Collection of Japanese Route Information Reference Expressions Using Maps as Stimuli

Yoshiko Kawabata[1], Mai Omura[1], Hikari Konishi[2], Masayuki Asahara[1], Johane Takeuchi[3]
[1]National Institute for Japanese Language and Linguistics, Japan,
[2]Tecca LLC,
[3]Honda Research Institute
Correspondence: masayu-a@ninjal.ac.jp

Abstract

We constructed a database of Japanese expressions based on route information for language-based direction instructions to autonomous driving systems. Using 20 maps as stimuli, we requested descriptions of routes between two points on each map from 40 individuals per route, collecting 1600 route information reference expressions. We determined whether the expressions were based solely on relative reference expressions by using landmarks on the maps. In cases in which only relative reference expressions were used, we labeled the presence or absence of information regarding the starting point, waypoints, and destination. Additionally, we collected clarity ratings for each expression using a survey.

1 Introduction

Accurately conveying route information in a language is challenging because it comprises details regarding the starting point, waypoints, and destination. Utilizing surrounding landmarks is crucial for effectively conveying positional information, movement direction, and distance. In languages where case elements such as subjects tend to be omitted (e.g., Japanese), it is difficult to generate clear expressions of route information.

By collecting Japanese route information reference expressions using maps as stimuli, this study aims to shed light on important perspectives for conveying route information for language-based direction instructions to autonomous driving systems. For each of the 20 maps, two starting and ending point patterns were established, resulting in 40 stimuli each. Through crowdsourcing, we sought to gather 40 expressions of route information references per map, articulated solely using specific or relative location information. Following the determination of whether the collected language expressions comprised only specific or relative location information, we annotated the inclusion of starting point, waypoint, or destination information. Furthermore, the clarity ratings for each expression were obtained through surveys.

2 Related Research

Early work on the analysis of direction-giving conversations was conducted by Psathas and Kozloff (1976), who identified three stages: situating, specifying information and directions, and concluding. They pointed out that during the initial directions, a) starting point, b) destination, c) mode of travel, d) time of travel, and e) membership categorization by the parties of each other are important. In a related study, Clark and Wilkes-Gibbs (1986) introduced a collaborative model to make definite references in conversations. In their model, speakers initiate the process by presenting a noun phrase, which is iteratively refined by participants until a mutually accepted version is reached, thus minimizing joint effort. In such dialogues, there is often a significant amount of coreference information regarding locations. Additionally, Levinson (2004) focused on the relationship between language and cognition and explored the role of space in cognitive diversity. Moreover, Lakoff (1987) proposed a SOURCE-PATH-GOAL Schema, identifying four structural elements: a) SOURCE, b) DESTINATION, c) PATH, and d) DIRECTION. Barclay and Galton (2008) constructed 'scene corpus' in which spatial expressions were collected by showing virtual objects. Shelton and McNamara (2004) performed three experiments based on the fictional environments described by Taylor and Tversky (1992). They evaluated the mental costs associated with switching between route and survey perspective stimuli. Their experiments utilized eight patterns of maps rotated at 45° increments. In contrast, this study collects route expressions by showing a real-world map instead of focusing on positions.

Proceedings of the 4th Workshop on Spatial Language Understanding and Grounded Communication for Robotics (SpLU-RoboNLP 2024), pages 30–35

3 Data Collection Methodology

3.1 Collection of Route Information Reference Expressions

Route information reference expressions were collected using Yahoo! Crowd Sourcing. Participants were provided with maps, as shown in Figure 1, and were asked to describe any route information starting from ■ and ending at ★ (for screening) or ● (for main task).

During the task, the goal was to collect relative reference expressions. Participants were presented with a rotated version of the original map as stimuli and given the following instructions:

- Use "front, back, left, and right" as if the participants were initially located at the ■ mark on the map.
- Do not use "up, down, left, and right."
- Do not use "east, west, south, and north."

Figure 1: Example map used as stimuli (rotated 30°)

Expression collection was conducted in two stages: a screening survey and main survey. In the screening survey, maps rotated at angles of 30°, 120°, 210°, and 300° were used, as shown in Figure 1. Data were collected from 400 participants for each map for a total of 1600 responses. An example survey screen is shown in Appendix Figure 2. Participants were compensated with 10 yen equivalent PayPay points per response for the screening survey. The screening survey was conducted from 08:01 November 2, 2023, to 03:40 November 3, 2023. Among the participants, those who rated the clarity of expressions in the subsequent survey with a score of 3.0 or higher (206 individuals) were selected for the main survey. A clarity assessment of the screening survey results was conducted from 08:06 November 17, 2023, to 09:55 November 17, 2023.

In the main survey, 40 stimuli were used, each consisting of 20 types of maps with two starting and ending point patterns. A total of 1600 expressions were collected from 10 participants for each of the 160 variations of stimuli, which included four types of rotations for the 40 stimuli. Participants were compensated with 50 yen equivalent PayPal points per response for the main survey. The main survey was conducted from 14:02 November 17, 2023, to 23:55 November 19, 2023. Because we use a real-world map, there are multiple routes to the starting point and destination. In this context, we also attempt to collect expressions regarding which target points on the map are easiest to explain.

3.2 Classification of Collected Expressions

All 1600 expressions collected in the main survey exclusively consisted specific location information expressions and relative location information expressions, without the use of absolute location information such as east, west, south, and north.

However, five expressions were identified as inappropriate route information reference expressions because they mistakenly recognized incorrect marks on the map as the starting and ending points.

Additionally, 29 expressions were identified as inappropriate route information reference expressions because they recognized the starting and ending points in reverse (labeled as "W").

Subsequently, the following classifications were assigned (examples are based on Figure 1 as stimuli):

- X: Detailed description of the starting point
- Y: Description of points along the route
- Z: Detailed description of the destination

For the determination of the starting point (X), if there was a clear indication of the starting point such as "facing [location]," "standing in front of [location]," "between [location A] and [location B]," or "leaving from [location]," it was classified as starting point present (X). Additionally, if the starting point was explicitly stated as "from current location" or "from ■," it was also classified as starting point present (X).

For the determination of the route (Y), the presence of verbs indicating movement such as "turn left," "turn right," "go straight," "turn," or "go around" was used.

For the determination of the destination (Z), if explicit words indicating the destination such as "goal," "destination," or "arrive" were identified, it was classified as destination present (Z). Even if explicit words were not present , they were classified as destination present (Z) if there was a specific description of the destination. However, if there was no specificity regarding the destination, such as "after a while, on the left side," "beyond the plaza," or "passed by," it was classified as destination absent.

These classifications were set as multi-labels.

3.3 Clarity Rating of Expressions

The clarity of expressions was assessed through a survey using Yahoo! Crowd Sourcing for 1600 expressions collected in both the screening and main surveys. For the screening survey, 887 individuals who provided expressions consisting only of specific and relative location information were recruited, with 216 participants providing ratings for the screening survey data and 605 participants providing ratings for the main survey data.

A survey screen example is shown in Appendix Figure 3. Seven expressions were randomly presented for each map and ratings were made on a 6-point scale from 0 (difficult to understand) to 5 (easy to understand). Participants were compensated with 2 yen equivalent PayPal points per response. The clarity rating survey for the main survey collected ratings from 35 individuals per expression. The main survey was conducted from 17:01 December 14, 2023, to 13:10 December 16, 2023.

The expression and impression rating collection for this study was approved by the ethical review board of the National Institute for Japanese Language and Linguistics.

4 Data Statistics

Table 1 presents the individual aggregations based on the presence or absence of each aspect (TRUE for presence and FALSE for absence) along with the average clarity for each aspect.

In cases where the starting point and destination were mistaken (W), they tended to be less clear than the correct ones. A t-test was conducted, assuming unequal variances between the two samples. The mean clarity rating for W=FALSE was 2.79, whereas that for W=TRUE was 2.07. The difference between the means was significant ($t = 10.11$, $p < 0.001$, two-tailed).

Table 1: Clarity by Aspect (Individual)

Misidentification (W)	W=FALSE	W=TRUE
Count	1571	29
Clarity (Average)	2.79	2.07
Starting Point (X)	X=FALSE	X=TRUE
Count	599	1001
Clarity (Average)	2.72	2.81
Route (Y)	Y =FALSE	Y=TRUE
Count	24	1576
Clarity (Average)	2.00	2.79
Destination (Z)	Z=FALSE	Z=TRUE
Count	51	1549
Clarity (Average)	2.43	2.79

Explanations for the starting point (X) were frequently omitted, with 37.5% (599/1600) of the expressions lacking such an explanation. Those without a starting-point explanation exhibited slightly lower clarity than those with an explanation. A t-test assuming unequal variances between the two samples showed a significant difference ($t = -4.54$, $p < 0.001$, two-tailed), with a mean clarity rating of 2.72 for X=FALSE and 2.82 for X=TRUE.

The absence of route explanations (Y=FALSE) comprised 1.5% (24/1600) of the expressions, with an average clarity rating of 2.00, which was notably lower than the average clarity rating of 2.79 for expressions with route explanations (Y=TRUE). A t-test with unequal variances revealed a significant difference between the two groups ($t = -8.94$, $p < 0.001$, two-tailed).

Similarly, the absence of explanations for the destination (Z) was observed in 3.2% (51/1600) of the responses. These expressions had a lower clarity average of 2.43 compared to those with a destination explanation (clarity average of 2.79). A t-test assuming unequal variances between the two samples revealed a significant difference ($t = -5.78$, $p < 0.001$, two-tailed), with a mean clarity rating of 2.43 for Z=FALSE and 2.79 for Z=TRUE.

Hence, it is evident that all three aspects - starting point, route, and destination - play crucial roles in effectively conveying route information. Addi-

tionally, it i s noteworthy that explanations for the starting point (X) are often omitted, with 37.5% (599/1600) of the expressions lacking a starting point explanation. This observation underscores the challenge of referencing a starting point, which contributes to the complexity of comprehensively conveying route information.

Table 2: Clarity by Aspect (Combinations)

W	X	Y	Z	Clarity	Count
F	F	F	T	1.75	8
F	F	T	F	2.42	25
F	F	T	T	2.76	558
F	T	F	F	1.93	2
F	T	F	T	2.21	12
F	T	T	F	2.52	22
F	T	T	T	2.85	944
T	F	T	T	2.07	8
T	T	F	T	1.87	2
T	T	T	F	2.20	2
T	T	T	T	2.08	17
Total				2.78	1600

Table 2 presents the aggregation results based on the combination of aspects. From these combinations, it was observed that both route (Y) and destination (Z) explanations were predominant, accounting for 93.9% ((558+944)/1600). There was a difference in clarity depending on the presence of starting point (X) explanations; those with a starting point explanation (clarity average of 2.85) were clearer than those without (clarity average of 2.76).

Appendix Figure 4 shows one of clearest examples. This example was the clearest explanation (avg. 3.74) of the route when showing the map in 13-a. These expressions included all information regarding Starting Point(X), Route(Y), and Destination(Z). Instead of describing complex alleys, the route chosen explains the main streets, selecting landmarks along the route such as Hareza Tower (ハレザタワー), the elevated highway, and Nitori (ニトリ) as reference points.

5 Conclusions

This study aimed to collect clear route information reference expressions using both specific and relative location information for language-based direction instructions to autonomous driving systems. Through crowdsourcing, expressions describing route information on maps were gathered. Each expression was manually annotated to determine whether it contained information regarding the starting point, route, and destination. The data revealed a tendency for the starting point information to be missing from expressions. Additionally, clarity ratings were collected through crowdsourcing, indicating the importance of including information regarding routes and destinations.

Existing studies, particularly map task corpora, have focused on analyzing interactions involving the transmission of new and old information based on map tasks; they lack attention on clear route information explanations using maps. This study, however, uses maps of real locations instead of virtual environments. Generating clearer route information reference expressions from map information requires collecting diverse expressions along with their clarity ratings, and examining what contributes to clarity. This study differs from previous research because it collected route information reference expressions.

In future research, we plan to annotate reference expression information. Relative reference expressions are inherently based on information from three or more points and are known to be abstracted as according to the double-cross model (Freksa, 1992). However, utterances that contain information from three or more points cannot be obtained (Kawabata et al., 2023). Data collected in this study include information from three or more points. Therefore, the data should be annotated based on the double-cross model.

The data gathered in this study, including the map information, classifications, and clarity ratings, are accessible to the public at https://github.com/masayu-a/HRI-JP-RIRE-DB.

Acknowledgments

We would like to express out gratitude for the support and collaboration provided by the National Institute for Japanese Language and Linguistics (NINJAL) in the joint research project, "Evidence-based Theoretical and Typological Linguistics." Additionally, we are deeply thankful for the joint research partnership between the Honda Research Institute and NINJAL. This research is partially supported by JSPS KAKEN JP22K13108 and JP19K13195.

References

Michael Barclay and Antony Galton. 2008. A scene corpus for training and testing spatial communication systems. In *AISB 2008 convention communication, interaction and social intelligence*, volume 1, page 26.

Herbert H Clark and Deanna Wilkes-Gibbs. 1986. Referring as a collaborative process. *Cognition*, 22(1):1–39.

Christian Freksa. 1992. Using orientation information for qualitative spatial reasoning. In *Theories and methods of spatio-temporal reasoning in geographic space*, pages 162–178. Springer.

Yoshiko Kawabata, Mai Omura, Masayuki Asahara, and Johane Takeuchi. 2023. Spatial information annotation based on the double cross model. In *Proceedings of the 37th Pacific Asia Conference on Language, Information and Computation*, pages 137–144, Hong Kong, China. Association for Computational Linguistics.

George Lakoff. 1987. *Women, fire, and dangerous things: What categories reveal about the mind*. University of Chicago press.

Stephen C Levinson. 2004. *Space in Language and Cognition: Explorations in Cognitive Diversity*. Cambridge University Press.

George Psathas and Martin Kozloff. 1976. The structure of directions. *Semiotica*, 17(2):111–130.

Amy L Shelton and Timothy P McNamara. 2004. Orientation and perspective dependence in route and survey learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(1):158–170.

Holly A. Taylor and Barbara Tversky. 1992. Spatial mental models derived from survey and route descriptions. *Journal of Memory and Language*, 31(2):261–292.

6 Appendix

[Please describe the route from the marker "■" to the marker "★" on the map using surrounding landmarks.]

- Provide an expression of 30 characters or more and less than 200 characters.
- Describe using the perspective of you initially being at the ■ mark on the map, considering front, back, left, and right.
- Refrain from using up, down, left, and right in reference to the map layout.
- Avoid using north, south, east, and west.

Figure 2: Example of the survey screen: Collection of route information reference expressions

[Please assess the clarity of the document describing the route from the marker to the destination, after viewing the map in the following link.]

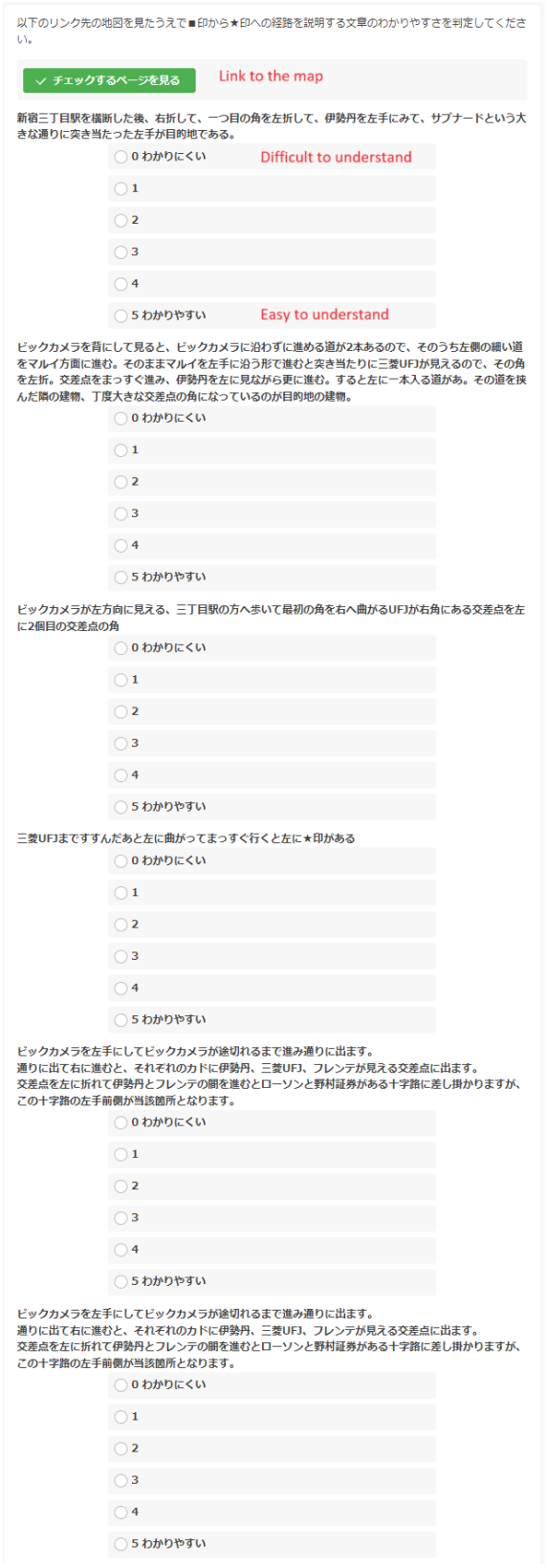

Figure 3: Example of Survey Screen: Clarity Rating Survey

(Example) 大通りをハレザタワー方面にしばらく進むと高速が高架になっている大通りにでます。その大通りを渡らずに右折してください。高架に沿ってしばらく進むと右手にニトリが見えてきます。ニトリを過ぎてすぐ右手に隣接しているのが目的地です。

(Label) W=FALSE, X=TRUE, Y=TRUE, Z=TRUE, Clarity=3.74

(English Translation) Proceed along the main street towards Hareza Tower (ハレザタワー) for a while until you reach a main street with an elevated highway. Turn right without crossing that main street. Continue along the elevated highway for a while, and you will see Nitori (ニトリ) on your right. The destination is immediately adjacent to Nitori(ニトリ) on the right.

Figure 4: Example of collected of route information reference expressions (13-a)