A Cognitive Model of the Factors Controlling the Characteristics of Shiritori Word Sequences

Jumpei Nishikawa (nishikawa.jumpei.16@shizuoka.ac.jp)
Kosuke Sasaki (sasaki.kosuke.19@shizuoka.ac.jp)
Junya Morita (j-morita@inf.shizuoka.ac.jp)
Department of Informatics, Graduate School of Integrated Science and Technology, Shizuoka University,
3-5-1, Johoku, Naka-ku, Hamamatsu, Shizuoka, 432-8011, Japan

Abstract

Recent advances in large language models (LLMs) have renewed interest in understanding the cognitive mechanisms underlying human language use. In this study, we focus on the Japanese word game "Shiritori" as a simple task related to language use, and aim to clarify the cognitive factors involved in its characteristics. To this aim, we model the execution of Shiritori based on the basic memory mechanisms of a cognitive architecture and investigate the characteristics that appear in Shiritori word sequences through simulations with different parameter settings. The results show that word sequences with different properties were obtained depending on the levels of lexical activation, inhibition, and semantic association. To complement the simulation findings, we also conducted a preliminary human evaluation in which participants rated the robot-generated word sequences. By constructing a model that can flexibly control Shiritori behavior, we explore its potential applications as a stimulus for research on human–robot interaction and language acquisition support.

Keywords: Cognitive architecture; Cognitive modeling; ACT–R; Shiritori

Introduction

Recent advances in large language models (LLMs) have renewed interest in understanding the cognitive mechanisms underlying human language use. While LLMs achieve impressive performance by learning from large-scale data, cognitive science takes a complementary approach: identifying a small set of general principles that govern human behavior. One such approach is cognitive modeling, which seeks to explain and predict cognitive processes by implementing them as computational programs.

Cognitive architectures provide a theoretical and computational framework for building cognitive models. Developed based on psychological and cognitive theories, these architectures offer structures and parameters that allow researchers to isolate and analyze the cognitive factors involved in task performance. ACT–R (Adaptive Control of Thought–Rational, Anderson, Boyle, & Reiser, 1985) is one such architecture, where memory processes are controlled by activation values. These values are calculated based on mechanisms such as base-level activation (reflecting learning and forgetting) and spreading activation (reflecting contextual or associative influences).

In this study, we focus on Japanese word game "Shiritori" as a simple task related to human conversation. Shiritori involves players taking turns to utter a noun that starts with the final sound of the previous word. The task is intuitive and widely understood, yet requires coordinated use of phonological awareness, lexical retrieval, and associative processes—making it well suited for cognitive modeling.

We aim to clarify the cognitive factors underlying the characteristics of word sequences that emerge during Shiritori by constructing a computational model using ACT–R. Specifically, we simulate Shiritori sessions with models that vary in their memory retrieval parameters, and examine how these settings affect the resulting word sequences. By doing so, we explore how word familiarity and semantic association influence behavior in this structured linguistic task. In addition, we conducted a preliminary human evaluation to assess how the generated sequences are perceived by human observers, providing an initial validation of the model's behavioral plausibility.

Related Studies

Shiritori and Language Learning

Shiritori is a Japanese word game in which participants take turns producing nouns that begin with the final mora[1] of the preceding word. For instance, after a player answers "リンゴ" (meaning apple), the next player continues with "ゴマ" (meaning sesame seeds) [2]. A player loses if they use a word that has already been used in the game or a word that ends with a specific character. While this game is typically played by multiple participants, some vocabulary learning materials designed for young children incorporate a version of Shiritori that can be played alone.

Shiritori is frequently used in speech-language pathology, including in the treatment of aphasia and intervention for autism. Several studies have demonstrated the use of Shiritori in assessments of children with autism during therapeutic interventions (Oishi, 1994). The conditions necessary for playing Shiritori have been identified through cross-sectional studies involving typically developing young chil-

[1] A type of phonological unit.Japanese is classified as a mora language, where each mora generally has a consistent duration, playing a crucial role in the rhythm and timing of speech (Port, Dalby, & O'Dell, 1987).

[2] In Japanese, a single character generally corresponds to a single mora. For example, the word リンゴ (apple) consists of three characters, and Japanese speakers perceive it as being composed of three distinct sounds (morae): リ (ri) – ン (n) – ゴ (go). The clear correspondence between characters and sounds in Japanese provides an advantage for linguistic and cognitive modeling.

dren (Takahashi, 1997). It has been suggested that phonological awareness, specifically the ability to segment sounds into phonemes, is essential, as well as the presence of a phonologically indexed mental lexicon. Furthermore, the acquisition of kana characters has been shown to be effective in facilitating phonological indexing of vocabulary.

Cognitive Modeling

A cognitive architecture serves as a foundation for modeling cognitive processes that occur in individual tasks. Models utilizing cognitive architectures enable the construction of models that distinguish various factors required for task completion. Among the various cognitive architectures that have been developed, this study focuses on a model based on ACT-R (Anderson, 2007).

ACT-R provides a modular framework for simulating human memory, attention, and decision making, integrating both symbolic and subsymbolic components. This hybrid architecture enables models that capture the multifaceted nature of language use. It is particularly suitable for modeling Shiritori, which involves both rule-based phonological constraints and graded processes such as semantic association and word familiarity.

Numerous studies have been conducted on language use utilizing ACT-R. Models have been developed to account for the acquisition of irregular verbs in English learning (Taatgen & Anderson, 2002) and noun learning in young children (Van Rij, Van Rijn, & Hendriks, 2010). Research has also examined brain function disorders, including studies that explain errors in sentence comprehension in aphasia using ACT-R parameters (Mätzig, Vasishth, Engelmann, Caplan, & Burchert, 2018).

A model that plays Shiritori using ACT-R has also been developed (Nishikawa & Morita, 2022). This model simulates the relationship between memory activation values and phonological awareness. The results demonstrated that an increase in phonological awareness leads to a higher number of continuous Shiritori turns and that engaging in Shiritori enhances phonological awareness. These findings are consistent with the aforementioned study by Takahashi (1997).

Limitations of Previous Studies and Purpose of this Study

As demonstrated in the preceding discussion, Shiritori has been utilized as a tool for investigating language development, and models and systems incorporating Shiritori as a task have been developed within research on human language use. However, previous studies have not clarified the cognitive factors underlying the diversity of behaviors observed during Shiritori. While Takahashi (1997) experimentally demonstrated the effects of providing hints, it focused only on a restricted form of Shiritori, where participants responded to a prompt with a single word. Furthermore, the prior modeling of Shiritori (Nishikawa & Morita, 2022) primarily examined errors related to phonological processing, without conducting a detailed analysis of the characteristics of the words appearing in Shiritori word sequences, such as sequence length and semantic properties.

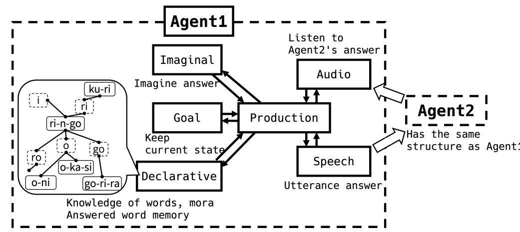

Figure 1: Overview of the Shiritori model

Given these limitations, this study aims to model the sequential process of playing Shiritori in an interactive setting and to identify the cognitive factors that characterize the resulting word sequences. To complement the simulation-based findings, we also performed a preliminary human experiment in which participants evaluated robot-generated Shiritori sequences. This enabled us to examine how variations in modeled retrieval parameters may influence human impressions.

Model

Model Structure

In a previous study (Nishikawa & Morita, 2022), a Shiritori model utilizing the cognitive architecture ACT-R (Anderson, 2007) has already been implemented. This model allows for the adjustment of its behavior by configuring ACT-R parameters. Since this study focuses on "Shiritori" as a simple task related to language use, we specifically examine parameters related to the familiarity of words appearing during Shiritori and the associative retrieval of words corresponding to the conversational context. Here, the previous model represents errors in Shiritori associated with immature phonological awareness by linking ACT-R's general memory retrieval mechanisms—particularly its partial match retrieval based on memory similarity—to phonological awareness. However, since this study does not focus on phonological errors, we have disabled the corresponding partial match retrieval parameter settings.

An overview of the model constructed in this study is shown in Figure 1. This model includes agents (represented by the dashed-line areas), who take turns responding with words to continue the Shiritori sequence. The boxes within each agent represent the respective modules of ACT-R. In the following sections, we describe the declarative memory settings and retrieval parameters within the ACT-R modular structure that are considered to characterize the Shiritori word sequences.

Declarative memory in the Model The declarative module of ACT-R is used to model the memory required for performing Shiritori. In ACT-R's declarative module, memory consists of structural elements called chunks. This model includes three types of chunks: "word chunks", which link a word's phonological representation and written form; "mora

Table 1: Declarative memory of the model

(a) Word chunk

word	sound
リンゴ	“ɾʲiŋ o”
ゴリラ	“goɾʲiɾa”
クリ	“kɯᵝɾʲi”
…	…

(b) Mora chunk

mora	sound
リ	“ɾʲi”
ゴ	“go”
ク	“kɯᵝ”
…	…

(c) Word–mora associations

word	mora	position
リンゴ	リ	head
リンゴ	ゴ	tail
ゴリラ	ゴ	head
…	…	…

chunks”, which represent the phonological representation and notation of individual morae; and “word-mora association chunks”, which indicate the morae that constitute a word.

Table 1 presents examples of these chunks. Words and morae are represented using Japanese characters (katakana), while their phonological representations are described using the International Phonetic Alphabet (IPA), enclosed in double quotation marks. In declarative memory, these three types of chunks are connected as a network. The speech bubbles extending from the declarative module in Figure 1 illustrate this structure. The agent retrieves words by navigating this network, searching for candidate words that connect to the previous word through shared morae.

Model Parameters The modules in ACT-R are regulated by numerical parameters[3]. In this study, we focus on parameters related to the retrieval of declarative memory, specifically activation values (Equation 1). Activation values are defined as the sum of multiple components corresponding to learning and forgetting (known as base-level activation), contextual influences (known as spreading activation), and noise added at each retrieval instance . In this study, we emphasize base-level activation to represent the familiarity of words used in Shiritori sequences and spreading activation to capture associative retrieval during the interaction.

$$A_i = B_i + S_i + \varepsilon_i \tag{1}$$

The first parameter is the effect of learning and forgetting (base-level activity value), which is used to represent word familiarity. In Equation 2, B_i represents the base-level activation of chunk i. This value is computed based on the number of times the chunk has been retrieved n, the elapsed time L since the chunk was first referenced, the decay rate d, and an offset value β_i [4]. The parameters d and β_i are each set by a single numerical value and exert a uniform influence across all chunks. The familiarity of individual chunks (words) can be adjusted by modifying n (the number of times a chunk has been retrieved) and L (the elapsed time since the chunk was first referenced).

$$B_i = \ln\left(\frac{n}{1-d}\right) - d \times \ln(L) + \beta_i \tag{2}$$

[3]For details, refer to the ACT-R manual (Bothell, 2020).

[4]In the specific implementation within ACT-R, the decay rate d is set using the `:bll` parameter, while the offset value β_i is configured using the `:blc` parameter.

In Equation 3, S_i represents spreading activation, which reflects the influence from chunks maintained in other modules and corresponds to contextual effects. This study focuses on this parameter to represent word association. S_i is the spreading activation value assigned to chunk i during a retrieval process. k denotes the module from which activation spreads, as determined by parameter settings[5]. j refers to an activation source, which is a value shared between the chunk held by module k and the retrieval cue. If chunk i contains the activation source j as one of its elements, the activation amount set for module k is divided by the number of activation sources j, yielding W_{kj}. The strength of the association between activation source j and chunk i, denoted as S_{ji}, is defined by Equation 4. Here, S represents the maximum association strength. The second term corresponds to the Fan effect(Anderson, 1974)[6]. It is calculated based on the total number of occurrences of value j across all of declarative memory ($slots_j$) and the number of slots in chunk i that contain j ($slotsof_{ji}$).

$$S_i = \sum_k \sum_j W_{kj} S_{ji} \tag{3}$$

$$S_{ji} = S - \ln\left(\frac{1 + slots_j}{slotsof_{ji}}\right) \tag{4}$$

The activation value, computed as the sum of these components, influences the retrieval time of a chunk, as described in Equation 5 or Equation 6. Equation 5 represents the retrieval time for chunk i with activation value A_i. Here, F is the latency exponent value, and f is the latency factor value, both of which regulate retrieval latency. The default values were used in this study. If no chunk is found in response to a retrieval request, or if the most highly activated chunk falls below the retrieval threshold, Equation 6 is applied. In this equation, τ represents the retrieval threshold , which defaults to 0. In this study, we set it to -5 to balance the activation value calculations. Given these considerations, observing changes in activation values throughout the simulation allows us to infer variations in the accessibility of knowledge within ACT-R.

$$RT = Fe^{-(f * A_i)} \tag{5}$$

$$RT = Fe^{-(f * \tau)} \tag{6}$$

Simulation

Simulation Settings A simulation was conducted in which two agents with the aforementioned settings played Shiritori. For the vocabulary used by the models, we utilized 2,436

[5]For example, activation spreading from the goal module can be controlled using the `:ga` parameter, while activation from the imaginal module can be specified using the `:imaginal-activation` parameter.

[6]The Fan effect explains that the more pieces of information associated with a given concept, the longer it takes to retrieve (or recall) that information. This is attributed to the activation value being distributed across multiple connections.

words from the associative vocabulary table of young children and elementary school students (National Institute for Japanese Language and Linguistics, 1981). This table was constructed based on a task where participants were asked to list as many words as possible within 40 minutes in response to a stimulus word (e.g., "animal"). Participants were categorized into the following nine age groups: 3- to 6-year-olds, first- to fourth-grade elementary school students, and adults.

For each chunk corresponding to a word in this vocabulary, base-level activation were set to represent word familiarity. Using the associative vocabulary table (which includes information such as "46 three-year-olds responded with 'elephant' to the stimulus word 'animal'"), the values for L (elapsed time) and n (retrieval count) in Equation 2 were calculated. In this model, L was set assuming an adult (e.g., for words responded to by three-year-olds, L was set to $21-3$ years $= 67,648,000$ seconds). Furthermore, the retrieval count n for word w (denoted as n_w) was computed using Equation 7. Here, i represents participant age, and N_{wi} denotes the number of participants aged i who responded with word w. To simulate the frequency of retrieval, this time we made the simple assumption that each individual refers the word once per day and multiplying by 365.

$$n_w = \sum_i (N_{wi} \times 365) \tag{7}$$

Additionally, spreading activation was introduced to represent word associations during Shiritori. Each vocabulary item in the model was linked to a category based on the associative vocabulary table of young children (National Institute for Japanese Language and Linguistics, 1981), where stimulus words were categorized. To implement contextual effects and word associations during Shiritori, the category of the previous word was placed in the goal module, with the parameter `:ga = 1`, allowing the model to incorporate semantic associations and contextual influences during word retrieval. Furthermore, to examine the impact of association strength, the maximum associative strength S in Equation 4 was manipulated at three levels: 0 (spreading activation off), 35 (adjusted so that the effects of B_i in Equation 1 and S_i are approximately equal), and 100 (an extremely high value).

Using the three models, we conducted ten Shiritori runs for each, with an upper limit of three minutes per session. The Shiritori sessions ended either due to retrieval failure (failure to recall a word) or rule violations, such as responding with a word ending in "ン" or repeating a previously used word.

Results Figure 2 shows a histogram of the base-level activation values of the vocabulary in the model, calculated using the method described in the previous section. The histogram reveals that there is a large number of vocabulary items with low base-level activation values, and the frequency gradually decreases as the activation values increase. This indicates that the associative task successfully captured a relationship between a small set of well-known words and a larger set of somewhat less familiar words. Additionally, the fact that the smallest bin has an exceptionally high frequency suggests that many words were contributed by only a few young participants. For example, a word like "Yattāman"[7] was given by a three-year-old in response to the stimulus word "animal."

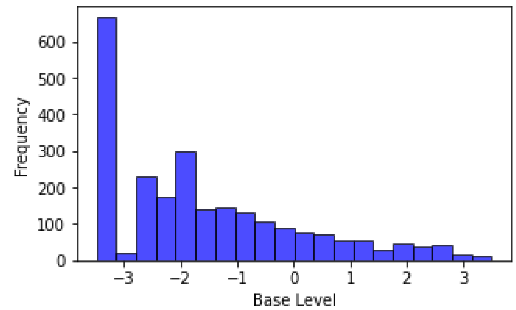

Figure 2: Histgram showing the base-level activation value distribution for word–mora association chunks

Table 2: Termination outcomes for the Shiritori simulations

	$S=0$	$S=35$	$S=100$
Time-limit	3	0	0
Retrieval-failure	7	6	4
Response by previously used word	0	2	1
Response by a word ends with "ン"	0	2	5

Table 2, Figures 3 and 4 summarize the characteristics of the Shiritori sessions conducted by the three models. First, regarding Shiritori length (Figure 3), no significant differences were observed among the three models. However, in both the number of category transitions during Shiritori ($Mean_{S=0} = 10.1$, $SD_{S=0} = 5.8$, $Mean_{S=35} = 2.7$, $SD_{S=35} = 1.8$, $Mean_{S=100} = 3.2$, $SD_{S=100} = 1.9$) and the ratio of cat-

[7]The name of a hero from a Japanese anime series.

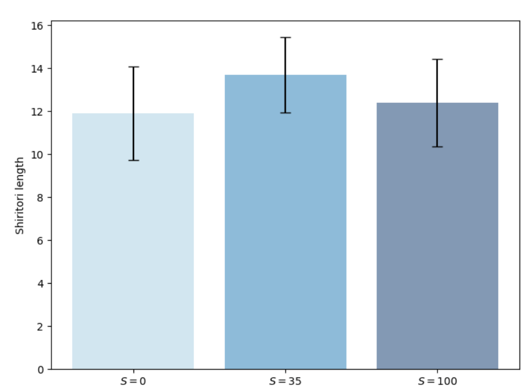

Figure 3: Number of uttered words before termination across sessions

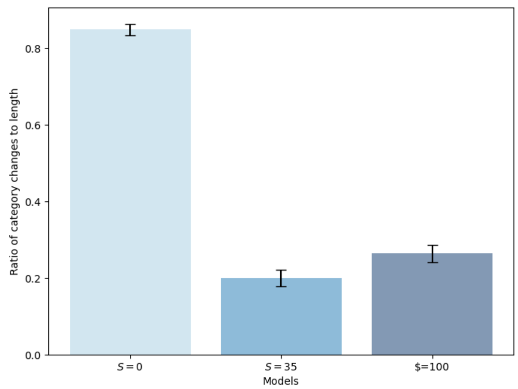

Figure 4: Proportion of category changes relative to sequence length

egory transitions to Shiritori sequence length (Figure 4), the models with $S = 35$ and $S = 100$ exhibited fewer transitions. This aligns with the model's intended behavior, indicating that word associations influenced retrieval, resulting in Shiritori being played within the same category. Table 3 presents a notable example from the $S = 0$ model, showing the Shiritori sequence with the highest number of category transitions. The word column presents Japanese words that appeared in the Shiritori word sequences, written in katakana, with their corresponding English translations in parentheses. The category column indicates the category of each word, extracted from the associative vocabulary table data(National Institute for Japanese Language and Linguistics, 1981). Since category information was not considered, a variety of high base-level activation words from different categories were retrieved. Table 4 shows an example of the longest-lasting Shiritori sequence with the fewest category transitions, generated by the $S = 35$ model. This sequence progressed entirely within the "animals" category, featuring relatively uncommon animal names (e.g., skunk, pit viper).

Furthermore, focusing on the termination conditions of Shiritori (Table 2), retrieval failure was the most frequent cause of termination across all conditions. Additionally, in the $S = 100$ model, termination due to responding with a word ending in "ン" occurred the most frequently. The model in this study was based on the settings of previous research (Nishikawa & Morita, 2022), where a re-retrieval process was included if a retrieved word violated Shiritori rules. However, in the $S = 100$ model, the relative influence of spreading activation on the overall activation value (Equation 1) became excessively large. As a result, even after the re-retrieval process, the model exhibited a tendency to fixate on specific words, leading to rule violations. This phenomenon can also be confirmed through the example sequence shown in Table 5. In this sequence, the early progression follows a pattern similar to the lowest-category-transition sequence (Table 4). However, after retrieving "コアラ", the model weighted same-category animal words more heavily when searching for a

Table 3: Shiritori sequence showing maximum category transitions (18) at $S = 0$.

	word	category
1	リンゴ (apple)	fruit
2	ゴリラ (gorilla)	animal
3	ラクダ (camel)	animal
4	ダブルデッカー (Double decker)	vehicle
5	アサガオ (morning glory)	flower
6	オートバイ (motorcycle)	vehicle
7	イヌ (dog)	animal
8	ヌイグルミ (cuddly toy)	tool
9	ミミズ (earthworm)	insect
10	ズック (Indoor shoes)	footwear
11	クツシタ (socks)	footwear
12	タイコ (drum)	instrument
13	コオロギ (cricket)	insect
14	ギター (guiter)	instrument
15	アリ (ant)	insect
16	リス (squirrel)	animal
17	スズメ (sparrow)	bird
18	メダカ (Japanese killifish)	plant
19	カキ (persimmon)	plant
20	キク (chrysanthemum)	flower
21	クツ (shoes)	footwear
	Stopped by time-limit	

word starting with "ラ", ultimately retrieving "ライオン", which violates the Shiritori rule, resulting in termination. Additionally, in three out of ten runs of the $S = 100$ model, this exact sequence was generated, further supporting this interpretation.

Preliminary Human Evaluation

The model described thus far was implemented in a communication robot, and an experiment was conducted to evaluate human impressions of Shiritori word sequences. While the experiment collected data on various factors, including robot movements, this paper focuses specifically on the evaluation of Shiritori word sequences. Participants watched four videos of robots playing Shiritori and evaluated their impressions. The videos were designed based on two levels of Shiritori settings (with ($S = 35$) vs. without spreading activation ($S = 0$)) and two levels of robot motion settings.

Participants rated the videos on a 7-point Likert scale. For naturalness, they responded to the prompt: "In the video you just watched, how natural did the Shiritori word sequence and robot motion appear?" Four naturalness-related items were included: (a) Words spoken by a single robot, (b) Posture or movement of a single robot, (c) Word-to-word connections between two robots, (d) Posture-to-posture connections between two robots. In addition, seven impression-related items were included: (a) Complexity, (b) Regularity, (c) Predictability, (d) Intentionality, (e) Intelligence (f) Mindedness, (g) Unity, (h) Activeness. These items were

Table 4: Longest Shiritori run with minimal category changes (1) at $S = 35$.

	word	category
1	リンゴ (apple)	fruit
2	ゴリラ (gorilla)	animal
4	ラクダ (camel)	animal
4	ダックスフンド (dachshund)	animal
5	ドラネコ (stray cat)	animal
6	コウモリ (bat)	animal
7	リス (squirrel)	animal
8	スカンク (skunk)	animal
9	クマ (bear)	animal
10	マンモス (mammoth)	animal
11	スイギュウ (Water buffalo)	animal
12	ウマ (horse)	animal
13	マムシ (pit viper)	animal
14	シシ (lion)	animal
	Stopped by retrieval-fail	

Table 5: Example of Shiritori concluding with a word ending in “ン” (ライオン) under $S = 100$.

0	リンゴ (apple)	fruit
1	ゴリラ (gorilla)	animal
2	ラクダ (camel)	animal
3	ダックスフンド (dachshund)	animal
4	ドラネコ (stary cat)	animal
5	コアラ (koala)	animal
	Stopped by word with “ン”:ライオン (lion)	

adapted from prior studies on perceived agency (Terada & Yamada, 2017) and impression formation in multi-robot systems (Ichijo, Munekata, & Ono, 2015; Mizumaru, Sakamoto, & Ono, 2017). A total of 172 participants (including 72 females, 4 participants who did not disclose their gender, $Mean_{\text{age}} = 45.34, Std_{\text{age}} = 9.68$) were recruited through crowdsourcing. No personal or identifying data were collected in this experiment.

Figure 5 presents notable results from the experiment. To analyze participant evaluations of robot intelligence and the naturalness of words spoken by a single robot, data from the two robot motion conditions were combined ($N = 344$), and a one-factor repeated-measures ANOVA was conducted with Shiritori settings ($S = 0$ vs. $S = 35$) as the independent variable. For perceived intelligence, the $S = 35$ condition received significantly higher ratings ($F(1,343) = 18.41, p < .01$). This suggests that spreading activation (associative retrieval) led to the appearance of relatively unfamiliar words, which may have contributed to the impression of intelligence. In contrast, for word naturalness, the $S = 0$ condition received significantly higher ratings ($F(1,343) = 16.43, p < .01$). This indicates that words influenced by spreading activation may have been perceived as inappropriate in relation to the robot's appearance and the other familiar words, leading to a lower sense of naturalness.

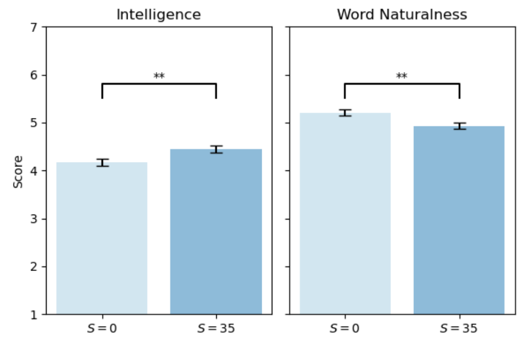

Figure 5: Human assessment of the Shiritori word sequence. Left: perceived naturalness of spoken words. Right: perceived intelligence of the robot.

Summary and Discussion

This study aimed to enhance our understanding of human language use by focusing on Japanese word game “Shiritori” as a simple task related to language use, and aim to clarify the cognitive factors involved in its characteristics. Through simulations using a cognitive model of Shiritori built on a cognitive architecture, we demonstrated that different Shiritori word sequences emerged depending on two key factors: base-level activation settings (word activation and suppression) derived from an associative vocabulary table and spreading activation settings (semantic associations) utilizing word category information. The results of the human evaluation revealed that while Shiritori word sequences incorporating associative effects received lower ratings for naturalness, they contributed to a more intelligent impression of the robot.

Future work requires further validation of the constructed model's behavior. This can be achieved through participant experiments evaluating the generated Shiritori sequences and comparisons with studies on the characteristics of Shiritori in human language acquisition (Tanaka, Ushiyama, Shimizu, & Goma, 2024). Developing a controllable model that regulates the characteristics of Shiritori by adopting a human cognition-based approach has the potential to contribute to HAI research and language acquisition support studies, serving as a source of insight and application in these fields.

References

Anderson, J. R. (1974). Retrieval of propositional information from long-term memory. *Cognitive Psychology*, *6*(4), 451-474. doi: https://doi.org/10.1016/0010-0285(74)90021-8

Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* Oxford University Press.

Anderson, J. R., Boyle, C. F., & Reiser, B. J. (1985). Intelligent tutoring systems. *Science*, *228*(4698), 456–462.

Bothell, D. (2020). ACT-R 7.21+ reference manual [Computer software manual].

Ichijo, T., Munekata, N., & Ono, T. (2015). Study for a joint attention system using communication activity of multiple robots. In *Hai sympsium 2015* (p. G-17).

Mätzig, P., Vasishth, S., Engelmann, F., Caplan, D., & Burchert, F. (2018). A computational investigation of sources of variability in sentence comprehension difficulty in aphasia. *Topics in cognitive science*, *10*(1), 161–174.

Mizumaru, K., Sakamoto, D., & Ono, T. (2017). Perception of the space emerged by overlap of utterance of multiple robots. In *Hai sympsium 2017* (p. G-15).

National Institute for Japanese Language and Linguistics. (1981). *Tables of vocabulary obtained from japanese children by association method*. TOKYO SHOSEKI CO.,LTD.

Nishikawa, J., & Morita, J. (2022). Cognitive model of phonological awareness focusing on errors and formation process through shiritori. *Advanced Robotics*, *36*(5-6), 318–331.

Oishi, N. (1994). Language problems observed in learning disabilities. *The Japanese journal of communication disorders*, *11*(2), 57–63.

Port, R. F., Dalby, J., & O'Dell, M. (1987). Evidence for mora timing in japanese. *The Journal of the Acoustical Society of America*, *81*(5), 1574–1585.

Taatgen, N. A., & Anderson, J. R. (2002). Why do children learn to say “broke” ? a model of learning the past tense without feedback. *Cognition*, *86*(2), 123–155.

Takahashi, N. (1997). Yoji no kotobaasobi no hattatsu : “shiritori” o kanon isuru joken no bunseki. *The Japanese Journal of Developmental Psychology*, *8*(1), 42–52.

Tanaka, S., Ushiyama, M., Shimizu, S., & Goma, H. (2024). Young childeren's language development resulting from playing “shiritori”. *Japanese Journal of Learning Disabilities*, *33*(2), 187-195. doi: 10.32198/jald.33.2_187

Terada, K., & Yamada, S. (2017). Mind-reading and behavior-reading against agents with and without anthropomorphic features in a competitive situation. *Frontiers in Psychology*, *8*, 1071.

Van Rij, J., Van Rijn, H., & Hendriks, P. (2010). Cognitive architectures and language acquisition: A case study in pronoun comprehension. *Journal of Child Language*, *37*(3), 731–766.