interactive language modeling and why it is hard to build conversational machines that go beyond current query-response systems.

Jupyter notebooks: live tutorial, walkthrough and exercise

Using Python-based interactive notebooks, participants will use a range of tools to visualize, explore, and model interactional phenomena in conversational speech data using sequentially-grounded clustering and classification methods (built with *convokit*, *scikit-talk*, *avgn*, *convplot*). The walkthrough is concluded with a live demonstration of how findings on interactional infrastructure might be evaluated in existing dialogue systems or implemented in a social robot (virtual Furhat).

Discussion: roundtable and wrap-up

We conclude with a discussion of what it takes to build more interactive human language technologies and how to get there. We also touch upon some societal and ethical issues that emerge alongside the rise of “conversational AI”.

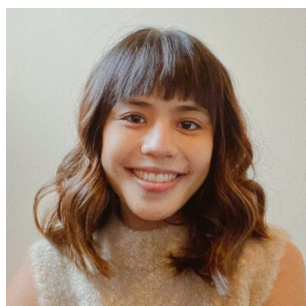
Learning outcomes

Cross-disciplinary introduction to the study of human interaction with a focus on language technological applications. Hands-on training in processing, analyzing and extracting features in real-world conversational speech data. Primer in interactive language processing, computational social science, and human-computer interaction.

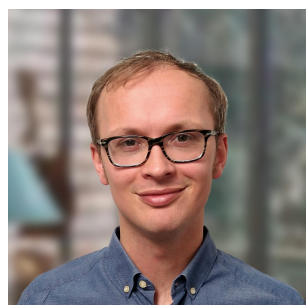
Instructors



Andreas Liesenfeld is a postdoctoral fellow with a PhD in computational linguistics (NTU Singapore, 2019) and expertise in data science and conversational AI. His work focuses on understanding how people interact with each other and with machines in the real world, and has appeared in venues like ACL, INTERSPEECH, LREC and SIGDIAL.



Ada Lopez is a Linguistics and Communication Sciences research master student specializing in Language and Speech Technology at Radboud University. Her research interests include studying the use of computational modeling to further understand language processing and conversations.



Mark Dingemans is Associate Professor in Language and Communication and head of the Elementary Particles of Conversation project. His work on interactive repair and conversational structure has revealed a number of striking candidate pragmatic universals that we study using cross-linguistic and computational approaches.

More information: www.markdingemans.net/elpaco/tutorial/

References:

- Dingemans, M. and Liesenfeld, A., 2022. From text to talk: Harnessing conversational corpora for humane and diversity-aware language technology. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics (ACL).
- Kopp, S. and Krämer, N., 2021. Revisiting human-agent communication: The importance of joint co-construction and understanding mental states. *Frontiers in Psychology*, 12, p.597.