An extension of ISO-Space for annotating object direction

Daiki Gotou **Hitoshi Nishikawa** **Takenobu Tokunaga**
Department of Computer Science
Graduate School of Information Science and Engineering
Tokyo Institute of Technology
{gotou.d.aa@m, hitoshi@c, take@c}.titech.ac.jp

Abstract

In this paper, we extend an existing annotation scheme ISO-Space for annotating necessary spatial information for the task placing an specified object at a specified location with a specified direction according to a natural language instruction. We call such task *the spatial placement problem*. Our extension particularly focuses on describing the object direction, when the object is placed on the 2D plane. We conducted an annotation experiment in which a corpus of 20 situated dialogues were annotated. The annotation result showed the number of newly introduced tags by our proposal is not negligible. We also implemented an analyser that automatically assigns the proposed tags to the corpus and evaluated its performance. The result showed that the performance for entity tags was quite high ranging from 0.68 to 0.99 in F-measure, but not the case for relation tags, i.e. less than 0.4 in F-measure.

1 Introduction

Understanding spatial relations in natural language dialogue is an important issue, particularly in situated dialogues (Kruijff et al., 2007; Kelleher and Costello, 2008; Coventry et al., 2009), as in the following interaction between a worker and a client in a moving setting.

client : Place the refrigerator next to the sink.

worker: Like this? (with an appropriate action)

client : Well, turn it to this side. (with an appropriate gesture)

Assuming a computer agent as a worker being asked to put things at specified places, the agent has to be able to interpret the client's instructions through identifying the target object to move, the location at which the target object should be placed and often the direction of the target object itself. We call this kind of task *the spatial placement problem*, namely the task placing an specified object at a specified location with a specified direction according to a natural language instruction. As a necessary first step to realising a computer agent that is capable of dealing with the spatial placement problem, the present paper proposes an annotation scheme to represent spatial relations by extending an existing scheme.

In order to represent spatial relations, Mani et al. (2008) proposed an annotation scheme that annotates spatial objects and relations between them. However, their scheme does not handle object direction. In the above example, the Mani's scheme annotates the spatial relation "next to" between the two objects "the refrigerator" and "the sink", but does not annotate the direction of the refrigerator specified by "to this side". As long as using their annotation scheme, the annotated corpus lacks the information of the object direction. When taking a machine learning approach with the annotated corpus to deal with the spatial placement problem, annotating object directions in the corpus is indispensable.

To tackle this problem, we extend an existing annotation scheme so that it can describe the spatial direction of objects in addition to the spatial relations between objects. Based on the proposed scheme, we annotate an existing dialogue corpus, and construct an analyser that extracts the spatial information necessary for solving the spatial placement problem. The effectiveness of the proposed scheme is evaluated through the annotation result and the performance of the analyser.

In what follows, we briefly survey previous studies that deal with spatial information in natural language processing (section 2), then describes *the spatial placement problem* in detail which is the main

Proceedings of the 12th Workshop on Asian Language Resources,
pages 1–9, Osaka, Japan, December 12 2016.

objective of the present study (section 3). The rest of the paper describes the proposed annotation scheme (section 4) and its evaluation through automatic tagging using the proposed scheme (section 5). Finally we conclude the paper and argue the future work in section 6.

2 Related work

The past studies related to our proposal can be categorised into three groups in terms of their focal issues: (1) studies on the annotation scheme to annotate spatial information in corpora, (2) studies on the corpus construction including spatial information, and (3) studies on systems that can manipulate various objects according to natural language instructions in virtual or real spaces.

SpatialML proposed by Mani et al. (2008), which was mentioned in the previous section, is an annotation scheme to annotate spatial information in text corpora. SpatialML focuses on capturing geographic relations such as the distance and the relative spatial relation between two entities. For example, given the phrase "a town some 50 miles south of Salzburg in the central Austrian Alps", SpatialML annotates "town", "Salzburg", and "the central Austrian Alps" with a geographic location tag, and "some 50 miles" and "south of" with the distance and spatial relation tags between the two locations. However, SpatialML has no way to represent the direction of an object itself, i.e. which direction the object faces to.

Pustejovsky et al. (2011) introduced annotating events that cause changes in spatial relations into their annotation scheme ISO-Space. One of the significant characteristics of ISO-Space is describing changes in spatial relations according to temporal progression. For instance, changes in the object location through a motion event are annotated with the event path tag. In the sentence "The [depression $_{se1}$] was [moving $_{m1}$] westward at about 17mph (28 kph) and was expected to continue that motion for the next day or two.", the event path tag "EventPath($ep1$, source=$m1$, direction=WEST, moving_object=$se1$)" will be annotated in terms of a motion event $m1$ and a moving object $se1$ that are also annotated in the sentence. ISO-Space, however, does not have a tag for representing the direction of an object itself neither.

Since ISO-Space has an advantage over SpatialML that it can represent events and changes in spatial relations, we extend the ISO-Space scheme by introducing tags that describe object intrinsic direction, namely the direction that the object faces to. This kind of tags play an important role in the spatial placement problem as we saw in the previous section.

There have been several attempts of constructing corpora related to the spatial placement problem. The REX corpus (Tokunaga et al., 2012) and the PentoRef corpus (Zarrieß et al., 2016) are the examples of this sort. Both corpora were collected through situated dialogues in which dialogue participants jointly solved geometric puzzles such as Tangram and Pentomino. The main goal of the dialogues is placing puzzle pieces in the right places, thus, these tasks are the typical spatial placement problem.

These corpora come with the visual information that is updated during the course of dialogues, thus they include the spatial information of the objects. However, the transcribed utterances were not annotated with spatial information corresponding to the object direction. To our knowledge, there is no corpus that is linguistically annotated with spatial information including both object location and direction. Our attempt compensates for these missing information in the corpora for the spatial placement problem.

Winograd's SHRDLU is the first and seminal working system that is capable of dealing with the spatial placement problem (Winograd, 1972). SHRDLU could understand natural language questions and instructions on a virtual block world, and could manipulate various kinds of blocks to change the state of the block world. More recently, Tellex et al. (2011) realised a SHRDLU-like system in the real environment. They proposed Generalised Grounding Graphs to infer corresponding plans to linguistic instructions. They collected possible expressions of the instruction through crowdsourcing to construct a corpus which is used to train the inference model. However, SHRDLU nor the Tellex's system do not care about the direction of manipulated objects. As we saw in our moving example, understanding the object direction is crucial in some applications, which is the motivation of this study.

3 Spatial placement problem

The spatial placement problem is a task to place an specified object at a specified location with a specified direction as instructed in natural language. In this paper, we assume that there are multiple objects on the

2D plane, and *the worker* is asked to place the objects at specified locations according to instructions by *the instructor*. Therefore, the worker needs to infer the location to place the object, and it also needs to infer the direction that the object faces to. The spatial placement problem can be broken down into the following three steps.

1. Identifying the object
   The worker needs to identify the object to be manipulated in the instructional utterance. This task is regarded as the reference resolution in a multimodal setting (Iida et al., 2010; Prasov and Chai, 2010).
2. Deciding the specified location
   The worker needs to decide the location where the target object should be placed. This is also considered as resolving referring expression, but the referent is a location instead of an object. The spatial referring expressions include expressions such as "next to a triangle", "about one meter right of the bed", and "the centre of the room". Those expressions often specify the location in terms of the spatial relations between the target object and the other objects, often called *the reference object or landmark* (Coventry and Garrod, 2004).
3. Deciding the specified direction
   After identifying the target object and its location, the worker needs to decide the direction of the object. For instance, given the instruction "Turn the desk left.", the worker needs to decide the direction that the desk faces to. We assume that objects have their own intrinsic coordinate system (Levinson, 2003), namely they have a front side of their own. Thus, the worker needs to infer the object's front side and place the object so that its front side faces to the appropriate direction.

4 Extending annotation scheme

This section exemplifies annotation with ISO-Space and our extension using the following sequence of instructions.

(1) Move the small triangle under the square.

(2) Rotate it so that the right angle comes down.

4.1 Annotating the objects

Following the ISO-Space scheme, we annotate physical objects with the Spatial Entity tag. Thus expressions referring to objects in the current example are annotated as Spatial Entity as shown in Figure 1. The Spacial Entity tag can have attributes for describing information represented by modifiers in the referring expression.

Move [the small triangle $_{se1}$] under [the square $_{se2}$].
Rotate [it $_{se3}$] so that the right angle comes down.

Spatial Entity($se1$, type=TRIANGLE, mod=SMALL)
Spatial Entity($se2$, type=SQUARE)
Spatial Entity($se3$)

Figure 1: Annotation example (objects)

4.2 Annotating the specified location

The location is annotated by the Location tag in ISO-Space as in Figure 2. The reference to the location "under the triangle", where the object "the small triangle" should be placed, is annotated as Location. In instruction (1), this location is specified in terms of the relative spatial relation "under" to the reference object "the square". The expression "under" is annotated as Spatial Signal as it implies the spatial relation between these two objects. The attribute of the Spatial Signal $sig1$ indicates its type of spatial relation,

namely DIRECTIONAL in this case. Note that the DIRECTIONAL type stands for the spatial relation between two objects and it does not represent the direction of the object itself that we mainly concern in this paper. The Qualitative Spatial Link $qsl1$ represents the relation among the spatial relation (relType) with its surface string (trigger), the reference object (figure) and the location (ground).

Move [the small triangle $_{se1}$] [[under $_{sig1}$] [the square $_{se2}$] $_{loc1}$].
Rotate [it $_{se3}$] so that the right angle comes down.

Spatial Entity($se1$, type=TRIANGLE, mod=SMALL)
Spatial Entity($se2$, type=SQUARE)
Spatial Entity($se3$)
Location($loc1$)
Spatial Signal($sig1$, type=DIRECTIONAL)
Qualitative Spatial Link ($qsl1$, relType=LOWER, trigger=$sig1$, figure=$se2$, ground=$loc1$)

Figure 2: Annotation example (locations)

4.3 Annotating the specified direction

Until this moment, we annotated the current example solely with the ISO-Space scheme. To annotate the direction of objects, e.g. "it (the small triangle)" in our current example, we introduce the following new tags: *Direction Signal*, *Direction Link*, *Part* and *Part Link*, which are underlined in the current annotated example in Figure 3.

Direction Signal and Direction Link are analogous to Location Signal and Qualitative Spatial Link in ISO-Space. Expressions implying the object direction such as "so that the right angle comes down" are annotated with the Direction Signal tag, which a counterpart of the Location Signal tag for describing location. As the location is often described in terms of some spatial relation to the reference object, the object direction is often described by mentioning a part of the object. As a device for interrelating the Direction Signal tag with the reference part of the object, we introduce the Part tag and Part Link tag. The former is annotated to expressions describing a part of the object (e.g. "the right angle") with various attributes, and the latter relates the reference part to the entire object. These tags enable the Direction Link tag to describe the object direction by specifying the spatial direction of the reference part of the object. The Direction Signal tag has an attribute dirType that indicates *the frame of reference* (Levinson, 2003); The ABSOLUTE frame adopts an absolute coordinate system such as east-west-north-south, while the RELATIVE frame uses a reference object to indicate a direction.

ISO-Space uses the Motion tag to describe object movement that causes the change of object location. When an object rotates by itself, its location could remain the same even it changes its direction. Therefore we allow the Motion tag to describe movements that cause the object direction as well as the object location. The Move Link tag relates the movement and its related elements.

4.4 Annotation experiment

To argue the efficacy of our proposal, we have conducted an annotation exercise using an existing dialogue corpus. The annotation target is the REX corpus, a Japanese dialogue corpus in which two participants jointly solve the Tangram puzzle on the computer simulator (Tokunaga et al., 2012). The goal of the puzzle is arranging the seven pieces into a given goal shape. Both participants share the same working area where the puzzle pieces are arranged, but play different roles. One was given the goal shape but not a mouse to manipulate the pieces, while the other was given a mouse but not the goal shape. Due to such asymmetric task setting, the participant with the goal shape mostly played as an instructor and the other played as a worker. Thus this task can be considered as a typical spatial placement problem. The following is an excerpt of a dialogue[1] and Figure 4 shows a screenshot of the Tangram puzzle simulator in which the goal shape "bowl" is shown on the left.

[1]Although the corpus is a Japanese corpus, we use examples of its English translation in the rest of the paper for the convenience of readers who do not understand Japanese.

[Move $_{m1}$] [the small triangle $_{se1}$] [[under $_{sig1}$] [the square $_{se2}$] $_{loc1}$].
[Rotate $_{m2}$] [it $_{se3}$] [so that [the right angle $_{p1}$] comes down $_{ds1}$].

Spatial Entity($se1$, type=TRIANGLE, mod=SMALL)
Spatial Entity($se2$, type=SQUARE)
Spatial Entity($se3$)
Location($loc1$)
Spatial Signal($sig1$, type=DIRECTIONAL)
Qualitative Spatial Link ($qsl1$, relType=LOWER, trigger=$sig1$, figure=$se2$, ground=$loc1$)
Part($p1$, partType=APEX, mod=RIGHT_ANGLE)
Part Link($pl1$, trigger=$p1$, source=$se3$)
Direction Signal($ds1$, dirType=ABSOLUTE, direction=LOWER)
Direction Link($dl1$, trigger=$ds1$, source=$p1$)
Motion($m1$, motionClass=MOVE)
Motion($m2$, motionClass=ROTATE)
Move Link($ml1$, motion=$m1$, object=$se1$, goal=$loc1$)
Move Link($ml2$, motion=$m2$, object=$se3$, dirSignal=$ds1$)

* Our proposal elements are underlined.

Figure 3: Annotation example (directions)

instructor: *sore wo hidari ni suraido sasete hamete kudasai.*
(Slide it leftward and fit it (to them).)
worker : *hai.*
(I see.)
instructor: *de, heikousihenkei wo 45 do kaiten sasete.*
(Then, rotate the parallelogram by 45 degrees.)

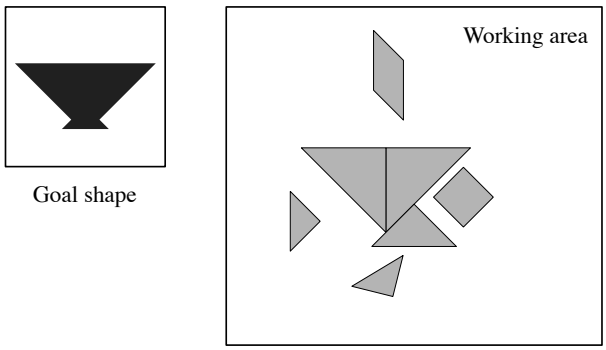

Figure 4: Screenshot of the Tangram puzzle

In this annotation experiment, we annotated the corpus with tags in Table 1 in which the underlined elements are newly introduced in our proposal. Although the ISO-Space scheme provides more than these tags, we used a minimum tag set necessary for describing the location and direction of objects for solving the spatial placement problem.

We annotated 20 dialogues with the tags listed in Table 1. The total number of utterances by the instructors was 2,020 including 360 instructional utterances. Among these 360 instructions, 60 (16.7%) of them mentioned the object direction. Table 2 shows the distribution of annotated tags in number. The table shows the number of the Directional Signal tag is comparable to that of the Spatial Signal tag, which is used for indicating spatial relations. According to this preliminary investigation, information of the object direction is not negligible in the spatial placement problem.

5 Evaluation

In order to evaluate feasibility of automatic tagging with our proposal, we implemented a system that assigns the tags shown in Table 1. Thus the goal of the system is assigning the tags to given utterances as shown in Figure 3. Given an instructional utterance, the task of the system is twofold:

entity tag	description
Spatial Entity	an entity that is not inherently a location, but one which is identified as participating in a spatial relation
Location	an inherently grounded spatial entity
Motion	an inherently spatial event, involving a change of location and direction of an object
Part	a reference to a part of an object

relation tag	description
Qualitative Spatial Link	the spatial relationship between two spatial objects
Spatial Signal	an expression representing a spatial relation
Move Link	the relation between an object changing its location or direction and its goal location or direction
Part Link	the relation between an object and its part
Direction Link	the relation between a Direction Signal instance and a reference part of an object
Direction Signal	an expression representing an object direction

* Our proposal elements are underlined.

Table 1: Annotated tags

entity tag	number	relation tag	number
Spatial Entity	357	Qualitative Spatial Link	270
Location	112	Move Link	1,420
Motion	315	Part Link	126
Part	66	Direction Link	101
		Spatial Signal	81
		Direction Signal	62
Total	850		2,060

* Our proposal elements are underlined.

Table 2: Distribution of annotated tags

1. identifying the spans to be assigned the entity tags in Table 1, and the Spatial Signal and Direction Signal tags, and
2. identifying the rest of the relations in Table 1 by linking the spans identified in step 1.

In the following subsections, each of the steps is described in more detail.

5.1 Identifying spans for entity tags

Considering the span identification for a certain tag as a sequential labelling problem, we adopt the IOB2 model (Tjong et al., 1999) to identify the tagged span. We employed the CRF++[2] implementation to conduct sequential labelling. We prepared the labelling program for each tag and ran them in parallel. Thus each tag has its own I-O-B labels. Table 3 shows correct labelling for the instruction "Rotate it so that the right angle comes down.".

tag	Rotate	it	so	that	the	right	angle	comes	down
Spatial Entity	O	B	O	O	O	O	O	O	O
Location	O	O	O	O	O	O	O	O	O
Motion	B	O	O	O	O	O	O	O	O
Part	O	O	O	O	B	I	I	O	O
Spatial Signal	O	O	O	O	O	O	O	O	O
Direction Signal	O	O	B	I	I	I	I	I	I

Table 3: Example of entity tag labelling

Given an instructional utterance for tagging, the system firstly applies the Japanese morphological analyser MeCab[3] to the input utterance to divide it into a sequence of words, then further applies the sequential labelling to the word sequence. As features for the labelling, the surface string, the part of speech, the script type (alphabet vs. digits) of the target word and its neighbouring two words in both

[2]https://taku910.github.io/crfpp/
[3]http://taku910.github.io/mecab/

sides, and the already assigned tags of the previous two words are used. Figure 5 depicts a set of features used for sequential labelling, in which the enclosed information is used for labelling the i-th word "the".

feature \ input	Rotate	it	so	that	the	right	angle	comes	down
index			$i-2$	$i-1$	i	$i+1$	$i+2$		
surface	Rotate	it	so	that	the	right	angle	comes	down
POS	verb	pron	conj	conj	det	adj	noun	verb	adv
script	alph	alph	alph	alph	alph	alph	alph	alph	alph
Spatial Entity	O	B	O	O					
Location	O	O	O	O					
Motion	B	O	O	O					
Part	O	O	O	O					
Spatial Signal	O	O	O	O					
Direction Signal	O	O	B	I					

Figure 5: Features for labelling "the"

5.2 Identifying relations

To decide the relation between tagged spans, we first constructed every pair from the set of tagged spans identified in the previous step, then we classified them into one of the relation tags listed in Table 1 except for Spatial Signal and Direction Signal since they have been already identified as the spans in the previous step. As we can see in Figure 3, the Qualitative Spatial Link and Move Link represent a trinary relation. The trinary relation is represented by two binary relations. For instance, the Move Link $ml2$ relates three spans $m1$ ("Rotate"), $se3$ ("it") and $ds1$ ("so that the right angle comes down") in Figure 3. We identify two Move Link relations between $m1$ and $se3$, and that between $m1$ and $ds1$ for this trinary relation.

Each pair is represented in terms of the features shown in Table 4 and is used for training the classifier implemented with LinearSVC[4].

feature	description
tag pair	a pair of tags assigned to two spans
distance	distance between two spans in Japanese characters
utterance length	length of the utterance in Japanese characters
number of spans	total number of spans in the utterance
pos	a quadruple of parts of speech of two adjacent word of each span
case	a pair of the case markers following each span

Table 4: Features for relation identification

5.3 Results and discussion

For both two subtasks, entity tag labelling and relation identification, we conducted the 10-fold cross validation using the corpus described in 4.4.

tag	Precision	Recall	F-measure	Total
SE	0.91	0.83	0.87	357
Motion	1.00	0.99	0.99	351
Location	0.91	0.76	0.83	112
Part	0.98	0.79	0.87	66
Spatial Signal	0.80	0.59	0.68	81
Direction Signal	0.98	1.00	0.99	62

Table 5: Result of labelling

Table 5 shows that the accuracy of the entity tag labelling is quite high. This is probably due to a very limited domain of the corpus. We should apply the proposal to corpora of broader and more complicated domains to confirm the current result.

There are two main reasons of the labelling errors: the insufficient annotation guidelines and the preprocessing errors. The number of the former is 52 and that of the latter is 138.

[4]http://scikit-learn.org/stable/

According to our annotation guidelines used in 4.4, only entities involved in some spatial relations were annotated. However, the analyser extracted all entities even though they had no relations with other entities. Those extracted entities were considered as the false positive instances, thus having caused errors. We should have annotated all entities regardless whether they had relations with others or not.

The errors due to the preprocessing are mainly caused by the erroneous segmentation of the Japanese morphological analyser. In this experiment, the utterances were automatically divided into a sequence of words by the Japanese morphological analyser, and thus the segmentation error causes a serious damage to the labelling phase.

Currently we apply the sequential labelling for each tag in parallel and independently. That means each labelling program does not utilise the previous two labels of other tag type. In Figure 5, for instance, when deciding the label of the Spatial Entity tag for "the", the system uses two previous Spatial Entity labels O and O but does not use labels of other tags. The performance of entity tag labelling could be further improved if the labels of other tags were also used.

Table 6 shows the result of relation identification. We calculated precision, recall and F-measure for two cases: "Gold" (using manually annotated entity tag labels) and "Estimated" (using the results of automatic labelling). The row "No Link" in the table denotes that there is no relation between the given pair of spans. Due to its dominant number of instances, the classifier might be over-tuned to the No Link class. In contrast to entity tag labelling, there is much room for improvement in relation identification. Such low performance might be attributed to the insufficient size of the corpus we used in the experiment. We need further experiments with a larger corpus.

	Gold			Estimated			
relation	P	R	F	P	R	F	Total
Qualitative Spatial Link	0.39	0.17	0.23	0.42	0.17	0.24	270
Move Link	0.41	0.39	0.40	0.43	0.27	0.33	1,420
Part Link	0.41	0.10	0.17	0.36	0.09	0.14	126
Direction Link	0.48	0.31	0.38	0.47	0.22	0.30	101
No Link	0.67	0.74	0.71	0.67	0.83	0.74	3,463

Table 6: Result of relation identification

6 Conclusion and Future Work

In this paper, we defined *the spatial placement problem* as a task placing an specified object at a specified location with a specified direction according to a natural language instruction. As a first step for tackling this problem, we proposed an extension of the existing annotation scheme ISO-Space for annotating the object direction in text corpora. To evaluate the efficacy of the proposed annotation scheme, we conducted an annotation experiment in which a corpus of 20 situated dialogues for solving the Tangram puzzle was annotated. The annotation result showed the number of newly introduced tags by our proposal is not negligible.

We implemented an analyser that automatically assigns the proposed tags to the corpus and evaluated its performance. The results showed that the performance of entity tag labelling was quite high but not the case for relation identification. The good performance of the entity tag labelling might be due to a very limited domain of the corpus. We need to conduct experiments with the corpora of broader and more complicated domains to confirm the current result. In contrast to entity tag labelling, the performance of relation identification was very poor, less than 0.4 in F-measure. This might be due to insufficient training data and over-tuning to the negative instances. We need to continue the evaluation with larger corpora of more complicated domains.

In the real setting of the spatial placement problem, the instructor uses other modalities that language, such as gesture and visuals in the instruction. The REX corpus that we used in the experiments has participant eye gaze and mouse operations on top of the transcribed utterances. Investigating the effectiveness of these kind of multimodal information in the spatial placement problem is one of the future research directions.

References

Kenny R. Coventry and Simon C. Garrod. 2004. *Saying, Seeing, and Acting*. Psychology Press.

Kenny R. Coventry, Thora Tenbrink, and John Bateman. 2009. Spatial language and dialogue: Navigating the domain. In Kenny R. Coventry, Thora Tenbrink, and John Bateman, editors, *Spatial Language and Dialogue*, pages 1–7. Oxford University Press.

Ryu Iida, Shumpei Kobayashi, and Takenobu Tokunaga. 2010. Incorporating extra-linguistic information into reference resolution in collaborative task dialogue. In *Proceedings of 48th Annual Meeting of the Association for Computational Linguistics*, pages 1259–1267.

John D. Kelleher and Fintan J. Costello. 2008. Applying computational models of spatial prepositions to visually situated dialog. *Computational Linguistics*, 35(2):271–307.

Geert-Jan M. Kruijff, Hendrik Zender, Patric Jensfelt, and Henrik I. Christensen. 2007. Situated dialogue and spatial organization: what, where and why? *International Journal of Advanced Robotic Systems*, 4(1):125–138.

Stephen C. Levinson. 2003. *Space in Language and Cognition*. Cambridge University Press.

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 2008)*, pages 410–415.

Zahar Prasov and Joyce Y. Chai. 2010. Fusing eye gaze with speech recognition hypotheses to resolve exophoric references in situated dialogue fusing eye gaze with speech recognition hypotheses to resolve exophoric references in situated dialogue. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 471–481.

James Pustejovsky, Jessica L. Moszkowicz, and Marc Verhagen. 2011. Using ISO-Space for annotating spatial information. In *Proceedings of the International Conference on Spatial Information Theory*.

Stefanie Tellex, Thomas Kollar, Steven Dickerson, Matthew R. Walter, Ashis Gopal Banerjee, Seth Teller, and Nicholas Roy. 2011. Understanding natural language commands for robotic navigation and mobile manipulation. In *Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence*, pages 1507–1514.

Erik F. Tjong, Kim Sang, and Jorn Veenstra. 1999. Representing text chunks. In *Proceedings of 9th Conference of the European Chapter of the Association for Computational Linguistics (EACL 1999)*, pages 173–179.

Takenobu Tokunaga, Ryu Iida, Asuka Terai, and Naoko Kuriyama. 2012. The REX corpora: A collection of multimodal corpora of referring expressions in collaborative problem solving dialogues. In *Proceedings of the Eigth International Conference on Language Resources and Evaluation (LREC 2012)*, pages 422–429.

Terry Winograd. 1972. Understanding natural language. *Cognitive Psychology*, 3(1):1–191.

Sina Zarrieß, Julian Hough, Casey Kennington, Ramesh Manuvinakurike, David DeVault, Raquel Fernandez, and David Schlangen. 2016. PentoRef: A corpus of spoken references in task-oriented dialogues. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, pages 125–131.