Spatial Information Annotation Based on the Double Cross Model

Yoshiko Kawabata[1] **Mai Omura**[1] **Masayuki Asahara**[1] **Johane Takeuchi**[2]
[1]NINJAL, Japan
[2]Honda Research Institute Japan Co., Ltd.
{kawabata,mai-om,masayu-a}@ninjal.ac.jp
johane.takeuchi@jp.honda-ri.com

Abstract

Spatial ML and ISO-Space have been proposed as methods for describing spatial information appearing in language. Though they are effective for describing proper position information and absolute references (such as cardinal directions), they are not suitable for describing relative references (such as front, back, left, and right) that are inherent in dialogues as entities in space. Resolving the ambiguity of relative references cannot be done only with directed edges (ordered pairs), and it is necessary to maintain two or more directed edges as frames, including the orientation of the entities. The double cross model is used in the field of spatial logic as a method for resolving the ambiguity of location information. In this study, we attempted to represent relative references in dialogue using the double cross model and report our findings.

1 Introduction

When we want someone to take us somewhere, we express the destination not only in terms of latitude and longitude, but also in various words. In dialogues aimed at sharing location information, relative positions based on one's own position and orientation are expressed in words to exchange location information that both parties know. However, when expressing the relative positions of multiple landmarks or spatial entities through language, the simple directed edge structure (ordered pairs) may not be enough.

For example, to express "Please proceed to the right front with Tokyo Tower behind you" using only directed edges, it is necessary to maintain two edges: one expressing the relative position between the Tokyo Tower (landmark) and the speaker (spatial entity), and another expressing the speaker's (spatial entity) current position and direction of movement (orientation) with the speaker's orientation. **To accurately describe location information, it is essential to define the relative position of these two edges and maintain this information as a single location frame.**

In the field of spatial logic, the double cross model is a location information frame that uses two crosses to indicate direction, and expresses the orientation and the position of the third landmark or spatial entity in relation to the two crosses by placing the location or spatial entity at the centre of the two crosses.

This paper proposes to use the double cross model to represent language expressions in dialogues that involve sharing location information. Furthermore, this study demonstrates how to express direction and orientation, time and space measurements, and mereotopological relations.

2 Representation of Spatial Information Frames

In this section, we describe the main concepts of the spatial information frame based on Spatial ML (Mani et al., 2008) and ISO-space (Pustejovsky and Yocum, 2014; Pustejovsky, 2017). In our discussion, we also refer to papers related to Spatial Logic (Renz and Nebel, 2007).

2.1 Formalisation of Spatial Information

We annotate the following expressions:

- Landmarks: Places where specific location information is defined, such as latitude and longitude, or street addresses.
- Spatial Entities: Entity that is located in space. The speaker can also be considered one.
- Signals: Expressions indicating location information, direction information, distance information, etc.

Three types of reference expressions are considered for the formalisation of spatial information.

- Intrinsic: The inherent direction and position of a location.
- Absolute: Deictic reference based on a bird's-eye view – cardinal directions (north, west, east, south).
- Relative: Deictic reference based on the viewpoint of the entity (front, back, left, right). It is assumed that the referring entity has an orientation.

In general, there are 'intrinsic' information defined by latitude and longitude or address number , 'absolute' deictic references based on bird's-eye view and 'relative' deictic references based on the viewpoints of entities when resolving ambiguity in location information.

The project-based model (Figure 1) (Ligozat, 1998) represents the configuration and orientation of entities in a space or location as directions (projections) without a range centred around a subject. The project-based model is represented as directed edge structures (ordered pairs).

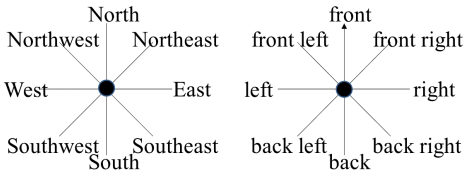

Figure 1: Project-based model (absolute, relative)

When language expressions include only 'absolute' deictic references based on the 'intrinsic' location information, the above directed graph model can be used to identify the location information. However, in practice, entities exist in space, and speakers and listeners use 'relative' deictic reference expressions from their respective viewpoints to exchange location information.

The double cross model allows the formalisation of location information that naturally represents the relative position of spatial entities from the viewpoint of the observer.

2.2 Formalisation of Topological Information

The preceding explanation of spatial information is primarily based on topological information. This is because the expressions of movement in several languages incorporate changes in topology.

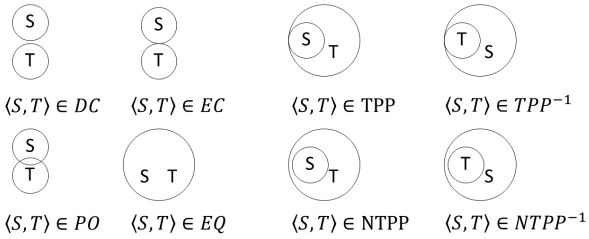

Figure 2: Topological information (RCC8)

The size of locations and entities in space must be considered when defining topological information (mereotopological relation). Based on the states of the whole-part relations and boundaries, the RCC8 (Region Connection Calculus Relations) (Randell et al., 1992) defines the following relations:

- `DC`: Disconnected
- `EC`: External connection
- `PO`: Partial overlap
- `EQ`: Equal
- `TPP`: Tangential proper part (source is contained inside target and touches its boundary)
- $\texttt{TPP}^{-1}$: Inverse of TPP (target is contained inside source and touches its boundary)
- `NTPP`: Non-tangential proper part (source is completely contained inside target without touching its boundary)
- $\texttt{NTPP}^{-1}$: Inverse of NTPP (target is completely contained inside source without touching its boundary)

We also use the RCC8-based annotation for the topological information. Below is an example of a description with the `type=topology` for the case of containment (`NTPP`). Here, `ID` is the identifier of mention in the text. `type=landmark` denotes that the expression type is landmark. `src` is the source of RCC8 relation, and `tgt` is the target of RCC8 relation.

[公園](ID:T412) のある [マンション街](ID:T413) の
[An apartment complex](ID:T413) with a park(ID:T412)

$$\left\{ \begin{bmatrix} \text{id} & \text{T412} \\ \text{type} & \text{landmark} \end{bmatrix}, \begin{bmatrix} \text{id} & \text{T413} \\ \text{type} & \text{landmark} \end{bmatrix}, \begin{bmatrix} \text{type} & \text{topology} \\ \text{src} & \text{T412} \\ \text{tgt} & \text{T413} \\ \text{rcc8} & \text{NTPP} \end{bmatrix} \right\}$$

If two mentions refer to exactly the same location, use `EQ`.

えっと [アニバーサリー](ID:T110)ってゆう [お店](ID:T111)
well, at [a store](ID:T111) named ['Anniversary'](ID:T110)

$$\left\{ \begin{bmatrix} \text{id} & \text{T110} \\ \text{type} & \text{landmark} \end{bmatrix}, \begin{bmatrix} \text{id} & \text{T111} \\ \text{type} & \text{landmark} \end{bmatrix}, \begin{bmatrix} \text{type} & \text{topology} \\ \text{src} & \text{T110} \\ \text{tgt} & \text{T111} \\ \text{rcc8} & \text{EQ} \end{bmatrix} \right\}$$

In this study, DC is not used for placement in the double cross model (default is DC). Additionally, TPP^{-1} $\langle S, T \rangle$ and NTPP^{-1} $\langle S, T \rangle$ are not used, as they are represented by TPP$\langle T, S \rangle$ and NTPP$\langle T, S \rangle$, respectively.

Regarding the identification of location information, the problem is that relative directional expressions cannot be adequately expressed due to the emphasis on whether they are adjacent or separated in the topological space. Both topology and annotations that take into account relative orientation are necessary.

3 Double Cross Model-based Annotation

3.1 Double Cross Model

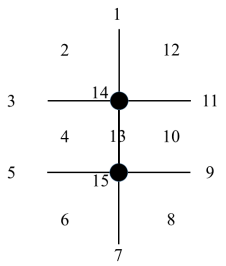

Figure 3: Double cross model

The double cross model (Figure 3) (Freksa, 1992) represents the relative position or orientation of the remaining element based on the configuration of the two preceding elements by placing two out of three elements in the space or location and denoting them as 14 and 15, as in the figure.

An example of a location description in a location frame is shown in Figure 4. In this study, we define the double cross model only in terms of a list of unordered hashes and do not insist on a specific notation for describing it.

It should be noted that when using absolute reference (e.g., cardinal directions), it is not necessary to use the double cross model. However, in this study, we consistently use a subset of the double cross model for description.

Here, we describe the information currently included in our annotation scheme:

- `id`: Identifier for the mention in the text. It is assigned in the upstream process as a location expression or a cue expression. Additional spatial entity representations are added and the dialog participants are denoted as `S0` (speaker) and `H0` (hearer).
- `type`: Type of the mention in the text. It can be one of the following:
  - `landmark`: A location expression.
  - `se`: A spatial entity expression, including the speaker and listener.
  - `signal`: A cue expression.
- `slot`: The slot number in the double cross model.
- `dir`: The direction that the entity is facing (specified by the slot number in the double cross model). This direction indicates where the entity is facing, not where it is moving.

In the example in Figure 4, ‘ Tokyo Tower ’ (T1), ‘ a tofu shop ’ (T3), and the hearer (H0) are placed in slots 15, 12, and 14 of the double cross model, respectively, indicating that the listener (H0) is facing direction 1.

3.2 Isomorphism

The original double cross model is based on three elements of origin, relatum, and reference. The origin is located at position 15, and the relatum is located at position 14. The reference can be located at all positions. Scivos and Nebel (2001) named 15 positions, as shown in the graph part of Figure 4 [1].

[1] {left, straight, right} × {forward, perpendicular, centre, line, back}.

[東京タワー](ID:T1) を背にして [右前方に](ID:T2) に [豆腐屋](ID:T3) があります
With [the Tokyo Tower](ID:T1) at your back, there is [a tofu shop](ID:T3) located [in the front right direction](ID:T2).

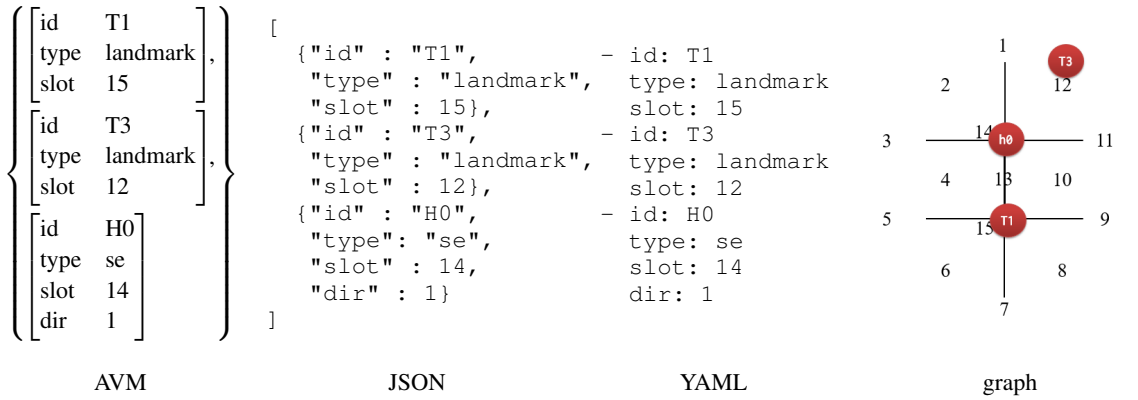

Figure 4: An example of spatial information frame by double cross model (AVM, JSON, YAML)

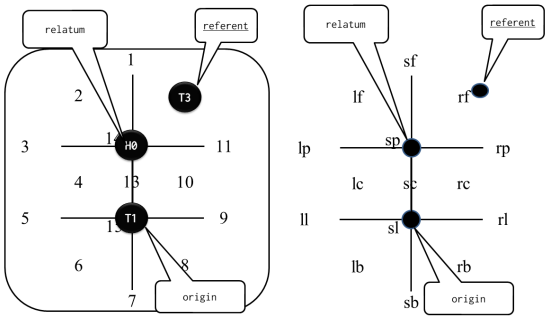

Figure 5: ⟨ origin, relatum, reference ⟩ in double cross model

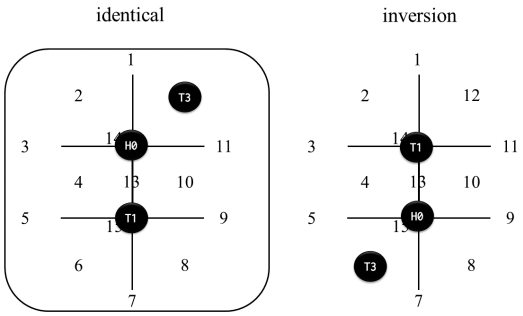

Figure 6: Isomorphism: inversion

operation	origin, relatum; referent
identical	a,b;c
inversion	b,a;c
homing	b,c;a
inverse homing	c,a;b
short cut	a,c;b
inverse short cut	c,a;b

Table 1: 3-permutation of origin, relatum and referent

In a three-element double cross model, there are cases where there are 3-permutation different expressions with the same structure as in Table 1. For example, the previous description has expressions such as 'inversion', 'homing', and 'short cut', as shown in Figures 6, 7, and 8 respectively (Zimmermann and Freksa, 1996).

An 'inversion' is obtained by swapping the positions of the 'origin' and 'relatum' (Figure 6). If the 'origin' is at position 15 and the 'relatum' is at position 14, this is equivalent to a 180-degree rotation.

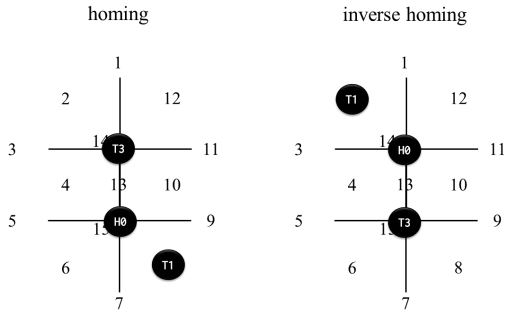

Figure 7: Isomorphism: homing, inverse homing

'Homing' refers to the act of mapping the 'relatum' to the 'origin', the 'referent' to the 'relatum', and the 'origin' to the 'referent' (Figure 7). 'Inverse homing' is the result of applying inversion to the homing mapping.

'Short cut' is a transformation that swaps the

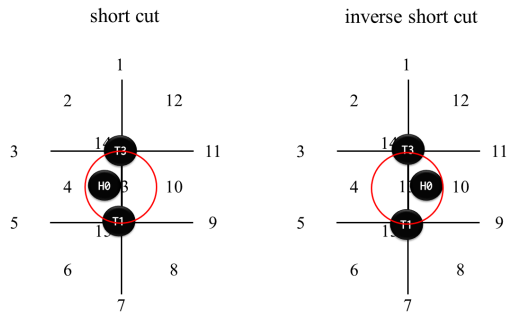

Figure 8: Isomorphism: short cut, inverse short cut

relatum and referent (Figure 8). In short cut, the circle with the diameter of line segment 14 and 15 is important. If the origin is placed at 15 and the relatum is placed at 14, performing short cut with referent at 2 or 12 will place the new referent inside the circle. If the referent is at 3 or 11, the new referent will be placed on the circle. If the referent is at 4 or 10, the new referent will be placed outside the circle.

Annotators are allowed to choose the representation that is easiest for them from the different expressions of the 3-permutation. When performing machine processing, it is necessary to implement the appropriate spatial logic unary operations (inversion, homing, short cut).

3.3 Distal

The double cross model indicates only relative direction and does not specify distance. There are also absolute distance expressions (Hernández et al., 1995; Clementini et al., 1997) and relative distance expressions (Isli and Moratz, 1999) for distance information.

The absolute distance expression specifies the distance or time required to reach between two points. In the following example, we define `type=distance`, describe T13 as `src`, T16 as `tgt`, and specify the distance between them as 20m in `absdist`.

[ジングウマエコウバン](ID:T13) から [メイジジングウ](ID:T14) のほうに [２０メートル](ID:T15)[行った歩道](ID:T16) で [待ってて](ID:T17)
Please [wait](ID:T17) on [the sidewalk](ID:T16) [20 meters](ID:T15) towards [Meiji Shrine](ID:T14) from [Jingumae Police Box](ID:T13).

$$\left\{ \begin{bmatrix} \text{id} & \text{T13} \\ \text{type} & \text{landmark} \\ \text{slot} & 15 \end{bmatrix}, \begin{bmatrix} \text{id} & \text{T14} \\ \text{type} & \text{landmark} \\ \text{slot} & 1 \end{bmatrix}, \begin{bmatrix} \text{id} & \text{T16} \\ \text{type} & \text{landmark} \\ \text{slot} & 14 \end{bmatrix}, \begin{bmatrix} \text{id} & \text{T15} \\ \text{type} & \text{signal} \end{bmatrix}, \begin{bmatrix} \text{type} & \text{distance} \\ \text{src} & \text{T13} \\ \text{tgt} & \text{T16} \\ \text{absdist} & 20\text{m} \end{bmatrix} \right\}$$

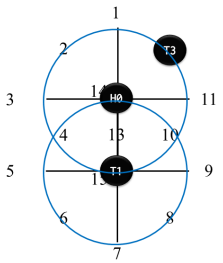

Figure 9: Relative distance in the double cross model

Relative distance is represented by the relative value of the distance between position 15 (origin) and position 14 (relatum). This distance is used as the radius to create circles with centres at positions 15 and 14, which are used as reference circles. The relative distance is thereafter expressed as a value relative to the distance between these circles (Figure 9). Though we had decided on an annotation method for relative distance, it did not appear in the data to be tagged.

4 Annotation Data

Here, we report our annotation of actual Japanese dialogues that involve sharing location information. The dialogue involved an experimental setting where the experimenter shared location information with a robot while viewing images from a 360-degree camera. The images used were of Omotesando and Toyosu in Tokyo. The robot role (experimental coordinator) engaged in dialogue with the participant of the experiment only using controlled speech templates until the location information was identified. The analysis fo-

Subject	Turns	EQ	TPP	NTPP	EC	PO	RCC-8 Total	Double-cross model	Complete
S01	146	106	14	6	29	1	156	0	15/20
S02	158	40	24	6	41	0	111	5	20/20
S03	159	34	1	8	23	1	67	7	18/20
S04	178	20	8	5	25	0	58	2	17/20
S05	144	35	4	1	13	0	53	0	19/20
S06	232	111	13	28	18	0	170	3	17/20
S07	114	28	19	25	25	0	97	1	19/20
S08	129	50	8	3	8	0	69	0	19/20
S09	183	26	5	2	6	0	39	3	18/20
S10	125	20	8	5	25	0	58	0	19/20
Total	1568	470	104	89	213	2	878	21	181/200

Table 2: Statistics: RCC-8 labels and double-cross model frames

cused on the experimental participant's dialogue with the robot role (experimental coordinator).

In this study, we collected voice data for 10 subject participants in the autumn of 2022. Each participant performed 20 sessions. We received ethical approval from the Institutional Review Board of HRI for the experiment. The voice data was transcribed, and mentions such as landmarks and spatial entities were labelled. Subsequently, topological information based on RCC-8 was annotated as co-reference information for the location data, and frames were described for the double cross model among the expressions corresponding to DC (disconnected) in RCC-8 that could be described by the double cross model.

Table 2 shows the statistics. The column of Turns shows the number of turns in robot dialogues, which is a unit of speech that starts with one speaker and ends when the next speaker begins. The columns of EQ, TPP, NTPP, EC, PO, and RCC-8 Total show the numbers of coreference annotation based on RCC-8. Note that EQ was assigned only to the nearest elements for each equivalence class. The column of double-cross model shows the number of frames by double-cross model. The column of Complete shows how many times each collaborator was able to share their location information out of 20 total sessions. Note that each session had a time limit of 210 seconds.

In the table, EQ represents normal coreference. Due to the tendency to repeat or paraphrase the same landmark or spatial entity when sharing location information, EQ appears frequently. TPP and NTPP are part-whole relationships. They are used when referring to a shop inside a building. They are distinguished by whether they face outside or not, with TPP indicating that they do and NTPP indicating that they do not. EC is primarily used to express things that are adjacent to each other. Though PO is used to represent partially overlapping objects, it is limited in its usage because there are few landmarks or spatial entities that partially overlap in urban areas.

The language expressions that should be described using the double cross model were only 21 utterances from 5 out of the 10 participants. It was found that experiment participants were not able to express relative positions appropriately in situations where they could not say 'north, south, east, or west,' and tended to rely only on building features or adjacency relationships to express relative positions.

Figure 10 shows the experiment time for 10 subject participants with 20 sessions each. The numbers below the bar graph indicate the number of sessions that required the double cross model. The white bars represent successful sessions, whereas the black bars represent failed sessions. The average time for the 17 sessions with the double cross model expression was 84.8 seconds (SD: 46.6), whereas the average time for the 183 sessions without the double cross model expression was 89.6 seconds (SD: 56.8). The t-test showed that the time for the sessions with the double cross model was significantly shorter. The percentage of incomplete sessions among the 17 sessions with the double cross model representation was 1 session (5.8%), whereas among the 183 sessions without the double cross model representation, there were 18 incomplete sessions (9.8%). This suggests that the double cross model repre-

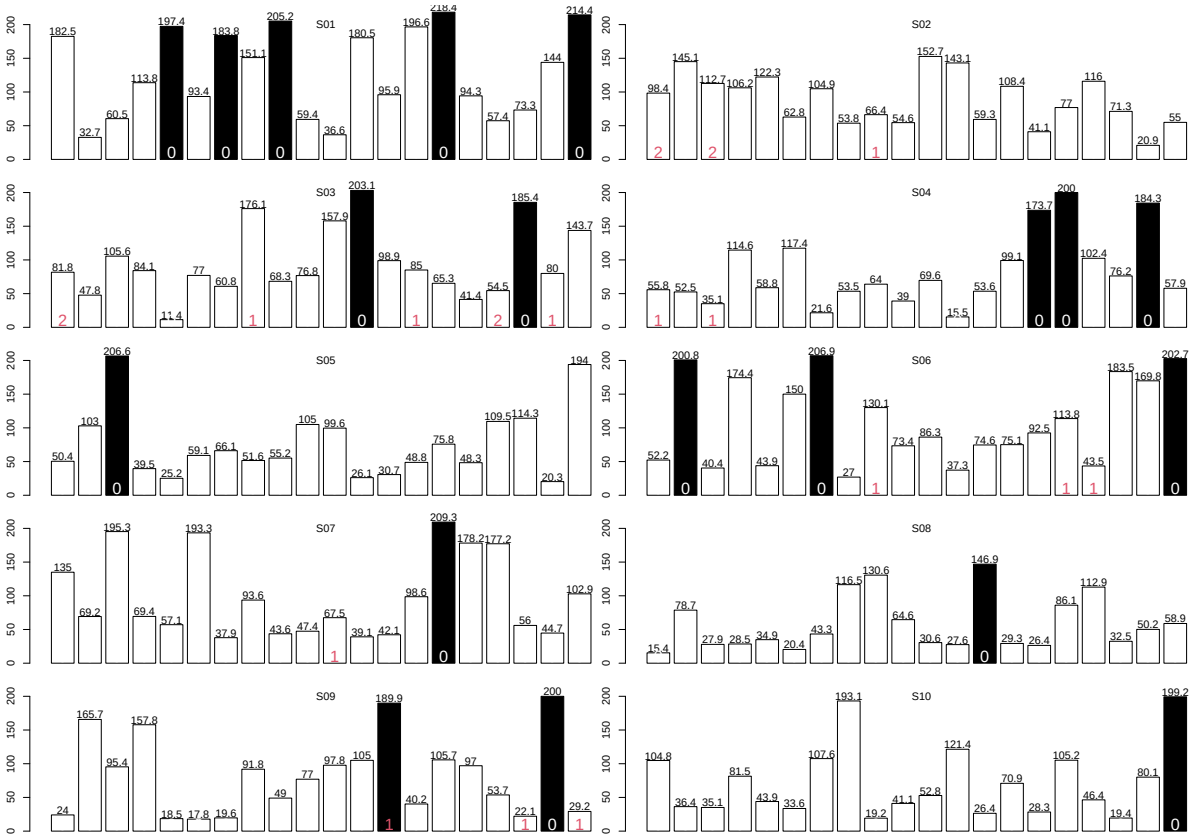

Figure 10: Results of the experiment time for 10 subject participants with 20 sessions each

sentation is effective for resolving location ambiguity.

In the world of spatial logic, language expressions that represent three-point relations, as described in the frame of the double cross model, are efficient. In this study, though we analysed the language expressions spoken by the participants, we believe that research on speech generation in location pointing tasks, which includes movement, is more important in practice. For example, research may focus on language generation models that express detailed movement instructions unambiguously, such as those for taxi dispatch or autonomous driving.

5 Conclusions

Through this study, we propose the use of the double cross model as a frame-based annotation method for spatial reference expressions. Traditional spatial reference expression annotations have mainly focused on expressions of same-reference, part-whole relationships, and adjacency relationships based on topological information. This study shows that for non-connected landmarks or spatial entities, three or more points are required to express direction, and suggest using frames to describe them. In the description of the frames, we also mentioned isomorphic relationships and suggested using those that are easier for annotators to tag. Furthermore, we proposed describing absolute and relative distances in the frame for those that are far apart.

We conducted an experiment using 360-degree camera images from Tokyo (Omotesando and Toyosu) to share location information, and collected speech data from 10 participants. We performed co-reference information annotation based on RCC-8 and relative position information annotation based on the double cross model for the speech data. From the annotation results, we found that language expressions based on the same reference, adjacency, and part-whole relationships were common, and the need for frame description using the double cross model was limited. We found that in sessions with relative spatial expressions that require the use of the double cross model, the achievement rate and time to achieve location information sharing were higher and shorter, respectively. Thus, relative position expressions are important in resolving location

ambiguity.

From the perspective of spatial logic, resolving the ambiguity of location information using only the same reference relation, adjacency relation, and part-whole relation is inefficient. Using the relative position information of three distant objects is more efficient for disambiguating location information than only using the same reference relation, adjacency relation, and part-whole relation of two objects. From this perspective, using relative position expressions is important in describing location information through language. As research on generative Artificial Intelligence has been thriving in recent years, it is important to be able to verbalise relative position information of three points based on still images, videos, and map information, especially in tasks such as autonomous driving. This study will serve as fundamental research for this purpose.

Acknowledgements

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

Eliseo Clementini, Paolino Di Felice, and Daniel Hernández. 1997. Qualitative representation of positional information. *Artificial intelligence*, 95(2):317–356.

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.

Daniel Hernández, Eliseo Clementini, and Paolino Di Felice. 1995. Qualitative distances. In *Spatial Information Theory A Theoretical Basis for GIS*, pages 45–57, Berlin, Heidelberg. Springer Berlin Heidelberg.

Amar Isli and Reinhard Moratz. 1999. Qualitative spatial representation and reasoning: algebraic models for relative position. Technical report.

G É Ligozat. 1998. Reasoning about cardinal directions. *Journal of Visual Languages & Computing*, 9(1):23–44.

Inderjeet Mani, Janet Hitzeman, Justin Richer, Dave Harris, Rob Quimby, and Ben Wellner. 2008. SpatialML: Annotation scheme, corpora, and tools. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco. European Language Resources Association (ELRA).

James Pustejovsky. 2017. *ISO-Space: Annotating Static and Dynamic Spatial Information*, pages 989–1024. Springer Netherlands, Dordrecht.

James Pustejovsky and Zachary Yocum. 2014. Image annotation with ISO-space: Distinguishing content from structure. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 426–431, Reykjavik, Iceland. European Language Resources Association (ELRA).

David A Randell, Zhan Cui, and Anthony G Cohn. 1992. A spatial logic based on regions and connection. In *Principles of Knowledge Representation and Reasoning: Proceedings of the 3rd International Conference*, pages 165–176.

Jochen Renz and Bernhard Nebel. 2007. Qualitative spatial reasoning using constraint calculi. In *Handbook of spatial logics*, pages 161–215. Springer.

Alexander Scivos and Bernhard Nebel. 2001. Double-crossing: Decidability and computational complexity of a qualitative calculus for navigation. In *Spatial Information Theory*, pages 431–446, Berlin, Heidelberg. Springer Berlin Heidelberg.

Kai Zimmermann and Christian Freksa. 1996. Qualitative spatial reasoning using orientation, distance, and path knowledge. *Applied intelligence*, 6(1):49–58.