

COGNITIVE FOUNDATIONS OF REAL-TIME LANGUAGE COMMUNICATION: TOWARD A THEORETICAL FRAMEWORK OF BEHAVIORAL LINGUISTICS

MUNEO KITAJIMA ¹ (PRESENTER) MAKOTO TOYOTA ² JÉRÔME DINET ³
KATSUKO T. NAKAHIRA ¹

¹NAGAOKA UNIVERSITY OF TECHNOLOGY, JAPAN
MKITAJIMA@KJS.NAGAOKAUT.AC.JP

²T-METHOD, JAPAN

³UNIVERSITÉ DE LORRAINE, FRANCE





Muneo Kitajima is currently a professor emeritus at Nagaoka University of Technology. His recently published book "Memory and Action Selection in Human-Machine Interaction" (2016) proposes a unified theory of action selection and development by integrating PDP, Two Minds, and layered structure of human action. The theory provides a comprehensive view of how our brain functionally works in our daily life. His current interest is to understand the implications of the theory to development of skill of adaptive problem solving, the important skill for survival.

Project HP: <http://oberon.nagaokaut.ac.jp/ktjm/organic-self-consistent-field-theory/index.html>

HP: <http://oberon.nagaokaut.ac.jp/ktjm/index.html>

email: <mailto:mkitajima@kjs.nagaokaut.ac.jp>

- This paper explores the cognitive processes involved in real-time language communication during conversations. It emphasizes the dynamic interplay between speakers and listeners, highlighting how both verbal and nonverbal cues contribute to effective communication.
 - The purpose of this paper is to construct a theory of behavioral linguistics on the cognitive architecture, Model Human Processor with Realtime Constraints (MHP/RT), that explains how language is generated in real-time, addressing the limitations of traditional models based solely on “Skinner’s behavioral psychology for verbal development with System 1” or “Chomsky’s linguistic theory grounded in the development of formal grammar through System 2”, and regarding current natural language being formed through the interaction of Chomsky’s grammatical system built upon Skinner’s foundation.
 - The proposed theory of behavioral linguistics offers a comprehensive framework for understanding real-time language generation, integrating insights from cognitive psychology and behavioral economics. By framing language use as an action controlled by bounded rationality, this paper highlights the cognitive constraints that influence how individuals communicate. This perspective encourages a more nuanced understanding of conversational competence, suggesting that effective communication relies not only on linguistic knowledge but also on the ability to anticipate and adapt to the dynamic nature of interactions.
 - The implications for improving communication in various contexts, such as education and therapy, underscore the relevance of this research in enhancing everyday interactions and reducing misunderstandings.
-

Skinner's Analysis of Verbal Behavior

- Humans engage in conversations with close friends and family, enriching daily life and fostering fulfillment.
- Conversation roles of speaker and listener alternate, maintaining dialogue flow, where listeners prepare responses while speakers convey information.
- The speaker uses both linguistic (auditory) and nonverbal (visual) cues to communicate effectively.
- Skinner's analysis reveals challenges in explaining generative verbal behavior, indicating limitations in relying solely on conversational information.

Introduction of Dual-Process Theory

- Complex information processing may be involved in verbal behavior; Speakers adjust speech based on listeners' states, with content selection occurring unconsciously.
- Concepts of bounded rationality and satisficing principle highlight decision-making processes in verbal communication.
- Kahneman's dual-process theory underpins the cognitive basis of bounded rationality, providing a foundation for behavioral economics.

Proposal of Behavioral Linguistics

- The proposed theory of behavioral linguistics examines real-time language generation through the cognitive framework of Model Human Processor with Realtime Constraints (MHP/RT).
- In MHP/RT, linguistic behavior arises from two processes: System-2-Before-Event-Mode (conscious preparation) and System-1-Before-Event-Mode (unconscious adjustments).
- Verbal behavior is a kind of routine cognitive skill modeled by Goals, Operators, Methods, and Selection rules (GOMS), enabling instant responses to conversational changes.

The main objective of the paper is to propose a theory of behavioral linguistics to explain the real-time generation of language in everyday conversation.

Nature of Smooth Conversation

- This study focuses on *smooth conversation* (dialogue) between two individuals.
- In this conversation, both *verbal* and *nonverbal* communication occur.

Verbal Communication

Information is transmitted between participants through auditory information conveyed by sound waves.

Nonverbal Communication

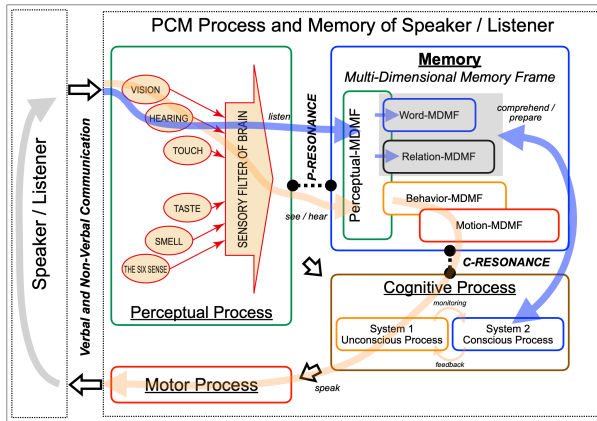
Information is transmitted through information from each modality conveyed via the five senses (vision, hearing (including filler information such as interjections), touch, smell, and taste).

Possible Conversation States

- This study focuses on verbal communication to delve deeply into the conversation itself.
- In verbal communication, the observable states of speakers can be divided into “speaking” and “not speaking.”
- The possible states are:
 1. *One party is speaking, while the other is not;*
 2. *Neither party is speaking;*
 3. *Both parties are speaking simultaneously.*

The Targeted Conversation

- This study examines smooth conversation. In such conversations, both parties concentrate on the conversation and aim to achieve its purpose.
- Each should be performing verbal behaviors efficiently, without wasting time on unnecessary thought.
- Therefore, this study assumes the first situation, i.e., 1) *One party is speaking, while the other is not.*
- This means that conversation participants have the roles of “speaker” and “listener,” and the conversation proceeds through alternating speakers.
- In this study, the timing of speaker change—that is, the moment when the event of speaker alternation occurs—is considered to be the instant when the speaker changes to the listener.



PCM Processes and Memory:

- *PCM Processes and Memory* shown above is based on the Model Human Processor with Realtime Constraints (MHP/RT) [1, 2], by which environmental information is taken into the body via sensory nerves, processed in the brain, and then acted upon by the external world via motor nerves.
- PCM processes use memory, modeled as Multi-dimensional Memory Frame (MDMF), consisting of Perceptual-, Behavior-, Motor-, Relation-, and Word-Multi-dimensional Memory Frame, abbreviated as P-, B-, M-, R-, and W-MDMF, respectively.

Features of MHP/RT and MDMF:

- P-MDMF overlaps with B-, R-, and W-MDMF; this unique configuration supports spreading activation from P- to M-MDMF, which connects perception with bodily movements.
- Perceptual information *resonates* with information in MDMF, called P-Resonance.
- In the cognitive process, conscious process (System 2) utilizes W- and R-MDMF via C-Resonance, and unconscious process (System 1) utilizes B- and M-MDMF via C-Resonance.
- The motor sequences are expressed according to M-MDMF.

PCM Processing and Memory during Conversation:

- Regardless of whether a conversation participant is acting as a speaker or listener, information taken into the brain via sensory organs activates the P-MDMF through P-resonance.
- The activation propagates within the MDMF, undergoing processing via the following two pathways.
 - 1) Conscious information processing by System 2 is executed via C-resonance between information propagated from the P-MDMF to W- and R-MDMF. – *indicated by the blue arrow*
 - 2) Unconscious information processing by System 1 occurs via C-resonance with information propagated along the P-, B-, and M-MDMF. This process connects to motor processes, resulting in physical actions reflecting the active state of M-MDMF. – *indicated by the yellow arrow*

Speakers and listeners execute the PCM process using the Multi-Dimensional Memory Frame during conversation to perform the following tasks.

Speaker's Tasks

- **Utterance:** Continuously speaks words based on the selected content;
- **Adjustment:** Adjusts one's speech content while observing the listener's reaction to one's own utterances or while listening to one's own utterances.

Listener's Tasks

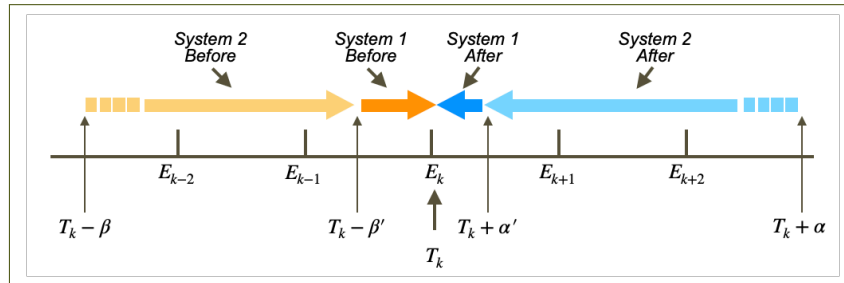
- **Understanding:** Comprehends the content of the other person's utterances. Here, both verbal and nonverbal information is utilized;
- **Nonverbal responses:** Exhibits unconscious nonverbal reactions (eye contact, facial expressions, gestures, interjections, etc. while listening to the other person's utterances and observing accompanying actions (nonverbal information);
- **Preparation:** Deliberates on what to say after the speaker alternation while listening to the other person's utterances.

As shown in the figure, the PCM process is a cycle. For example, when a speaker hears their own utterance, notices a mispronunciation, and corrects it, the process connects as follows: Motor (utterance) ⇒ Perception (listening) ⇒ Cognition (detecting the mispronunciation, deciding on a correction method) ⇒ Motor (utterance) ⇒ ... In this way, the PCM process runs as a continuous cycle, but by breaking it at consciously recognized events, it can be perceived as a sequence of events. When considering a conversational behavior, a representative event is speaker alternation. Alternatively, one may become conscious of their own previous utterances while speaking, as if noticing a slip of the tongue. This too is an event occurring during conversation. By focusing on such events, the unbroken PCM cycle can be captured through the following four processing modes.

The experience associated with an individual's activity is characterized by a series of events that are consciously recognized serially. Let $E(T_k)$ denote the event that occurred at time T_k . The experience is then defined as a series of events along the timeline as follows:

$$\dots \rightarrow E(T_{k-1}) \rightarrow E(T_k) \rightarrow E(T_{k+1}) \rightarrow \dots$$

Considering the way System 1 and System 2 are involved in individual events, *four processing modes* can be defined as shown in the following figure.



Let us focus on an event that occurs at time T_k . For an “event $E(T_k)$ ” that should occur at time T_k , there exist System 2 conscious processes and System 1 unconscious processes related to $E(T_k)$ before that time T_k . Also, for the “executed event $E(T_k)$ ” at time T_k , there exist unconscious processes of System 1 and conscious processes of System 2 involving $E(T_k)$ after that time T_k . MHP/RT's System 1 and System 2 operate before and after the event $E(T_k)$ in one of four processing modes for this event.

Before the Event ($T < T_k$)

$E(T_k)$ reflects the result of P-Resonance between the MDMF and the perceptual and cognitive systems, and is generated by the activities of System 1 and System 2 cognitive processes in the time period before T_k . The different time bands of processing activities result in two processing modes before the event $E(T_k)$. The two processing modes are:

- **System-2-Before-Event-Mode:** In the time range of $T_k - \beta \leq t < T_k - \beta'$, MHP/RT plans for future events to occur. There is enough time to think carefully.
- **System-1-Before-Event-Mode:** In the time range of $T_k - \beta' \leq t < T_k$, the action selections smoothly generate the immediate event. Here, a series of action selections is executed through feedforward processing led by System 1. During this time, System 2 evaluates the results of the action selections in a timely manner. If it determines that the system is likely to deviate from the expected trajectory or has already deviated, it issues instructions to System 1 for trajectory correction.

Here, $\beta > \beta'$, $150\text{msec} < \beta' < T_k - T_{k-1}$, and β ranges from seconds to hours and months.

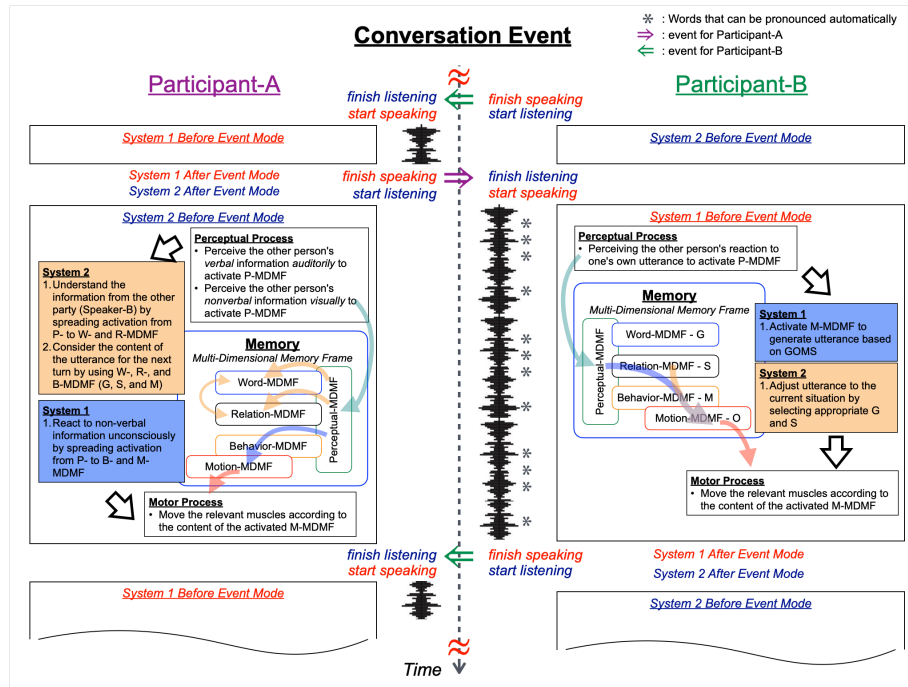
After the Event ($T > T_k$)

The result of $E(T_k)$ is stored in the MDMF. Actions occur by integrating the resonances that emerge through interacting with the environment prior to the event, and after the actions are taken, they are bundled and collected. The MDMF is updated to reflect the results of $E(T_k)$ by the activities of System 1 and System 2 during the time period after T_k . The different time bands of processing activities result in two processing modes after the event. The two processing modes are:

- **System-1-After-Event-Mode:** In the time range of $T_k < t \leq T + \alpha'$, to perform better for the same event that may be encountered in the future, the connection between the incoming perceptual information and the output motor content is adjusted unconsciously.
- **System-2-After-Event-Mode:** In the time range of $T_k + \alpha' < t \leq T_k + \alpha$, the event is reviewed and reflected upon. The results are stored and used in the next System-2-Before-Event-Mode before a similar event occurs.

Here, $\alpha' > 150\text{msec}$, and ranges from seconds to months. In these two modes, action selection results for $E(T_k)$ would be reflected in the network connections of the respective MDMF.

The PCM process and memory during a conversation between conversation participant A and B



The left area shows the process when conversation participant A is listening to conversation participant B's utterance.

The right area shows the process when conversation participant B is speaking.

The processes described in these areas are explained in the next slide.

What Participant A Does while Listening

(a) Understanding the Content of Utterance

System 2-1: Green and yellow arrows indicate propagation of activation. The speaker's utterance is incorporated as auditory information to activate the P-MDMF. This activation propagates to the W- and R-MDMF, enabling understanding of the utterance content in System-2-Before-Event-Mode, represented as an activation pattern within the MDMF.

(b) Preparation for Utterance

System 2-2: Propagation of activation within the MDMF occurs between the W- and R-MDMF. This is indicated by the yellow arrows connecting them. This process is performed by System-2-Before-Event-Mode, utilizing the knowledge modeled in *GOMS*.

(c) Nonverbal Reaction

System 1-1: Green, blue, and red arrows indicate the propagation of activation for this process. Path-1: the listener perceives nonverbal information through the five senses, thereby activating the P-MDMF. Path-2: the P-MDMF overlaps with the W- and R-MDMF, activated by speech understanding and preparation. In sum, P-MDMF activation occurs through these two pathways. The listener exhibits nonverbal behavior via System 1 reflecting the propagation of activation from the P- to B- and M-MDMF.

What Participant B Does while Speaking

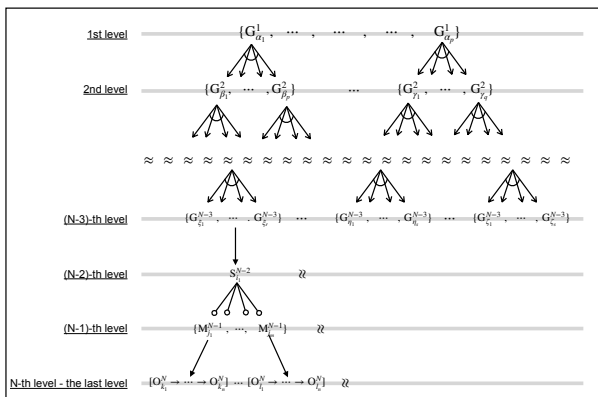
(d) Generation of Utterance

System 1-1: Yellow arrows indicate the propagation of activation while selecting *M* suitable for the current situation by applying *S* to the candidates in the B-MDMF, placed in a standby state during speech preparation while listening. Then along the red arrow activation is propagated to the M-MDMF to produce speech, executed by System-1-Before-Event-Mode. Upon completion, the utterance becomes conscious as an event; System-1-After-Event-Mode and System-2-After-Event-Mode are executed.

(e) Adjustment of the Contents of Utterance

System 2-1: Green and blue arrows indicate the propagation of activation for this process. Not only the speaker's *own* utterance but the listener's nonverbal responses activate P-MDMF; then the activation propagates to W-, R-, and B-MDMF. The activation pattern, triggered by one's own utterance and the listener's nonverbal responses, is processed in a timely manner via System 2 mediated by C-resonance for checking discrepancy between this pattern and the activation pattern of the goal structure having been activated during speech preparation. When discrepancies are detected, *G* and *M* are re-selected. The utterance is then continued via System-1-Before-Event-Mode (red arrow).

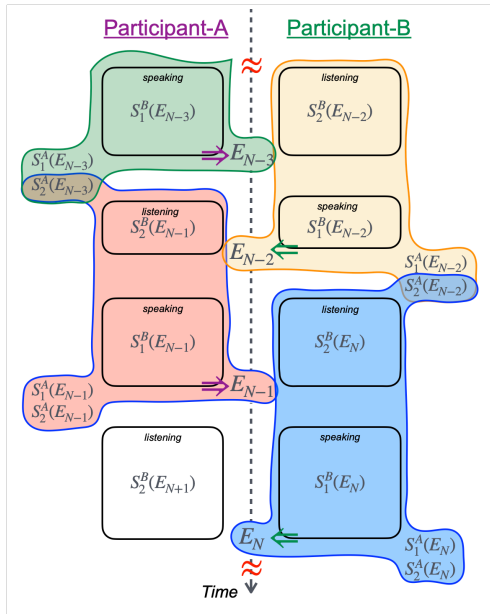
GOMS Connection Structure (GOMS-CS) [3]



- $O = [O_{R_1}^N \rightarrow \dots \rightarrow O_{R_n}^N]$: Operator sequence at the N -th layer
- $M = M_{j_1}^{N-1}$: Method at the $(N - 1)$ -th layer
- $G_i = G_{\xi_i}^{N-3}$: Goal at the $(N - 3)$ -th layer; Above the $(N - 3)$ -th layer, a hierarchical goal structure develops.
- $S_{i_1}^{N-2}$: Selection rule at the $(N - 2)$ -th layer if there are multiple methods, $\{M_{j_1}^{N-1}, \dots, M_{j_m}^{N-1}\}$, that can achieve G_i

The goals located at the top level $G_i^1 = G_{\alpha_i}^1$ are expanded into a set of goals, $G^2 = \{G_{\beta_1}^2, \dots, G_{\beta_p}^2\}$, at the second layer, and G_i^1 is achieved by the achievement of all goals contained in G^2 .

- Smooth conversation is executed as a routine goal-oriented task. Therefore, this study represents the knowledge utilized during speech using the *GOMS* model, which consists of goals, operators, methods, and selection rules.
- This figure illustrates the connection structure of G, O, M, and S. Layers 1 through $(N - 3)$ correspond to the goal structure, layer $(N - 2)$ to selection rules, layer $(N - 1)$ to methods, and layer N to operators [4].
- In the *GOMS* model, the goal structure G is stored in the W-MDMF, and the selection rules S , which determine the appropriate method to apply based on the situation, are stored in the R-MDMF. These are objects consciously manipulated by System 2.
- The methods M , which are pointers to operator sequences, are stored in the B-MDMF. These are linked to the R-MDMF. The operators O are stored in the M-MDMF, and the operator sequences defined by the methods are passed to the motor process, where actions are executed [3].



Four Processing Modes in Conversation

For the utterance termination event E_i , the four processing modes of MHP/RT are:

- $S_2^B(E_i)$: System-2-Before-Event-Mode
- $S_1^B(E_i)$: System-1-Before-Event-Mode
- $S_1^A(E_i)$: System-1-After-Event-Mode
- $S_2^A(E_i)$: System-2-After-Event-Mode

Four Processing Mode for E_N (B Finishes Utterance)

1. $S_1^B(E_{N-1}) \leftrightarrow S_2^B(E_N)$: While listening to A's utterance in $S_1^B(E_{N-1})$, execute one's own utterance preparation in $S_2^B(E_N)$.
 $\Rightarrow E_{N-1}$ occurs when A finishes utterance.
2. $S_1^B(E_N)$: Execute one's own utterance.
 $\Rightarrow E_N$ occurs when B finishes utterance.
3. $S_1^A(E_N)$: Adjust the connection state of MDMF nodes unconsciously.
4. $S_2^A(E_N)$: Reflect consciously on E_N . This will cause changes in the activation patterns in the MDMF.

MDMF Used for Utterance Preparation and Generation

- $\dots, S_2^A(E_{N-4}), S_2^A(E_{N-2}) \mapsto S_2^B(E_N)$:
 In utterance preparation, the MDMF is employed, reflecting the outcome of conscious reflection in $S_2^A(E_{N-2})$ on the event E_{N-2} of one's previous utterance. The MDMF of B used during utterance preparation in $S_2^B(E_N)$ reflects the experience E_{N-4}, E_{N-6}, \dots with A up to this point.
- $S_2^B(E_N) \mapsto S_1^B(E_N)$: The MDMF prepared so far is used to generate verbal behavior in $S_1^B(E_N)$.

Summary: Processes Shaping Smooth Conversation

Participant_A : $S_1^B(E_{N-1}) \leftrightarrow$

Participant_B : $S_2^A(E_{N-2}) \mapsto S_2^B(E_N) \mapsto S_1^B(E_N)$

Conversation In Smooth Mode

When an utterance is being executed in System-1-Before-Event-Mode, System 2 monitors the utterance content in real time to verify that the method is being executed correctly. If no issues are detected, processing by System 1 continues.

Transition to Intermittent Conversation Mode

When the utterance is judged to have been executed improperly, the method could be re-selected and, if necessary, the goal could be re-established. System 2 interrupts the utterance executed in System-1-Before-Event-Mode and, after the cognitive processing in System-1-After-Event-Mode and System-2-After-Event-Mode, the next utterance is executed starting with System-2-Before-Event-Mode. *In this situation, events occur during the speaker's turn without any speaker change.*

Modification of Prepared Actions

Conversation participant B begins utterance in $S_1^B(E_N)$ after the turn-taking event E_{N-1} . However, during the process leading up to the next turn-taking event E_N , an interruption by System 2 occurs. The reflection of the interruption event is performed in System-2-After-Event-Mode, and the prepared actions are modified to reflect the active state of MDMF at that time by referencing the goal structure deployed at the W- and R-MDMF levels.

What Happens in B during Utterance Modifications

The i -th interruption event occurring during the process leading up to E_N is denoted as $E_{N,i}$. By the time E_N occurs, the following have taken place:

Participant_B :

$$\begin{array}{rcccc}
 S_2^B(E_{N,1}) \rightarrow S_1^B(E_{N,1}) & \rightarrow & E_{N,1} & & \\
 & & & \rightarrow & S_1^A(E_{N,1}) \rightarrow S_2^A(E_{N,1}) \rightarrow \\
 & & & & \vdots \\
 S_2^B(E_{N,i}) \rightarrow S_1^B(E_{N,i}) & \rightarrow & E_{N,i} & & \\
 & & & \rightarrow & S_1^A(E_{N,i}) \rightarrow S_2^A(E_{N,i}) \rightarrow \\
 & \dots & \rightarrow & \dots & \rightarrow \\
 & \dots & \rightarrow & E_N &
 \end{array}$$

What Happens in A while Listening

When B is speaking in the intermittent conversation mode, changes in B's utterance goals interfere with listener A's ability to consistently execute utterance understanding of B's utterance content in $S_2^B(E_{N+1})$. This is because the goal structure activated by B in $S_2^B(E_N)$ —which was active at the start of the utterance—smoothly connects to the MDMF activated by A in $S_2^B(E_{N-1})$ while A is processing the utterance. However, as the utterance progresses, this structure gets updated, forcing A to follow accordingly.

Synchronization Among Conversation Participants

Both Participants' MDMFs Continuously Updated

- The utterance content of B in event E_N reflects the content of B 's MDMF activated in $S_2^B(E_N)$; Including an evaluation of the results of A 's utterance in event E_{N-1} and B 's own utterance in event E_{N-2} .
 - The content of A 's utterance reflects the content of A 's MDMF activated in $S_2^B(E_{N-1})$.
- ⇒ The MDMF possessed by each speaker shifts as the conversation progresses, with the activated domain changing based on the evaluation of both the content of the other's utterances and the results of their own utterances.

Condition for Sync in Verbal Behavior

- The condition for a smooth conversation to proceed, where speakers alternately process each other's utterances via their respective System-1-Before-Event-Mode without intervention from System 2, can be summarized as follows.

**The activation pattern of MDMF via $S_2^B(E_{i-1})$
 ≈ The activation pattern of MDMF via $S_2^B(E_i)$**

- This indicates that the activation pattern of one's own MDMF that triggers event E_i related to one's own utterance significantly overlaps with the activation pattern of the other's MDMF that triggers event E_{i-1} related to the other's utterance.

Synchronization in Human-VR Conversation

Sync Felt in Immersiveness

- The synchronization of conversations involving turn-taking between humans differs in nature from the synchronization that occurs when users interact with VR systems.
- Dinet et al. [5] identified weak synchronization between the user and the system as the condition for users to interact with multimodal systems with a sense of immersion.

Weak Synchronization for Immersiveness

- Weak synchronization is achieved by designing the specific VR content of the interaction at T_N based on cognitive processing in $S_2^B(E_N)$, $S_1^B(E_N)$, $S_1^A(E_N)$, $S_2^A(E_N)$ regarding the interaction event E_N where the user utilizes the system.
- To do so, it is necessary to appropriately estimate the propagation of activation within the MDMF and the updates to the MDMF during the period $[T_N - \beta, T_N + \alpha]$.

Balance Between Goals and Methods

The shape of the GOMS-CS is characterized by the following numbers:

- \hat{G} : Total number of goals is denoted,
- \hat{M} : Total number of methods,
- \hat{S} : Total number of selection rules,
- \hat{O} : Total number of operators,
- \hat{N} : Average depth of the hierarchy, and
- \hat{C} : Upper bound on the number of nodes.

Due to the finite processing capacity of the brain, the GOMS-CS becomes either System-2-After-Event-Mode or System-1-After-Event-Mode dominant [3].

Four Cases of the GOMS-CS Shapes

- Case 1: $\hat{M} \gg \hat{G}$;
System-1-After-Event-Mode is dominant.
- Case 2: $\hat{G} \gg \hat{M}$;
System-2-After-Event-Mode is dominant.
- Case 3: $\hat{M} \gg \hat{G} \sim 0$;
Actions occur almost exclusively under System-1-After-Event-Mode.
- Case 4: $\hat{G} \gg \hat{M} \sim 0$;
Actions occur almost exclusively under System-2-After-Event-Mode.

Understanding Conversational Behavior

Behavioral Linguistics Perspective

- **Case 1: Real World Conversation ($\hat{M} \gg \hat{G}$):** Corresponds to a smooth conversation mode. The prerequisite for establishing this mode is that the participants share a set of conversational methods. Actions driven by System 1 become predominant.
- **Case 2: Real World Conversation ($\hat{G} \gg \hat{M}$):** Each speaker's turn is short, requiring frequent speaker changes to maintain mutual understanding and continue the conversation. When a group consists of both direct communities and indirect societies and/or communication occurs through structural language, System-2-After-Event-Mode becomes the dominant behavioral ecology.

Other Perspectives (Not Realistic)

- **Case 3: Skinner's Perspective:** Involves situations where the goal is extremely limited. This reduces to stimulus-response behavior that can be executed without cognitive processing. This aligns with the understanding of verbal behavior based on Skinner's behavioral psychology.
- **Case 4: Chomsky's Perspective:** Represents a situation where the methods become extremely limited. The goals determine the details of the utterance. This aligns with the Chomskyan understanding of linguistic behavior, which posits that goals are symbols and that linguistic actions arise through the manipulation of these symbols.

CONCLUSION — WHAT THIS PAPER HAS DONE

Proposed a behavioral linguistics theory to explain real-time language generation in everyday conversation.

- This approach was based on the MHP/RT [2, 1] and the dual-process theory of cognition [6, 7, 8].
- The analysis modeled fluent conversation as a PCM cycle alternating between the speaker and the listener, describing in detail how the MDMF and GOMS are used for understanding, preparing, and producing utterances.

Differentiated between fluid and intermittent conversation modes.

- This approach offers a richer framework than models based solely on Skinner [9] or Chomsky [10] by integrating behavioral psychology and behavioral economics.
- Our paper argued that real-time linguistic behavior cannot be fully captured by static grammatical models or by purely associative accounts of verbal behavior. Instead, it requires a dynamic framework that explains how linguistic choices emerge from moment-to-moment cognitive constraints, environmental cues, and interactive demands.
- By positioning language production within the PCM cycle, the theory emphasized that utterances are not pre-constructed entities retrieved whole from memory but are assembled progressively through rapid iterations of perception, interpretation, planning, and articulation.

Highlighted the role of prediction and anticipation in conversation.

- Speakers and listeners continuously forecast each other's intentions, adapt to turn-taking cues, and adjust their linguistic formulations based on cognitive load and situational incentives. This predictive loop is shaped not only by linguistic competence but also by heuristics, biases, and cost-benefit evaluations, central themes in behavioral economics.
- As a result, language use is portrayed as an activity controlled by bounded rationality, optimized under real-time processing constraints rather than idealized grammatical rules.

Suggested that a behavioral-linguistic theory grounded in cognitive architecture offers a more comprehensive explanation of conversational competence.

- The theory bridged gaps between psycholinguistics, cognitive psychology, and behavioral science, providing a unified account of how humans understand and produce language in everyday interaction.
- When we understand that speech is built moment-by-moment using limited attention and working memory, we become more aware of why misunderstandings happen. This can help people speak more clearly, listen more actively, and manage turn-taking more smoothly in conversations.

From an applied perspective, several implications can be drawn for our daily lives. Research to realize these implications must be conducted in the future:

1) Improved Learning and Teaching

A theory that explains how language is processed in real time can improve language teaching. Teachers can design exercises that match how the brain naturally organizes and retrieves language, making learning more intuitive and efficient.

2) Reduced Communication Stress

Knowing that hesitations, pauses, or “ums” are natural results of cognitive processing—not signs of incompetence—can help people feel less anxious when speaking. This is especially helpful for public speaking, second-language use, or social anxiety.

3) Better Human–Computer Interaction (HCI)

If we understand how humans generate language under time pressure, we can design voice assistants, chatbots, and AI systems that interact more naturally. The theory can guide systems to adapt to human pacing, prediction patterns, and conversational rhythms.

4) More Effective Teamwork and Decision-Making

In workplaces, communication failures are often cognitive failures. Understanding how System 1 and System 2 influence what we say can help people catch biases, avoid rushed judgments, and communicate more thoughtfully in meetings or negotiations.

5) Insights for Therapy and Mental Health

Speech disruptions often reflect cognitive overload, stress, or emotional pressure. A behavioral model of real-time language can help psychologists better understand how anxiety, ADHD, or trauma affect communication—and help people manage these effects.

6) Conflict Prevention & Smoother Social Interactions

Recognizing that people often speak using quick, automatic processing (System 1) can make us more tolerant of minor errors or emotional reactions in others. It encourages patience and gives a more compassionate understanding of how real conversations work.

- [1] Muneo Kitajima. ***Memory and Action Selection in Human-Machine Interaction***. Wiley-ISTE, 2016.
- [2] Muneo Kitajima and Makoto Toyota. **Decision-making and action selection in Two Minds: An analysis based on Model Human Processor with Realtime Constraints (MHP/RT)**. *Biologically Inspired Cognitive Architectures*, 5:82–93, 2013.
- [3] Muneo Kitajima, Makoto Toyota, Jérôme Dinet, and Katsuko T. Nakahira. **Transforming Conscious Goals into Unconscious Actions in Real-world Interactions: Real-world Use of Behavioral Ecological Memes via GOMS**. *International Journal On Advances in Intelligent Systems*, 18(3 & 4):173–186, 2025.
- [4] Stuart K. Card, Thomas P. Moran, and Allen Newell. ***The Psychology of Human-Computer Interaction***. Lawrence Erlbaum Associates, Hillsdale, NJ, 1983.
- [5] Jérôme Dinet and Muneo Kitajima. **Immersive interfaces for engagement and learning: Cognitive implications**. In *Proceedings of the 2015 Virtual Reality International Conference, VRIC '18*, pages 18/04:1–18/04:8, New York, NY, USA, 2018. ACM.
- [6] Daniel Kahneman. **A perspective on judgment and choice**. *American Psychologist*, 58(9):697–720, 2003.
- [7] Daniel Kahneman. ***Thinking, Fast and Slow***. Farrar, Straus and Giroux, New York, NY, 2011.
- [8] Jonathan St. B. T. Evans. **Dual-processing accounts of reasoning, judgment, and social cognition**. *Annual Review of Psychology*, 59(Volume 59, 2008):255–278, 2008.
- [9] B. F. Skinner. ***Verbal Behavior***. Appleton-Century-Crofts., 1957.
- [10] Noam Chomsky. ***Language*, 35(1):26–58, 1959.**