Where is my Lily?! An Investigation into Spatial Language Implicature

Lena Hong (lenahong@stanford.edu)
Stanford University

Jesik Min (jesikmin@stanford.edu)
Stanford University

Mathieu Rolfo (senorsquirrel@cs.stanford.edu)
Stanford University

Nathan James Tindall (ntindall@stanford.edu)
Stanford University

Abstract

We present a corpus analysis of spatial language use, examining the ways that humans reason about the relation of objects in space. We modify a multiplayer web experiment framework in order to yield a corpus of semantic and referential data in a two-player spatial language game. The data was analyzed using natural language processing in order to determine the most common prepositions used. A qualitative analysis of the data was also performed. A rational speech act (RSA) model was successfully deployed that emulates human behavior for a subset of the data observed: "just inside" and "inside."

Keywords: spatial language; spatial implicature; pragmatics; RSA model; multi-player web experiment

Background

Recently, there has been some effort to construct quantitative models of a pragmatic approach to spatial language. It allows coarse fixed meanings such as spatial language to gain context-specific refinements (Grice, Cole, & Morgan, 1975; Horn, 1984).The rational speech act (RSA) framework offers a way to make useful quantitative predictions about pragmatic reasoning using Bayesian inference (Frank & Goodman, 2012). The RSA framework has been shown to successfully model linguistic behavior for many conversational implicatures as exhibited by both adults and children (Bergen, Levy, & Goodman, 2014; Kao, Wu, Bergen, & Goodman, 2014; Potts, Lassiter, Levy, & Frank, 2015). Although RSA models have been successful in describing semantic inferences about language in context, previous work on the pragmatics of spatial knowledge was fairly limited in its scope, only exploring two spatial terms "in and "near" (Ullman, Xu, & Goodman, 2016).

Contribution

Given that spatial reasoning has a far larger number of constructions than have previously been studied, we were interested in exploring the ways in which it was used. We implemented a paradigm to investigate spatial language understanding on an unrestricted lexicon, relying on a multi-player web experiment framework allowing users to choose their own linguistic formulation of the spatial relationship between objects. We performed qualitative analysis of the data produced by the experiment, and produce a novel RSA model that emulates a subset of the data. In particular, we studied the utterances which contain "inside" and "just inside" to observe in which context people use "just" and how the RSA model captures such difference.

Methodology

Multi-player Web Experiment Framework

We have extended an existing paradigm for conducting real time multi-agent experiments on the web in order to gather data (Hawkins, 2015). The experiments are conducted using the Mechanical Turk Platform, where anonymous participants are paid to complete labor or research tasks. Once recruited, the participants enter a waiting room where they are paired with another participant with whom they will complete the entire experiment.

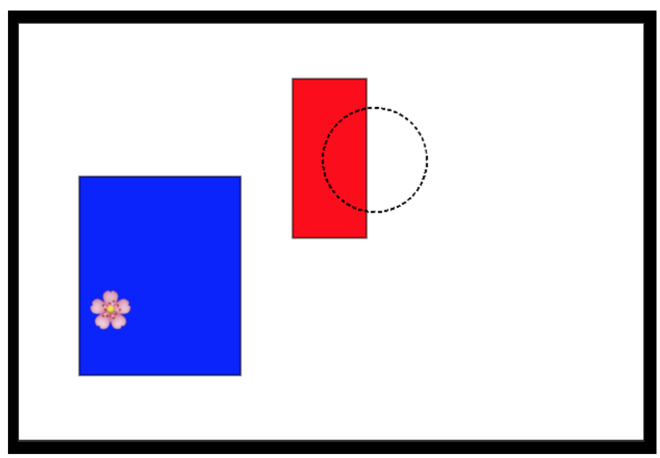

Figure 1: An example trial, as it appears to the speaker

One participant is assigned the speaker role while the other is assigned the listener role: the roles do not change during the experiment after they have been assigned. The interface for the experiment features a message box through which the participants can have bidirectional communication with each other. Additionally, a "map" of a city is displayed, see 1. The maps were designed to represent a red sector, a blue sector, and a victory plaza, similar to stimuli used for previous research on spatial language (Ullman et al., 2016). The speaker also is shown a pink lily (pictured inside the blue sector in 1. The listener is not shown the lily.

The speaker is instructed to “Send messages to tell the listener where the lily is” while the listener is instructed to “Click as closely as possible to the location of the lily on the map.” The listener is only able to click after the speaker has sent a message. After clicking, the listener is shown the true location of the lily, and the speaker is shown the location where the listener guessed. This gives both the players feedback about the other’s actions.

After performing some test trials and observing the data collected, we noted that many participants were being pedantically specific in their specifications and were unhappy with the length and lack of compensation for the task. To address these points, we gave the speaker a “target” for the first three trials by indicating a region around the lily where they must make the listener click, attempting to depress the tendency to be extremely specific. The target was removed after the first three trials in order to prevent the strategy of picking the easiest-to-describe spot within the target radius even if it is a technically incorrect description of the flowers location.

Upon clicking, the current state of the map, the location of the plaza, the lily, each of the sectors, the listener’s guess, and the messages that were sent are stored. The feedback information is displayed to the participants as described above, and after a delay the next stimulus is displayed to the speaker and the listener, which also clears the chat history.

Each pair was assigned 30 trials in this manner. The maps used for the trials were sampled from a set of 50 maps that were fixed across the entire experiment, and presented in a random order for each pair. The maps were randomly generated and the lily given an even distribution over the map. Some hand pruning was performed in order to favor lily locations that were near other objects, since spatial language is more relevant in the scenario when there are multiple competing relations between the lily and the other objects on the map. Since the areas do not always take up a large portion of the map as they vary in size, the majority of lily locations are somewhat near a sector or the plaza, though a few trials leave the lily “out in the open.”

Experiment

Mechanical Turk Platform

We recruited participants through the Mechanical Turk Platform. Due to time and budget constraint, we ran 9 batches and collected the data of 241 rounds. There were 9 unique listeners and 9 unique speakers.
See the collected data at `https://git.io/v19vT`.

Data Processing Pipeline

After collecting data by running the multi-player web experiment using the Mechanical Turk Platform, we processed the collected data to retrieve some useful information. We first implemented a part of speech tagger (or POS Tagger) in order to extract the frequency of all the prepositions from the raw data set. We also implemented the Object Parser that extracts pairs of objects that are associated with prepositions in a given sentence. Raw data, which was originally in CSV or json format, was fed into our NLP pipeline and went through our POS Tagger and Object Parser which are based on Stanford Core NLP and Python NLTK. The output was stored in json or simple text format. See Figure 2.

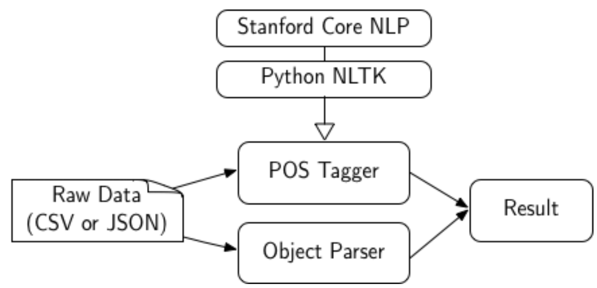

Figure 2: NLP pipeline flowchart

1) POS Tagger The POS Tagger uses Python NLTK to tag part of speech of each word in a given sentence. In detail, we use nltk.word_tokenize(sentence) and nltk.pos_tag(wordList) function to tokenize and tag part of speech respectively.
The POS Tagger not only extracts frequencies of every preposition appearing in the data set but also shows ten most frequent prepositions. Moreover, as shown in Figure 4, the POS Tagger generates a word bubble that visualizes frequency of each preposition as a set of bubbles such that the radius of each word bubble is proportional to the frequency of the corresponding word (or preposition in this case).

2) Object Parser The Object Parser matches each preposition with a pair of objects in a given sentence. For example, if we encounter a sentence, “Lily is inside the red rectangle,” we want to parse the relationship, “(Lily, inside, rectangle).” In addition to this prepositional relationship, as we observe many color modifiers in our data as well, the Object Parser also matches a color modifier with the object modified by the color in a given sentence. For example, we want to extract “(red, rectangle)” from the sentence above. See Figure 3.

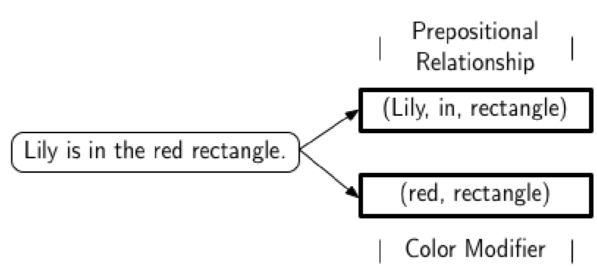

Figure 3: Object Parser

The code of the entire data processing pipeline can be found here:
`https://github.com/jessemin/POS-Object-Tagging`

Empirical Data Analysis

After processing the collected data using the POS Tagger and Object Tagger, we could view the prepositions people used during the experiment in descending order. In Figure 4, we can see that the preposition “of” was used most frequently with 267 occurrences. Although the preposition “of” does not have any direct spatial implication, it was reported as the most frequent one because “of” is used with many other prepositions as in the cases such as “inside of.” The next most frequent spatial preposition used was “in,” but similarly it was primarily used in conjunction with other prepositions, e.g. “in between.” The most common bigrams are visualized in figure 5 (the most common bigram was “of the,” but it was excluded to prevent it from distorting the data. We also observed that the frequency of “near”, another spatial phrase covered in Ullman et al., 2016, is only 4. In our data set, the next most frequent preposition with meaningful spatial implication was “inside” with 47 occurrences in total (see Figure 4). Thus, we focused our analysis on “inside” rather than “in.”

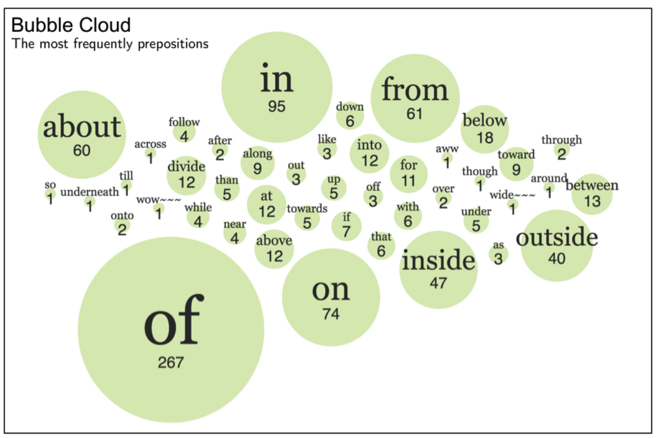

Figure 4: Word bubble visualization of prepositions in data

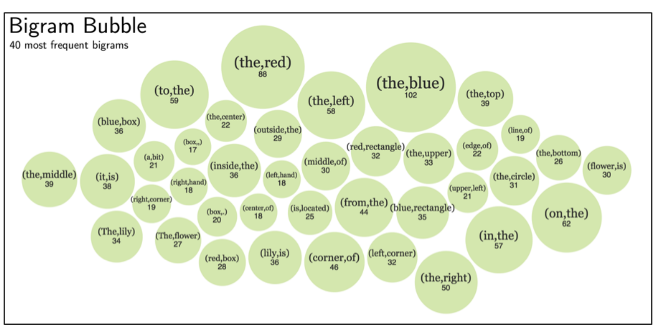

Figure 5: Word bubble visualization of bigrams in data

Qualitative Analysis

We observed that there were some interesting aspects of the speakers dialogue that emerged. First, for one of the most common terms used across all speakers, “inside, many of them chose to put an adverb in front of it. For instance, one speaker who used the term “inside six times across 30 rounds used it with an adverb four times. Expressions “Completely inside and “mostly inside were used to describe to which extent the surface area of the Lily was overlapping the object being described. These expressions were followed by a simple term to refer to the object. On the other hand, “Directly inside and “just inside were used by a speaker when the lily was completely inside and close to the inner edge of the object. Specifically, the phrase “just inside was used across four speakers. These expressions were always followed by a noun phrase to refer to the parameter of the object, such as “the left corner of the circle, “bottom of circle, “the right edge of the blue box, and “the border of the circle.

Secondly, two individuals each utilized the face of a clock and fractions in their explanations of the Lilys location. The speaker who used the clock directions used expressions such as “at six oclock on the circle and “between 4 and 5 o clock just inside the border of the circle. This individual had no back and forth dialogue with the listener at all. This result may indicate that the speakers initial explanation was always sufficiently clear to the listener due to the speakers use of the clock directions. However, it is difficult to tell from just one speaker-listener pair. The speaker who used fractions would use expressions such as “inside the blue $\frac{1}{3}$ from the left or “$\frac{1}{4}$ from the top to describe a location in terms of its distance from an objects inner or outer edge. This speakers explanations were among the shortest of those of other speakers and this individual used simple prepositional phrases–no adverbs in front of the prepositions. The dialogue log shows that the listener understood the speakers phrases well and the succinct explanations were sufficiently clear to the listener.

A final observation was that participants sometimes used concrete measurements that focused on the absolute, rather than relative, position of the lily. For example, one user reported “just above the middle of the blue rectangle about half an inch outside the border.” Interestingly, the absolute description was almost always followed or preceded by a relative description, demonstrating that people use a mixed framework in order to increase clarity. This is particularly interesting in an online game setting where there is no guarantee that “an inch” corresponds to the same physical distance for each user due to differing screen resolutions, etc.

An interesting conclusion from these phenomena is that different speakers tend to converge on communication styles, reusing particular constructions or patterns repeatedly. Future work could further investigate this taxonomy.

On the whole, participants erred on the side of being extremely specific, as it was in their best interest. From another perspective, there was no utterance length cost in this scenario, unlike in normal conversation. This demonstrates that while this was a free-form natural language communication platform, it was still different from conversation outside of an experiment. Attempts to deflate this tendency by giving speakers a “target radius” and thus leeway in finding the lily's location seem ineffectual. Additionally, a few of the users seemed confused about the experimental setup, being unclear that they were looking at a "map" of a city, and that the circle was a "plaza," (or even that the plaza was visible to the listener as well, perhaps because it had dotted lines on the map). This did not seem to be an issue for Ullman et al., 2016, where they included a key on the map in order to help listeners identify the referents of phrases. For these reasons, more tweaking of the experimental framework is needed in order to remove irregularities in the data.

RSA Model in WebPPL

Given the wide variance in utterance structure and the extremely verbose descriptions of the lily's locations used by speakers, it was not feasible to model the whole corpus. We selected to focus on the modifier "just," applied to the prepositions "inside" and "outside." We used the rational speech act model outlined in (Ullman et al., 2016) and developed in (Frank & Goodman, 2012). The goal of this language model is to accurately predict the location of the lily given a particular utterance that a speaker has chosen to use. There are three levels of processing to be considered: a pragmatic listener, who reasons while listening about the intentions of an informative speaker, who reasons while speaking about a literal listener.

As an illustrative example, assume the (pragmatic) speaker produces the sentence "The lily is in the blue quarter." First, a literal agent would characterize a location as possible or impossible based on the truth value of the sentence. Then, using this metacognitive knowledge, the speaker would produce an utterance from which the literal listener could guess the lily's location. Finally, the pragmatic listener updates her belief about the location of the lily, using the knowledge that the speaker chose the utterance that was spoken. We can formalize this as the following:

$$P_{L_0}(s|u) \propto \delta_{[\![u]\!](s)} P(s)$$
$$P_{S_1}(u|s) \propto P_{L_0}(s|u) P(u)$$
$$P_{L_1}(s|u) \propto P_{S_1}(u|s) P(s)$$

We describe the components of the algorithm below.

- s refers to the (x, y) coordinates of the lily. We build the model on a single hard-coded world for simplicity; we could also include the location of all objects in the world. We scaled down the world from 500x600 in data to 50x60 in the model for computational efficiency.
- u refers to the utterance. In our model, there were 12 possible utterances that were constructed by combining one of {inside, just inside, outside, just outside} with one of {blue, red, plaza}.
- δ refers to a truth-value indicator function given the state and utterance. We set a radius for "just" to match experimental data.
- $P(s)$ refers to the prior of the lily's location, which we assume as uniformly distributed over the world.
- $P(u)$ refers to the prior over utterances. It is modeled as uniform over all possible utterances.

This is essentially the modeling framework used in (Ullman et al., 2016), except that we did not provide an epsilon tolerance in the pragmatic model. Additionally, to up-scale from WebPPL outputs to the experimental map, we added noise uniformly to the pixel location by up to five points in either direction.

The entire model and supporting code can be found at `https://github.com/ntindall/reference\_games/blob/master/analysis/spatialReference/model.wppl`.

Model Results

Our model successfully captures the distinction between "inside" and "just inside" that we hoped to replicate from the experimental data. As you can see from our appendix plots, the locations predicted from our model when told that the lily was "just inside" are far closer to the edges than when told that the lily was "inside," both for the rectangular and spherical world landmarks. These values roughly correspond to our data from the experiment: further work might improve the model by tweaking definitions of "just" for rectangular and circular objects, as we observe a larger distance between click location and borders for "just inside" in the experimental data. We also include plots for "outside plaza" and "just outside plaza." These were not the main focus of our study, but they produce interesting results. The values for "outside" are scattered across the map, in line with the logical definition. However, they primarily concentrate in areas not covered by other terms: they are not inside or near the blue or red rectangles. This is what we would expect from a successful pragmatic model. Similarly, "just inside plaza" avoids entering the red rectangle. This is likely due to the comparatively high probabilities of "inside red" and "just inside red" in the listener's speaker model.

Conclusion

In conclusion, we found that our model produced results that aligned with the subset of the experimental natural spatial language we obtained. However, we did not expect the speech of the reference game to be as semantically rich as we expected, and so were not able to cover as much of the data as we had hoped. In future versions of this experiment, we suggest that researchers either limit the length of utterances, or constrain the set of potential utterances. Additionally, the large number of different worlds we deployed made our data fairly sparse in modeling listener behavior - we could not compare data for a single map to our model, because we had fewer than five data points for a given map and spatial construction. Another avenue for further investigation is improving the model of the pragmatic speaker. Here, we followed Ullman et al. in addressing the question, "Given an utterance, where should you click?" However, one could also focus on, "Given the lily location, what should you say?" Other researchers could attempt to build a speaker model that takes in a lily location, and produces a distribution over utterances similar to the experimental data gathered here. There are many other interesting dimensions of spatial language that we hope to explore in the future.

Acknowledgments

We thank Robert X.D. Hawkins and Professor Noah Goodman for assistance with this study and for their comments that greatly improved the paper.

References

Bergen, L., Levy, R., & Goodman, N. D. (2014). Pragmatic reasoning through semantic inference. *Unpublished manuscript*.

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*(6084), 998–998.

Grice, H. P., Cole, P., & Morgan, J. L. (1975). Syntax and semantics. *Logic and conversation*, *3*, 41–58.

Hawkins, R. (2015).
Behav Res, *47*.

Horn, L. (1984). Toward a new taxonomy for pragmatic inference: Q-based and r-based implicature. *Meaning, form, and use in context: Linguistic applications*, 11–42.

Kao, J. T., Wu, J. Y., Bergen, L., & Goodman, N. D. (2014). Nonliteral understanding of number words. *Proceedings of the National Academy of Sciences*, *111*(33), 12002–12007.

Potts, C., Lassiter, D., Levy, R., & Frank, M. C. (2015). Embedded implicatures as pragmatic inferences under compositional lexical uncertainty. *Journal of Semantics*, ffv012.

Ullman, T. D., Xu, Y., & Goodman, N. D. (2016). The pragmatics of spatial language. In *Proceedings of the thirty-eighth annual conference of the cognitive science society.*

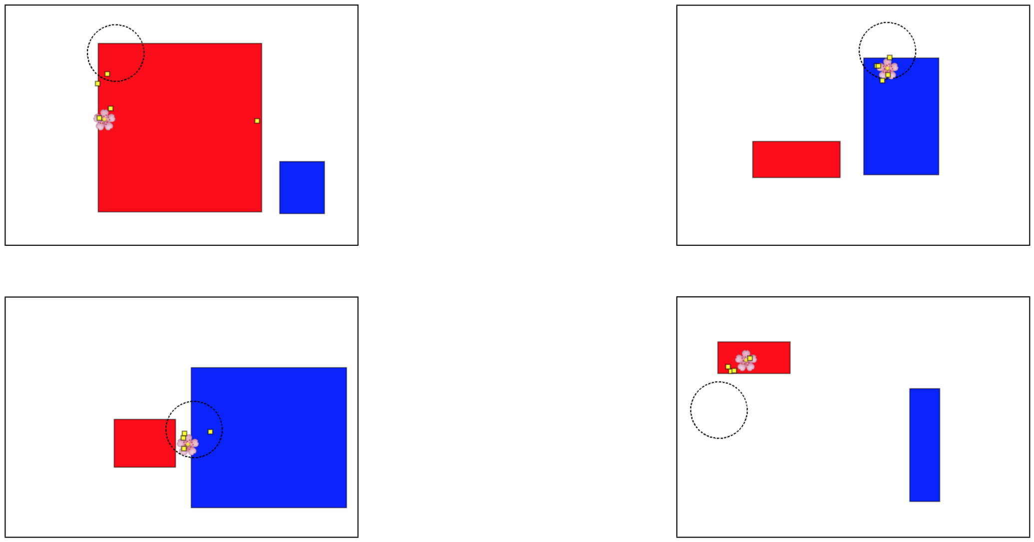

Figure 6: Data fromm experimental trials conducted on Mechanical Turk with click locations indicated with yellow squares.

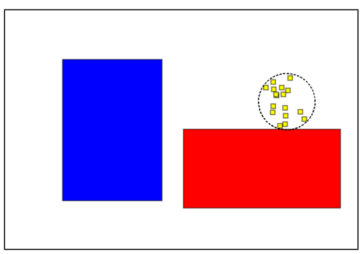

Figure 7: Model output for “Inside plaza”

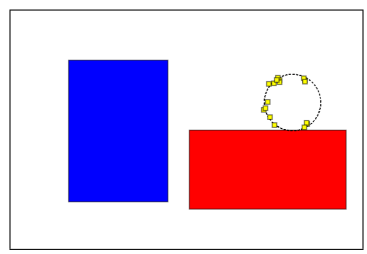

Figure 8: Model output for “Just inside plaza”

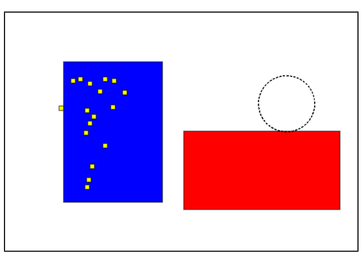

Figure 9: Model output for “Inside blue”

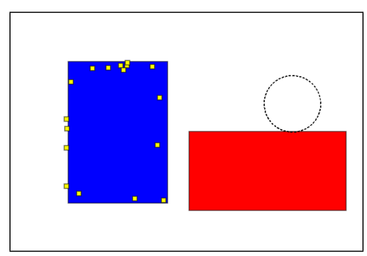

Figure 10: Model output for “Just inside blue”

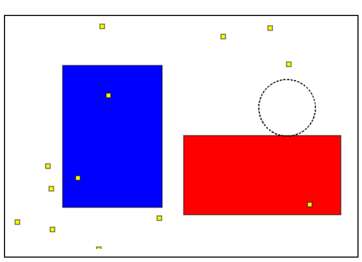

Figure 11: Model output for “Outside plaza”

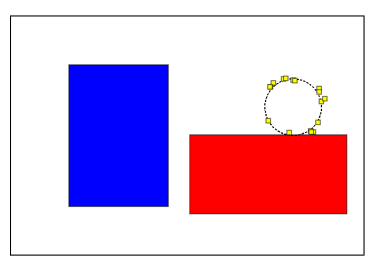

Figure 12: Model output for “Just outside plaza”

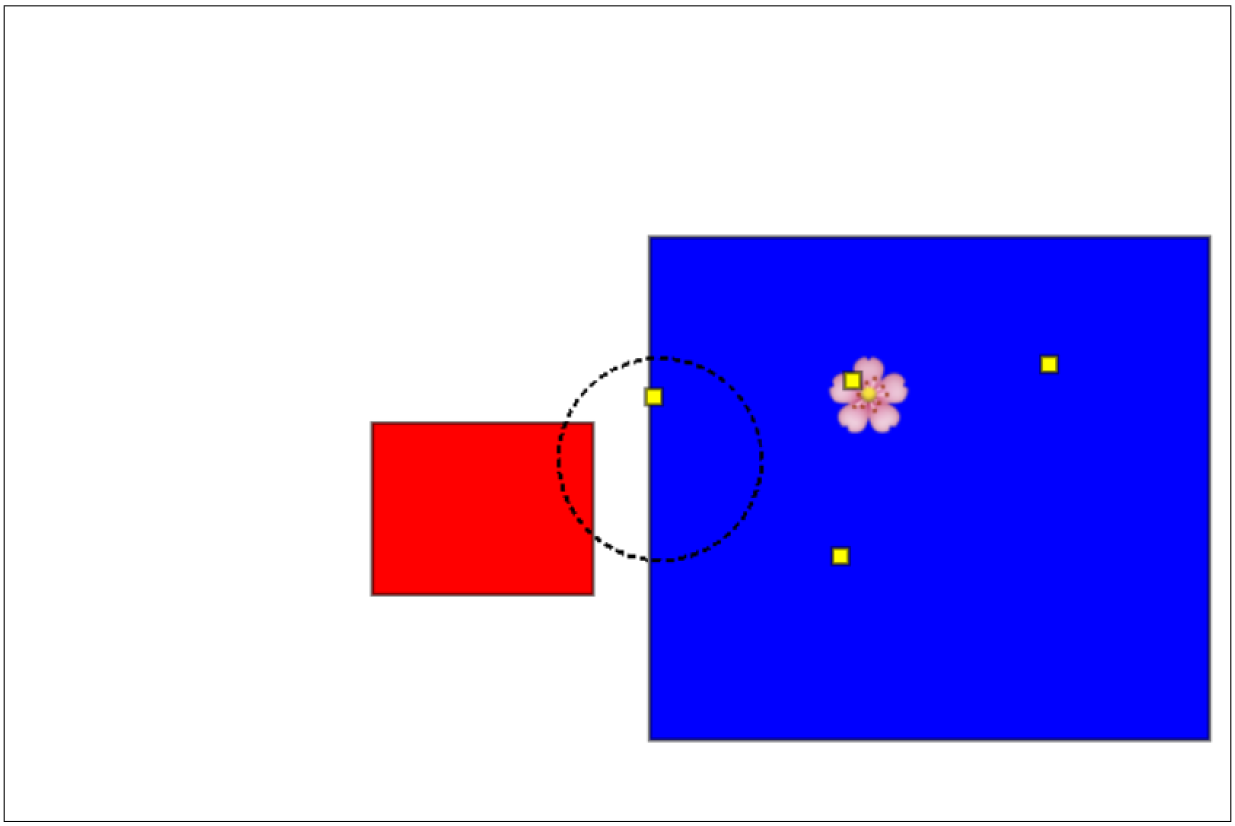

Figure 13: Utterances corresponding with listener clicks (yellow squares): “the lily is inside the blue top 1/3 towards the middle,” “The flower is on the upper right hand part of what you would consider to be the middle of the blue square, “it is located in the large blue square it is about a third of the way down from the top and inside about a third of the way from the left side,” “From the center of the blue box move up a bit then to the left a bit,” “Follow the topmost point of the circle to the right into the blue building for about two lillies wide”