# Vittorio Tantucci, Carlotta Sparvoli\*

# AI, be less ‘stereotypical’: ChatGPT’s speech is conventional but never unique

**Abstract**: Can AI reproduce human interaction? It can, but only stereotypically. While it can simulate (and even exaggerate) dialogic engagement, its lexicon is less diverse, and the speech acts it realizes are more repetitive and less varied (we took directives as an example). Most importantly, AI struggles to represent ‘conversational uniqueness’, that is ways to interact that define the specificity of a particular conversation and are not entirely conventional. We discovered this after analyzing dialogic resonance (the re-use of an interlocutor’s construction), recombinant creativity (the creative reformulation of an interlocutor’s construction), relevance acknowledgement (the acknowledgement of what an interlocutor said) and other variables from a sampled section of the CallHome Corpus of Chinese telephone conversations. After feeding ChatGPT with speakers’ demographics and contextual information, we asked it to reproduce telephone interactions among Chinese family members. We then fitted a conditional mixed effects Bayesian regression comparing the two datasets. We found that AI over-generalizes human dialogue. It provides a stereotypical way of conversing but shows scarce flexibility in including ‘atypical’ and unconventional utterances, which are, in turn, constitutive of human interactions that occur in real life.

**Keywords:** AI interaction; human interaction; dialogic resonance; directives; recombinant creativity

## 1 Introduction

How stereotypical is AI language? A bit too much, maybe. It is very successful at matching highly conventionalized expressions and speech acts in context. It is also very good at simulating recurrent forms of conversational engagement. However, it is scarce at contemplating ‘conversational uniqueness’, that is, all those linguistic functions and structures that people do not use conventionally but that are distinctive of specific interactions in real life.

Two key aspects of humans’ communicative behavior are of interest to this paper: one is dialogic engagement (showing interest in what an interlocutor is saying), and the other is dialogic creativity (contributing with something new to an ongoing conversation) (cf. Tantucci et al. 2023). We tested ChatGPT’s ability to reproduce these two components of human interaction. We did so by analyzing human vs AI’s dialogic resonance (Du Bois 2014; Tantucci and Wang 2024), the ability to re-use (and often recombine creatively) the linguistic expressions of an interlocutor. We also controlled for relevance acknowledgement, i.e. strategies that are used to remark that what an interlocutor said is relevant or interesting (e.g. Tantucci 2023; Tantucci and Wang 2024). We found that AI is very good (perhaps even too good) at imitating humans’ ability to show interest in what an interlocutor has said. Its levels of resonance and relevance acknowledgement are even higher than humans. However, ChatGPT is not very good at avoiding being repetitive and stereotypical. Humans engage in conversation trying to avoid making the same lexical choices and performing the same forms of speech acts in context. This is something AI is still unable to perform in a human-like manner. In particular, we found that in the context of telephone conversations among family members AI’s lexical diversity is much lower than humans’. We also focused on speech acts’ diversity, directives in particular. We found that AI’s reproduced dialogues are mostly limited to recommendations, perhaps due to a bias towards a stereotypical telephone exchange between a parent and her son/daughter.

All in all, AI’s ability to simulate human dialogue is geared towards over-generalization. What AI simulates is normative communicative behavior, even when asked to reproduce human dialogues on a large scale. This leads to a stereotypical account of human communication that iteratively repeats itself and overlooks the uniqueness and dispersion of humans’ real conversations.

Our paper is structured as follows: Section 2. touches on pragmatics research that has been conducted on ChatGPT to date. Section 3. Is devoted to dialogic engagement and creativity and the way we operationalized them for contrasting AI and Human communication. Section 4 discusses directive speech acts, as this study will also compare how varied the illocutionary force of AI can be compared to human conversation. In Section 5 we discuss our methods, our statistical analysis and the ensuing results. We discuss the relevance and the implications of our results in Section 6 and formulate our conclusions in Section 7.

## 2 ChatGPT pragmatics

ChatGPT, a widely known conversational AI model, is designed to improve its performance through Reinforcement Learning from Human Feedback (RLHF), a process where human input refines its behavior in complex and nuanced situations. Since 2020, studies addressing the question of how pre-trained Large Language Models (LLMs) can decode pragmatic reasoning have been growing in number (Pandia et al. 2021; Ruis et al. 2022; Hu et al. 2022; Barattieri di San Pietro 2023, among others). Various methodologies have been recently deployed to evaluate AI’s pragmatic competencies. Experimental designs often involve presenting AI models with scenarios that require nuanced understanding of context, implicature, and indirectness.

Ruis et al. (2022) investigated whether LLMs have the ability to interpret language in context, make inference, and understand *scalar implicatures*. To this end they used zero-shot prompting, assigning a question without any prior task-specific training or fine-tuning, relying solely on AI’s pre-trained knowledge and general understanding of language to generate a response. They designed a protocol based on binary implicature resolution, of the type: Utterance: "Want to stay for a nightcap?"; Response: "I’ve gotta get up early", implicature: "No". The study evaluated four categories of widely used LLMs, namely, pre-trained, fine-tuned on conversations, fine-tuned with benchmark-level instructions, and fine-tuned on example-level instructions. While the former three performed close to random, LLMs instruction-tuned at the example-level (that is, Open AI instructional model) perform significantly better, suggesting that certain fine-tuning strategies are far better at inducing pragmatic understanding in models. Cong (2024) instead focused on LLM’s ability to understand *manner implicatures*, triggered by features such as negation, modality, and causality; to explain LLMs’ behavior, the author constructed three metrics including LLMs-surprisals, embedding vectors’ similarities, and natural language prompting. The study suggests that pre-training paradigms are not particularly effective in boosting LLM’s competence in manner implicature.

Comparative analyses between AI-generated responses and human interactions are also common, providing insights into the strengths and limitations of AI in replicating human conversational behavior. Similar to Ruis et al. (2022), also Hu et al. (2022) use zero-shot prompting to evaluate LMs on pragmatic interpretation tasks. However, Hu et al’s study considers a broader range of pragmatic phenomena and utterance types, and prompts LLM to select from multiple interpretations, enabling a detailed error analysis and comparison to humans. Their study highlights that large models show high accuracy, align with human error patterns, but struggle with tasks involving social expectation violations, suggesting that, while pragmatic behaviors can emerge without explicit mental state representations, it still remains limited in some areas. Seals and Shalin (2023) note the importance of pragmatic considerations in AI applications, particularly in conversation summarization and other conversational AI tasks. They emphasize that AI made significant strides in semantics and syntax but there is a pressing need to address pragmatic limitations to achieve more human-like interactions. Role-play tasks in second language pragmatics research have been used to assess AI’s ability to engage in context-sensitive exchanges (Sydorenko et al. 2024). Jones and Bergen (2024) evaluated GPT-4 in a Turing test setting, where participants conversed with either a human or the AI without knowing which. GPT-4 was identified as human 54% of the time, highlighting its advanced conversational abilities while also pointing out areas where it still falls short of human interaction. Chen et al. (2024) examined 148 AI-generated dialogues and 82 human-authored ones across various speech acts and found that ChatGPT performs comparably to human participants in most pragmalinguistic and sociopragmatic features and even surpasses humans in syntactic diversity and formal appropriateness. Similar to Jones and Bergen (2024), human evaluators were often unable to distinguish AI-generated responses from those written by humans, suggesting that ChatGPT could serve as a viable tool for expanding research on pragmatic competence and enhancing language education.

Concerning human-AI interactions, Diedrichsen’s (in this volume) analysis show how the concept of common ground intersects with contemporary AI technologies and human-machine interaction, showcasing two case studies from ChatGPT and Microsoft's CoPilot. Dynel (2023), based on a corpus of posts from a community on Reddit focused on discussions related to ChatGPT, examines users’ attention as they conduct linguistic experiments with ChatGPT, suggesting that interacting with AI mirrors and potentially increases users' metapragmatic awareness, which is reflected in ChatGPT’s responses, showing its human-assisted training.

## 3 Dialogic engagement and creativity

To assess AI’s ability to simulate humans’ engagement and creativity in dialogue, we controlled for several pragmatic variables: dialogic resonance, recombinant creativity and relevance acknowledgement.

### 3.1 Dialogic resonance and recombinant creativity

Dialogic resonance refers to a form of alignment in dialogue where speakers imitate and/or creatively recombine constructions used by their interlocutors (Du Bois 2014; Tantucci and Wang 2022a). Constructions, understood as holistic pairings of form and meaning, are central to the usage-based approach in Cognitive Linguistics (Langacker 1987; Goldberg 1995, 2006) and Pragmatics (Schmid 2020; Tantucci 2015, 2017a, 2017b, 2020, 2021). They are naturally acquired through interaction, as speakers frequently encounter and reuse forms with similar semantic or morphosyntactic properties, enhancing their ability to categorize meaning and generate novel forms. While traditional studies in the usage-based approach emphasize individual processing, recent research highlights the dialogic enactment of constructions involving multiple interlocutors, leading to new models of dyadic cognition (Arundale 2010; Weigand 2018; Tantucci 2023; Kecskes 2023). In this framework, utterances are dynamically recalibrated and re-conceptualized across turns during dialogue (Dingemanse 2020). Dialogic constructions, central to theories like Emergent Grammar (Hopper 2011) and Dialogic Syntax (Du Bois 2014), emerge through creative re-elaborations of forms and meanings during interaction (Tantucci and Wang 2021). This process involves speakers adapting to dialogic stimuli in real-time, showcasing language as recombinant and interactive rather than fixed or idiosyncratic (Tantucci 2023). Resonance is a key conversational mechanism that allows interlocutors to reuse and recombine elements of others’ utterances, creating morphosyntactic, semantic, and pragmatic analogies that drive the conversation forward (Fischer 2008; Gentner and Christie 2010). This process not only facilitates shared categorisation and innovation but also fosters interactional engagement by textually acknowledging the relevance of prior speech, functioning like spoken “citations” (Tantucci 2023). It varies cross-culturally (Tantucci and Wang 2021), it is partly inhibited in Autism (Hobson et al. 2012; Du Bois et al. 2014; Tantucci and Wang 2023), it becomes increasingly sophisticated around the fourth year of age in child acquisition (Tantucci and Wang 2022b) and it has been used to shed light on engagement across social grades (Tantucci and Wang 2024) and in Doctor-Patient interaction (Tantucci and Lepadat 2024).

In the example below, a mother M resonates with her daughter D talking about group projects and individual responsibilities in an academic setting:

D: 现在他叫我们自己做了。

Xiànzài tā jiào wǒmen zìjǐ zuò le.

now 3SG ask 1PL self do-ASP

‘Now he asks us to do it ourselves.’

D: 就是每个人做自己的。

Jiùshì měi gèrén zuò zìjǐ de.

exactly each-CLF person do one’s.own POSS

‘Exactly, everyone does their own.’

M: 每一个人做这样的。

Měi yī gèrén zuò zhèyàng de.

every one-CLF person do like.this NOM

‘Every person does something like this.’

(Callhome 0022)

In the exchange, M gives importance to what D has said as she actively re-uses some of her linguistic material in her utterance. She does not merely repeat what she heard but contributes to establishing a dialogic construction across her and her daughter’s utterances that bears functional and formal analogies. In fact, D’s expression 每个人做自己的 *měi gèrén zuò zìjǐ de* ‘everyone does their own’ is resonated by M in the new form 每一个人做这样的 *měi yī gèrén zuò zhèyàng de* ‘every person does something like this’. Resonance occurs when some words are repeated verbatim (每 *měi*, 个 *ge*, 人 *rén*, 做 *zuò*, 的 *de*) but also when some new structural pairing of form and meaning emerges as a result of that. In this case, the more schematic form [每个人做 MOD 的] is co-constructed by both speakers, as given in Table 1.

Table 1: Diagraph: [每个人做 MOD 的]

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 每 | 个 | 人 | 做 | MOD | 的 |
| D | 每 | 个 | 人 | 做 | 自己 | 的 |
| M | 每 | 个 | 人 | 做 | 这样 | 的 |

Table 1 represents a diagraph, a syntactic structure that emerges overt attempts of one speaker to resonate with another (cf. Du Bois and Giora 2014: 354). In fact, despite the specificity of each speaker’s utterance, a more schematic construction emerges dialogically as a sign of conversational engagement of M with D’s speech. The annotation of resonance in dialogue involves a two-step process (cf. Dialogic Categorization Model, Tantucci 2023):

1. Identifying any sort of lexical repetition (including interjections) across utterances.
2. Identifying the structure (if any) that emerges as a result.

As said, a number of words were repeated from D to M (step i.). Secondly, the emergent structure [每个人做 MOD 的] can be identified due to the modifier (MOD) 这样 *zhèyàng’* ‘like this’ replacing the modifier 自己 *zìjǐ* ‘on their own’ preceding the structural particle 的 *de*, normally used to refer to an ellipted generic Object, in this case 事 *shì* ‘things’. This structure bears meaning. It expresses a generic statement of *how anyone does things in a similar manner in a certain situation.*

Dialogic resonance is important because it can be quantified and measured on a large scale as a formal indicator of dialogic engagement. In the case of (1), the resonance value can be counted as 6, namely the number of internal constituents of an emerging dialogic construction. Now, resonance involves engagement but can be more or less creative. In the case of (1), not much so, as only one constituent has been innovatively recombined: the replacement of 自己 *zìjǐ* with 这样 *zhèyàng*. The degree of recombinant creativity here is thus only 1.

Let’s consider a second example from our dataset:

D: 哦, 那你下回来, 下回你写信的时候, 把他的地址告诉我一下, 好不好?

Ō, nà nǐ xià huílái, xià huí nǐ xiěxìn de shíhòu, bǎ tā de dìzhǐ gàosù wǒ yīxià, hǎobùhǎo?

INTJ then 2SG next return next time 2SG write letter SUB time BA 3SG POSS address tell 1SG

a.little alright

‘Oh, when you come back next time, the next time you write a letter, can you let me know

his address, okay?’

F: 哦, 下一次我告诉你.

Ō, xià yīcì wǒ gàosù nǐ.

INTJ next one.time 1SG tell 2SG

‘Oh, I’ll let you know next time.’

(Callhome 003)

In this case, a father resonates with his daughter over the phone. In particular, he overtly engages with the construction 下回你写信 *xià huí nǐ xiěxìn* ‘next time you write a letter’ in the new form 下次我告诉你 *xià yīcì wǒ gàosù nǐ* ‘next time I will tell you’. Across turns, there is the verbatim repetition of 下 *xià* ‘next’ and the structural emergence of the new construction [下 ADVTemp SubjPp Verb Obj] whereby the Temporal adverbial 次 *cì* ‘occasion’ is replaced with a synonymous adverbial 回 huí ‘occasion’, the personal pronoun subject 你 *nǐ* ‘you’ is recombined as 我 *wǒ* ‘I’, the verb 写 *xiě* ‘write’ is now given as 告诉 *gàosu* ‘inform’ and 信 *xìn* ‘letter’ in the object position is replaced by 你 *nǐ* ‘you’. The diagraph in Table 2 illustrates the analogies that grant the formation of this dialogic construction.

Table 2 : Diagraph [下 ADVTemp SubjPp Verb Obj]

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 下 | ADVTemp | SubjPp | Verb | Obj |
| D | 下 | 回 | 你 | 写 | 信 |
| F | 下 | 次 | 我 | 告诉 | 你 |

Here, the two speakers jointly realize ad hoc a pairing of form and meaning semantically expressing *what one will do next time*. The internal constituents of this are 5, hence the value of dialogic resonance. What critically differs from (1) is the degree of recombinant creativity involved, which is now quite high: 4. Recombinant creativity is a key indicator of cognitive efforts made to engage with the linguistic material of an interlocutor and one we will control for in the comparison across human and AI speech.

### 3.2 Relevance acknowledgement

When people tell things to each other, they expect their talk to be acknowledged somehow. No one likes a silent reaction after having told a story or given some new information. Overtly acknowledging that what we hear is interesting underpins epistemic reciprocity (Culpeper and Tantucci 2021; Tantucci et al. 2022; Culpeper et al. 2025), i.e. the social expectation that the efforts made to inform someone of something will be acknowledged and treated as relevant for an ongoing interaction. There is more than one strategy to respond to show that what we heard is relevant or interesting. Resonance is a decisive way to demonstrate that the linguistic material we heard is valuable for continuing a conversation. People also often make use of grammaticalized backchannels and pragmatic markers. These are all highly entrenched and, in most cases, semantically bleached (e.g., Lehmann, 2015; Bybee 2003). Due to their high degree of frequency and conventionalization, backchannels have been shown to express relevance acknowledgement as baseline forms of relevance acknowledgement (RA) (Tantucci et al. 2022). Conventional backchannels, such as *Yeah, Definitely*, or *Oh*, function to acknowledge what has just been said in ways specific to the interaction (Heritage 2012). These backchannels often occur at the left periphery of sentences or conversational turns, aligning with intersubjective engagement. Relevance acknowledgement also occurs propositionally, that is when B responds to A with a propositional, creative strategy that discursively acknowledges A’s talk, e.g. *it’s interesting that you said that, I wouldn’t have thought so*, and so on. In example (3) below the form 哎呦 *Āiyō* is used to express surprise (cf. Heritage, 2012), accompanied by mild disapproval in a conversation between a mother and her pregnant daughter are discussing about whether buying clothes for her future child:

D: 我不买衣服.

Wǒ bù mǎi yīfú.

1SG NEG buy clothes

‘I’m not buying clothes.’

M: 哎哟, 那你干脆去买一个吧.

Āiyō nà nǐ gāncuì qù mǎi yī gè ba.

INTJ then 2SG simply go buy one-CLF SFP

‘Oh, then you might as well go buy one.’

(Callhome 497)

We annotated RA by counting the number of words used at the left-hand side of an utterance to express that what an interlocutor said is important for the current turn.

### 3.3 Sentence final particles (SFP)

Beyond resonance, recombinant creativity, and relevance acknowledgement, speakers constantly align with one another using interjections, backchannels, pragmatic markers, and other conventionalised constructions that demonstrate their awareness of their interlocutors’ stances, emotions, and other cues (Tantucci 2017a, 2021). World languages show a clear tendency to develop more or less grammaticalised markers at the end of utterances to facilitate intersubjectivity (e.g Traugott, 2012, 2016; Heine 2023). This refers to the expression of additional meaning focused on a specific or general addressee’s likely reaction to what is being communicated (Tantucci 2021: 33). For example, peripheral constructions like *isn’t it* or *you know what I mean* positioned at the end of a turn functions as a marker of expected agreement. Such constructions establish intersubjective expectations, ensuring the plausibility of the statement for the hearer.

Languages such as Mandarin and other Southeast Asian languages have evolved grammaticalized systems of sentence-peripheral particles (SFP) to engage with the listener’s potential reactions. Consider the Mandarin sentence-final particle 吧 *ba* in example (3), transforming the directive illocutionary force of a request, command, or disposition into a collaborative joint action with the addressee (Tantucci, 2017b). In the exchange, the speaker adds 吧 *ba* at the end of the sentence to mitigate the potentially face-threatening force of a direct command like *go and buy one!* Instead, 吧 softens the directive, framing it as a collaborative action (Tantucci, 2021).

## 4 A focus on directives

Directive speech acts occupy a pivotal position in both naturalistic and AI interactions. They are acts where the Speaker’s illocutionary force aims to prompt the Addressee to take action (Searle, 1976: 11). They are inherently dialogical and surface in various expressions such as commands, requests, advice, and norms. Directives can have the same locutionary (the propositional content of an utterance) and illocutionary (the actual goal of an utterance) content as in direct forms such as (4). This is where an explicit request is made.

M: 你有空, 你跟他们写两封信啊。

Nǐ yǒu kòng, nǐ gēn tāmen xiě liǎng fēng xìn a.

2SG have free.time, 2SG with 3PL write two-CLF letter SFP

‘When you have time, you write two letters to them.’

D: 我是没有空啊  (…)

Wǒ shì méiyǒu kòng a

1SG emphatic not.have free.time SFP

‘I truly don't have time, …’

             (Callhome 0027)

They can also be realized indirectly (e.g. Ervin-Tripp 1976; Blum-Kulka et al. 1989, 1995; Kohnen 2004; Craven and Potter 2010), as in the case of (2), in Section 3.1, where the directive force is conveyed more subtly, using the tag question *hǎobùhǎo* 好不好 ‘alright?’, a conventional means of seeking permission or approval.

The interpretation of how illocutionary force is attached to an utterance has generated diverse scholarly perspectives. From a Gricean viewpoint, the illocutionary force arises during the communicative act when the speaker’s intention is conveyed through the utterance, and the listener infers that intention based on context and shared knowledge. This inference process aligns with the Cooperative Principle and its associated maxims, particularly the maxim of relation (or relevance). The effective recognition of illocutionary force thus relies on the interlocutors’ ability to interpret meaning beyond the literal content, ensuring that the communicative act achieves its intended purpose. A key aspect of directives is the conversational background shared by the Speaker and the Addressee, which encompasses their common knowledge, beliefs, and presuppositions, including their roles in context (Clark 1996; Stalnaker 2002). Within this framework, the use of directives enables the Speaker to provide the Addressee with reasons for taking action (Kissine 2013).

The illocutionary force is contextually determined, thereby closely linked to power dynamics, shared objectives, and common cultural backgrounds and linguistic conventions (Bach 2005; Wilson 2006) and it may be contingent upon factors such as politeness (Brown and Levinson 1987), face considerations. In a socio-cognitive approach (Kecskes 2008, 2010) the construction of meaning by the interlocutors also draws on co-constructed knowledge, emerging in the process of interaction.

There are several reasons why we decided to focus on AI’s ability to simulate directives. For one, they are the speech act ChatGPT is most exposed to, as it normally responds to orders or requests made by a user so that something can get done on his/her behalf. Secondly, in the context of the Callhome conversations, parents talk with their daughters or sons living abroad. This is an occasion where recommendations and requests of various kinds are very likely to occur due to the geographical distance between them. Directives often necessitate a response, which has been discussed in cognitive linguistics and conversation analysis (e.g. Ruiz de Mendoza Ibáñez and Pérez-Hernández 2002; Heinemann 2006; Curl and Drew 2008; Takahashi 2014). In the context of our dataset, a stereotypical response to a directive takes place using a commissive speech act. This is what happens in the simulated AI interaction in (5), wherein the modal 会 *huì* in a futurity reading combines with the sentence-final particle (SFP) *de* 的 (cf. Tantucci 2021), expressing certainty, in a typical form of reassurance (Sparvoli and Saillard 2024).



F: 你应该多放松，不要给自己太多压力。

Nǐ yīnggāi duō fàngsōng, búyào gěi zìjǐ tài duō yālì.

            2SG should more relax, must.not to yourself too much pressure

             ‘You should relax more, do not put too much pressure on yourself.’

D:   我知道了，我会多注意放松的。

Wǒ zhīdào le, wǒ huì duō zhùyì fàngsōng de.

1SG know SF, 1SG FUT more pay attention relax SF

‘I understand, I will try to relax more.’

(AI 00010)

Intuitively, this is a very ‘obvious’ way to respond. D here plays the part of the ‘good daughter’ as she diligently affiliates with her father’s recommendation. Alignment is marked through D’s resonance and relevance acknowledgement (我知道了 *wǒ zhīdào le* ‘I understand’) of F’s speech. A resonating construction emerging here is [SubjPp AUXMod 多放松], emerging from F’s construction 你应该多放松 *nǐ yīnggāi duō fàngsōng ‘*You should relax more*’* and D’s response 我会多注意放松 *wǒ huì duō zhùyì fàngsōng ‘* will try to relax more*’.* This is given in the diagraph in Table 3.

Table 3: Diagraph: [SubjPp AUXMod 多放松]

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | SubjPp | AUXMod | 多 | 放松 | IF |
| D | 你 | 应该 | 多 | 放松 | Directive |
| F | 我 | 会 | 多 | (注意)放松 | Commissive |

However, addressees often do not comply with directives in human interaction. This can lead to a breakdown in communication or result in a reproach that carries an ironic undertone, as illustrated in (6) from our naturalistic dialogic dataset:



D: 你别在那费事了。

nǐ bié zài nà fèi shì le

2SG do.not at that take.trouble SFP

‘Do not waste your time on that.’

M: 我-我一看你-你们都扎。

Wǒ Wǒ yì kàn nǐ  nǐmen dōu luàn.

1SG 1SG  as.soon.as look 2PL all prick

‘I… I look at you… and you all poke.’

(Callhome 247)

Here, the negative imperative 别 *bié* ‘don’t’ is employed as a prohibitive, with a similar force as the negative modal 不要 *búyào* in (5). Chinese directives are often expressed through via modalized proposition, including deontic modals, such as 应该 *yīnggāi* ‘should’ in (5), or other items conveying circumstantial or goal-oriented necessity, such as 必须 *bìxū* and 得 *děi* ‘must, have to’ (Sparvoli 2019). Furthermore, when the directive carries the nuance of advice, the mitigated form 最好 *zuìhǎo* ‘(you) better’, as seen in (7), is frequently used as an alternative of the more common 要 *yào* ‘must’.



M: 我得买个缝纫机。

Wǒ děi mǎi gè fèngrènjī.

1SG have.to buy CLF sewing.machine

‘I need to buy a sewing machine.’

D: 哎，我刚给说，你最好是去买一个。

Āi, wǒ gāng gěi shuō, nǐ zuìhǎo shì qù mǎi yīgè.

INTJ 1SG just to tell 2SG have.better emphatic go buy 1-CL

‘Oh, I just told you that you’d better go buy one.’

(Callhome 0497)

We will explore the distribution of these forms in our naturalistic and AI dialogic datasets in Section 5. Before that, we need to present the methodology used for data retrieval.

## 5 Data retrieval, annotation, and analysis

For this study, we randomly sampled 5 dialogic exchanges from the Callhome Mandarin Chinese Corpus of Telephone conversations. The Callhome Mandarin Chinese Corpus[[1]](#footnote-1) was developed by the Linguistic Data Consortium (LDC), it consists of spontaneous telephone conversations between native Mandarin speakers. It contains informal conversations over the phone among family members. Audio recordings are typically 5-10 minutes long and all together include 250,000 words.

### 5.1 Human interaction

We selected the first 240 turns of each and annotated them based on Resonance, Recombinant Creativity, Relevance Acknowledgment, Gender, Speakers’ ID, and Directives. We conducted an interrater reliability test involving three annotators to evaluate resonance, RC, RA values. Each annotator independently assessed 20% of the dataset, which was randomly sampled. To measure reliability, we used the Intraclass Correlation Coefficient (ICC) for numeric data and Fleiss’ Kappa for directives. The results showed substantial agreement among the annotators, with excellent reliability for degrees of resonance (ICC = 0.86) RC (ICC = 0.83) and RA (0.89).

Our goal was to model engagement and creativity, so our dependent variable was Resonance, which was annotated based on the Dialogic Categorisation Model (DCM, Author 2023, cf. Section 3). We then controlled for recombinant creativity (RC), relevance acknowledgement (RA) and sentence-final particles (SFP). We also distinguished between resonance co-occurring with those features in speaker B’s utterance and B’s resonance as a response to those features present in A’s preceding turn. In the model, those are given as SFP(lag), RC(lag), RA(lag), respectively. We this fixed a mixed effects Bayesian regression model (e.g. Natarajan et al. 2000) which produced the results given in Table 4.

Table 4: Mixed effects linear Bayesian regression of Resonance in Human dialogues[[2]](#footnote-2)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Random effects | | | | | | | |
| Child ID | (5 Levels) |  |  |  |  |  |  |
|  | Estimate | Est.Error | l-95% CI | u-95% CI | Rhat | Bulk\_ESS | Tail\_ESS |
| sd(Intercept) | 0.14 | 0.16 | 0.00 | 0.52 | 1.01 | 2529 | 2729 |
| CHAT ID:Speaker | (10 Levels) |  |  |  |  |  |  |
| sd(utterance) | 0.10 | 0.09 | 0.00 | 0.31 | 1.01 | 2797 | 2509 |
| Regression coefficents | | | | | | | |
|  | Estimate | Est.Error | l-95% CI | u-95% CI | Rhat | Bulk\_ESS | Tail\_ESS |
| Intercept | 1.48 | 0.20 | 1.11 | 1.87 | 1.00 | 4724 | 3489 |
| SFP | -0.02 | 0.13 | -0.27 | 0.23 | 1.00 | 7343 | 4857 |
| RA | -0.74 | 0.25 | -1.24 | -0.24 | 1.00 | 7914 | 4407 |
| RC | 1.36 | 0.09 | 1.18 | 1.54 | 1.00 | 7930 | 4259 |
| SFP(lag) | 0.10 | 0.13 | -0.15 | 0.36 | 1.00 | 8632 | 4270 |
| RA(lag) | 0.06 | 0.09 | -0.11 | 0.23 | 1.00 | 7434 | 4867 |
| RC(lag) | -0.31 | 0.25 | -0.80 | 0.17 | 1.00 | 7084 | 4338 |

The Estimate column in the Random Effects section indicates the standard deviation of the intercepts for variability at the conversation level (CHATID = 0.14, [l=0.00, u=0.52]) and the speaker level nested within conversations (CHATID:Speaker = 0.10, [l=0.00, u=0.31]). This reflects slight variability in baseline resonance across conversations and minor variability between speakers within conversations.

The Estimate column in the Fixed Effects section provides resonance estimates for each predictor, along with the variability associated with these coefficients (Estimate Error) and the 95% credible intervals (l = lower, u = upper bounds). The intercept indicates a baseline resonance value of 1.48. From the model, we can see that in human interaction, relevance acknowledgement (RA) is a negative predictor of resonance RA (β = -0.74, [l=-1.24, u=-0.24]). This is not surprising and quite in line with what was found in Tantucci and Lepadat (2024), confirming that resonance and marked acknowledgement of what one has said are two often alternative strategies of conversational engagement. Quite differently, recombinant creativity (RC) (β = 1.36, [l=1.18, u=1.54]), shows a very strong association with resonance. This means that interlocutors’ engagement is boosted by creativity and their proactive re-combination of their interactants’ words and constructions. While resonance is well predicted by RA and RC in single utterances, across turns, there is much more variance and uncertainty: SFP(lag) (β = 0.10, [l=-0.15, u=0.36]), RC(lag) (β = 0.06, [l=-0.20, u=0.33]), and RA(lag) (β = -0.31, [l=-0.80, u=0.17]).

All in all, in the context of our data, the most decisive predictors in single turns were relevance acknowledgement (RA, competing with resonance) and recombinant creativity (RC, boosting resonance). Much more uncertain were predictions across turns: B’s resonance as a result of A’s previous use of RA and RC. This is clearly captured in the plots in Figure 1.

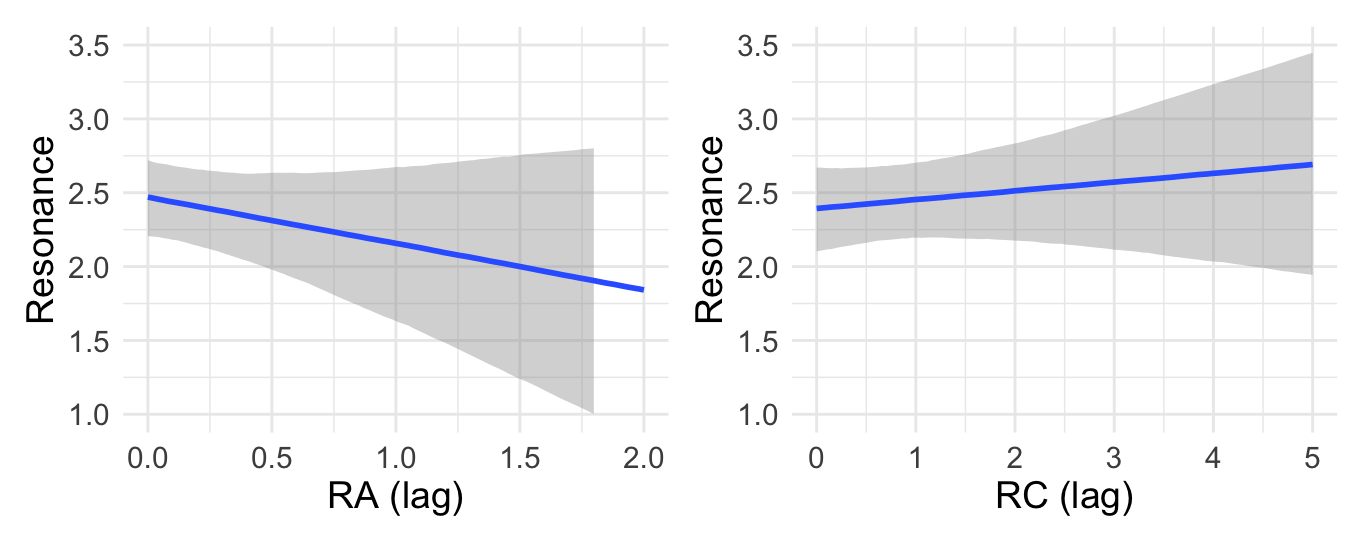
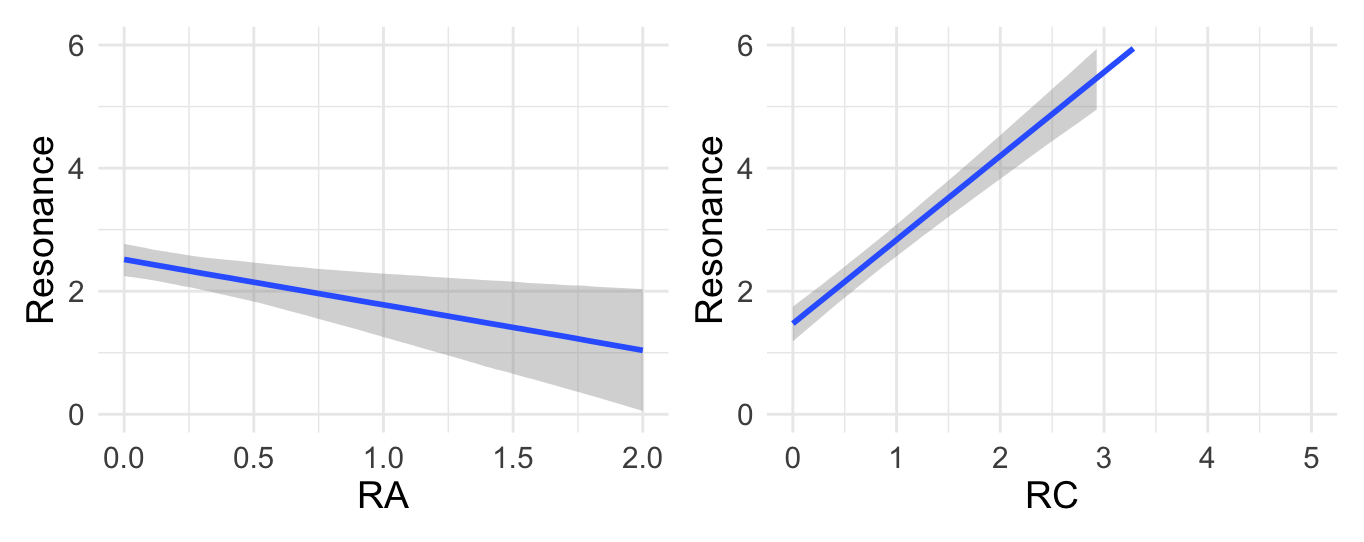


Figure 1: Human relevance acknowledgement and recombinant creativity as predictors of resonance

Three important insights can be gathered with this visualization:

1. Speakers tend to alternate RA and resonance in interaction. They either engage with one or the other. The higher the values of RA, the stronger the speakers’ efforts to go beyond schematic backchannels and use more propositional strategies to acknowledge that what they heard is relevant and interesting. This has a negative effect on resonance as speakers either opt to engage by re-using their peers’ linguistic material or by overtly expressing that what they heard is relevant.
2. In human interaction, recombinant creativity (RC) is a key engagement booster. The more a respondent makes an effort to be creative with the words of his/her interlocutor, the higher the chances that her overall resonance values will be higher. This means higher conversational engagement.
3. The most important insight for this paper has to do with credible intervals (the grey areas around the blue regression lines). These indicate uncertainty. The wider their scope, the less certain the trend. Why are these important? They indicate that we can confidently predict points i. and ii. (The grey areas are quite thin). At the same time, they are pretty fat in the second row. This means that, from our sample, it is hard to predict speaker’s B’s resonance as a result of A’s RA and RC. Put simply, this pertains to conversational uniqueness, it is not an effect that we could expect in every single telephone interaction among Chinese family members.

### 5.2 AI interaction

At this point, we wanted to see if AI could simulate resonance, recombinant creativity, relevance acknowledgement and sentence-final particles in the same way humans enact them in conversation. We uploaded on ChatGPT our dataset, including sociodemographic information (gender and age) but without all of our annotated variables (e.g., resonance, RC, RA, SFP, directive types and so on). This was our prompt:

*This is an annotated series of interactions among family members talking for a bit over the phone. Speakers may be younger (Y) or older (O), males (M) or females (F). Every chat includes a specific Chat.ID. Produce an alternative set of conversations and export another CSV file with the same number of rows.*

We newly annotated the AI-simulated conversations based on the same scheme we used for human exchanges (cf. Sections 3 and 4), leading to stable values of inter-rater agreement: (ICC = 0.88) RC (ICC = 0.84) and RA (0.88). We then fitted mixed effects Bayesian linear regression as we did for human data. The results are reported in. Table 5.

Table 5: Mixed effects linear Bayesian regression of Resonance in AI dialogues

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Random effects | | | | | | | |
| Child ID | (5 Levels) |  |  |  |  |  |  |
|  | Estimate | Est.Error | l-95% CI | u-95% CI | Rhat | Bulk\_ESS | Tail\_ESS |
| sd(Intercept) | 0.11 | 0.09 | 0.01 | 0.34 | 1.00 | 1532 | 2346 |
| CHAT ID:Speaker | (10 Levels) |  |  |  |  |  |  |
| sd(utterance) | 0.06 | 0.05 | 0.00 | 0.18 | 1.00 | 1996 | 2379 |
| Regression coefficents | | | | | | | |
|  | Estimate | Est.Error | l-95% CI | u-95% CI | Rhat | Bulk\_ESS | Tail\_ESS |
| Intercept | 0.36 | 0.09 | 0.20 | 0.54 | 1.00 | 2538 | 2039 |
| SFP | -0.02 | 0.05 | -0.11 | 0.07 | 1.00 | 7837 | 4019 |
| RA | -0.19 | 0.07 | -0.32 | -0.06 | 1.00 | 7406 | 4089 |
| RC | 1.90 | 0.05 | 1.81 | 2.00 | 1.00 | 6143 | 4156 |
| SFP(lag) | -0.01 | 0.05 | -0.11 | 0.08 | 1.00 | 6261 | 4043 |
| RA(lag) | -0.15 | 0.07 | -0.28 | -0.02 | 1.00 | 6712 | 3944 |
| RC(lag) | -0.01 | 0.05 | -0.11 | 0.08 | 1.00 | 6261 | 4043 |

We remarkably found very similar levels of resonance. With the negative effects associated with RA (β = -0.19, [l=0.07, u=-0.32]) and highly positive ones in combination with RC (β = 1.90 [l=1.81, u=2.00]). This means that AI could simulate – on a relatively large scale – the same degrees of conversational engagement and creativity found in humans. In fact, being ‘over-engaging’ runs the risk of extra politeness (cf. Leech 2014; Culpeper and Tantucci 2021), which could lead to mock politeness or sarcasm (cf. Culpeper et al. 2017). What is striking is that ChatGPT could do so without any explicit instruction about what resonance, RC or RA are or how one would express them in dialogue. Put simply, AI did an extraordinary job at independently learning how humans express engagement and the way they do it creatively in the specific context of our data. What is yet even more striking is that AI ‘eliminated uncertainty’, it reduced the credible intervals of all predictions, including the ones we deemed unreliable in Table 4, e.g. the ones across A and B’s turns as per the second row in Figure 1. We can easily capture this in Figure 2 below, which includes the predicted values of resonance based on RA and RC in single turns (first row) and the ones based on RA and RC from A to B (the lag values).

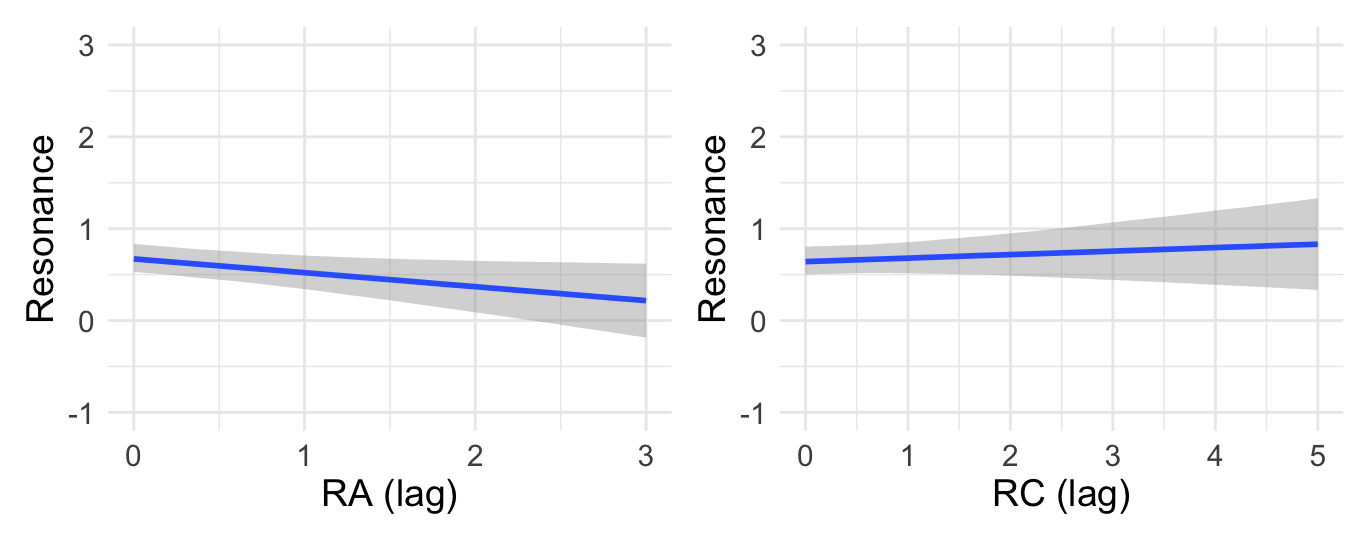
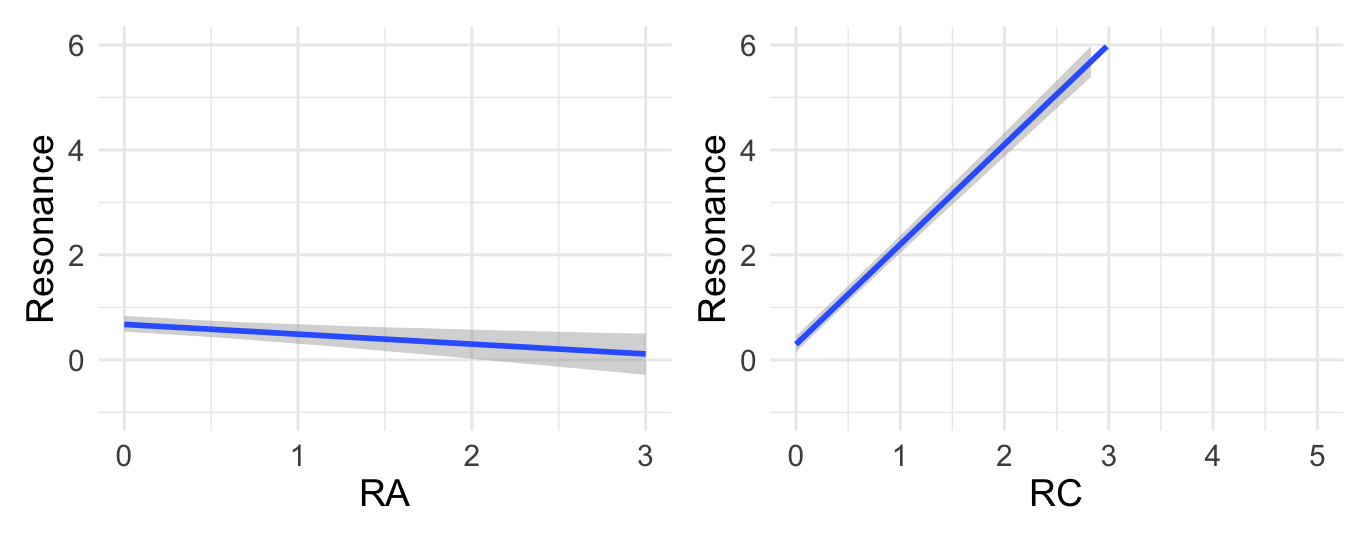
****

Figure 2: AI relevance acknowledgement and recombinant creativity as predictors of resonance

The predicted values or AI resonance are very similar to humans’, but the key here is the ‘grey areas’, the uncertainty of the predictions. These are now much thinner, even in the second row, which was scarcely reliable in human data. This only means one thing: overgeneralization. AI learned the interactional style of the Callhome dialogues and created a stereotypical version of those. This meant producing a set of dialogic exchanges that could always represent that context as predictably as possible. This also meant ‘getting rid of conversational uniqueness’, that is of all the conversational elements that would make each Callhome dialogue distinctive and specific to each dyad of human speakers. In the AI exchange below, a daughter (D) and her father (F) are talking about the living conditions of a mutual friend:



D: 哦，也不太好的。

Ō, yě bù tài hǎo de.

INTJ also NEG too good SFP

‘Oh, it’s also not very good.’

F: 哦。

Ō.

INTJ

‘Oh.’

D: 到现在还没写。

Dào xiànzài hái méi xiě.

until now still NEG write

‘He still hasn’t written until now.’

F: 哎，他没有给你写啊?

Āi, tā méiyǒu gěi nǐ xiě a?

INTJ 3SG NEG to 2SG write SFP

‘Ah, he didn’t write to you?’

D: 没有.

Méiyǒu.

NEG

‘No.’

(AI 0003)

In the exchange in (9), all the conditions that involve conversational engagement are present, and there is resonance e.g. [哦] *ō*, [Subj 没写], RA and SFP at sentence peripheries. These are all features that distinctively characterize Chinese telephone conversations among family members, which are here simulated in the form of a ‘most plausible’ exchange in that context. Human interaction is yet not always a standard version of itself, in some cases, all of those features may be absent, and indeed, the absence of regularities is a key component of any unique human conversation. The mother in (10) offers suggestions (e.g., *contacting the office*) to ensure a message is received. The final lines reflect frustration, as the daughter explains they received nothing.

F: 南京邮电学院, 要么, 外室那个什么.

Nánjīng yóudiàn xuéyuàn, yàome, wài shì nàge shénme.

Nanjing post.telecom university or external office that-CLF what

‘Nanjing University of Posts and Telecommunications, or that external office, whatever it

is.’

F: 意见室可能比较好.

Yìjiàn shì kěnéng bǐjiào hǎo.

suggestions room maybe relatively good

‘The suggestions office might be better.’

F: 问意见室肯定能收到的.

Wèn yìjiàn shì kěndìng néng shōudào de.

ask suggestion room definitely can receive SFP

‘If you ask the suggestions office, it will definitely be received.’

F: 问意见室.

Wèn yìjiàn shì.

ask suggestions room

‘Ask the suggestions office.’

F: 写那个名字啊.

Xiè nàge míngzì a.

write that-CLF name SFP

‘Write down the name.’

F: 写魏济老-.

Xiě Wèi Jì lǎo-.

write Wei Ji old-

‘Write Wei Ji, Old—.’

D: 没有，反正我没有收到。

Méiyǒu，fǎnzhèng wǒ méiyǒu shōudào

NEG，anyway 1SG NEG receive-ASP

‘No. Anyway, I didn’t receive it either.’

(Callhome 0022)

F’s turns do not alternate with D and all occur in the form of bare imperatives. In D’s turn there is no (immediate) resonance, no RC, no RA. As far as our data is concerned, this is a ‘conversational outlier’ as it includes very little of Callhome conversations’ typical behavior. These are the kind of exchanges that were almost absent in AI data, precisely because they are not ‘typical’ enough.

### 5.3 Lexical diversity

Another element where AI did not perform well is lexical diversity, “the range of vocabulary displayed” in written and spoken discourses (Durán et al. 2004: 220). We calculated the Type-Token Ratio (TTR) for Human and AI datasets. TTR is a standard measure of lexical diversity, computed as the ratio of unique words (types) to the total number of words (tokens) in a text (cf. Bottini and Le Foll 2024 on related measures of lexical complexity). We plotted the density distribution of TTR values in Figure 3.

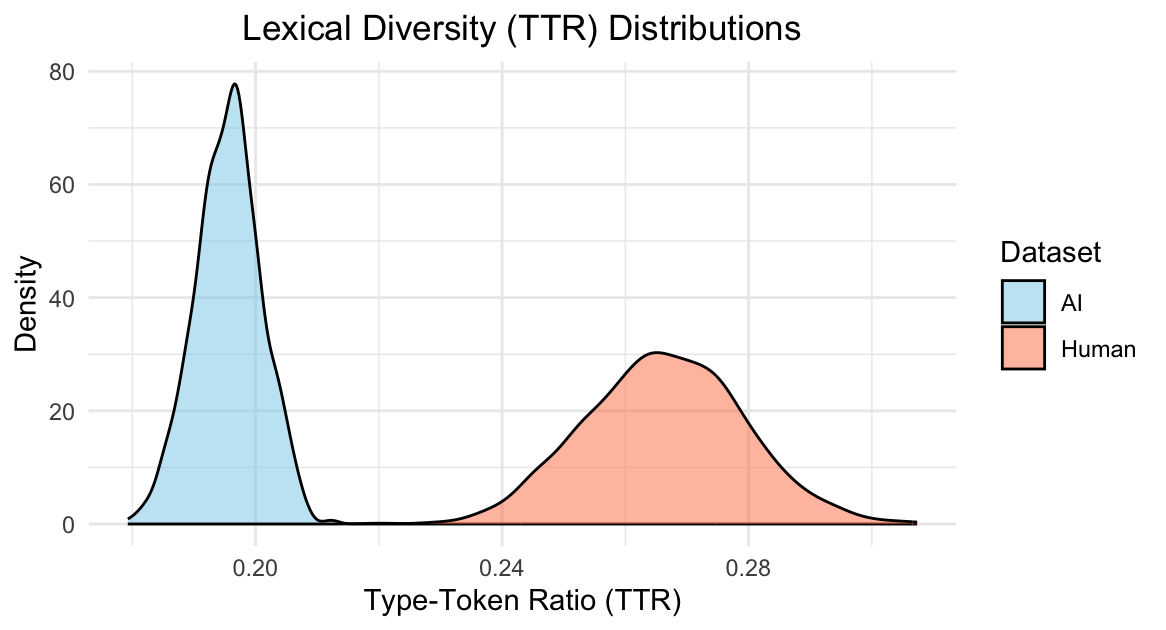
****

Figure 3: Human vs AI lexical diversity

As the blue area suggests, AI-generated interactions have a much lower TTR distribution, clustered around 0.20 (20% of the total words in the dataset are unique). This entails less lexical variation as ChatGPT iteratively focused on the most prototypical lexicon of the Callhome, disregarding of ‘lexical outliers’ that may occur in highly specific conversations. Human dialogues exhibit higher TTR values, peaking around 0.26, indicating greater diversity in word choice. This shows once more that AI conversation is geared towards

### 5.4 Directives’ diversity

The occurrence of directives shows key differences between human (naturalistic) and AI-generated dialogues in terms of distribution and directive frequency, as shown in Table 6. The first observation is that AI-generated dialogues exhibit a significantly higher proportion of directive language (19.2% relative to total tokens). Conversely, in the naturalistic dialogues from our dataset, only 3.8% of all tokens are directives; despite the smaller ratio, their range is more nuanced, covering a cluster of directive classes, whereas in the AI-generated dialogues, these types of utterances revolve around two main communicative functions: Recommendations and Prohibitives.

Table 6: Directives’ types in Human and AI interactions

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Directive types | Naturalistic dialogues | | | AI dialogues | | |
| no. | % total tokens | % directives | no. | % total tokens | % directives |
| Requests | 16 | 1.3% | 35.6% | 2 | 0.8% | 4.1% |
| Clarification Requests | 6 | 0.5% | 13.3% | 0 | 0.0% | 0.0% |
| Request refusals | 3 | 0.3% | 6.7% | 0 | 0.0% | 0.0% |
| Reassurance Requests | 1 | 0.1% | 2.2% | 0 | 0.0% | 0.0% |
| Recommendations | 5 | 0.4% | 11.1% | 32 | 12.5% | 65.3% |
| Prohibitives | 8 | 0.7% | 17.8% | 13 | 5.1% | 26.5% |
| Permissives | 6 | 0.5% | 13.3% | 2 | 0.8% | 4.1% |
| **Total directive tokens** | **45** | **3.8%** | **100.0%** | **49** | **19.2%** | **100.0%** |
| Non-directive tokens | 1155 | 96.3% |  | 206 | 80.8% |  |
| **Total tokens** | **1200** | **100.0%** |  | **255** | **100.0%** |  |

A second observation concerns the directive types being predominant in the two repertoires, namely, Requests for naturalistic dialogues (35.6% of directives) and Recommendations in the AI-generate dataset (65.3% of directives). In the context of our data, requests are performed through imperatives, necessity modals, and direct questions. In both datasets, occurrences of indirect requests are found, uttered via a possibility modal, see (2) in Section 4. As expected, mitigation strategies are widely at play: 58% of the tokens in the naturalistic repertoires are mitigated by items as conditional politeness formulas as 你有空 *Nǐ yǒu kòng* ‘when you have time’, as in (4), question tags, as in (2), and the use of the delimitative aspect, 一下 *yíxià* ‘a bit’, as in (10); however, in the AI-generated dialogues, mitigation markers are found only in 31% of the directive tokens.

M: 也应该的, 应该呢, 松一下

yě yīnggāi de, yīnggāi ne, sōng yīxià

also should SFP, should SFP, relax a.little

**‘**It should also be, should it. Relax a bit.’

(Callhome 0022)

Importantly, requests which are pervasive in the naturalistic repertoire, also correlate with Permissive utterances (13.3% of directives, compared to only one occurrence in the AI repertoire). In human dialogues, requests often initiate negotiation among participants, as shown in the excerpt below. This includes seeking clarification on requests (13.3% of directives) and even refusals of requests (6.7% of directives). As highlighted in the following section, no negotiation between participants is observed in the AI-generated repertoires.

|  |  |  |
| --- | --- | --- |
| M: | (…) 我跟你讲, 有空跟谢嘉魁写一封信哦. | Let me tell you, when you have time, write a letter to Xie Jiakui. |
| D: | 现只是没时间, 呃  我跟你讲, 根本没时间. &=laugh | Right now, I just don't have time, uh.  I'm telling you, I really don't have time. (laughs) |
|
| M: | 哎, 以后, 等你放假了. 有 | Oh, later, when you're on vacation. Yes. |
| D: | 好的,有空,等我放假,现在已经放假了  但是我有一门课,我们七个人做 project@s 结果&=laugh,结果全都 incomplete@s | Okay, when I have time, when I'm on vacation, now I'm already on vacation.  But I have a course, seven of us are doing projects, and (laughs) as a result, all of them are incomplete. |
|
| F: | 你写信不是挺快吗,十分钟就可以写完 | Don't you write letters pretty quickly? You can finish in ten minutes. |
| D: | 哎, 现在没有, | Oh, not now. |
| M: | 哎, 不, 不用, 等他那个,放假有空了,再写好了 | Oh, no, no need, wait until she's on vacation and has time, then write it. |
| D: | 对, 对, 对, 呃哈, | Right, right, right, uh |

(Callhome 055)

Recommendations in the AI-generated dialogues are typically combined with a deontic proposition, expressing a duty for the addressees, with the intent of persuading and advising them on a course of action for their interests and well-being. In the AI dataset, recommendations mostly manifest as stereotypical paternal advice such as 注意身体 *zhùyì shēntǐ* ‘pay attention to your health’ and 保暖  *bǎonuǎn ‘*keep warm’. When addressing their parents, the polite form 您 *nín* ‘you’ is used, which is highly unlikely in a family conversation. Recommendations are also found as indirect directives, as in (11), where the advice to value stability is expressed as an adage. Notably, indirect directives are also more frequent in AI (33%, no. 16), compared to human dialogue (11%, no. 5), potentially as a strategy to mimic non-confrontational and supporting behaviors. In the case of (11) below, a daughter is seeking advice from her father regarding her intention to change jobs.

F: 年轻人要多尝试，不过稳定也很重要。

Niánqīng rén yào duō chángshì, bùguò wěndìng yě hěn zhòngyào

young people must more try, but stability also very important

‘Young people must explore more options, but maintaining stability is equally important.’

(AI 0004)

Prohibitives are also present in the two datasets: 8 occurrences in the naturalistic dialogues (0.7% of tokens, 17.8% of directives) and 13 occurrences (5.1% of tokens, 26.5% of directives) in the AI generated; the latter uses Prohibitives slightly more frequently. Notably, in the latter repertoire, Prohibitives typically instantiate in the same tokens including a recommendation and are followed by a Commissive speech acts (see Section 4). The correlation between Prohibitives and Commissives is observed in both the repertoires, with dialogues patterns including a negative deontic, as *búyào* 不要 ‘must not’, followed by a commissive statement by the addressee.

Another important difference in the pragmatics of the two repertoires is the distribution of Commissives (promises, offers), which in the naturalistic data amounts to 4 occurrences (0.4% of all tokens), whereas in the AI-generated dialogues they reach 37 occurrences (15% of all tokens), reflecting a preference for conclusive and action-oriented utterances; in the AI dataset, Commissives are mainly expressed via commissive future (Sparvoli and Saillard 2024) in the *huì …* (*de*) construction, occurring in 24 tokens (65% of Commissives). Importantly, a similar instance is also found in the naturalistic dataset (12); however, in this case, the Prohibitives-Commitment sequence is closed with a reassurance request: 不会的, 是吧? *Búhuì de, shìba*? ‘It won’t happen, right?’ tinging the exchange with an emotional nuance that cannot be found in the AI dataset. In the case below, the daughter is asking her father how he is handling the grandmother’s passing.

D: 算了, 你不要, 不要太伤心了.

Suànle, nǐ búyào, búyào tài shāngxīn le.

forget.it, 2SG NEG.must NEG.must too sad SFP

‘Let it be, don't be too sad.’

F: 这不会的.

Zhè búhuì de.

this NEG.FUT SFP

‘It's fine.’ (Literally: ‘This won’t happen)

D: 不会的, 是吧?

Búhuì de, shì ba?

NEG.FUT SFP, right SFP

‘It's fine, right?’ (Literally: ‘It won’t happen, right?)

(Callhome 003)

The two repertoires vary not only for the different distribution of Commissives (pervasive in AI). A difference is also found in the strategies employed, namely, mostly commissive future in AI and more mixed construction in naturalistic dialogues, including a type of affirmative resonance, as in (13), wherein the addressee endorses the suggested course of action by echoing the very words uttered by the addressor.

D: (…)你写信现在还可以写到这,但包裹先不要寄啊.

Nǐ xiě xìn xiànzài hái kěyǐ xiě dào zhè, dàn bāoguǒ xiān búyào jì a.

2SG write letter now still can write arrive here, but parcel first NEG.must send.

'You can still write letters up to this point, but don't send the parcel yet.'

F: 哦,包裹先不寄.

Ó, bāoguǒ xiān bù jì.

INTJ, parcel first NEG send.

'Oh, I don't send the parcel yet.'

(Callhome 0027)

In summary, AI-generated dialogues are more directive-heavy, focusing on action-oriented language (such as Prohibitives co-occurring with Commissives). This higher proportion of directives suggests a task-oriented conversational style, with less flexibility for casual or exploratory interactions.

Less directive overall, human talk is more geared to mitigation patterns, displays greater interpersonal negotiation (as shown by the prominence of Requests), and exhibits a broader and more balanced distribution of directive types that reflect social and emotional dynamics. This is evidenced by a higher proportion of requests, clarifications, permissions, refusals, and reassurance-related exchanges.

The AI dataset frequently provides stereotypical Recommendations, though within a limited vocabulary. It appears less inquisitive about the addressor’s intention or adaptive in maintaining conversational flow, avoiding clarifications, refusals, and negotiations, possibly due to design limitations or politeness constraints, favoring directiveness over conversational nuance.  
The narrower range of directive types in AI dialogues may indicate avoidance of socially sensitive acts (such as refusals and reassurance requests), favoring an over-use of commissive responses to endorse the addressee’s Recommendations. To this end, AI dialogues heavily rely on the Commissive future, whereas in naturalistic dialogues, commissive speech acts are performed using a broader range of constructions.

## 6 Conclusions

This study shed new light on the pragmatics of AI. It tested ChatGPT’s ability to simulate human interaction in the context of telephone conversations among Chinese family members. AI showed outstanding accuracy at reproducing humans’ ability to engage in conversation using resonance, relevance acknowledgement, sentence-final particles and recombinant creativity. However, it did so at the expense of conversational uniqueness. It produced a set of highly stereotypical conversations that would match how a ‘most predictable dialogue under certain conditions would occur’. At the same time, it has lost sight of the lexical and illocutionary diversity that characterizes human communication: people don’t want to just be ‘proper’ they also want to avoid being boring and repetitive. Even more importantly, AI rarely included less conventional ways to converse, which, in turn, are necessarily present in most human communication. This led to a highly stereotypical representation of human dialogue that exaggerated some of its features (e.g. engagement, relevance acknowledgement, recombinant creativity, the predominant use of recommendations and commissives, and so on). These results bear relevance for our current understanding of AI’s representation of human communication and shed light on what is perhaps the most compelling limitation of its immense computational power: a constant over-generalization of human behavior.

## References

Arundale, Robert B. 1999. An alternative model and ideology of communication for an alternative to politeness theory. *Pragmatics* 9(1). 119–153

Bach, Kent. 1997. The Semantics-Pragmatics Distinction: What It Is and Why It Matters. In Eckard Rolf (ed.), *Pragmatik. Linguistische Berichte*, 33-50. Wiesbaden: VS Verlag für Sozialwissenschaften. <https://doi.org/10.1007/978-3-663-11116-0_3>

Barattieri di San Pietro, Chiara, Federico Frau, Veronica Mangiaterra & Valentina Bambini. 2023. The pragmatic profile of ChatGPT: Assessing the communicative skills of a conversational agent. *Sistemi Intelligenti* 35(2). 379-399. <https://doi.org/10.1422/108136>

Blum-Kulka, Shoshana, Julian House & Gabriele Kasper (eds.). 1989. *CrossCultural Pragmatics. Requests and Apologies*. Norwood: Ablex.

Brown, Penelope & Stephen C. Levinson. 1987. *Politeness: Some Universals in Language Usage*. Cambridge: Cambridge University Press.

Cialdini, Robert B. & Noah J. Goldstein.  2004. Social influence: Compliance and conformity. *Annual Review of Psychology* 55(1). 591-621.

Clark, Herbert H.  1996. *Using Language*. Cambridge: Cambridge University Press.

Cong, Yan. 2024. Manner implicatures in large language models. *Scientific Reports* 14. 29113. <https://doi.org/10.1038/s41598-024-80571-3>

Craven, Alexandra & Jonathan Potter. 2010. Directives: Entitlement and contingency in action. *Discourse Studies* 12(4). 419–442.

Culpeper, Jonathan, Michael Haugh & Valeria Sinkeviciute. 2017. (Im)politeness and mixed messages. In Jonathan Culpeper, Michael Haugh & Dániel Z. Kádár (eds.), *The Palgrave handbook of linguistic (im)politeness*, 323–355. London: Palgrave Macmillan. <https://doi.org/10.1057/978-1-137-37508-7_14>

Culpeper, Jonathan & Vittorio Tantucci. 2021. The principle of (im) politeness reciprocity. *Journal of Pragmatics* 175. 146-164.

Culpeper, Jonathan, Vittorio Tantucci & Eleanor Field. 2025. Impoliteness reciprocity online. *Journal of Pragmatics, 242*, 216–236. https://doi.org/10.1016/j.pragma.2025.04.011

Curl, Traci S. & Paul Drew. 2008. Contingency and Action: A Comparison of Two Forms of Requesting. *Research on Language and Social Interaction* 41(2). 129–153.

Diedrichsen, Elke. 2025. (This Volume). Common ground in artificial intelligence applications. *Intercultural pragmatics.*

Dingemanse, Mark. 2020. Resource-rationality beyond individual minds: The case of interactive language use. *Behavioural and Brain Sciences* 43. 23–24.

Du Bois, John W. & Rachel Giora. 2014. From cognitive-functional linguistics to dialogic syntax*. Cognitive Linguistics* 25(3). 351-357.

Du Bois, John W. 2014. Towards a dialogic syntax. *Cognitive Linguistics* 25(3). 359–410.

Dynel, Marta. 2023. Lessons in linguistics with ChatGPT: Metapragmatics, metacommunication, metadiscourse and metalanguage in human-AI interactions. *Language & Communication* 93. 107-124. <https://doi.org/10.1016/j.langcom.2023.09.002>

Ervin-Tripp, Susan. 1976. Is Sybil there? The structure of some American English directives. *Language in Society* 5(1). 25–66.

Heine, Bernd. 2023. *The grammar of interactives*. Oxford: Oxford University Press.

Heinemann, Trine. 2006. ‘Will you or can't you?’: Displaying entitlement in interrogative requests. *Journal of Pragmatics* 38(7). 1081-1104.

Heritage, John. 2012. Epistemics in action: Action formation and territories of knowledge. *Research on language & social interaction* 45(1). 1-29. https://doi.org/10.1080/08351813.2012.646684

[Hu](https://arxiv.org/search/cs?searchtype=author&query=Hu,+J), Jennifer, [Sammy Floyd](https://arxiv.org/search/cs?searchtype=author&query=Floyd,+S), [Olessia Jouravlev](https://arxiv.org/search/cs?searchtype=author&query=Jouravlev,+O), [Evelina Fedorenko](https://arxiv.org/search/cs?searchtype=author&query=Fedorenko,+E) & [Edward Gibson](https://arxiv.org/search/cs?searchtype=author&query=Gibson,+E). 2022. A fine-grained comparison of pragmatic language understanding in humans and language models. <https://doi.org/10.48550/arXiv.2212.06801>

[Jones](https://arxiv.org/search/cs?searchtype=author&query=Jones,+C+R), Cameron R. & [Benjamin K. Bergen](https://arxiv.org/search/cs?searchtype=author&query=Bergen,+B+K). 2024. People cannot distinguish GPT-4 from a human in a Turing test. https://doi.org/10.48550/arXiv.2405.08007

Kecskes, Istvan. 2008. Dueling contexts: A dynamic model of meaning. *Journal of Pragmatics* 40(3). 385-406.

Kecskes, Istvan. 2010. The paradox of communication: A socio-cognitive approach. *Pragmatics and Society* 1(1). 50-73.

Kecskes, Istvan. 2023. *The socio-cognitive approach to communication and pragmatics*. Cham: Springer.

Kissine, Mikhail. 2013. *From Utterances to Speech Acts*. Cambridge: Cambridge University Press.

Kohnen, Thomas. 2004. Let mee bee so bold to request you to tell mee”: Constructions with let me and the history of English directives. *Journal of Historical Pragmatics* 51. 159–173.

Natarajanm Ranjini & Robert E. Kass. 2000. Reference Bayesian methods for generalized linear mixed models. *Journal of the American Statistical Association* 95(449). 227-237.

Pandia, Lalchand, Yan Cong & [Allyson Ettinger](https://arxiv.org/search/cs?searchtype=author&query=Ettinger,+A). 2021. Pragmatic competence of pre-trained language models through the lens of discourse connectives.<https://doi.org/10.48550/arXiv.2109.12951>

Pickering, Martin J. & Simon Garrod. 2021. *Understanding dialogue: Language use and social interaction*. Cambridge: Cambridge University Press.

Ruis, Laura, Akbir Khan, Stella Biderman, Sara Hooker, Tim Rocktäschel & Edward Grefenstette. 2022. Large language models are not zero-shot communicators. <http://dx.doi.org/10.48550/arXiv.2210.14986>

Ruiz de Mendoza Ibáñez, Francisco José & Lorena Pérez Hernández. 2002. Metonymy and the grammar motivation, constraints and interaction. *Language & Comunication* 21(4). 321-357. <https://doi.org/10.1016/S0271-5309(01)00008-8>

Sandler, Morgan, Hyesun Choung, Arun Ross and Prabu David. 2024. A linguistic comparison between human and ChatGPT-generated conversations. <https://doi.org/10.48550/arXiv.2401.16587>

Seals, S. M. & Valerie L. Shalin. 2023. Discourse over discourse: The need for an expanded pragmatic focus in conversational AI. https://doi.org/10.48550/arXiv.2304.14543

Searle, John. R. 1976. A Classification of Illocutionary Acts. *Language in Society* 5(1). 1-23.

Sparvoli, Carlotta & Claire Saillard. 2024. Futurity reading of *hui.* Distributional features from a corpus-based study. Paper presented at the 12th Conference of the European Association of Chinese Linguistics, Università Roma3, 19-20 September.

Sparvoli, Carlotta. 2019. Modality in the general linguistic investigations carried out in China before 1949. In Barbara Meisterernst (ed.), *New perspectives on Aspect and Modality in Chinese Historical Linguistic*s (Frontiers in Chinese Linguistics Series 5), 135-157. Singapore: Springer.

Sperber, Dan & Deirdre Wilson. 1995. *Relevance: Communication and Cognition*. Oxford (UK): Blackwell.

Stalnaker, Robert. 2002. Common ground. *Linguistics and Philosophy* 25(5-6). 701-721.

Sydorenko, Tetyana, Judit Dombi, Ameeta Agrawal, Steven L. Thorne, [Jung In Lee](https://www.taylorfrancis.com/search?contributorName=Jung%20In%20Lee&contributorRole=author&redirectFromPDP=true&context=ubx) & [Yufei Tao](https://www.taylorfrancis.com/search?contributorName=Yufei%20Tao&contributorRole=author&redirectFromPDP=true&context=ubx). 2024. Spoken dialogue systems and ChatGPT for second language pragmatics research. In Karim Sadeghi (ed.), *The Routledge Handbook of Technological Advances in Researching Language Learning*, 378-391. Abingdon, Oxon: Routledge.

Takahashi, Hidemitsu. 2014. A usage-based analysis of indirect directives in English (1): A preliminary quantitative survey. *The Annual Report on Cultural Science* 143. 99–135.

Tantucci, Vittorio. 2015. Traversativity and grammaticalization: The aktionsart of 过 guo as a lexical source of evidentiality. *Chinese Language and Discourse* 6(1). 57-100.

Tantucci, Vittorio. 2017a. From immediate to extended intersubjectification: A gradient approach to intersubjective awareness and semasiological change. *Language and Cognition* 9(1).88-120.

Tantucci, Vittorio. 2017b. An evolutionary approach to semasiological change: Overt influence attempts through the development of the Mandarin 吧-ba particle. *Journal of Pragmatics* *120*. 35-53.

Tantucci, Vittorio. 2021. *Language and social minds: The semantics and pragmatics of intersubjectivity*. Cambridge University Press.

Tantucci, Vittorio. 2023. Resonance and recombinant creativity: Why they are important for research in Cognitive Linguistics and Pragmatics. *Intercultural Pragmatics* 20(4). 347-376.

Tantucci, Vittorio. (forthcoming). *Language and (Creative) Imitation: Dialogic resonance in Pragmatics and Grammar*. Cambridge University Press.

Tantucci, Vittorio, & Aiqing Wang. (2021). Resonance and engagement through (dis-) agreement: Evidence of persistent constructional priming from Mandarin naturalistic interaction. *Journal of Pragmatics*, *175*, 94-111.

Tantucci, Vittorio & Aiqing Wang. 2022a. Dynamic resonance and explicit dialogic engagement in Mandarin first language acquisition. *Discourse Processes*, *59*(7), 553-574.

Tantucci, Vittorio & Aiqing Wang. 2022b. Resonance as an applied predictor of cross-cultural interaction: Constructional priming in Mandarin and American English interaction. *Applied Linguistics* 43(1). 115-146. <https://doi.org/10.1093/applin/amab012>

Tantucci, Vittorio & Aiqing Wang. 2023. Dialogic priming and dynamic resonance in autism: creativity competing with engagement in Chinese children with ASD.  *Journal of autism and developmental disorders*, *53*(6), 2458-2474.

Tantucci, Vittorio & Aiqing Wang, 2024. [British conversation is changing : Resonance and engagement in the BNC1994 and the BNC2014](https://www.research.lancs.ac.uk/portal/en/publications/british-conversation-is-changing(44194ead-46ba-44a2-958d-8e4c4f15bc46).html). *Applied Linguistics*. <https://doi.org/10.1093/applin/amae040>

[Tantucci, Vittorio](https://www.research.lancs.ac.uk/portal/en/people/vittorio-tantucci(78af78d6-3f4e-4b2c-8efd-d3140da63a7d).html) & Carmen Lepadat. 2024. [Verbal engagement in doctor–patient interaction: Resonance in Western and Traditional Chinese Medicine.](https://www.research.lancs.ac.uk/portal/en/publications/verbal-engagement-in-doctorpatient-interaction-resonance-in-western-and-traditional-chinese-medicine(1c33437a-7ede-48ee-a64e-f32716e2884d).html)  *Journal of Pragmatics* 230. 126-141.

Tantucci, Vittorio, Aiqing Wang & Jonathan Culpeper. 2022. Reciprocity and epistemicity: On the (proto) social and cross-cultural ‘value’of information transmission. *Journal of Pragmatics* 194. 54-70.

Ten Bosch, Louis, Nelleke Oostdijk, & Jan Peter de Ruiter. 2004. Durational aspects of turn-taking in spontaneous face-to-face and telephone dialogues. In [Petr Sojka](mailto:sojka@informatics.muni.cz?Subject=TSD%212004%21Proceedings), [Ivan Kopecek](http://www.fi.muni.cz/~kopecek/) & [Karel Pala](http://www.fi.muni.cz/usr/pala/) (eds.), *Text, Speech and Dialogue: 7th International Conference, TSD 2004, Brno, Czech Republic, September 8-11, 2004, Proceedings*, 563-570. Berlin Heidelberg: Springer.

Terkourafi, Marina. 2015. Conventionalization: A new agenda for im/politeness research. *Journal of Pragmatics* 86. 11-18.

Traugott, Elizabeth Closs. 2012. Intersubjectification and clause periphery. *English Text Construct* 5(1). 7-28.

Traugott, Elizabeth Closs. 2016. On the rise of types of clause-final pragmatic markers in English. *Journal of Historical Pragmatics* 17(1). 26-54.

Weigand, Edda. 2018. Dialogue: The key to pragmatics. In Edda Weigand & Istvan Kecskes (eds.), *From pragmatics to dialogue, 5*–28. Amsterdam/Philadelphia: Benjamins.

Wilson, Deirdre. 2006. Relevance and the role of attention in communication. In (eds.) *Attention, Information and Interaction*, 219-236. Cambridge: Cambridge University Press.

**Bionotes**

**Vittorio Tantucci** is Senior Lecturer of Linguistics and Chinese Linguistics at Lancaster University, UK, Editorial Board Meber of Journal of Pragmatics, Intercultural Pragmatics and Frontiers in Human Neuroscience. His publications focus on usage-based intersections of interactional pragmatics, computational and cognitive linguistics. His recent and forthcoming books include *Language and Social Minds: The Semantics and Pragmatics of Intersubjectivity* (CUP, 2021); *Different Slants on Grammaticalization* ([Benjamins, 2023](https://www.sciencedirect.com/science/article/pii/S0378216625000943" \l "bib69); co-edited with Sylvie Hancil) and *Language and (Creative) Imitation: Dialogic resonance in Pragmatics and Grammar* (CUP, forthcoming).

**Carlotta Sparvoli** is Associate Professor at Ca’ Foscari University of Venice, and previously held positions at the University of Bologna and University College Cork (Ireland), where she directed the MA in TCSOL. Her research focuses on modality and the acquisition and teaching of L2 Chinese. She has authored numerous papers on the expression of modality in both modern and Classical Chinese and has served on the Executive Boards of the International Association of Chinese Linguistics (IACL) and the European Association of Chinese Linguistics (EACL), as well as President of the Italian Association (AILC).

1. https://ca.talkbank.org/access/CallHome/zho.html. [↑](#footnote-ref-1)
2. Model diagnostics indicate reliable estimates, with Rhat values of 1.00 for all parameters, signaling good convergence. Bulk\_ESS and Tail\_ESS values are sufficiently high, ensuring stability in posterior sampling. [↑](#footnote-ref-2)