

Categorisation, Typicality & Object-Specific Features in Spatial Referring Expressions

Adam Richard-Bollans
University of Leeds, UK
mm15alrb@leeds.ac.uk

Lucía Gómez Álvarez
TU Dresden, Germany
lucia.gomez_alvarez@tu-dresden.de

Anthony G. Cohn
University of Leeds, UK
a.g.cohn@leeds.ac.uk

Abstract

Various accounts of cognition and semantic representations have highlighted that, for some concepts, different factors may influence category and typicality judgements. In particular, some features may be more salient in categorisation tasks while other features are more salient when assessing typicality. In this paper we explore the extent to which this is the case for English spatial prepositions and discuss the implications for pragmatic strategies and semantic models. We hypothesise that object-specific features — related to object properties and affordances — are more salient in categorisation, while geometric and physical relationships between objects are more salient in typicality judgements. In order to test this hypothesis we conducted a study using virtual environments to collect both category and typicality judgements in 3D scenes. Based on the collected data we cannot verify the hypothesis and conclude that object-specific features appear to be salient in both category and typicality judgements, further evidencing the need to include these types of features in semantic models.

1 Introduction

Various accounts of cognition and semantic representations have highlighted that, for some concepts, different factors may influence category and typicality judgements (Smith et al., 1974; Rips, 1989). In particular, some features may be more salient in categorisation tasks while other features are more salient when assessing typicality. In this paper we explore the extent to which this is the case for English spatial prepositions and discuss the implications for pragmatic strategies and semantic models. We hypothesise that object-specific features — related to object properties and affordances — are more salient in categorisation, while geometric and physical relationships between objects are more salient in typicality judgements. In order to

test this hypothesis we conducted a study using virtual environments to collect both category and typicality judgements in 3D scenes. Based on the collected data we cannot verify the hypothesis and conclude that object-specific features appear to be salient in both category and typicality judgements, further evidencing the need to include these types of features in semantic models.

The spatial prepositions we analyse are those considered to have a functional component as well as those prepositions that seem to act as their geometric counterpart. For the ‘functional’ prepositions, object affordances and functional relationships, such as *support* and *location control*, appear to be salient (Garrod et al., 1999; Coventry et al., 2001) compared to the geometric counterparts where geometric features and relative positions of objects appear to be more salient. In English, we consider the functional prepositions to be: ‘in’, ‘on’, ‘over’ and ‘under’; and their respective geometric counterparts to be: ‘inside’, ‘on top of’, ‘above’ and ‘below’. We also consider ‘against’ to be a functional preposition (Talmy, 1988), without a clear geometric counterpart.

To clarify the distinction between category and typicality judgements, we suppose that a category decision is when an entity is labelled with a category or concept and, though priming and context may certainly be factors, the judgement is not made in direct comparison with another entity. For example, a categorisation judgement occurs when an agent is asked whether ‘the apple is *in* the bowl’. In order to reply, the agent judges the membership of the instance in the relevant category and the wider context plays a relatively minor role.

Typicality usually refers to the extent to which an entity is a good example of a concept — how similar it is to some ideal conceptual representation. In our study we ground the notion of typicality in comparison and preference — an entity, x , is more typical of a category than entity y if, when

x is compared with y , people in general pick x as a better category member. For example, when requested to pick ‘the apple *in* the bowl’, an agent must compare a set of candidate objects for how well they fit the description.

By assessing typicality in this way we distinguish the notion of typicality from graded category membership. The typicality data that we collect does not arise from graded membership judgements where study participants are asked to assign a value of how well the concept fits the category, e.g. in (McCloskey and Glucksberg, 1978; Hampton, 1997), but instead from tasks in which participants select the best fitting instance from a given description.

In existing models of spatial language (and semantic models more generally), it is generally assumed that the underlying semantics of categorisation and typicality are essentially the same. However, as we will discuss in Section 4, appropriately distinguishing categorisation and typicality judgements is important when generating and processing referring expressions.

2 Background

Following various criticisms of definitional representations of concepts in human cognition (Wittgenstein, 1953), Rosch’s Prototype Theory (Rosch, 1978) provided an approach based on family resemblance which does not presuppose that concepts have necessary and sufficient conditions for making category judgements. By relying on a degree of resemblance to some prototypical notion of a concept, category membership in this account may be treated as a matter of degree. With such an account it becomes appealing to conflate the notions of categorisation and typicality — the more an entity resembles a prototype the more likely it is to be labelled with the category *and* the more typical it is. There are however various accounts of concept analysis which suggest that category and typicality judgements are fundamentally different.

Smith et al. (1974) consider the influences of category decisions and propose a model to account for experimental findings. Central to the model is the ‘characteristic feature assumption’, which supposes that features vary in the extent to which they define a concept. Smith et al. suppose that there is a distinction in types of features — ‘defining’ features which strongly influence category judgements and ‘characteristic’ features which strongly

influence typicality judgements — and give the example of ‘robin’ to illustrate this. For the concept ‘robin’, ‘have wings’ is an important defining feature relating to the categorisation of an entity as ‘a robin’, while ‘perches in trees’ is a characteristic feature which relates to how typical an entity is of ‘a robin’.

Rips (1989) argued that categorisation of some entity is more than simply a judgement of how similar the entity is to some typical representation of the category. Rips provides support for this in an experiment where participants are asked to imagine an object of a given size and are asked which of two concepts, A , B say, the object is more similar to and which the object is most likely to be. For the given concept pairs, e.g. ‘pizza’ and ‘quarter’, the hypothetical object may be considered more similar to one of the concepts yet more likely to be the other. In the case of the pizza and quarter, a round object with a three inch diameter is regarded as more similar to a quarter as pizzas are rarely so small, but more likely to be a pizza as the size of a quarter is generally fixed and is less than 3 inches.

This issue is explored further in (Osherson and Smith, 1997) where it is again argued that concept membership and typicality are distinct phenomena. Based on a study in (Berlin and Kay, 1991), Osherson and Smith argue that the notions differ using the seemingly extreme example of the concept ‘red’, even though it may be hard to imagine distinct defining and characteristic features for a concept with such simple semantics. However, the noted difference in judgements arises as people recognise particular wavelengths of light as being unambiguously red yet less typical than prototypical red. This difference in judgements preserves a monotonic relationship between category and typicality judgements i.e. if an entity, x , is a better category instance than y , then it is not less typical than y .

We believe that such monotonicity results offer a trivial case which may be explained without any fundamental modification of the underlying semantics or how they are processed. In the simple case of the colour red, we may represent the semantics in both categorisation and typicality judgements by considering the *dominant wavelength* — an instance is more or less typical and a better or worse category instance based on the similarity of the dominant wavelength to prototypical red and the distinction in category and typicality judgements

is explained via a threshold which is applied in category judgements.

However, in the case of spatial language, the semantics are more complex and are influenced by a variety of geometric, functional and object-specific features. We believe that as a result, the relationship between typicality and categorisation may in fact be non-monotonic — there may be instances of a spatial preposition, i_1, i_2 , such that i_1 is a better category member but less typical than i_2 .

2.1 A note on terminology

Regarding the names of the objects being discussed we use *figure* (also known as: target, trajector, referent) to denote the entity whose location is important e.g. ‘the **bike** next to the house’ and *ground* (also known as: reference, landmark, relatum) to denote the entity used as a reference point in order to locate the figure e.g. ‘the bike next to the **house**’. We call potential figure-ground pairs *configurations*.

2.2 Object-Specific Features

It is apparent that various object properties and affordances influence the categorisation and usage of spatial prepositions (Coventry et al., 1994; Feist and Gentner, 1998). For example, the animacy of the figure object may influence a decision to use ‘in’ or ‘on’¹ (Feist and Gentner, 1998). Following (Coventry et al., 1994), we call these features ‘object-specific’ features.

We believe that these types of features provide a source of disagreement between category and typicality judgements and below for each of the functional prepositions we describe the object-specific features that are explored in this paper.

In As ‘in’ expresses a notion of *containment*, the ability of the ground to contain the figure is often salient whether or not the ground does in fact contain the figure in a geometric sense. Therefore, whether the ground is a type of container appears to be salient for ‘in’ (Coventry et al., 1994; Feist and Gentner, 1998; Richard-Bollans et al., 2019) and this may be considered a salient object-specific feature.

Over/Under ‘over’ and ‘under’ appear to have a ‘covering’ sense which is closely related to the functions of the figure and ground (Coventry et al.,

¹People were found to prefer ‘in’ when describing an inanimate figure (a coin) and ‘on’ when describing an animate figure (a firefly).

2001; Tyler and Evans, 2001; Mori, 2019). For example, a covering object like a lid may exhibit this sense of ‘over’ when covering a container. Therefore, whether or not the figure is a covering object or the ground is a type of container may be salient object-specific features.

There is also a non-covering sense where a specific functional interaction exists between part of the figure and ground. For example, a tap may be ‘over’ a sink if only the spout of the tap is above the sink. Similarly, an object may be ‘under’ a lamp when the object is not under the lamp in a geometric sense but the light from the lamp shines on the object. These specific functional interactions rely on particular properties of the figure or ground and so we consider them to be object-specific features.

Relating to the functional interactions of the figure and ground, an intermediary object between the figure and ground may serve to block any functional interaction, as studied in (Coventry et al., 1994), and diminish the effect of any object-specific features which are present.

Against ‘against’ is commonly used to denote contact between two objects and as argued in (Herskovits, 1987) is more applicable in situations where the ground object is fixed and the figure is mobile. For example, one may describe a chair as being ‘against a wall’ but it would be odd to describe a wall as being ‘against a chair’.

On ‘on’ is ubiquitous in the English language and is applied to many situations where usually at least one of the following hold: the figure is supported by the ground, the figure is above the ground or the figure is in contact with the ground. As a result, it is not clear that there are particular properties of figure or ground objects at table-top scales which create strong preferences for ‘on’.

As discussed above, the preposition ‘in’ is often preferred when the ground object is a container. ‘on’ is therefore used less frequently in these scenarios (Feist and Gentner, 1998), even though the physical relationships between the objects often fulfil the requirements for ‘on’. As a result, whether or not the ground is a container appears to be a salient object-specific feature for ‘on’.

Finally, ‘on’ may be used to denote attachment of the figure to the ground. It is therefore plausible that, similarly to ‘against’, ‘on’ is more applicable in situations where the ground object is fixed relative to the figure.

3 Hypothesis

Our main hypothesis is that categorisation and typicality judgements may differ for spatial prepositions in the following manner: given configurations c_1, c_2 and preposition P , participants may be more likely to categorise c_1 with P yet more likely to select c_2 as a better instance of P . We hypothesise that this may arise in part because particular features are salient in category judgements which become less salient in typicality judgements. Note that under this hypothesis the relationship between categorisation and typicality is non-monotonic, making this in a sense stronger than the findings related to ‘red’ discussed by (Osherson and Smith, 1997).

In general, we expect that object-specific features will be more salient in categorisation while geometric and physical relationships, such as *containment* or *support*, will be more salient in typicality judgements.

Furthermore, we believe that this particular distinction may be more pronounced for the functional prepositions than for their geometric counterparts. This would be somewhat a corollary of the assumption that functionality in general is more salient for the former, for which tentative evidence is provided in (Richard-Bollans et al., 2020a) for these prepositions and (Coventry et al., 2001) in the case of ‘over/above’ and ‘under/below’.

4 Generating & Understanding Referring Expressions

In this section we discuss the motivation for this investigation and the possible implications of a significant distinction between category and typicality judgements. In particular, we consider the ramifications for the field of Referring Expression Generation and Comprehension (REG/C), where noun phrases are used to identify entities e.g. ‘the box under the table’.

Humans often prefer brief ambiguous descriptions over lengthy unambiguous descriptions (Rohde et al., 2012), and expressions involving spatial prepositions often fulfil this desire for brevity. For example, rather than referring to objects based on elaborate visual attributes like ‘the yellow cup with two pink dots on it’, humans often refer to objects using simple locative expressions, say ‘the cup next to the stapler’. We also see many examples of expressions containing spatial prepositions in the SemEval-2014 corpus (Dukes, 2014) and the

HuRIC corpus (Bastianelli et al., 2014), both of which consider natural language commands given to robots.

To explain why a distinction in category and typicality judgements is important in these scenarios, suppose we have a speaker and listener, intended referent, r , set of distractor objects² and suppose that the speaker is generating a description, D .

When the speaker generates an utterance in order to refer to r , there are various semantic and pragmatic considerations they must make. A naive model of such a speaker may simply find a concept within its vocabulary which is most suitable for r . This would clearly be a flawed strategy, however, as there may be other entities in the scene which better fit the concept — an expression may be true but not satisfy the speaker’s communicative goals and so pragmatics must be considered. A more refined speaker model may find a description which best distinguishes r from the distractor objects in O , similar to the algorithm of (Dale, 1989) which aims to maximise the *discriminatory power* of a description while minimising superfluous information.

It appears that the speaker must model how well an object fits a description compared to other objects (typicality judgement) and how well a description fits an object (category judgement). Further, it is apparent that humans will reason recursively about possible intentions of speakers and possible interpretations of listeners (Goodman and Frank, 2016), making these judgements also necessary for listeners.

Let us consider a more concrete example. Suppose we have a scenario as in Figure 1 where a bowl, b , is on a table and there is one cube, c_{red} , in it and one cube, c_{blue} , next to and not touching it. It seems plausible that humans are more likely to categorise the configuration (c_{blue}, b) with the preposition ‘near’ than the configuration (c_{red}, b) even though when comparing the configurations humans may agree that c_{red} is more ‘near’ the bowl than c_{blue} .

Suppose a speaker gives an utterance ‘the cube near the bowl’ in order to refer to c_{blue} which an agent must interpret. As may be expected, semantic models, e.g. (Platonov and Schubert, 2018), are likely to assign a better score for ‘near’ to (c_{red}, b) than (c_{blue}, b) . If the system has a crude strategy

²A distractor object is an object in the scene which the speaker is not intending to refer to.

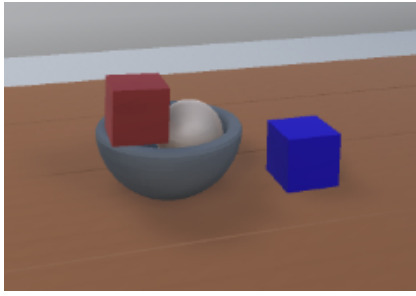


Figure 1: An example of possible confusion when not accounting for object-specific features.

for interpretation which simply selects the configuration with the highest semantic score, then such a system would erroneously select c_{red} .

Many systems with more sophisticated pragmatic strategies, e.g. (Golland et al., 2010), have been developed which aim to take into account and reason with the possible utterances available to the speaker. In this case, such a system may correctly select c_{blue} if it recognises that other better utterances would have been available to the speaker if the intended referent was c_{red} . ‘the cube in the bowl’ would be a clear example of such an utterance which seems to clearly identify c_{red} over c_{blue} . However, supposing that our hypothesis is correct, we contend that this marked distinction is not simply a matter of typicality and the fact that it would be unusual to categorise (c_{red}, b) with ‘near’ is more salient than any distinction in typicality between the two configurations. As c_{red} is not ‘in’ the bowl in an ideal sense we can imagine that a semantic model based simply on the physical relationships between the objects may provide a more marked distinction between the configurations for ‘near’ than for ‘in’. In this case, ‘the cube in the bowl’ wouldn’t necessarily seem like a better utterance to identify c_{red} than ‘the cube near the bowl’.

Moreover, even if typicality on its own works well for REC, understanding and modelling the differences between categorisation and typicality is important for producing *natural* utterances in REG. For instance, suppose the speaker creates an utterance where r is more typical for the concept, C say, in D than any of the distractor objects. If categorisation is not aligned with typicality, such a strategy may produce unusual utterances where, though r is typical for C , it is uncommon to categorise r with C . Such unnatural utterances may trigger unwanted conversational implicatures and be a source of confusion. For example, the utter-

ance ‘the ball on the bowl’ in the context of Figure 1 would be an unusual way to describe the ball as the preposition ‘in’ is often used with objects such as bowls. From this unconventional usage of ‘on’ a listener may imply that for some reason ‘in’ was not suitable e.g. if the speaker is actually referring to another unseen ball. The issue of producing natural utterances has been recognised by others in the field (Krishnaswamy and Pustejovsky, 2019) and is an important challenge to overcome if we are to develop more sophisticated REG systems.

With regards to existing models of spatial language, it is generally assumed that the underlying semantics of categorisation and typicality are essentially the same. For example, in the PRAGR mechanism proposed in (Mast et al., 2016) a pragmatic strategy is presented which aims to maximise both the *acceptability* and *discriminatory power* of a description. Acceptability is calculated using similarity to a prototype based on physical relationships while the discriminatory power is calculated considering the acceptability of the description for the referent compared to other distractor objects.

5 Data Collection

In order to investigate typicality measures and categorisation judgements in detail, we conducted a study which is described below. Collected data, details of the framework and results of the analysis can be found in the Leeds research data repository.³ The latest version of the data collection environment and code for analysis can be found in the GitHub repository.⁴

5.1 Environment & Tasks

The data collection framework is built on the Unity3D⁵ game development software, which provides ample functionality for the kind of tasks we implement. Two tasks were created for our study — a Categorisation Task and a Typicality Task. The former allows for the collection of categorical data while the latter provides typicality judgements.

In the Categorisation Task participants are shown a figure-ground pair (highlighted and with text description, see Figure 2) and asked to select *all* prepositions in the list which fit the configuration. Participants may select ‘None of the above’ if they deem none of the prepositions to be appropriate.

³<https://doi.org/10.5518/873>

⁴<https://github.com/alrichardbollans/spatial-preposition-annotation-tool-unity3d>

⁵<https://unity.com/>

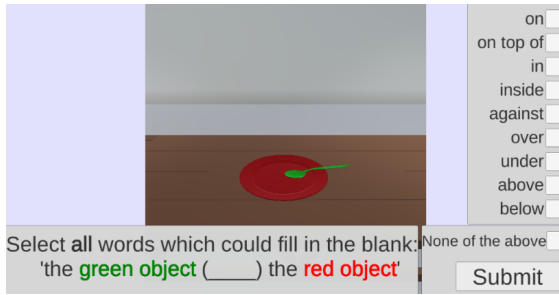


Figure 2: Categorisation Task

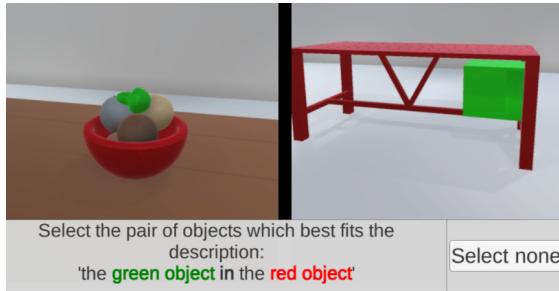


Figure 3: Typicality Task

Often categories and concepts are viewed as antagonistic entities; for example work on Conceptual Spaces is often concerned with comparison of categories, e.g. partitioning a feature space (Douven et al., 2013). We believe however that the vagueness present in spatial language is so severe that it is not clear that a meaningful model distinguishing the categories is possible. For this reason, participants in our studies are asked to select *all* possible prepositions for a configuration rather than a single best-fitting preposition.

In the Typicality Task participants are given a description and shown two configurations, see Figure 3. Participants are asked to select the configuration which *best fits* the description. Again, participants can select none if they deem none of the configurations to be appropriate.

In order to minimise differences in the tasks that may elicit different conceptualisations of objects in the scenes, the phrasing of the descriptions is the same in both the Categorisation Task and Typicality Task e.g. both tasks use the definite determiner ‘the’ and objects are referred to by their colour.

5.2 Scenes

We hypothesise that object-specific features strongly influence category decisions while the geometric ideals associated to the prepositions are more salient in typicality decisions. Scenes are therefore created for each preposition which vary

the degree to which these object-specific features are present and also vary how similar the relational aspects of the configurations are to the geometric ideals associated with the given preposition. For example for the preposition ‘in’, we have a scene where the ground is a container and the figure is not very well contained in it and a scene where the ground is not a type of container but the figure is well contained in it. In this case, if the hypothesis is correct, we would expect a preference for categorisation in the former and a preference for typicality in the latter.

We have created 18 virtual 3D scenes each containing a single highlighted figure-ground pair. Four scenes each were created for ‘in’, ‘on’, ‘over’ and ‘under’ and these scenes were also shared with their respective geometric counterparts: ‘inside’, ‘on top of’, ‘above’ & ‘below’. Two scenes were created for ‘against’. In the Typicality Task, participants compare scenes/configurations associated with the preposition given in the description.

5.3 Study

The study was conducted online and participants from the university were recruited via internal mailing lists along with recruitment of friends and family.⁶ Each participant performed first the Categorisation Task on 6 randomly selected scenes and then the Typicality Task on 15 randomly selected scenes, which took participants roughly 5 minutes. 30 native English speakers participated providing 180 annotations in the Categorisation Task and 447 annotations in the Typicality Task.

As the study was hosted online, we first asked participants to show basic competence. This was assessed by showing participants two simple scenes with an unambiguous description of an object. Participants are asked to select the object which best fits the description. If the participant makes an incorrect guess in either scene they are taken back to the start menu.

6 Results

In this section we use the collected data to test the hypothesis and conclude that we cannot verify the hypothesis. We propose that this is because object-specific features are in fact salient in both category and typicality judgements and provide examples to support this.

⁶University of Leeds Ethics Approval Code: MEEC 19-030. Participants were recruited without incentive.

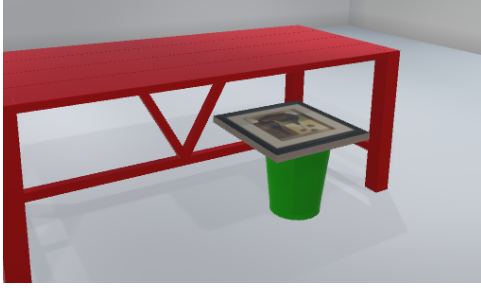


Figure 4: The bin under the table

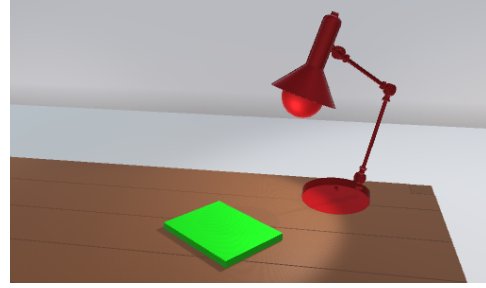


Figure 5: The notepad under the lamp

6.1 Comparing Categorisation & Typicality

To analyse the collected data, we consider pairs of tested configurations for each preposition and evaluate the degree to which category and typicality judgements differ. For each pair, (c_1, c_2) , we first decide whether c_1 is a genuinely better category instance than c_2 or is more typical than c_2 . To do this we simply use a hypothesis test with significance level 10% and null hypothesis that the given configurations are equally likely to be labelled with the preposition (in the category case) or equally likely to be selected (in the typicality case). In the category case the p-value is calculated using the one-tailed version of Fisher’s exact test. In the typicality case the p-value is simply: $p = \sum_{k=C_{1,2}}^N \binom{N}{k} \times 0.5^N$, where N is the number of times the pair is tested and $C_{1,2}$ is the number of times c_1 is selected over c_2 . In 22 out of the 49 given pairs, one of the configurations is a significantly better category instance or is more typical than the other.

Considering the somewhat trivial case, similar to the case of ‘red’ discussed in Section 2, where two entities are both unambiguous cases of a concept but one of the entities is more typical than the other; there is one instance of this in our dataset. For the preposition ‘under’ the configurations shown in Figures 4 & 5 were both always labelled with the preposition, out of seven tests in the former and ten tests in the latter, but the (bin, table) configuration in Figure 4 is significantly more typical than the (notepad, lamp) configuration in Figure 5. As previously discussed, however, this is an unsurprising result.

Regarding the main hypothesis of this paper, there are no pairs of configurations in our dataset where one of the configurations is a significantly better category member and the other is significantly more typical. Moreover, in only nine pairs is there any possible disagreement where one of

the configurations is more often labelled with the preposition and the other configuration is more often selected in the Typicality Task — in all but one of these cases neither configuration is a significantly better category member or is significantly more typical. We therefore cannot conclude that our hypothesis is correct and it appears that the notions of categorisation and typicality do not significantly vary due to object-specific features — these appear to be both defining and characteristic features.

Clearly we have only tested a small number of features and there are a vast array of salient features for each preposition for which the hypothesis may still be correct. However, our results suggest that object-specific features are salient in both categorisation and typicality judgements — in some cases the object-specific features appear to have a stronger influence than the physical relationships. Interestingly, there is some tentative evidence that this extends in general to the geometric counterparts and suggests that these prepositions are not purely spatial — supporting findings in (Dobnik and Ghanimifard, 2020).

6.2 Importance of Object-Specific Features

In the following we provide some examples from our dataset which highlight the importance of object-specific features.

On For the preposition ‘on’, the (mug, pencil) configuration in Figure 6 is both a significantly better category member and is significantly more typical than the (pear, bowl) configuration in Figure 7. Regarding the physical relationships, (pear, bowl) appears to be a better example of ‘on’ than (mug, pencil). If we consider the usual salient features for ‘on’:

- The pear is fully *supported* by the bowl, while the mug is leaning on both the pencil and the table

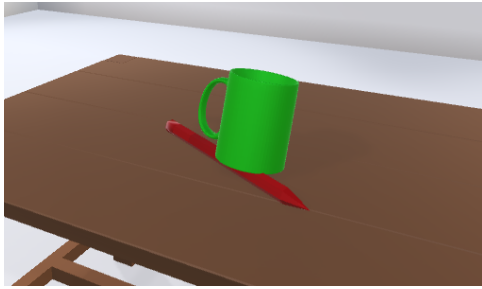


Figure 6: The mug on the pencil



Figure 8: The pear in the bowl

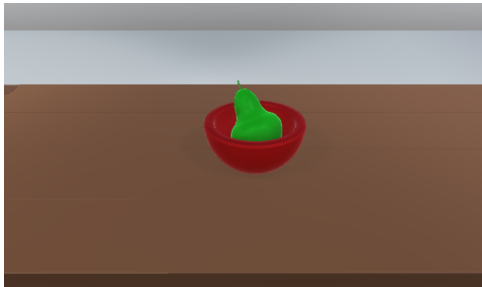


Figure 7: The pear in/on the bowl

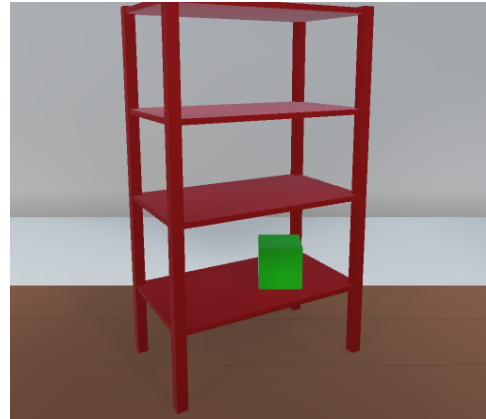


Figure 9: The cube on/in the shelf

- There is a high degree of *contact* between the pear and the bowl compared to the mug and pencil
- The entirety of the pear is *above* some part of the bowl, while the bottom of the mug is level with the bottom of the pencil

We therefore believe this result is not due to the physical relationships of (mug, pencil) better representing ‘on’ than (pear, bowl) and that this result arises primarily because ‘on’ is generally not used for containers. One may have expected this result if the objects in the experiments were named — ‘on the bowl’ sounds strange while ‘on the pencil’ seems more plausible. It is therefore even more surprising given that the objects were not named in a way that influences the decisions.

In/Inside For the prepositions ‘in’ and ‘inside’, the (pear, bowl) configuration in Figure 8 was more likely to be selected in the Typicality Task than the (cube, shelf) configuration in Figure 9. If we measure containment simply as the degree to which one object is contained in the bounding box or convex hull of another, as is common in grounded semantic models e.g. (Chang et al., 2014; Platonov and Schubert, 2018), then the cube would be fully contained by the shelf whereas the pear is not even partially contained by the bowl.

It therefore appears that the role of the bowl as

a type of container is influencing typicality judgements for both these prepositions.

Over/Above For the preposition ‘over’, the (tap, sink) configuration in Figure 10 is both a significantly better category member and is significantly more typical than the (lid, pan) configuration in Figure 11. The same is true for the preposition ‘above’, though in this case the results are not significant.

Again, the physical relationships of (lid, pan) appear to better capture the geometric meanings of ‘over’ and ‘above’ than (tap, sink). There is also some functional interaction between the objects in both cases — lids are used to cover pans and sinks are placed below taps to catch the water. The preference for the (tap, sink) configuration is therefore



Figure 10: The tap over the sink

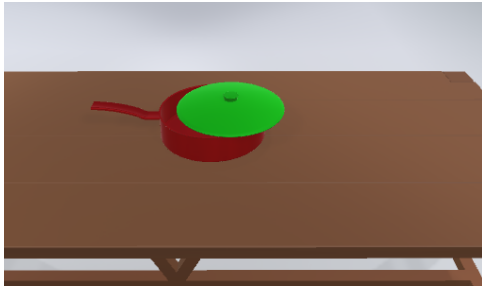


Figure 11: The lid over the pan

somewhat surprising.

However, the lid does not appear to be properly fulfilling its functional role, as it is not fully covering the container part of the pan. This may explain the preference for (tap, sink) and further highlight the importance of considering functional interactions based on usual object usages.

7 Discussion

Regardless of the hypothesis, these results highlight the importance of including object-specific features in semantic models of spatial language. With the possible exception of (Platonov and Schubert, 2018), these types of features are rarely included in semantic models and many systems are developed in block-world type environments, e.g. (Spranger, 2013; Mast et al., 2016; Perera et al., 2018), where these types of features are not needed.

One approach to improving semantic models may be to incorporate information from knowledge bases such as ConceptNet (Speer and Havasi, 2012) or AfNet (Varadarajan and Vincze, 2012). For example, from ConceptNet one can determine that lids are used for covering and that bowls are containers. Another approach is to leverage affordance detection systems, e.g. (Do et al., 2018), which use information from the scene to predict object affordances.

In (Rodrigues et al., 2020) some object-specific features, e.g. whether or not the ground is a type of container, are taken to distinguish separate polysemes.⁷ By leveraging previous work on modelling the polysemy of spatial prepositions (Richard-Bollans et al., 2020b), it may be possible to incorporate object-specific features into a semantic model by using these features to distinguish polysemes. In order to carry this out, further

⁷A word is said to exhibit *polysemy* where the word has multiple related senses. Each of these senses is called a *polyseme*.

work identifying salient object-specific features for each preposition would be beneficial and a larger dataset is needed which provides more instances of prepositions representing a greater variety of object-specific features.

Acknowledgments

The first author is supported by an EPSRC Studentship; the final author is partially supported by an Alan Turing Institute Fellowship, and also by the EC AI4EU project under grant agreement 825619. We would also like to thank the reviewers for their constructive feedback.

References

- Emanuele Bastianelli, Giuseppe Castellucci, Danilo Croce, Luca Iocchi, Roberto Basili, and Daniele Nardi. 2014. HuRIC: a Human Robot Interaction Corpus. In *LREC*, pages 4519–4526.
- Brent Berlin and Paul Kay. 1991. *Basic color terms: Their universality and evolution*. Univ of California Press.
- Angel Chang, Manolis Savva, and Christopher D. Manning. 2014. *Learning Spatial Knowledge for Text to 3D Scene Generation*. In *Proc EMNLP*, pages 2028–2038. Association for Computational Linguistics.
- Kenny R. Coventry, Richard Carmichael, and Simon C. Garrod. 1994. *Spatial prepositions, object-specific function, and task requirements*. *Journal of Semantics*, 11(4):289–309.
- Kenny R. Coventry, Mercè Prat-Sala, and Lynn Richards. 2001. *The Interplay between Geometry and Function in the Comprehension of Over, Under, Above, and Below*. *Journal of Memory and Language*, 44(3):376–398.
- Robert Dale. 1989. *Cooking up referring expressions*. In *Proceedings of 27th Annual Meeting of the association for Computational Linguistics*, pages 68–75. Association for Computational Linguistics.
- Thanh-Toan Do, Anh Nguyen, and Ian Reid. 2018. *AffordanceNet: An End-to-End Deep Learning Approach for Object Affordance Detection*. In *Proceedings of 2018 IEEE international conference on robotics and automation*. IEEE.
- Simon Dobnik and Mehdi Ghanimifard. 2020. *Spatial Descriptions on a Functional-Geometric Spectrum: the Location of Objects*. In *Proc German Conference on Spatial Cognition*, pages 219–234. Springer.
- Igor Douven, Lieven Decock, Richard Dietz, and Paul Égré. 2013. *Vagueness: A Conceptual Spaces Approach*. *Journal of Philosophical Logic*, 42(1):137–160.

- Kais Dukes. 2014. [SemEval-2014 Task 6: Supervised Semantic Parsing of Robotic Spatial Commands](#). In *Proceedings of the 8th International Workshop on Semantic Evaluation*, pages 45–53.
- Michele I Feist and Dredre Gentner. 1998. On plates, bowls, and dishes: Factors in the use of English IN and ON. In *Proc 20th annual meeting of the cognitive science society*, pages 345–349.
- Simon Garrod, Gillian Ferrier, and Siobhan Campbell. 1999. In and on: investigating the functional geometry of spatial prepositions. *Cognition*, 72(2):167–189.
- Dave Golland, Percy Liang, and Dan Klein. 2010. A Game-Theoretic Approach to Generating Spatial Descriptions. In *Proc EMNLP*, pages 410–419.
- Noah D. Goodman and Michael C. Frank. 2016. [Pragmatic language interpretation as probabilistic inference](#). *Trends in Cognitive Sciences*, 20(11):818–829.
- James A. Hampton. 1997. [Conceptual combination: Conjunction and negation of natural concepts](#). *Memory & Cognition*, 25(6):888–909.
- Annette Herskovits. 1987. *Language and spatial cognition*. Cambridge University Press.
- Nikhil Krishnaswamy and James Pustejovsky. 2019. [Generating a Novel Dataset of Multimodal Referring Expressions](#). In *Proceedings of the 13th International Conference on Computational Semantics - Short Papers*, pages 44–51. Association for Computational Linguistics.
- Vivien Mast, Zoe Falomir, and Diedrich Wolter. 2016. [Probabilistic reference and grounding with PRAGR for dialogues with robots](#). *Journal of Experimental & Theoretical Artificial Intelligence*, 28(5):889–911.
- Michael E. McCloskey and Sam Glucksberg. 1978. [Natural categories: Well defined or fuzzy sets?](#) *Memory & Cognition*, 6(4):462–472.
- Souma Mori. 2019. [A cognitive analysis of the preposition over: image-schema transformations and metaphorical extensions](#). *Canadian Journal of Linguistics*, 64(3):444–474.
- Daniel Osherson and Edward E Smith. 1997. [On typicality and vagueness](#). *Cognition*, 64(2):189–206.
- Ian Perera, James Allen, Choh Man Teng, and Lucian Galescu. 2018. [Building and Learning Structures in a Situated Blocks World Through Deep Language Understanding](#). In *Proc 1st International Workshop on Spatial Language Understanding*, pages 12–20.
- Georgiy Platonov and Lenhart Schubert. 2018. [Computational Models for Spatial Prepositions](#). In *Proc 1st International Workshop on Spatial Language Understanding*, pages 21–30.
- Adam Richard-Bollans, Brandon Bennett, and Anthony G. Cohn. 2020a. [Automatic generation of typicality measures for spatial language in grounded settings](#). In *Proceedings of 24th European Conference on Artificial Intelligence*.
- Adam Richard-Bollans, Lucía Gómez Álvarez, Brandon Bennett, and Anthony G. Cohn. 2019. [Investigating the dimensions of spatial language](#). In *Proceedings of Speaking of Location 2019: Communicating about Space*. CEUR Workshop Proceedings.
- Adam Richard-Bollans, Lucía Gómez Álvarez, and Anthony G. Cohn. 2020b. [Modelling the polysemy of spatial prepositions in referring expressions](#). In *Proceedings of 17th International Conference on Principles of Knowledge Representation and Reasoning*.
- Lance J. Rips. 1989. [Similarity, typicality, and categorization](#). In Stella Vosniadou and Andrew Ortony, editors, *Similarity and Analogical Reasoning*, 1st edition. Cambridge University Press.
- Edilson J Rodrigues, Paulo E Santos, Marcos Lopes, Brandon Bennett, and Paul E Oppenheimer. 2020. [Standpoint semantics for polysemy in spatial prepositions](#). *Journal of Logic and Computation*.
- Hannah Rohde, Scott Seyfarth, Brady Clark, Gerhard Jäger, and Stefan Kaufmann. 2012. Communicating with cost-based implicature: A game-theoretic approach to ambiguity. In *Proceedings of 16th Workshop on the Semantics and Pragmatics of Dialogue*, pages 107–116.
- Eleanor Rosch. 1978. Principles of categorization. In Eleanor Rosch and Barbara B. Lloyd, editors, *Cognition and Categorization*, volume 1, pages 27–78. Lawrence Erlbaum Associates, Hillsdale, NJ.
- Edward E Smith, Edward J Shoben, and Lance J Rips. 1974. [Structure and process in semantic memory: A featural model for semantic decisions](#). *Psychological review*, 81(3). Publisher: American Psychological Association.
- Robert Speer and Catherine Havasi. 2012. Representing General Relational Knowledge in ConceptNet 5. In *LREC*, pages 3679–3686.
- Michael Spranger. 2013. [Grounded lexicon acquisition - case studies in spatial language](#). *2013 IEEE Third Joint International Conference on Development and Learning and Epigenetic Robotics (ICDL)*, pages 1–6.
- Leonard Talmy. 1988. [Force Dynamics in Language and Cognition](#). *Cognitive science*, 12(1):49–100.
- Andrea Tyler and Vyvyan Evans. 2001. [Reconsidering prepositional polysemy networks: the case of over](#). *Language*, 77(4):724–765.
- Karthik Mahesh Varadarajan and Markus Vincze. 2012. [Afnets: The affordance network](#). In *Asian Conference on Computer Vision*, pages 512–523. Springer.

Ludwig Wittgenstein. 1953. *Philosophical investigations*. Macmillan.