Cognitive Mechanisms of Turn-Taking (Post-print)

Authors: Wenbo Yu, Liang Dandan

Date: 2026-02-15T09:57:47+00:00

Abstract

Turn-taking is a fundamental mechanism in the operation of conversation. The most commonly used method in turn-taking research is conversation analysis, which typically employs a descriptive approach to summarize linguistic features during the turn-taking

Full Text

The Cognitive Mechanism Underlying the Process of Turn-taking

YU Wenbo and **LIANG Dandan**

Abstract

Conversation is the fundamental medium through which human language is expressed, and turn-taking forms its primary ecological niche. While conversation analysis (CA) provides a robust framework for describing the principles and characteristics of turn-taking such as Sacks' insights into how turns are allocated, CA relies on "off-line" data from recordings or transcripts of conversations, which is less concerned with the mental processes of participants during conversation. In contrast, researches on psycholinguistics and neurolinguistics offer insights into how the brain functions during turn-taking among two or more interlocutors. Over the past decade, a large number of studies have investigated the cognitive processes involved in turn-taking. This article synthesizes the findings from these studies. While earlier research emphasized the role of semantic and grammatical cues in guiding turn-taking, recent work suggests that interlocutors also rely heavily on prosodic cues to identify transition-relevance places. Moreover, prosodic information may play an even more significant role than previously thought. Additionally, event-related potential (ERP) studies reveal that the brain begins preparing the content of the next turn as soon as

it gathers sufficient information, rather than waiting for the current speaker to finish. This finding supports the early planning model and aligns corpus-based analyses showing that turn-taking in conversation is rapid and seamless, which greatly enhances the efficiency of conversation. Conversations also contain rich pragmatic information. Research shows that interlocutors process this information in real-time, as reflected in EEG data capturing both time-domain and frequency-domain responses. Beyond summarizing these three cognitive processes, namely prosody, early planning, and real-time pragmatic processing, this article also explores future research directions. First, we note that many studies reviewed here rely on a "question-answer" experimental paradigm, which limits the ecological validity of findings. We recommend incorporating more online data capturing real-time neural activity during natural turn-taking. Second, the experimental materials used in psycholinguistic and neurolinguistic studies are often overly simplistic, failing to fully replicate real-world conversations. This highlights the value of integrating CA with experimental data for more comprehensive studies of conversation. Finally, we emphasize investigating the development trajectory of turn-taking abilities, particularly in infants and children with autism. We hope this article contributes to a deeper understanding of the cognitive process underlying turn-taking and provides valuable guidance for future research.

Keywords: turn-taking, transition-relevance place, psycholinguistics, neurolinguistics

1. Introduction

The turn constitutes the fundamental structural unit of everyday conversation, referring to the continuous speech produced by a speaker during any given period in a conversation, marked by the exchange of roles between speaker and hearer or by mutual silence (Liu 1992, 2004). Conversation Analysis (CA) represents the most commonly employed methodology in turn-taking research, which primarily involves obtaining audio/video recordings of natural conversation, transcribing these recordings, analyzing selected segments, and reporting research findings (Ma 2014). The most classic research on turn-taking derives from Sacks et al. (1974), who defined concepts such as the turn, the transition-relevance place (TRP), and proposed three fundamental principles governing turn-taking.

Qualitative research methods represented by CA emphasize induction and summary derived from linguistic phenomena, analyzing large quantities of conversational data to identify linguistic features characterizing the turn-taking process. This approach focuses on linguistic outcomes, emphasizing the description of turn-taking processes through discourse results (phonetic, semantic, and syntactic features). However, in real conversational scenarios, the brain must process linguistic information in real-time, involving multiple cognitive processing stages. Describing the process solely from linguistic outcomes cannot determine what "operations" the brain performs during turn-taking or the sequential order of these operations. Psycholinguistic and neurolinguistic research, by contrast,

emphasizes viewing linguistic outcomes as the result of coordinated activity across a series of cognitive stages, with the core objective of utilizing eye-tracking technology, event-related potentials (ERP), and functional magnetic resonance imaging (fMRI) to decompose seemingly holistic language processing into distinct cognitive components and identify relevant influencing factors.

Adopting a psycholinguistic (neurolinguistic) perspective, this study reviews cutting-edge research on turn-taking, summarizing three processing stages involved: (1) listeners primarily utilize prosodic cues to judge transition-relevance places; (2) listeners prepare speech production content in advance, consistent with the Early Planning Model; and (3) the brain processes pragmatic information in real-time.

2. Listeners Mainly Use Prosodic Cues to Judge Transition-Relevance Places

2.1 The Cueing Function of Prosodic Information

The alternation of turns is a process in which the listener anticipates or judges that the current turn is about to end and the next turn is about to begin based on various cues provided by the turn-constructional unit projected by the speaker' s discourse (Le 2016). Related research has found that syntax, intonation, and non-verbal information (such as body movements) can all serve as cues, and the more cues appear simultaneously, the greater the likelihood that the speaker intends to yield the turn (Duncan 1972; Oreström 1983). Generally, transition-relevance places tend to occur at positions where semantic and syntactic sequences are completed; meanwhile, grammatical projection and conversational projection share certain commonalities, with the former being viewed as the sedimentation and transformation of the latter (Auer 2005). For example, in English, predicate verbs occur relatively early, allowing listeners to predict subsequent verb structures after hearing a noun phrase, and subsequently to predict the nominal elements in object position (Wan 2018). Early research also found that listeners indeed rely on syntactic-semantic cues for prediction (De Ruiter et al. 2006; Magyari et al. 2014). However, recent scholarship has increasingly inclined toward the view that during turn-taking, the brain relies exclusively on prosodic information to judge transition-relevance places, noting that syntactic-semantic information alone is insufficient to predict turn-ending times (Stephens and Beattie 1986; Bögels and Torreira 2015). For instance, example (2) below includes all the syntactic-semantic content of (1); thus, according to the view that syntactic-semantic information plays the primary role, listeners' judgments of the ending time for sentence (2) should be consistent with those for sentence (1).

(1) So are you a student?

(2) So are you a student here at Radboud University?

Figure 1

Figure 1: Figure 1

In reality, however, subjects' judgments of the ending time for 96% of type (2) sentences were concentrated within the final few syllables before sentence end or immediately at sentence termination. Subsequently, Bögels and Torreira (2015) conducted splicing manipulations of long and short sentences at semantic and prosodic levels (as shown in Figure 1), finding that under "full replacement (short)" and "partial replacement (short)" conditions, subjects' judgments of turn transition points were concentrated at 400ms after the actual ending, significantly longer than the "original (short)" condition (102ms). In the long sentence version, for trials under the original sentence (2) condition, no subjects pressed the button to judge conversation completion after "student"; conversely, some subjects pressed at corresponding time points in the other two conditions ("full replacement (long)" and "partial replacement (long)"). This indicates that in predicting turn endings, listeners rely not only on syntactic-semantic information but also require prosodic cues—specifically, duration and pitch cues at intonational phrase boundaries.

Although the aforementioned studies confirm that listeners rely on certain linguistic cues to anticipate transition-relevance places, the experimental tasks predominantly employed button-press paradigms requiring subjects to judge when a speech stimulus ended. Due to the absence of authentic dialogue processes, these tasks diverge considerably from real turn-taking scenarios and likely fail to reflect the anticipatory mechanisms operative during turn-taking. On the other hand, Bögels and Torreira's (2015) research found that listeners primarily judged turn endings based on late prosodic indicators (the F0 and duration information of the final word in the intonational phrase under the "original (short)"condition in Figure 1, rather than prosodic information from words earlier in the phrase), implying that the anticipatory function of prosodic information serves primarily to signal turn completion rather than representing active prediction by the listener. Some scholars have further investigated how transition-relevance places are judged, proposing a two-stage model of transition-relevance place prediction.

2.2 The Two-Stage Model of Transition-Relevance Place Prediction

Researchers from the Levinson team at the Max-Planck Institute employed a list-completion paradigm, presenting subjects with images of three to five common objects and requiring them to orally produce the objects not named by their conversational partner after hearing the partner' s naming utterances. Within the partner' s sentences, the researchers designed early prosodic cues, late prosodic cues, and lexical cues. The experiment collected not only response time data for subjects' oral productions but also eye-movement indicators of subjects' gaze toward images. The results revealed that lexical cues not only shortened subjects' oral production response times but also facilitated earlier gaze toward the

unnamed objects, whereas late prosodic cues affected only response times without influencing eye-movement indicators. Based on these findings, Torreira et al. (2015) and Barthel et al. (2017) proposed the two-stage model, positing that listeners prepare conversational content through the speaker's semantic material, utilize late prosodic cues as markers of turn completion, and need only vocalize rapidly when the speaker' s turn concludes. Magnetoencephalography (MEG) research has similarly found that when individuals must wait for a signal to produce content, alpha-beta band (8–30Hz) power in the occipital cortex shows marked decreases while beta band (12–20Hz) power in the frontal cortex shows marked increases, respectively related to attentional resource allocation and the maintenance of sensorimotor activity and cognitive states, further supporting the applicability of the two-stage model during listener speech production.

Beyond the Levinson team' s findings, additional experimental evidence supports the two-stage model. Such research typically examines listeners' predictions of transition-relevance places throughout the entire turn-taking process, finding that linguistic information in conversation merely assists listeners in understanding conversational content but cannot help them predict when the speaker' s turn will end (Corps et al. 2018, 2019). Corps et al. (2018) established predictive and non-predictive textual conditions: in the former, conversational content (e.g., "Are dogs your favorite···?") could predict the target word (e.g., "animal"), whereas in the latter, content (e.g., "Do you enjoy going to the ···?") could not assist listeners in predicting the target word (e.g., "supermarket/dentist/beach"). Two experiments respectively employed button-press tasks identical to De Ruiter et al. (2006) and oral report tasks, finding that in button-press tasks, subjects' response times (the difference between button-press time and actual turn-ending time, reflecting how early subjects began preparing conversational content) and accuracy (the absolute value of the difference between button-press time and actual turn-ending time, reflecting precision of subjects' turn-ending predictions) showed no significant differences in main effects; whereas in oral report tasks, subjects' response times were earlier in the predictive condition, though accuracy differences remained non-significant. In another study (Corps et al. 2019), the authors employed longer textual content to investigate whether individuals need to predict transition-relevance places, establishing dialogue and monologue conditions. Results found only that when subjects heard predictive conversational content, response times were shorter, though accuracy still showed no significant improvement.

The two-stage model posits that interlocutors do not predict transition-relevance places through syntactic-semantic information but need only capture turn-ending projection information (primarily prosodic information) to initiate their turn. Currently, the two-stage model not only satisfactorily explains results from both conversation analysis and psycholinguistics but also aligns with the Early Planning Model discussed below. However, several issues regarding whether and how to anticipate transition-relevance places require attention. First, differences in experimental tasks may account for contradictory results among studies. Some studies follow the De Ruiter et al. (2006)

paradigm, presenting subjects only with the speaker' s utterance content and requiring prediction of turn-ending times, which diverges substantially from real turn-taking scenarios, whereas Corps et al. (2018) and the Levinson team' s tasks typically require subjects to complete dialogue tasks, such as orally producing unnamed objects, more closely approximating authentic conversational situations. Furthermore, the cognitive activities undertaken by subjects in the De Ruiter et al. (2006) paradigm may differ from those in actual turn-taking processes; thus, subsequent research must weigh the advantages and disadvantages of experimental paradigms to reach reliable conclusions. Second, regarding whether listeners need only project locally based on prosodic information, intermediate perspectives exist. For instance, Heldner and Edlund (2010) suggest that for turn-taking times exceeding 200ms, listeners do not anticipate but rather begin vocalizing only after the speaker' s turn ends, whereas for intervals under 200ms, listeners do predict turn endings. Some researchers propose that listeners can employ two methods to complete turn-taking: when syntactic-semantic information is rich, listeners tend to anticipate turn-ending time points, but when only syntactic or prosodic information is available, listeners less frequently utilize anticipatory mechanisms (Riest et al. 2015).

3. Interlocutors' Advance Preparation of Production Content

This section explores the question of when listeners begin preparing the content they intend to express (Levinson and Torreira 2015; Corps et al. 2018). Existing empirical research shows marked disagreement on this issue: some studies support the Late Planning Model, positing that listeners concentrate attention on understanding the speaker' s content and only begin planning their own utterance content when the turn is nearly complete; other research supports the Early Planning Model, proposing that listeners prepare desired content as quickly as possible based on the speaker' s verbal information, thereby ensuring rapid articulation when their own turn begins (Bögels and Levinson 2017).

3.1 Late Planning Model

Experimental evidence supporting the Late Planning Model derives primarily from dual-task experiments (Boiteau et al. 2014; Sjerps and Meyer 2015), which assume that speech planning inevitably consumes cognitive resources, thereby reducing performance levels in control tasks. Consequently, one can determine when speech planning begins by observing the temporal dynamics of performance decline in control tasks. Sjerps and Meyer' s (2015) experiment required subjects to simultaneously complete a finger-tapping task and a dialogue task. In the comprehension dialogue condition, subjects listened to two people conversing and judged whether the second person' s response was reasonable; in the production dialogue condition, subjects responded based on heard speech fragments; in the control condition, subjects performed no dialogue task. Beyond comparing finger-tapping performance, the experiment also recorded eye-

movement indicators. Results revealed that in the production dialogue condition, finger-tapping performance showed significant decline only approximately one second before the first speaker' s turn ended (compared to the comprehension dialogue condition). Eye-movement data showed that in the comprehension condition, subjects' gaze trajectories shifted according to both speakers' speech content, but in the production condition, subjects' gaze only shifted toward the image they needed to produce when the first speaker' s content was nearly complete. These results indicate that listeners only begin planning their production content when the speaker is nearly finished with their turn—that is, late planning. However, some scholars have questioned the dual-task paradigm, noting that the small number of images used and the fixed sentence structures between speaker and subject greatly reduced experimental difficulty; additionally, they argue that although data indicators show subjects only attended to the object to be named at late turns, this does not demonstrate that they did not begin planning earlier (Barthel et al. 2016).

3.2 Early Planning Model

The Early Planning Model originated from Stivers et al.' s (2009) corpus study, which collected large quantities of informal conversational speech from ten languages across five continents, extracting yes-no question-answer pairs and wh-question-answer pairs. Acoustic analysis of the corpus revealed that conversational partners in all languages attempt to avoid conversational overlap, and that turn intervals are brief across all ten languages. Statistical analysis of the frequency distribution of turn transition times found that the highest-frequency transition times fell within 200ms, with an overall mean of 208ms. However, speech production research indicates that producing even a single word requires at least 600ms, and depending on word frequency and priming effects, may require up to 1200ms (Indefrey and Levelt 2004). This suggests that turn-taking intervals are insufficient to support speakers "preparing-producing" a complete speech fragment; thus, it is reasonable to infer that listeners must begin preparing their own content before the speaker has finished their turn.

Bögels et al. (2015a) first proposed the Early Planning Model, with subsequent evidence including eye-movement research supporting this theoretical position (Barthel et al. 2016; Magyari et al. 2017). Bögels et al. (2015a) designed experimental and control groups in an ERP experiment. In the experimental group, subjects answered heard questions; in the control group, subjects memorized heard questions without answering. Both groups heard identical sentences divided into early and late conditions:

Early condition: *Which character, also called 007, appears in the famous movies?*
Late condition: *Which character from the famous movies, is also called 007?*

Researchers recorded ERP indicators while subjects listened to sentences, conducting time-domain and frequency-domain analyses using "007" and "movies"

as keywords. Results revealed that 500ms after the keyword "007" presentation, both conditions in the experimental group elicited positive-going components at parietal sites; while the control group showed similar positive components, the difference was significantly smaller than in the experimental group. Frequency-domain analysis similarly revealed analogous interactions: within 500ms–1500ms after the first keyword presentation, the experimental group showed marked decreases in alpha-band energy, though no interaction was found when the second keyword was presented. The researchers posit that the positive component elicited by the keyword relates to speech production, while alpha-band energy decrease signifies that subjects shifted more attention from comprehension to answer retrieval and production processes.

Barthel et al. (2016) employed the list-completion paradigm to investigate this issue. The first independent variable was whether the conversational partner' s sentence ending contained a verb. According to the Early Planning Model, after hearing the final noun spoken by the partner (knowing which objects the partner named), the listener can prepare the sentence to be produced; thus, for conditions where the ending contains a verb, subjects can utilize the verb' s duration to begin preparation, resulting in shorter response times than in conditions without a verb, with no eye-movement differences between conditions. Conversely, according to the Late Planning Model, listeners wait until all verbal content is spoken before planning production content; thus, no response time differences should exist between conditions, but because sentences with verb endings are longer, gaze shifts would appear later. The second independent variable was whether the partner' s sentence contained words cueing transition-relevance place information. In conditions containing words allowing prediction of the ending, listeners can predict sentence structure and components, effectively anticipating the transition-relevance place, resulting in significantly shortened eye-movement indicators and response times; otherwise, neither measure would differ significantly. Experimental results found that subjects showed shorter response times in the verb-ending sentence condition and looked at target objects earlier, supporting the Early Planning Model (Torreira et al. 2015; Barthel et al. 2017).

According to the Early Planning Model, while comprehending the speaker' s verbal content, listeners simultaneously prepare speech production. Due to divided cognitive resources, the planning process likely interferes with comprehension of the speaker' s production content. In experiments, researchers placed semantically expected or unexpected words at sentence endings in the early planning condition; ERP results found that unexpected words elicited N400 effects, and correlation analyses revealed that subjects with faster behavioral responses elicited smaller N400 effects, indicating that these subjects invested more resources in the planning process (Bögels et al. 2018). This finding indirectly supports the Early Planning Model and demonstrates that parallel "comprehension-production"processing exists in the listener's mind during turn-taking. Although experimental evidence supporting the Early Planning Model is more abundant, the interference of planning with comprehension violates the economy principle

of cognitive processing. Therefore, some scholars speculate that for listeners, there is no clear time point for when to begin preparing content; evidence supporting the Early Planning Model merely emphasizes that listeners can begin planning when sufficient information is gathered. If sufficient information only becomes available at the speaker' s content ending, experimental results would likely support the Late Planning Model. In other words, when listeners begin planning depends on the actual conversational situation.

4. Real-Time Processing of Pragmatic Information in Turn-Taking

Grice (1975) proposed the Cooperative Principle in language use, emphasizing that interlocutors follow certain social norms, transmitting and obtaining information within determined topics. Therefore, when focusing on turn-taking research, we must also emphasize listeners' processing and utilization of information regarding communicative intentions and speech acts.

4.1 Inference of Interlocutors' Cooperative Willingness

Previous cross-cultural research shows that turn-taking times hover around 200ms, with affirmative or prosocial responses typically appearing earlier than negative or non-prosocial responses (Stivers et al. 2009). Conversation analysis research has similarly found that accepting responses to proposals and invitations usually occur earlier, whereas rejecting responses tend to be delayed (Heritage 1984; Pomerantz and Heritage 2013). Kendrick and Torreira (2015), through acoustic analysis of 195 conversational fragments, found that only when turn-taking times exceeded 700ms did the proportion of negative responses increase substantially; this result was confirmed in behavioral experiments. Roberts and Francis (2013) presented conversational fragments to subjects, requiring them to judge the degree of listeners' willingness to respond to speakers' requests, finding that when turn-taking times exceeded 700ms, subjects' ratings showed marked decline.

ERP evidence not only demonstrates that turn-taking time length can affect speakers' expected judgments of listeners' responses (Bögels et al. 2015b), but also reveals the changing process of speakers' expectations during longer transition times (Bögels et al. 2020). For example, subjects first understood background information through textual materials (e.g., "The speaker is talking with a friend who is busy this week due to a new job"), then heard recorded materials (e.g., "Do you have time to host us next week?"). The independent variables were two lengths of turn-taking time (1000ms or 300ms) and answer type ("yes" or "no"). Results found that in the 300ms condition, negative answers elicited obvious N400 components compared to positive answers; since positive responses typically appear earlier, researchers believe this ERP component stems from conflict between turn-taking time and answer type (Bögels et al. 2015b). In subsequent research, Bögels et al. (2020) increased background

materials, strengthening subjects' expectations of answer type (positive expectation vs. negative expectation), finding that in longer transition times, the negative expectation condition elicited larger positive components after 300ms. Previous research indicates that when individuals expect upcoming stimuli or behaviors, negative-going ERP components accompany this expectation; the positive component appearing in the negative condition in this study suggests that subjects at this time (300ms) had not yet begun expecting the upcoming negative answer, further indicating that during turn-taking, listeners dynamically process speakers' cooperative willingness.

4.2 Recognition and Processing of Speech Acts

Speech acts refer to the behavior of language use itself or behaviors elicited in listeners, encompassing rich categories including, for listeners, directives, promises, and declarations. However, in daily conversation, speakers do not specifically emphasize their speech acts; thus, do listeners need to recognize the speech act corresponding to speech content, and when does this recognition occur? Employing ERP technology, scholars have discovered that listeners complete speech act recognition at very rapid speeds during turn-taking (Gisladottir et al. 2015, 2018).

In Gisladottir et al. (2015), the authors designed three speech acts: refusal, offer, and answer. Target sentences eliciting different speech acts were identical across conditions, but background contexts differed, as shown in Table 1 . Subjects listened to two people conversing and then judged what the second person' s behavioral response was. The experiment collected EEG signals, analyzing the first and last words of target sentences as critical stimuli.

Analysis of the first word revealed that compared to the answer condition, refusal and offer behaviors respectively elicited frontal positive components 200ms and 400ms after target word presentation. Analysis of the last word revealed that only the offer behavior elicited obvious late negative waves at posterior sites. The researchers posit that during turn-taking, listeners can rapidly identify speech acts, and specific behaviors elicit more complex subsequent processing mechanisms. In subsequent research (Gisladottir et al. 2018), the authors used identical materials to further analyze frequency-domain indicator changes, finding that 200ms before refusal behavior appeared, alpha-beta bands showed desynchronization phenomena with significantly decreased energy values, while refusal behavior also elicited decreased theta-band energy values. Typically, anticipatory mechanisms cause alpha-beta band energy decreases; the authors infer that background information guides subjects to guess upcoming speech acts. For instance, refusal behavior carries certain social-emotional information; in such cases, textual content guides subjects to pay attention to or expect upcoming socially dispreferred rejection information.

5. Prospects and Conclusion

5.1 Conducting Online Research on Dialogue Processes

The greatest characteristic of conversation research lies in the necessity of simultaneously including both (or multiple) parties in experiments and examining their interactive relationships. The studies reviewed here simplify turn-taking to "question-answer" processes, mostly requiring button-press responses to record subjects' behavioral and neural indicators, relatively neglecting the interactive process between conversational parties. Pickering and Garrod (2004) propose that people in dialogue automatically integrate private and common information rather than processing language in an egocentric manner. During integration, an interactive alignment phenomenon exists: during conversation, the listener' s linguistic representation automatically aligns with the speaker' s at multiple levels, aiming to make the listener' s representation converge with the speaker' s, thereby increasing the likelihood that the listener can accurately predict the speaker' s speech (Pickering and Garrod 2004, 2013; Sui et al. 2021).

In recent years, the role of low-frequency neural oscillations (cortical oscillation) in tracking speech envelope signals and capturing temporal information has received widespread attention (Arnal and Giraud 2012; Garrod and Pickering 2015). Existing research shows that most language syllables last approximately 200ms (Song and He 2005); thus, speech envelope information frequency is approximately 5Hz, similar to the oscillation frequency of brain cortex theta waves. Researchers further hypothesize that speech rate plays an important role in turn-taking (Wilson and Wilson 2005), with some studies indeed finding that listeners' speech rates during production are influenced by the speech rate of detected stimuli (Jungers et al. 2002; Jungers and Hupp 2009). Combining these experimental evidences, we can infer that the auditory cortex automatically tracks speech rate information and achieves alignment between conversational parties, thereby facilitating smooth turn-taking. Future research can continue from this angle, using real conversation scenarios as experimental blueprints.

5.2 Implementing Joint Research Between Conversation Analysis and Cognitive Neuroscience

Research employing psycholinguistic and neurolinguistic techniques and methods for turn-taking has gradually increased over the past decade, mostly centering on two aspects: listeners' expectations of turn endings and preparing conversational content. In fact, language processing issues involved in the conversational process extend beyond these. First, regarding adjacency pair forms, existing cognitive experiments predominantly employ "question-answer" structures as experimental materials; however, real language situations include various forms of adjacency pairs, such as wishes (A: Wish you a pleasant journey. B: Thanks.), suggestions (A: Let' s have a meeting this afternoon. B: No, I don' t have time.), etc. The reasons for differences in turn-taking times across different adjacency pair forms remain unknown, particularly when accounting

for positive/negative responses and speech acts; appropriate material design can more accurately restore real dialogue scenarios.

Second, conversation analysis methods can reveal numerous linguistic rules and phenomena in conversational processes. For example, in assertion-type conversations across Finnish, Japanese, and Mandarin Chinese cultures, overlapping turns demonstrate universal patterns: (1) affirmative preface + understanding statement; (2) independent stance repetition (Endo et al. 2018; Vatanen et al. 2020). However, such research is often static description or qualitative analysis. Subsequent research can adopt a language processing perspective, creating similar natural language environments to observe subjects' turn production.

5.3 Attention to the Developmental Trajectory of Human Turn-Taking Ability

Research on adult turn-taking mechanisms has achieved relatively abundant results, but for infants in the pre-linguistic stage, relevant research findings remain scarce. From a communication science perspective, information transmission occurs not only through speech signals; in many protoconversation processes, infants have already begun taking turns with caregivers in controlling the interaction. For example, although mothers dominate during nursing, they also adjust according to the infant's sucking speed and intensity, ultimately reaching a rhythm suitable for both parties. It is generally believed that turn-taking is a social communication skill that develops relatively early and is closely related to language ability (Tomasello and Farrar 1986; Ninio and Snow 1996), but its relationship with language remains unclear.

Meanwhile, turn-taking disorder (conversational disorder) represents a typical language deficit in autistic children. During conversation, they often struggle to provide valuable information, sometimes constantly repeating what others have just or previously said, or producing adjacency pair second parts that are unrelated to the topic or overly detailed (Baltaxe 1977; Adams et al. 2002; Tantucci and Wang 2022). Whether such turn-taking deficits in this population appear during infancy remains unknown; future research that pays more attention to children' s turn-taking behavior in the pre-linguistic stage will better serve screening and rehabilitation for language disorder groups.

5.4 Conclusion

Research on the cognitive processes of turn-taking from a language processing perspective can complement traditional conversation analysis, exploring the brain' s representational processes for this language phenomenon while obtaining turn-taking structures and characteristics. Through systematic review, this study identifies three clear processing stages in turn-taking: (1) listeners mainly utilize prosodic cues to judge transition-relevance places; (2) listeners prepare production content in advance, conforming to the Early Planning Model; and (3) the brain processes pragmatic information in real-time. We hope this article

promotes the development of related research.

References

Liu, Hong. 1992. Turn, non-turn and semi-turn. *Foreign Language Teaching and Research* (3): 17–24.

Liu, Hong. 2004. *Analysis of Conversational Structure*. Beijing: Peking University Press.

Ma, Chunyan. 2014. Analysis of conversational structure and gender construction in Chinese polylogue: A study based on data from TV talk shows. PhD diss., Zhejiang University.

Song, Yannan, and He Wei. 2005. Statistics analysis of the duration of standard Mandarin monosyllables. In *Proceedings of the 8th National Conference on Man-Machine Speech Communication*. Communication Acoustics Laboratory, Communication University of China.

Sui, Xue, et al. 2021. Perspective taking and its cognitive mechanism in language processing. *Advances in Psychological Science* (6): 990–999.

Wan, Quan. 2018. The minor sentence as a locus of grammar and interaction in Chinese. In *Interactive Linguistics and Chinese Studies* (Vol. 2), eds. Fang Mei and Cao Xiuling, 16–32. Beijing: Social Sciences Academic Press.

Le, Yao. 2016. The most relevant turn-projection units in Mandarin conversation. In *Interactive Linguistics and Chinese Studies* (Vol. 1), ed. Fang Mei, 49–74. Beijing: World Book Publishing Company.

Adams, Catherine, et al. 2002. Conversational behaviour of children with Asperger syndrome and conduct disorder. *Journal of Child Psychology and Psychiatry and Allied Disciplines* 43: 679–690.

Arnal, Luc H., and Anne-Lise Giraud. 2012. Cortical oscillations and sensory predictions. *Trends in Cognitive Sciences* 16 (7): 390–398.

Auer, Peter. 2005. Projection in interaction and projection in grammar. *Text* 25 (1): 7–36.

Baltaxe, Christiane A. M. 1977. Pragmatic deficits in the language of autistic adolescents. *Journal of Pediatric Psychology* 2: 176–180.

Barthel, Mathias, et al. 2016. The timing of utterance planning in task-oriented dialogue: Evidence from a novel list-completion paradigm. *Frontiers in Psychology* 7: 1858.

Barthel, Mathias, et al. 2017. Next speakers plan their turn early and speak after turn-final "go-signals" . *Frontiers in Psychology* 8: 393.

Bögels, Sara, and Francisco Torreira. 2015. Listeners use intonational phrase boundaries to project turn ends in spoken interaction. *Journal of Phonetics* 52:

46–57.

Bögels, Sara, and Stephen C. Levinson. 2017. The brain behind the response: Insights into turn-taking in conversation from neuroimaging. *Research on Language and Social Interaction* 50 (1): 71–89.

Bögels, Sara, et al. 2015a. Neural signatures of response planning occur midway through an incoming question in conversation. *Scientific Reports* 5: 12881.

Bögels, Sara, et al. 2015b. Never say no…: How the brain interprets the pregnant pause in conversation. *PLOS ONE* 10 (12): e0145474.

Bögels, Sara, et al. 2018. Planning versus comprehension in turn-taking: Fast responders show reduced anticipatory processing of the question. *Neuropsychologia* 109: 295–310.

Bögels, Sara, et al. 2020. Conversational expectations get revised as response latencies unfold. *Language, Cognition and Neuroscience* 35 (6): 766–779.

Boiteau, Timothy W., et al. 2014. Interference between conversation and a concurrent visuomotor task. *Journal of Experimental Psychology: General* 143 (1): 295–311.

Corps, Ruth E., et al. 2018. Early preparation during turn-taking: Listeners use content predictions to determine what to say but not when to say it. *Cognition* 175: 77–95.

Corps, Ruth E., et al. 2019. Predicting turn-ends in discourse context. *Language, Cognition and Neuroscience* 34 (5): 615–627.

De Ruiter, Jan P., et al. 2006. Projecting the end of a speaker' s turn: A cognitive cornerstone of conversation. *Language* 82 (3): 515–535.

Duncan, Starkey. 1972. Some signals and rules for taking speaking turns in conversations. *Journal of Personality and Social Psychology* 23 (2): 283–292.

Endo, Tomoko, et al. 2018. Agreeing in overlap: A comparison of response practices and resources for projection in Finnish, Japanese and Mandarin talk-in-interaction. *The Japanese Journal of Language in Society* 21 (1): 160–174.

Garrod, Simon, and Martin J. Pickering. 2015. The use of content and timing to predict turn transitions. *Frontiers in Psychology* 6: 751.

Gisladottir, Rosa S., et al. 2015. Conversation electrified: ERP correlates of speech act recognition in underspecified utterances. *PLOS ONE* 10 (3): e0120068.

Gisladottir, Rosa S., et al. 2018. Oscillatory brain responses reflect anticipation during comprehension of speech acts in spoken dialog. *Frontiers in Human Neuroscience* 12: 34.

Grice, Paul. 1975. *Logic and Conversation*. New York: Academic Press.

Heldner, Mattias, and Jens Edlund. 2010. Pauses, gaps and overlaps in conversations. *Journal of Phonetics* 38 (4): 555–568.

Heritage, John. 1984. *Garfinkel and Ethnomethodology*. Cambridge: Polity.

Indefrey, Peter, and Willem J. M. Levelt. 2004. The spatial and temporal signatures of word production components. *Cognition* 92: 101–144.

Jungers, Melissa K., and Julie M. Hupp. 2009. Speech priming: Evidence for rate persistence in unscripted speech. *Language and Cognitive Processes* 24 (4): 611–624.

Jungers, Melissa K., et al. 2002. Time after time: The coordinating influence of tempo in music and speech. *Cognitive Processing* 1 (2): 21–35.

Kendrick, Kobin H., and Francisco Torreira. 2015. The timing and construction of preference: A quantitative study. *Discourse Processes* 52 (4): 255–289.

Levinson, Stephen C., and Francisco Torreira. 2015. Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology* 6: 731.

Magyari, Lilla, et al. 2014. Early anticipation lies behind the speed of response in conversation. *Journal of Cognitive Neuroscience* 26 (11): 2530–2539.

Magyari, Lilla, et al. 2017. Temporal preparation for speaking in question-answer sequences. *Frontiers in Psychology* 8: 211.

Ninio, Anat, and Catherine E. Snow. 1996. *Pragmatic Development*. London: Routledge.

Oreström, Bengt. 1983. *Turn-taking in English Conversation*. Lund: Gleerup.

Pickering, Martin J., and Simon Garrod. 2004. Toward a mechanistic psychology of dialogue. *The Behavioral and Brain Sciences* 27 (2): 169–190.

Pickering, Martin J., and Simon Garrod. 2013. An integrated theory of language production and comprehension. *The Behavioral and Brain Sciences* 36 (4): 329–347.

Pomerantz, Anita, and John Heritage. 2013. Preference. In *The Handbook of Conversation Analysis*, eds. Jack Sidnell and Tanya Stivers, 210–228. Hoboken, NJ: Wiley-Blackwell.

Riest, Carina, et al. 2015. Anticipation in turn-taking: Mechanisms and information sources. *Frontiers in Psychology* 6: 89.

Roberts, Felicia, and Alexander L. Francis. 2013. Identifying a temporal threshold of tolerance for silent gaps after requests. *The Journal of the Acoustical Society of America* 133 (6): 471–477.

Sacks, Harvey, et al. 1974. A simplest systematics for the organization of turn taking for conversation. *Language* 50 (4