EFFECTS OF PERCEIVED GENDER AND ANTHROPOMORPHIZATION OF AI ON LINGUISTIC ALIGNMENT

A Thesis

Presented to the

Faculty of

San Diego State University

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

in

Linguistics

---

by

Travis Putman

Spring 2024

SAN DIEGO STATE UNIVERSITY

The Undersigned Faculty Committee Approves the

Thesis of Travis Putman:

Effects of Perceived Gender and Anthropomorphization of AI on Linguistic Alignment

Gabriel Doyle (Mar 15, 2024 11:37 PDT)

Gabriel Doyle, Chair
Department of Linguistics & Asian/Middle Eastern Languages

Robert Malouf
Department of Linguistics & Asian/Middle Eastern Languages

Yang Xu
Yang Xu (Mar 15, 2024 09:12 GMT+8)

Yang Xu
Southern University of Science and Technology

3/15/24

Approval Date

ABSTRACT OF THE THESIS

Effects of Perceived Gender and Anthropomorphization of AI on
Linguistic Alignment
by
Travis Putman
Master of Arts in Linguistics
San Diego State University, 2024

This thesis investigated the nature of social factors in the psycholinguistic phenomenon of alignment in humans' interaction with computer conversational agents using text-to-speech (TTS) audio prompts across two tasks, a Word List and Question-and-Answer task (Q&A). This study focused on the nature of priming-based and grounding-based accounts of alignment and alignment across multiple linguistic levels in communication. Participants were presented with pre-generated TTS prompts of low-frequency words and questions about known facts during these tasks. TTS prompts consisted of male and female voices with or without a same-gender avatar. In the Word List, pitch was manipulated across the TTS prompts into uncommon pitch patterns to elicit alignment behaviors from participants. Phonetic alignment was measured by Difference-in-Distance scores between the baseline and shadowed conditions across acoustic features, selected a priori, including pitch, F1-F4, and speech rate. In the Q&A task, TTS-generated questions were manipulated in the choice of lexical and syntactic alternations. Lexical and syntactic alignment was measured by the presence or absence of target structures in participants' utterances. Overall, participants did not converge to TTS voices, and the lack of alignment behaviors was not predictable by TTS gender, participant gender, or presence of avatar. Unexpectedly, participants did not converge, but rather diverged, from the TTS voice's pitch during the Word List task. In the Q&A task, most participants did not align across any of their utterances. However, two participants' utterances responded to the nature of their role in the task with utterances containing many target structures by way of rephrasing questions into statements. Together, these results argue against a purely priming-based mechanism of alignment and suggest that divergence behaviors, conversational role, and communicative intent are a key part of the joint-meaning making process in grounding-based accounts of alignment which must be accounted for in future work.

TABLE OF CONTENTS

PAGE

LIST OF TABLES

PAGE

LIST OF FIGURES

PAGE

CHAPTER 1

INTRODUCTION

In conversational interaction, the same or similar lexical items, phonetic realizations, syntactic structures, and morphological markers are used by participants in the same conversations (Pickering & Garrod, 2004). This process is referred to by many names such as alignment,[1] accommodation, and entrainment (Garrod & Anderson, 1987; Giles et al., 1991; Pickering & Garrod, 2004). These terms are used in various fields of study from linguistics, cognitive science, and psychology to refer to similar processes, yet are not often collapsed together. Each of these independent answers to the question of alignment not only have implications for theories of language, but for its role in human cognition.

Two of the leading models of the mechanism underlying these processes are priming-based and grounding-based. These competing theories argue the questions: Is the cognitive goal of this process to reach a common understanding and general affinity between interlocutors or is it automatic and unconscious where previously used linguistic features are easily accessible in the mind? The most prominent of these theoretical frameworks provides an answer based on an automatic priming-based mechanism, namely the Interactive Alignment Model (Pickering & Garrod, 2004). The IAM suggests that activation of linguistic representations percolates between levels of linguistic structure automatically through independent activation by way of previously used utterances activating mental representations or priming. Juxtaposed with the IAM is the process of grounding, a social and communicative goal-oriented approach to linguistic alignment (Clark & Brennan, 1991).

---

1 I adopt the use of alignment as a cover term for entrainment, imitation, adaptation, convergence, divergence, etc. I discuss the implications and meaning of this in Chapter 2.

Grounding approaches define alignment as a joint meaning-making process to fulfill some social goal between interlocutors in dialogue. Priming and grounding-based mechanisms of alignment have traditionally been characterized as two mutually exclusive approaches of explaining the same phenomenon. However, recent work by Rasenberg et al. (2020) suggests that they are fundamentally characterizing similar processes that are a part of the same continuum, one an implicit and the other an explicit mechanism.

One of the greater points of contention in explanations of linguistic alignment is the method in which the framework accounts for social factors. Priming and grounding accounts can be distinguished by the method in which social factors affect alignment processes. In a priming-based account, social factors affect the alignment and the extent of alignment implicitly and automatically; thus, depending on certain social factors, an interlocutor may pay increased attention to another interlocutor's speech, aligning without conscious intention. Meanwhile, grounding, and social/communicative goal-oriented approaches suggest that alignment is the result of successful joint efforts for interlocutors to establish a common ground; therefore, linguistic forms are chosen by the interlocutor to ameliorate the creation of this common ground. In grounding frameworks, interlocutors make (semi-)conscious decisions based on these social factors. Much of the previous literature on this distinction of social factors takes place in the realm of phonetic imitation, which I take to be a subset of phonetic alignment. The extent to which interlocutors align is different, depending on many of these social factors such as age, gender of speaker, gender of listener, and many others. Babel (2012) argues that some social factors such as the presence of a picture of a model talker, a speaker being shadowed in a word-shadowing task, and the attractiveness of a model talker, which is argued to be implicit to a speaker, are predicted by the IAM to affect alignment implicitly and automatically. In this case, the presence of a model talker and their perceived attractiveness should not affect the participant's need to fulfill some social goal due to the nature of the imitation task; thus, this effect can be categorized as being predicted by priming-based accounts of alignment. In fact, many effects of social factors on alignment follow from predictions set by the IAM. However, it does not follow that all effects of social factors are entirely predicted by a priming-based mechanism of alignment, and some can be characterized as grounding. In an alleged grounding effect on alignment, Shin and Doyle (2018) find that in-group interlocutors choose to diverge in pronoun usage from out-group

members in order to signal their membership to the in-group. In grounding approaches, building a common ground between interlocutors requires explicit social factors considered by the interlocutor unlike implicit effects in the IAM. If alignment processes are to be characterized on a continuum of priming vs. grounding, a full investigation of this union must also start by understanding the effects of social factors on alignment.

The presence of variability in alignment behaviors characteristic of either priming or grounding does not necessarily disprove the other. Instead, we can characterize effects of alignment on a continuum between priming and grounding (Rasenberg et al., 2020). Rather than between social or cognitive processes, Rasenberg et al. (2020) characterize alignment as pairs of behavior, where the behavior of interlocutor A is paired with the behavior from interlocutor B. They suggest an integrative framework of pairs of behavior, characterized through five dimensions time, sequence, meaning, form, and modality. They argue that assumptions of the underlying mechanism of alignment can be regarded in terms of these five dimensions and the effects of paired behavior matchings can be placed on a continuum between priming and grounding. However, a major point of contention has not been fully investigated, which is the effects of social factors on the extent of alignment.

A developing area of research in the effects of social factors in alignment has begun at the level of human and artificial intelligence conversational interactions. Humans have begun to converse with non-human interlocutors at an increasing rate as advanced Artificial Intelligence (AI) and language models have become prevalent. Like human-human dialogue, human interlocutors align to AI models as well conversation mediated through either text-based chatbot communication or voice-AI systems (Gessinger et al., 2021; Spillner & Wenig, 2021; Zellou et al., 2021a; Zellou et al., 2021b). Although these computer and AI systems are not human, humans still treat them as social agents, presumably applying the same, or similar, implicit, and explicit social expectations (Gambino et al., 2020; Nass & Moon, 2002). Furthermore, a developing vein of research has shown that the social factors that affect alignment are still prevalent in human-AI communication (Cohn et al., 2019; Zellou & Cohn, 2020a; Zellou & Cohn, 2020b; Zellou et al., 2021a; Zellou et al., 2021b) As many of these studies focus on the phonetic level of alignment, little work has been done on the effects of social factors in syntactic and lexical alignment in human-AI communication. To fully test the predictions of the mechanism underlying alignment, many linguistic features

must be measured. To the researcher's knowledge, there exists no prior study that considers multiple levels of alignment in human-AI communication.

Under the framework set by Rasenberg et al. (2020), I investigate the effects of social factors on multiple levels of alignment, phonetic, syntactic, and lexical in human-AI dialogue. I tested this with an experiment where humans interact with a simulated voice-AI in two tasks: the word list task, a variation on word shadowing tasks, where participants repeat a word list spoken by the AI, and the Q&A task, where participants answer questions asked by the AI. To manipulate the effects of social factors, this study focuses on the perceived gender and anthropomorphization of AI systems, two factors which have not been fully explored in the literature. I looked for evidence of alignment across phonetic, syntactic, and lexical levels, but found little evidence of convergence at F1-F4, speech rate, or lexical and syntactic alternations. Unexpectedly, the results show divergence in F0 and task-dependent features that interfere with potential alignment at the lexical and syntactic levels. These unexpected patterns require further investigation and exploration in future work. As well, these results show that by studying communication and language through human-AI conversations, we can better understand the underlying mechanism of psycholinguistic processes in dialogue.