Interpreting Spatial Scenes for Choice of Demonstratives:

A Psycholinguistic Contrastive Study Between Japanese and Chinese

SUGAYA, Yusuke

Abstract: Our visual-world eye-tracking study examined the semantic/pragmatic aspects of cognitive processing behind the utterance of demonstratives in Japanese and Mandarin Chinese. After theorizing on the conceptual structure and process of indexicals in terms of distance judgment, a psycholinguistic experiment scrutinized relevant hypotheses using the recorded responses and eye movements over the trials containing visual stimuli instantiating different schemas. Consequently, this research verified the following: (i) The line representing equal distance from the speaker and the hearer operates considerably in Japanese; however, only the distance from the speaker functions significantly in Chinese. (ii) The use of demonstratives elicits the processing of suggested components, such as competitors and standards, in both languages. (iii) Japanese speakers are more inclined than Chinese speakers to execute the processing in terms of the hearer for choosing an indexical (above all, *so*-).

Keywords: eye tracking, distance judgment, demonstratives

1 Introduction

Languages involve deictic functions where an object is identified in terms of its relative location with the speaker, either spatially or temporally, and then introduced into a linguistic context. This feature is typically inherent, for example, in the English words *you, come, this, yesterday, left, next*, and *west*, as these references can vary depending on the viewpoint from which an entity is considered. Demonstratives—the subject of this study—also known as indexicals, representatively include such a deictic sense, as well as the anaphoric use, which are assumed to emerge in any language, such as Japanese (*ko-, a-,* and *so-*) and Mandarin Chinese (*zhè* and *nà*).

Interestingly, there are several (basically, two or three) contrastive/competitive demonstratives in most languages (e.g., “this” and “that” in English), the choice of which is based on **distance judgments** as to whether a target is close to or far from a person (speaker/hearer). Regarding usage in space (e.g., *kono hon* (this book), *ano ki* (that tree)), which is undoubtedly the most radical or original use, a speaker needs to construe a spatial situation and comprehend a positional relationship between themselves and a target to make such a statement. In addition,

some languages, such as Japanese and Korean, seem to require another kind of judgment for an indexical—whether an object is posited on the speaker's or on the hearer's side. In sum, Japanese demonstratives work as a system of two kinds of axes and the choice among three options, whereas Chinese demonstratives work based on one axis and a selection between two options. Notably, the former axis, concerning "near" or "far," may conceptually overlap with the latter—"the speaker-territory" or "the hearer-territory" (as detailed in the following sections)—which may complicate the use of Japanese demonstratives.

General questions

This study deals with the *deictic* and *spatial* use of demonstratives in Japanese and Chinese, with their different systems and options. There are several general and specific questions that linguists should resolve in terms of within-language and cross-language. Broadly, one of the most relevant mysteries is related to dividing a space into one or two axes, specifically raising the following questions: (i) Is the distance evaluation same among languages? (ii) Where and how does a speaker draw a line between "near" and "far" in each language? (iii) How does a language with a relatively complex system of indexicals (i.e., Japanese) compromise the overlapping of those two axes and distinguish the three indexicals?

With respect to (i) and (ii), when applying the perspective of Saussurean's thesis about **arbitrariness** in words and in dividing conceptual fields to this, the processing of distance judgments, dividing physical spaces, and outcomes of demonstrative selection (e.g., proximal or distal) must differ among languages, including that between Japanese and Chinese. If so, a linguistic-cultural distinction, as well as individual differences, should be incorporated into the linguistic research of demonstratives. Regarding (iii), this issue has been traditionally studied in research on individual language or in Japanese linguistics, as clarified in Section 2. However, no model has predicted precisely the occurrence of demonstratives (or their selection) because of the complexity involved in this phenomenon. It is obvious that any current artificial intelligence (AI) cannot use these expressions, as employed by human beings, despite the highly developed visual recognition. Currently, the intractable complexity involved in the choice has forced researchers to make such a fruitless claim—different factors may affect the choice. Simply put, this remains a question to be addressed in future studies.

In addition, although this is beyond the study's scope, the deictic use of demonstratives in the spatial domain can be extended to anaphora and other domains (e.g., time). Because the parallelism between them is uncertain, we could not discuss how the outcome of this study can be applied to such developed usages.

2 The linguistic literature

As mentioned in the previous section, Japanese demonstratives *ko*-, *a*-, and *so*- are selected from two perspectives; therefore, their usage is more complicated than that in Chinese, as there are many prior studies on Japanese indexicals. Unfortunately, however, most Japanese linguistic studies on demonstratives tend to focus on their **anaphoric** uses (e.g., Kuno 1973, 1992, Kuroda 1979, Kinsui and Takubo 1992, Togo 2000, Iori 2007) and the relationship between their deictic and non-deictic senses (e.g., Kinsui 1999, Tsutsumi 2012, Hirata 2014, Tokimoto 2015), as most linguists have been interested in the reference to prior linguistic context or the **discourse deixis** rather than the **place deixis** (see Fillmore (1997) for their terms).

Nevertheless, there are many theories regarding one of the above questions or how to divide the spatial use of indexicals (e.g., Sakuma 1951, Shoho 1981, Kamio 1990), most of which have insisted that *ko*- refers to a target on the speaker territory, *so*- on the hearer territory, and *a*- on the outside of these territories. In addition, some researchers have compared different language demonstratives in a deictic sense (e.g., Hattori 1992, Shi 2011, Yoshida 2014, Sugimura 2017). As it must be usual in language that seemingly equivalent lexicons from distinct languages should show different functions that can be differently divided by lexicons according to language, demonstratives seem to be no exception. Thus, no linguist would consider that, for example, even *ko*- and *zhè*, commonly referring to a "close" object, have the same meaning and use.

So-forms and the role of the hearer

It appears that the peculiarity of Japanese indexicals is attributable to the existence of *so*- series, as Kinsui (1999) and Tsutsumi (2012) assumed that the primitive essence of the *so*-series was not its direct reference (as in the other two) but instead may originate from anaphoric use. Kinsui (1999) mentioned that, however, this has been unclear owing to insufficient evidence so far. In addition, setting aside the historical viewpoint and considering only the deictic use, it is not for the current Japanese native speakers that *so*- should be treated as a special class separated from *ko*- and *a*-.

Furthermore, Ogawa (2008) and Ogawa and Nozawa (2015) addressed the *so* series, insisting that there were two pairs (*ko–a* and *ko–so*) (cf. Mikami 1972), each divided in terms of distinct criteria. The former (*ko*- and *a*-) is the **ease of perception**, while the latter (*ko*- and *so*-) is the (relative) **prominence of the target**. These authors claim that, in sum, *so*- (or the *so–a* alignment) diverges from the *ko–a* alignment, is affected by psychological distance, and requires joint attention from the hearer. One of its advantages is that it allows for cognitive/phenomenological

flexibility in the selection.

Finally, it is necessary to discuss the **split** (or opposition) and **unification** (or fusion) types of hearers, concerning the positional relation between the speaker and the hearer (Shoho 1981). The former situates the hearer in contrast with the speaker and gives the same status to them (related to the choice between *ko-* and *so-*). The latter regards the speaker and hearer as a unified judge—considering the distance from it, a speaker chooses from *ko-* (proximal), *so-* (medial), and *a-* (distal). However, it is debatable whether *so-* can operate as a middle-range reference under the unification type (e.g., Sakata 1971, Yoshimoto 1992, Kinsui 1999).

Distance judgment and language differences

Few have focused on distance judgment per se, irrespective of the hearer, or on how speakers judge a target as close to or far from themselves, let alone its linguistic and cultural differences. For a language with a system of two-way demonstratives, such as English or Chinese, to evaluate in that term appears to be the primary and the only process for choosing an indexical. Evidently, however, the *psychological* relationship has a profound effect on that distance judgment; its example is the "affective *this*" (Lakoff 1972, Takubo 2010). Moreover, even if the Chinese language lacks the *so-* type of demonstrative, its speakers seem to consider the shared discourse field with the hearer, followed by *zhè* or *nà* (cf. Sanui 1988). Thus, considering the essential importance of the hearer in language or human life, this feature might apply to other languages with the same two-way system.

Breakaway from the traditional methodology

The prior studies, as introduced above, based on linguistic evidence, could be an answer to questions (i)–(iii) in Section 1, suggesting that the interpretation (or location) of the hearer and the psychological distance affect the distance judgment and the choice of demonstratives in any language. Radically, however, most of the studies above might be too linguistic and theoretical to answer the questions precisely, as those investigations have relied mostly on a small portion of linguistic expressions and researchers' individual intuitions based on their own linguistic experiences. From the standpoint of scientists studying objectively, such a fairly unscientific, authoritarian way of explanation without collecting any quantitative data cannot merely go beyond the level of theory (or hypotheses) nor accumulate plausible facts that can persuade researchers working in scientific spheres (e.g., cognitive science and psychology).

Furthermore, the questions raised here are psychological, as operating the system of demonstrative or related judgments must emerge only in the speaker's mind. The methodology that

focuses only on the consequence of that, or superficial linguistic utterances, as the scope of research, which is common in standard linguistics, does not approach the essence of language or human minds, leading to less predictiveness; instead, it continues merely the description (albeit convenient for language education), disturbed by the complexity of languages. For both reasons, this study avoids methodologically standing on the line of traditional Japanese linguistic conventions.

3 Theory

As presented in Section 4, our research acquired objective and quantitative data from a visual world eye-tracking experiment and analyzed it statistically to test the theory regarding the process that contributes to the choice of demonstratives. The present section will remark on the methodological background and put forth the theory and hypotheses of the topic, partly in relation to the literature in Section 2, which can be tested by a psycholinguistic experiment.

3.1 Background

The focus of this study is on **mental processing** rather than particular linguistic expressions. The speaker must implement some psychological processes before the occurrence of linguistic expressions, followed by the hearer executing some comprehensive processes in the mind to understand these expressions—not just intervening cords (i.e., linguistic sounds and letters) but also the entire process should be inclusively considered as the language. As pre-linguistic processing is invisible, many linguists have ignored the exploration of such a black box. However, cognitive scientific studies of language—not implying so-called cognitive linguistics—or psycholinguistics, have highlighted these aspects by utilizing the latest technology to measure physiological responses such as eye movements and brain waves. These studies uncovering psychological process can also contribute to revealing superficial outcomes as language expressions. This is, metaphorically speaking, similar to the idea that observing a chef's entire cooking process can clarify the reason for the tastiness of the meal presented to customers.

To explore the momentary processing of language, whether syntactic or semantic, eye-tracking is one of the most appropriate methods, enabling a higher **time resolution** of 60 to 300 (and more) hertz. Some psycholinguists conducting eye-tracking experiments have emphasized the **visual world paradigm** (e.g., Tanenhaus et al. 1995), in which analyzing *selective attention to visual stimuli* can reveal the cognition concerning different aspects of language processing. Consider this case: When one speaks of a past episode, one would attempt to exter-

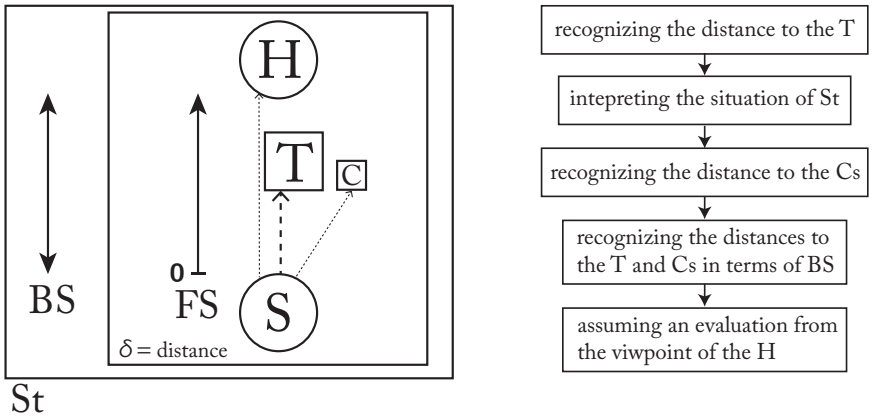

Figure 1 The assumed semantic/pragmatic construction of demonstratives and its components (on the left) and the hypothetical processing of distance evaluations (on the right).

nalize a stored mental image. As such, interpreting a spatial situation, whether real or merely a memory, to verbalize the construal can be considered a vital part of the language activity—obviously, this process is directly relevant to language or linguistic expressions.

3.2 Hypotheses

Clarifying the mechanism of distance judgment involved in most demonstratives, regardless of language, must be an essential key to answering research questions (i)–(iii) in Section 1. Sugaya (2015, 2020) has offered and substantiated the abstract structure and mental process of various kinds of evaluations (mainly responsible for adjective expressions). This could be partly applied to the current research, as the concepts "close to" or "far from" are expressed by adjectives in many languages.

The components of distance evaluation

The left side of Figure 1 shows the relevant components structurally arranged. Since each of the components shall be elaborated and illustrated in the sections that follow, let us now explain the figure as concisely as possible. First, the component that can be overtly expressed as a language is only a **target** (T) to be judged (e.g., "This *book* is interesting"), which is shown in a large square. The other components are covert in nature, including the **speaker** (S) who makes a judgment and an expression—persons are displayed by a circle. Evaluations cannot be made only with these two items, instead requiring some grounds for judgment. By comparing some **competitors** (Cs), whether they are real or memories, a T can be identified in terms of distance, or the **foregrounded scale** (FS), as scales are marked by arrows. Other scales named **backgrounded**

scales (BSs) may stipulate that FS—psychological distance (familiarity), accessibility, and prominence are the cases in point. Furthermore, a non-gradable, absolute condition called **standard** (St) may affect the choice of a demonstrative, exhibited by an enclosing square. For example, the norm concerning whether a T is in a space (e.g., on a table) might contribute to a selective decision in demonstratives. Finally, the speaker may recognize the existence of another judge evaluating a target from its viewpoint, or the **hearer** (H), who also plays the role in listening to and understanding S's speech.

Symmetric distance judgments from the perspective of the hearer

It is assumed that H would have the maximum impact on the choice of indexicals—especially the Japanese indexical system (cf. Shoho 1981, Ogawa 2008, Hirata 2013)—as another person (or judge) must be relevant and indispensable in a communicative or social scene. The status of H can be the same as that of S, or the relationship between S and H is completely symmetric and mutually replaceable, where the equidistance from both ($d_s = d_h$) would work as a boundary of the choice, thus equally dividing a space into two. This means that the sizes of the S and H fields (the circles with radius d) are the same ($f_s = f_h$), which may further lead to, in Japanese, the same size of regions that can include targets to which *ko-* and *so-* refer. In this case, the following conditions are induced: *ko-*; $(T \in f_s) \wedge (d_s < d_h)$, *so-*; $(T \in f_h) \wedge (d_s > d_h)$. Without any special criteria, this may be considered as the default or baseline in the case of split hearers. Notably, the fields are fairly flexible in size according to different situations or BSs. For example, these fields may be concerned with the scope of vision, the commonly attended space, or the range of movement. Moreover, in the case of Japanese, *a-* is selected when T is posited out of those fields, *a-*; $T \notin (f_s \vee f_h)$.

The processing for demonstratives

Additionally, the right-hand side of Figure 1 shows the assumed simplified online processing in meaning, based on the components introduced above, prior to demonstrative selection. This does not imply that all the components must be interpreted and considered; however, a speaker may encounter a complicated situation containing multiple components. For example, imagine the following complex situation in which an S has to choose a demonstrative: A man is sitting down across the table (St_1) from a woman in a small room (St_2). He is pointing at a book (T) (out of St_1) near the woman to refer to it, but the book lies further than some other books (Cs) from the woman, and some obstacles prevent the man from directly accessing T (BS).

Different factors appear to simultaneously affect the speaker's choice. The processing can

nonetheless incrementally progress in line, rather than being distributed in parallel. For instance, the following stepwise processes can be assumed: (1) the man comes up with an idea that he likes to refer to the book for some reason; (2) the book has to be evaluated in terms of distance for the utterance of an indexical; (3) he finds that it is not located on the current working space (table); (4) he cannot easily access the book because of the obstacles; (5) he notices that she is close to the reference target; (6) many closer books may make her feel that T is far away; (7) he thinks that T is far from both and so chooses *a-* (for example, '*ano* hon omoshiroi' ≈ '*that* book is interesting').

4 Experiment

To examine our hypotheses about the processing of utterances of demonstratives in Japanese and Chinese, we conducted an eye-tracking experiment, based on the visual world paradigm, for their native speakers. Their gaze data at a number of drawings, each of which exhibited a different condition, were recorded in addition to participants' responses to tasks. Eye-movement data were analyzed in terms of the areas of interest (AOIs) and times of interest (TOIs), which we set up in advance. Finally, the results of this experiment revealed the complex aspects of speakers' mental behaviors while uttering a demonstrative.

4.1 Methods

Participants

Twenty-one participants were recruited on our university campus. Among them, 11 (four female, eight male) participants, aged 19–43 years (M = 24.9, SD = 7.4), spoke Japanese as a native language. In addition, 10 (8 female, 2 male) participants speaking Mandarin Chinese, aged 23–27 years (M = 24.1, SD = 1.5), were all foreign students from China, studying in a Japanese graduate school, whose periods (years) of living in Japan varied (M = 2.10, SD = 1.56). After they had a Japanese or Chinese version of the experiment, and another experiment unrelated to this study, they received 500 yen for participation.

Working hypotheses

The theory shown in Section 3 can elicit several working hypotheses for each language. Regarding Japanese demonstratives, we assumed the following: (i) the distribution of "ko-" and "so-" would be primarily determined in terms of the bisector of a line connecting the speaker (S) and the hearer (H); (ii) as any deictic indexical (*ko-, a-, so-*) involves a distance evaluation, a

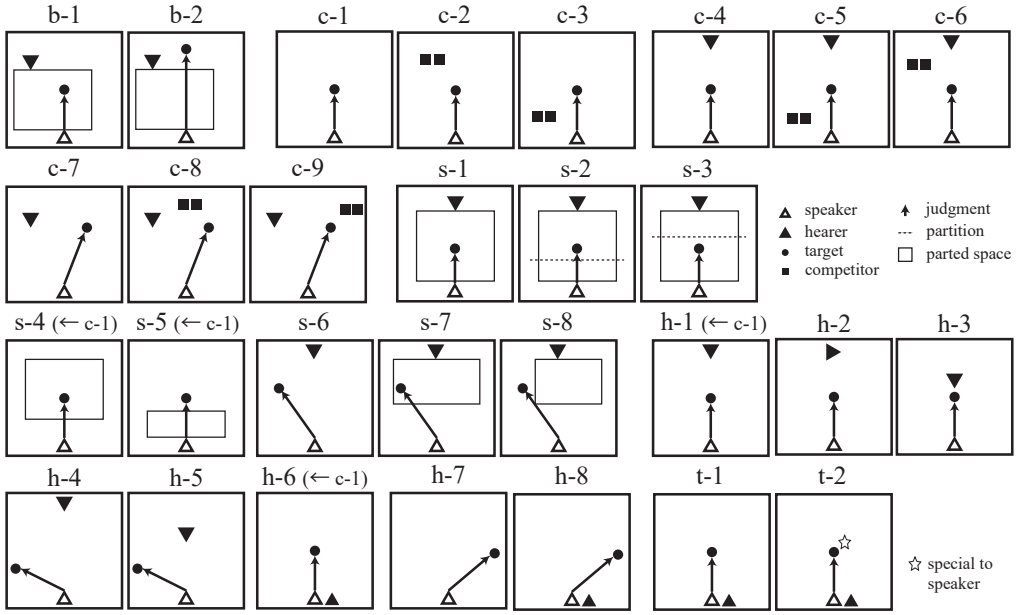

Figure 2 Schemas of stimuli eliciting different conditions for judging a target and then choosing a demonstrative among two or three options in Chinese or Japanese, respectively.

speaker (i.e., participants of the experiment) would have to process several environmental components—competitor (C), standard (St), backgrounded scale (BS), and the hearer—and consider the relationship of a speaker with a target (T); (iii) as the demonstrative *so-* is hearer based, speakers would need to consider the perspective of a hearer during its use, processing their positional relation to a hearer. Moreover, our hypotheses on Chinese demonstratives were that (iv) because the judgments are based solely on distance (i.e., closeness or farness) from S, they would not employ processing on H—at least, the degree must be lower than that of Japanese *ko-, a-, so-*. Further, (v) both *zhè* and *nà* also involve a proximity evaluation in pre-linguistic processes. Therefore, S would psychologically execute the processing concerning T, C, St, BS, and H, before making a statement including these demonstratives.

Graphics as stimuli

A total of 82 images were prepared for this experiment, and all were not photographs but drawings designed by Adobe Illustrator—a few examples are shown in Figure 3−5. One of the advantages of using illustrated images is that it allows us to easily arrange situations or items placed on a picture according to the purpose. This experiment investigated complicated criteria in the use of demonstratives in Japanese (i.e., *ko-, a-,* and *so-*) and Chinese (i.e., *zhè* and *nà*) and their cultural/linguistic contrasts by comparing these languages.

Figure 2 indicates abstract constructions, called **schemas** in this study, of all presented stimuli of images, depicting speakers' (participants') views. Although all schemas share the structure of a target being evaluated by a speaker for presenting a demonstrative, items are variously arranged according to the purpose. First, as illustrated in schemas *b*−1 and *b*−2, both have a face-to-face H and a parted space on/near which a T is located. As is portrayed on Figures 3 to 5, these schemas were specified as follows: a book, a cup, or shoes are positioned on/around a table; S points out T with their index finger; H is facing S across the table; and Ts are positioned in 16 different places—their positions are clarified by the results shown in Figure 4. These schemas and their instances are mainly concerned with hypothesis (i), while simultaneously exploring the significance of St and BS (a portion of hypotheses (ii) and (v)).

Next, schemas from *c*−1 through *c*−9 are associated with examining the effect of **competitors** on the judgment of the deixis. For example, *c*−1, *c*−2, and *c*−3 commonly contain T exactly at the same spot. However, in the latter two schemas, competitors are situated near or far from an S. If the results of an S's judgments differ from the case of *c*−1, they imply that competitors affect the speaker's feeling of distance. It is theoretically predicted that, despite evaluating the same T in the same place, say, a Chinese participant would choose *zhè* in the case of *c*−1 but *nà* in the case of *c*−3. Moreover, schemas *c*−4 to *c*−9 are related to the existence of H, or another person's viewpoint assumed by S, from which to evaluate T; for instance, it is expected that a Japanese speaker would prefer to select *so-* in the case of *c*−9 but *a-* in the case of *c*−8 to evaluate the same place of T.

The schemas from *s*−1 to *s*−8 were used to survey the effects of **standards** and **backgrounded scales**. The concept of a standard is briefly defined as a reason for a judgment and is related to absolute judgments regarding the occurrence of an event ($A \vee \neg A$). This broad character enables us to choose a perspective and create a setting from infinite options. One of the most typical Sts, and the easiest to make drawings, must be concerned with *accessibility* (accessible $\vee$ inaccessible), *reachability* (reachable $\vee$ unreachable), or *visibility* (visible $\vee$ invisible), each of which might conceptually overlap with one another; however, we utilized such perspectives for this research. When you view such a feature as gradable or scalar (e.g., "more accessible" and "less visible") rather than discrete, the perspective should be considered a BS instead of an St. For example, consider the schemas *s*−1 to *s*−3, where a partition prevents S from accessing T, and then see a drawing instantiated from *s*−2 on the left side of Figure 3, in which S evaluates an untouchable cup (T) placed over the transparent partition on the table. In addition, a distinguished region (e.g., a table) could be a standard per se, as we can judge whether T is placed in an area, and the schemas from *s*−4 to *s*−8 are the ones used to test the significance of this type of

standard.

Furthermore, *h*−1 through *h*−8 are schemas that investigate the relation of the **hearer** with Japanese and Chinese indexicals. Although it is true that some of the above schemas (including H) are also related to this point, these schemas can comparatively study the functions of H from different standpoints: (i) the existence of H, (ii) H's direction to T, (iii) H's distance from T, and (iv) H's position in relation to S. As for (ii), some drawings depict an H looking at a T; however, others look in the opposite direction, with other situations being the same (see schema *h*−2 and its instance on the right of Figure 3), the effects of which might make a participant choose a different demonstrative (e.g., *so-* → *a-*). Moreover, (iv) relates to two types of situations concerning the usage of Japanese demonstratives: the **split** and **unification** types (Shoho 1981). As explained in Section 2, the former is concerned with the scene in which S and H are clearly recognized as distinct judges (see schemas *h*−1 to *h*−5). The latter is related to the case in which they are fused and consequently regarded as one judge (see schemas *h*−6 to *h*−8).

Finally, schemas *t*−1 and *t*−2 are situated to check whether the nature of **target**, or particularly its relationship with S, would have an effect on demonstrative selections. Specifically, to substantiate the schema *t*−2, the instruction (see the top of Figure 3) was shown for the participants to suppose a targeted dog as their pet and then decide. As the position of T varies inside a room, we can compare the instances of *t*−1 and *t*−2 at several different locations.

Apparatus and procedure

The data of participants' eye movements were recorded using a screen-based eye-tracker (Tobii Technology's Tobii Pro Nano), with a sampling rate of 60 Hz, mounted on a 15.6-inch mobile monitor. In addition, the Tobii Pro Lab was used as a presentation and analysis tool, enabling precise manipulation over the entire experiment.

After filling out a form, completing an eye-tracker calibration process, and reading the general instructions, participants performed several rehearsals as opportunities to practice. Next, they started the main trials, where the same kind of sequence, as shown in Figure 3, was repeated. This experiment comprised eight sections, and at the beginning of each, some instructions and a situation description were indicated (see the top square of Figure 3). All the trials comprised three steps—(i) options presented, (ii) fixation, and (iii) a drawing presented—all of which were self-paced. After participants chose (an option including) a demonstrative by pressing a key during step (iii), they could proceed to step (i) for the following trial. This three-step loop lasted until each section ended.

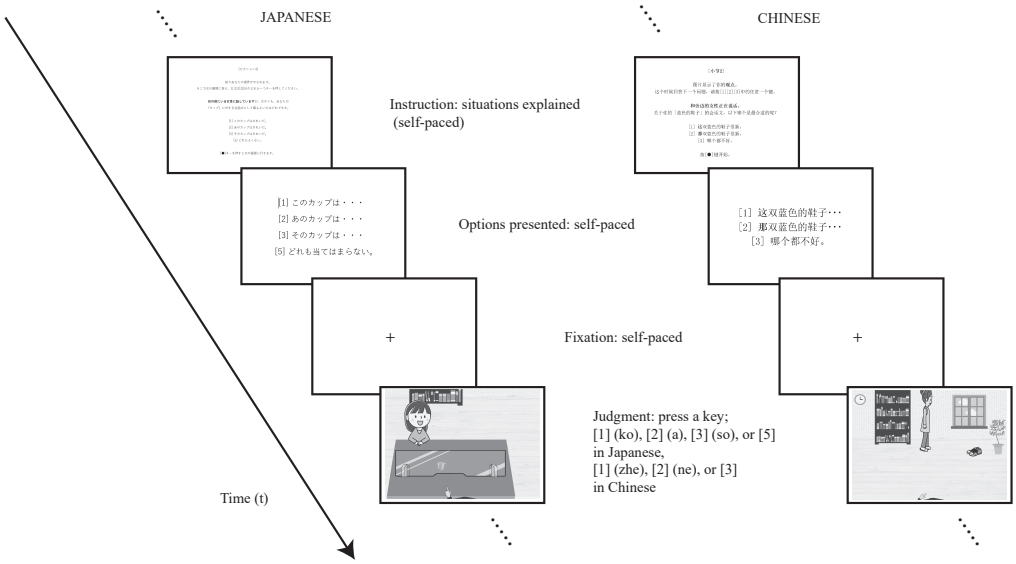

Figure 3 An example of the sequences in the experiment.

4.2 Results

This experiment provided us with a few kinds of data for analysis: responses (i.e., selections of demonstratives), response times (between a stimulus being present and a response), and gaze points (60 times per second) that showed visual behaviors, such as **fixations** and **saccades**.

As it is important to acquire information about where or which region a participant looks at in a visual image, some areas of interest (AOIs) were imposed on each image. Notably, items that were assigned an AOI included not only those standing for a semantic/pragmatic component, such as target and competitor, but also **distractors** irrelevant to such a component (a window, bookshelf, wall watch, plant, etc.). As these items would theoretically not be involved in any conceptual structure of demonstratives, they were treated as distractors, useful in comparison with the components suggested in this work. In addition, to reveal the online processing for a linguistic utterance (i.e., a demonstrative choice), we established times of interest (TOIs) every 500 ms from the onset in each trial. Owing to these two factors, we could be aware of when and to what degree a participant looked at an item responsible for a semantic/pragmatic component (e.g., C, St, and H).

4.2.1 Speaker's and hearer's territories for demonstratives

Schemas *b*−1 and *b*−2 and their instances (see Figure 4) tested two working hypotheses: (i) territories of use of demonstratives would be significantly divided by the bisector connecting a

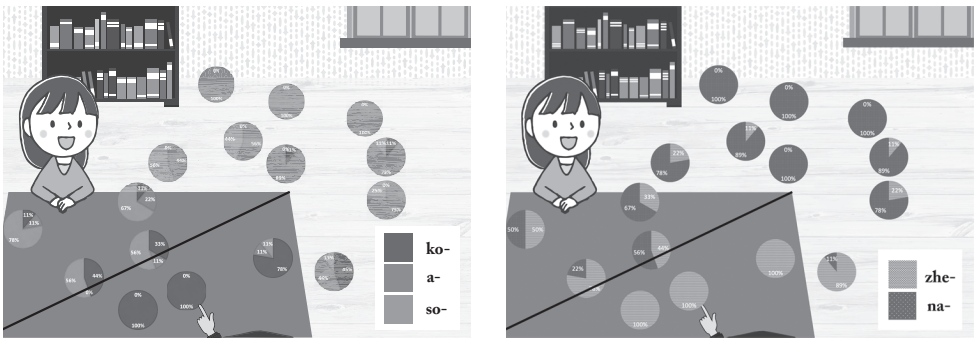

Figure 4 Distribution of demonstratives in a room situation: Regions are significantly divided by the bisector of a line connecting the speaker and the hearer and by the separated space (i.e., table).

speaker and a hearer (i.e., the line indicating the same distances from S and H); (ii) a judgment about whether T is posited on a parted area (or working space) could be seen as a standard that affects the selection of indexicals.

Figure 4 and Table 1 indicate the results, which support both the hypotheses in intuitive and statistical fashions, respectively. First, Figure 4 shows two pictures with the indexicals that Japanese/Chinese participants chose, at different loci, among *ko-*, *so-*, and *a-*, or among *zhè* and *nà*, respectively. Note that in these figures, diagonal lines (equal distances from S and H) are added to these original stimuli to clarify the bisector in this situation.

From the Japanese results, the following may be found: (i) *ko-* was used to refer to the targets that were both on the table and in the speaker-side field; (ii) *so-* was selected to refer to the targets in the hearer-side area over the bisector; and (iii) participants chose *a-* for the targets that were not on but near the table. However, the results of Chinese indicated that it was obvious that the further a T was, the more likely participants preferred *nà* rather than *zhè*; nevertheless, one might not be confident in the effects of the territory division or speaker/hearer-side fields on this demonstrative selection. In conclusion, the following can be understood from this figure: (a) the further a T is from the S, the likelier it is to be expressed by "far" demonstratives in both languages (i.e., *a-* and *nà*); (b) in Japanese, the bisector dividing the S and H fields plays a significant role in choosing between *ko-* and *so-*, and the standard in terms of parted space can function to make participants choose *a-*; (c) the effects of these two (in (b)) remain unknown in Chinese.

To confirm this statistically, Table 1 shows that the results of **multiple logistic regression analysis** exhibit almost the same conclusions as the consideration above. This statistical method of analysis measured the effects of several quantitative factors as independent variables—

distances from S and H and the positional relation of a T with a parted space—on choices of demonstratives as qualitative dependent variables. The overall results are shown in Table 1, which provides new insights. First, when considering the Japanese *ko-/so-* pair, the distance from the speaker must be significant for the selection (p <.001), and that from the hearer must be relevant for the judgment (p <.05). This is not the case in Chinese. While the distance from S is considered the primary aspect of indexical selection, the distance from the H does not demonstrate a significant difference in its effects on the choice between *zhè* and *nà*. Second, the absolute distinction concerning whether a T is put on a table seems to influence the preference of demonstratives in Japanese. As long as the results indicate, this is not true in Chinese; instead, this language seems to focus solely on the distance between S and T for the choice among its demonstratives.

4.2.2 Examining the effects of semantic/pragmatic components

In the preceding sections, we assumed that demonstratives share a complex semantic/pragmatic structure composed of multiple components, such as S, H, T, C, FS, BS, and St (see Figure 1 for their interrelationships). Thus, the working hypotheses are as follows: (i) a difference in any of these components could result in choosing a different indexical; (ii) a speaker would have to process some or all of these components mentally before the utterance—if so, they would gaze either frequently or lengthily at items relevant to the components. The data of the trials containing the prepared materials, explained in Section 4.1, were analyzed using a series of statistical methods, including association analysis and square tests, to validate the relevance of such components.

Table 1 Results of the multiple logistic regression analysis.

		estimate value	standard error	odds ratio	95%CI	p
Japanese (i)	intercept	-1.739	1.700	0.176	[0.005, 5.262]	
(*ko-* and *so-*)	distance from S	2.823	1.027	16.828	[2.664, 162.057]	**<.001**
	distance from H	-1.481	0.773	0.227	[0.039, 0.923]	**<.05**
Japanese (ii)	intercept	-4.828	1.716	0.008	[0.000, 0.195]	<.01
(*ko-* and *a-*)	distance from S	2.982	0.895	19.729	[3.935, 149.612]	**<.001**
	distance from H	-0.204	0.764	0.816	[0.154, 3.499]	
	on/around table	2.494	1.370	12.106	[1.041, 271.806]	**<.05**
Chinese	intercept	-4.036	1.065	0.018	[0.002, 0.135]	<.001
(*zhè* and *nà*)	distance from S	3.013	0.635	20.339	[6.330, 78.539]	**<.001**
	distance from H	-0.523	0.443	0.593	[0.235, 1.381]	
	on/around table	1.396	0.877	4.040	[0.785, 26.006]	

Competitors

The schemas *c*−1 to *c*−7 are, as stated in Section 4.1, selected for testing the roles of competitors, each of which was instantiated by different drawings. First, let us discuss how much time the participants took to gaze at the AOIs of each component. The ratio of gazing time of each component to the total trial time was calculated. In the cases that included T, C, S, H, and some distractors, Japanese participants looked more at components in the following order: T (*M* = 0.337, *SD* = 0.173), C (*M* = **0.191**, *SD* = 0.158), H (*M* = 0.055, *SD* = 0.110), S (*M* = 0.037, *SD* = 0.061), and distractors (*M* = 0.016, *SD* = 0.036). Within the range of error, this trend is almost the same as that in Chinese, as follows: T (*M* = 0.304, *SD* = 0.187), C (*M* = **0.228**, *SD* = 0.193), H (*M* = 0.069, *SD* = 0.141), S (*M* = 0.041, *SD* = 0.073), and distractors (*M* = 0.027, *SD* = 0.057). At least, the duration of focusing on Cs was longer than any distractor, suggesting that most participants, regardless of the language, considered the existence of competitors to a greater degree when producing a demonstrative.

However, the response data provide slightly different implications, especially in Japanese, as shown in Table 2. First, the analysis of association could not find any association rule in the results of the relevant trials [the existence of components (LHS) ⇒ the selection of demonstratives (RHS)]. Moreover, the probability of altering a choice owing to the change in schemas or the addition of components (e.g., *c*−1 → *c*−2) was low. This implies that the existence of the targeted component (competitors) did not affect Japanese participants' selection of the correct sentence fitting a situation. In addition, the results of the square test failed to reject the null

Table 2 The response data investigating the meaning of semantic/pragmatic components of demonstratives in Japanese and Chinese: Results of the analysis of association, calculating the probabilities of alternation, and the square test.

		association analysis					probabilities of alternation	square test
language	component	LHS	RHS	support	confidence	lift	0−1	p
Japanese	C	—	—	—	—	—	0.104	
	St	partition	*ko-/so-*	0.889	0.889	1	0.833	<.05
	H	opposite	*so-/a-*	0.972	0.972	1	0.201	<.05
		face-to-face	*ko(a)-/so-*	0.981	0.981	1	0.535	<.05
		unification	*ko(a)-/so-*	0.990	0.990	1	0.279	
	T	relation to S	*ko-/a-*	0.907	0.907	1	0.397	<.01
Chinese	C	competitor	*zhè/nà*	0.932	0.932	1	0.240	<.05
	St	partition	*zhè/nà*	0.960	0.960	1	0.296	<.05
	H	opposite	*zhè/nà*	0.958	0.958	1	0.174	<.05
		face-to-face	*zhè/nà*	0.972	0.972	1	0.069	
		unification	*zhè/nà*	0.967	0.967	1	0.138	
	T	relation to S	*zhè/nà*	0.911	0.911	1	0.488	<.01

hypothesis that the addition would not affect any difference from the control condition, which implied that competitors did not play a significant role for Japanese participants. However, this was not the case for the Chinese participants (see the lower row of Figure 2). As such, competitors had a profound effect on distance evaluation when choosing between *zhè* and *nà*.

Standard

The next component, the standard, defined as a two-valued judgment in a domain, can be responsible for the ground of the demonstrative choice. Again, if it can take indefinite values on a scale, we regard it as a backgrounded scale. Remember that the standards prepared for this experiment were (i) whether a T is situated in a region, (ii) whether an S can reach a T, and (iii) whether an S can view a T. Here, we focus on the data of only the standard (ii) because the standard (i) has been treated in Section 4.2.1, and (iii) is concerned with another component, the hearer, which will be dealt with in the following section.

The effect of *accessibility* on the judgment (St) was measured by adding a partition to the table in a drawing. In terms of the amount of gazing duration, any difference unfortunately did not work out according to the different schemas from *s*−1 to *s*−3. Although we had expected that participants would focus more on H in the case of *s*−2 than in the controlled condition *s*−1, they did not do so significantly in either language (*t*-test, $p > .05$). Regarding the response data, the effect of St was tremendous in Japanese, as most Japanese participants chose another indexical for the target, even in precisely the same locus. This is applied to Chinese participants to a much lesser degree, and the other two statistical methods supported the effects of St without linguistic differences.

Hearer

Third, one of the most important issues to resolve in this study is whether a speaker has to consider (and mentally process) the hearer when producing an indexical and its linguistic-cultural differences between languages with three-way demonstratives (e.g., Japanese *ko-, a-,* and *so-*) and two contrastive ones (e.g., Chinese *zhè* and *nà*). As a result of the eye-tracking data of all trials including an H, the use of Japanese *so-* called for the processing of (or focusing on) H—T (M = 0.334, SD = 0.184), H (M = **0.208**, SD = 0.191), compared to *ko-* (T (M = 0.412, SD = 0.165), H (M = **0.073**, SD = 0.109)), and *a-* (T (M = 0.420, SD = 0.225), H (M = **0.073**, SD = 0.127)). As such, the gazing data confirmed that the Japanese speakers were fully conscious of H or its location in their minds before the utterance of the deictic expression *so-*, compared to the others. However, although we had predicted that Chinese participants would draw less attention to H

during the selection, the data showed that the processing of H was evoked to a greater extent than expected by the use of *zhè* (T (*M* = 0.346, *SD* = 0.241), H (*M* = **0.139**, *SD* = 0.180)), and *nà* (T (*M* = 0.329, *SD* = 0.212), H (*M* = **0.137**, *SD* = 0.180)). Unlike Japanese *so*-, this language does not have any hearer-based indexical. Nevertheless, the distance evaluation—the judgment of whether T is close to or far from S—for the choice between *zhè* and *nà* seems to involve H in the entire process.

Furthermore, the data of responses also demonstrated the importance of the hearer for demonstrative productions in different ways from the data of viewing sight. From a comprehensive and contrastive perspective, Japanese participants were affected significantly more by the H in a stimulus than Chinese participants, a tendency shown in the previous discussion as well. This can be inferred because Table 2 shows that although the association analysis found every causal relation concerning the H, irrespective of the language, a few differences in behaviors—positions and directions—of the H were more inclined to influence the responses of Japanese participants in different scenes than those of Chinese participants. Specifically, when focusing on the rows whose LHSs are "face-to-face" in Table 2, their comparison convinces us that face-to-face interactions with H are considerably important and effective for native Japanese speakers, probably owing to the existence of the hearer-based demonstrative *so*-.

Target

Finally, this experiment presented a target, for comparison, that was instructed to be familiar with a speaker (participant), which clearly verified the effect of the *psychological* distance between T and S. In line with our expectation that people would want to use a "near" type of indexical (i.e., *ko*- and *zhè*) for a psychologically close target, participants of the experiment actually used those, according to any statistical analysis. In any language, at least in Japanese and Chinese, the mental proximity or intimacy—which can be seen as a BS—can get through a physically long distance. Such a psychological aspect must be a significant factor in the production of a demonstrative, which makes the usage of these expressions more complex and difficult.

4.2.3 Temporal variation in gazing and online processing

Additionally, this section copes with the data of the temporal change in gazing spots to approach the online processing contributing to the utterance of demonstratives. We have elucidated the components of the semantic/pragmatic structure of demonstratives or items to be processed by a speaker to make the choice. Another question concerns the processing order—how a speaker processes the semantic/pragmatic components demonstrated above. The combination of AOIs

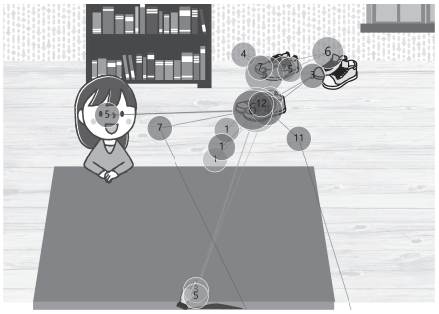

Figure 5 The example of a few participants' gaze plots in a stimulus instantiating *c*-9.

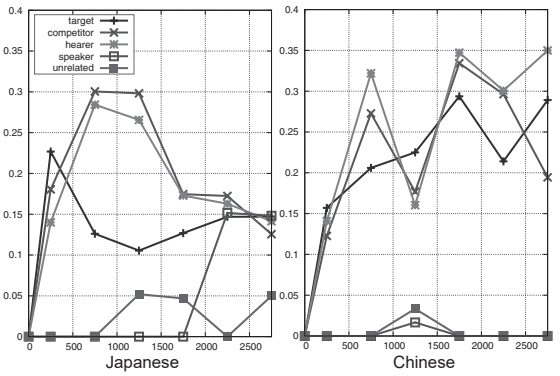

Figure 6 The temporal variation in gazing at objects in a trial including all of S, H, T, and C (0 ms = onset).

and TOIs can reveal this issue, as the eye-tracker recorded participants' gazing points 60 times per second.

Figure 5 denotes some examples of the temporal change in viewpoints over the drawing, an instance of the schema *c*−9, including a T, some Cs, the S, and the H. One can intuitively understand that, starting from the center, these participants looked at the items playing such roles from one to the next; this tendency was also observed in the others. Although we could not confirm anything new and clear from this, the data comprised many fixations and saccades, each of which was provided for stopping at the components and the movements between them, respectively. On this premise, Figure 6 displays the average of fixation durations of each language participants varying according to the change of time—the stimuli here are limited to the ones including all of S, H, T, and C. Notably, during the first 500 millisecond, they focused the most on targets, after which they equally looked at competitors and hearers to a great extent as environmental elements to make a judgment. In both languages, the figure does not make a difference between the behaviors of C and H, suggesting that this experiment cannot resolve the question regarding which is first processed by the speaker, C or H, to utter a demonstrative. Thus, we need to conduct an additional test to address this problem.

4.3 Discussion

This experiment revealed a new perspective for viewing demonstratives, highlighting the distance judgment involved in their utterances. Investigating the black box processing behind these expressions can also result in revealing part of their meaning and usages that traditional linguists have tried to explain. This study premises that the time to interpret a scene (image) for a

linguistic choice should be the term when the speaker executes some cognitive process of production. Based on this, a visual world eye-tracking experiment was conducted with Japanese and Chinese native speakers. In varying degrees, the results of the experiment supported most of the hypotheses in Section 3, related to the research questions presented in Section 1.

First, we examined the processing of the distance to a target, seen as the most fundamental for indexicals in any language, assuming that the speaker-/hearer-territories are bisected equally (cf. working hypotheses (i) and (iv) in Section 4.1). In the experiment with visual stimuli, schemas *b −1* and *b−2* and their concrete images tested the distance division and symmetry of S and H. As expected, the data indicated that the distances from both S and H affected the choices among *ko-* and *so-* in Japanese. Moreover, the distance only from S did the selections among *ko-* and *a-* in Japanese and *zhè* and *nà* in Chinese. This means that computing the S–T distance functions always for any indexical, and the choices among *ko-* and *so-* make the speaker consider the H–T distance as well. This is considered the default case that could be affected by the components in question. However, there is still a question about the choice of *nà* for a target in the hearer-territory, as the results could not exclude the possibility that a target in that area could make the speaker consider the hearer, as shown in Figure 4.

Second, with respect to the distance evaluation (see working hypotheses (ii) and (v) in Section 4.1), the semantic or pragmatic components offered by Sugaya (2015, 2020) appear to operate well, irrespective of language, for indexical selections. The experiment presented several drawings instantiating various schemas created for researching each component (see Figure 2) and recorded the response and eye movement data. Although these two types of data were incoherent in some cases, the existence of each component largely affected participants' gazes and responses at a significant level (cf. Section 4.2.2 for detailed descriptions). However, some questions remain unresolved in terms of the components used in the test: (i) the standard and background scales can be realized in considerably different ways, but this experiment offered only a few manifestations (e.g., partition and familiarity); (ii) the reason is unclear why competitors were not employed significantly even if participants gazed lengthily at them. Another experiment should be designed to resolve such issues.

In particular, from a contrastive perspective, the most outstanding divergence was in the interpretation of the hearer, which has been in recent focus (e.g., Ogawa 2008, Hirata 2013). The present experiment acquired more coherent and robust data regarding the role of the hearer, supporting hypotheses (iii) and (iv). System distinctions between languages obviously emerged in the results of the experiment, as the change in H's behavior and position had a significant effect on the Japanese speakers, compared to the Chinese participants. This suggests that Japanese

demonstratives, including *so-*, evoked more selective attention to H than did Chinese demonstratives. Considering that speakers did not look at H before choosing the others, however, the existence of the indexical *so-* might work uniquely on the results. Regarding the remaining questions, although the effects of the face-to-face H on indexical choices were outstanding in the Japanese *ko-/so-* series, it is difficult from the results to ascertain the effects of two different—opposition and fusion—cases of the hearer (cf. Section 2) on the meaning of demonstratives; in connection with this, another study can verify the **medial** *so-* in the latter situation.

Finally, the eye tracker recorded the processing sequence over the visual drawings. The results showed that, after processing a target, speakers seemed to take care of competitors or the hearer to a greater degree, with less differences between the two components. This tendency was the same as that observed in the Chinese participants. However, from the present results, we cannot draw other significant findings. Therefore, another eye-tracking experiment focusing on this dimension could test the theory and hypotheses in Section 3 to reveal the online processing leading to indexical expressions.

5 Conclusion

This study theorized the cognitive processing needed to make an utterance including a demonstrative, from the viewpoint of semantics and pragmatics. It then demonstrated the related hypotheses through an eye-tracking experiment. Specifically, we examined the theory concerning not only the basic processing of the distance from the speaker and hearer, but also the processing in terms of relevant components of distance judgments, the latter of which is connected with the flexibility in the usage of indexicals. The results of the experiment measuring eye movements over spatial scenes demonstrated these aspects and uncovered the cultural/linguistic contrasts between Japanese and Chinese, as specifically shown in the previous section.

Furthermore, the present research methodologically departed from earlier linguistic studies on this topic and others. We have coherently dealt with the hidden mental processes, rather than language expressions per se, and adopted an objective, scientific approach by conducting reproducible experiments testing the falsifiable hypothesis, instead of merely linguistic descriptions and explanations commonly seen in traditional linguist studies. Based on these, the current exploration is enabled to introduce a new perspective to view the topic and accumulate credible facts. As a final remark, this study implies that cognitive process of language is more likely to operate commonly in most languages than the surface of language. Investigating such universal aspects of language that allows for connections with neighboring spheres of science and

technology may lead to the broad elucidation of the mechanisms of human language and cognition.

Acknowledgments

The author would like to thank the two anonymous reviewers for their valuable comments. This research was partly supported by a grant from JSPS KAKENHI (Project number: 19K13159). The oral presentation using a portion of the experimental data of Japanese participants was conducted at the 163th Meeting of the Linguistic Society of Japan (LSJ).

References

Fillmore, Charles J. 1997. *Lectures on Deixis*. Stanford: CSLI Publications.

Hattori, Shirou. 1992. Kore, sore, are and this, that [kore, sore, are to this, that]. In S. Kinsui and Y. Takubo, eds., *Demonstratives [Shizishi]*, pages 47−53. Hitsuji Shobo.

Hirata, Miki. 2013. From distance to joint attention—the analysis of 'medial' so- in Japanese demonstratives using the concept of addressee's attention—[kyori kara kyodotyusi e: kikite no tyui wo motiita tyukyorishizi no sokei no saibunseki]. *Papers from the 13rd National Conference of the Japanese Cognitive Linguistics Association* 13: 461−472.

Hirata, Miki. 2014. An integrative analysis of deictic and non-deictic uses of Japanese demonstrative so- using the concept of attention [tyui gainen wo motiita sokei no tyokuzi youhou to hi-tyokuzi youhou no touitsuteki bunseki]. *Gengo Kenkyu* (146): 83−108.

Iori, Isao. 2007. *A Study of Cohesion in Japanse [Nihongo ni okeru Tekisuto no Kessokusei no Kenkyu]*. Tokyo: Kuroshio shuppan.

Kamio, Akio. 1990. *The Theory of Territory of Information: A Functional Analysis of Language [Zyouhou no Nawabari Riron: Gengo no Kinou-teki Bunseki]*. Tokyo: Taishukan.

Kinsui, Satoshi. 1999. On the relation between the deictic use and the non-deictic use of the Japanese demonstratives [nihongo no shizishi niokeru tyokuziyouhou to hi-tyokuziyouhou no kankei nitsuite]. *Natural Language Processing [Shizen Gengo Shori]*. 6(4): 67−91.

Kinsui, Satoshi and Yukinori Takubo. 1992. Japanese demonstratives seen from a dynamic model of discourse management [danwa kanri riron kara mita nihongo no shizishi]. In S. Kinsui and Y. Takubo, eds., *Studies on the Japanese Language: Demonstratives [Nihongo Kenkyu Siryou-shu: Shizishi]*, pages 123−149. Hitsuji Shobo.

Kuno, Susumu. 1973. Ko, so, a. In *Japanese Grammar Research [Nihon Bunpo Kenkyu]*, pages

185−190. Tokyo: Taishukan.

Kuno, Susumu. 1992. Ko, so, a. In S. Kinsui and Y. Takubo, eds., *Studies on the Japanese Language Demonstratives [Nihongo Kenkyu Siryoushu: Shizishi]*, pages 69−73. Tokyo: Hitsuji Shobo.

Kuroda, Shige-yuki. 1979. On (ko,) so, a [(ko) so a ni tsuite]. In *Japanese and English*, pages 41−59. Tokyo: Kuroshio shuppan.

Lakoff, Robin. 1972. Remarks on 'this' and 'that'. *Proceedings of the Chicago Linguistics Society* 9: 229−246.

Mikami, Akira. 1972. *An Introdauction to Modern Japanese Grammar [Gendai Gohou Josetsu]*. Tokyo: Kuroshio shuppan.

Ogawa, Noriko. 2008. A cognitive research on Japanese demonstratives: On the treatment of the listener in so-series [nihongo shizishi no ninchiteki kenkyu: so-kei shizishi niokeru kikite no itizuke saikou]. *Papers in Linguistic Science [Gengo Kagaku Ronshu]* 14: 57−88.

Ogawa, Noriko and Hazime Nozawa. 2015. Cognitive foundation of demonstratives and selection principles [shizishi no ninchi teki kiban to sentaku genri]. In M. Yamanashi, ed., *Studies in Cognitive Linguistics [Ninchi Gengogaku Ronko]*, vol. 12, pages 115−166. Tokyo: Hitsuji Shobo.

Sakata, Yukiko. 1971. On the functions of demonstratives ko, so, a [shizishi ko so a no kinou nitsuite]. *Tokyo University of Foreign Studies* 21: 125−138.

Sakuma, Kanae. 1951. *Modern Japanese Expression and Usage [Gendai Nihongo no Hyougen to Gohou]*. Tokyo: Kouseikaku.

Sanui, Tadayoshi. 1988. A pragmatic review on Chinese demonstratives [tyugokugo shizi-daimeishi no goyouron teki kentou]. *The Journal of Social Sciences and Humanities* 198: 1−19.

Shi, Sen. 2011. *A comperative study of demonstratives in Japanese and Chinese [Nittyu shizishi no taishou kenkyu]*. Ph.D. thesis, Hitotsubashi University, Graduate school of language and society.

Shoho, Isamu. 1981. The system of ko/so/a [ko, so, a no taikei]. In *Japanese Demonstratives [Nihongo no Shizishi]*, pages 51−122. Tokyo: National Institute for Japanese Language and Linguistics.

Sugaya, Yusuke. 2015. Inside the mental mechanism of evaluation processing. *Papers in Linguistic Science* 21: 153−182.

Sugaya, Yusuke. 2020. *Constructing Evaluations: The Meaning-Making Process of Adjectives*. Ph.D. thesis, Kyoto University.

Sugimura, Yasushi. 2017. On choice of Japanese demonstratives “kono,” “sono,” “ano” in the conversation between a speaker and a listener [nishakaiwa bamen ni-okeru nihongo no kono sono ano no sentaku]. *Japanese Language Education [Nihongo Kyouiku]* 82: 39−52.

Takubo, Yukinori. 2010. *The Structure of the Japanese Language: Inference and Knowledge Management [Nihongo no Kouzou: Suiron to Tisikikanri]*. Tokyo: Kuroshio shuppan.

Tanenhaus, Michael K., Michael J. Spivey-Knowlton, Kathleen M. Eberhard, and Julie C. Sedivy. 1995. Integration of visual and linguistic information in spoken language comprehension. *Science* 268(5217): 1632−1634.

Togo, Yuji. 2000. Discourse model and Japanese demonstatives ko/so/a [danwamoderu to nihongo no shizishi ko, so, a]. *The Integrated Human Studies* 7: 27−46.

Tokimoto, Shingo. 2015. An experimental description of deictic and nondeictic usages of Japanese indexicals, ko-, so-, and a-forms [ko-so-a siyou kozinsa no zikkenteki kijutu]. *Mathematical Linguistics* 30(1): 1−13.

Tsutsumi, Ryoichi. 2012. *An Overall Study of the Demonstratives of Modern Japanese [Gendai Nihongo Shizishi no Sougouteki Kenkyu]*. Tokyo: Koko Shupan.

Yoshida, Mitsunobu. 2014. Contrastive analysis of deictic and anaphoric demonstratives in Japanese, English, and German [nihongo doitsugo eigo no shizishi no hikaku ni kansuru iti-kousatsu]. *Studies in European and American Culture [Oubei Bunka Kenkyu]* pages 31−64.

Yoshimoto, Kei. 1992. The system of Japanese demonstratives ko/so/a [nihongo shizishi ko/so/a no taikei]. In S. Kinsui and Y. Takubo, eds., *Studies on the Japanese Language: Demonstratives [Nihongo Kenkyu Siryoushu: Shizishi]*, pages 105−122. Tokyo: Hitsuji Shobo.