Typicality and Object Reference

Margaret Mitchell (m.mitchell@jhu.edu)
Human Language Technology Center of Excellence, Johns Hopkins University
Baltimore, MD 21211 USA

Ehud Reiter (e.reiter@abdn.ac.uk)
Kees van Deemter (k.vdeemter@abdn.ac.uk)
Computing Science Department, University of Aberdeen
Aberdeen, Scotland AB24 3FX UK

Abstract

Does the typicality of an object affect how we identify it? When we produce initial reference to a visible object, we are influenced by a variety of factors, including what is visually salient (bottom-up influences) as well as our previous experiences with the object (top-down influences). In this study, we seek to understand how the top-down influence of *typicality* affects initial reference to an object. We use real world, everyday objects, and focus on the visual properties of SHAPE and MATERIAL. Our findings suggest that there is a tendency to select the atypical over the typical. But we have only begun to scratch the surface of understanding reference to real world objects. The annotated corpus from this study is made available for future research on modeling reference in visual domains.[1]
Keywords: referring expressions; description; reference; vision; typicality

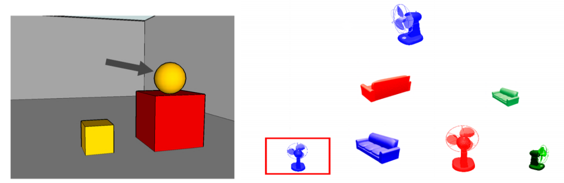
Introduction

I never saw a purple cow.
I never hope to see one,
But I can say this anyhow:
I'd rather see than be one.
— Gelett Burgess

When we identify an object for a hearer, we have a number of choices to make about what to mention. When the object is visible to both speaker and hearer, properties that help guide visual attention, such as color and size, are particularly informative (Treisman & Gelade, 1980; Wolfe, 2006). Properties that are salient to the discourse or relevant to the speaker and hearer's previous interactions also affect what we will mention and describe (Clark & Wilkes-Gibbs, 1986; Brennan & Clark, 1996; Clark & Krych, 2004).

We hypothesize that when we generate initial reference to an object for a hearer, our knowledge about objects of the same type is also likely to affect what we mention. In other words, our understanding of what is *typical* for an object category influences the selection of modifiers – the adjectives and longer descriptive phrases – that we produce when we first describe an object. This understanding of what is typical for an object category may stem from stored object prototypes (Rosch & Mervis, 1975; Rosch, C. Mervis, W. Gray, Johnson, & Braem, 1976) or mental representations of similar objects in previous situations (Yeh & Barsalou, 2006; Wu & Barsalou, 2009). Because of typicality, we know that *the purple cow* mentioned in the example above is remarkable.

Previous work on reference has paid little attention to the role of typicality. This is equally true for psycholinguistic work (Arnold, 2008) and for work on computational models of reference (Dale & Reiter, 1995; Krahmer, van Erk, & Verleg, 2003; Krahmer & van Deemter, 2012). The present study addresses what we believe to be a significant gap in our understanding of reference.



(a) A GRE3D3 scene. (b) A TUNA scene.

Figure 1: Example scenes from the GRE3D3 Corpus and the TUNA Furniture sub-corpus. Participants produce referring expressions such as *yellow ball on top of the red cube* (GRE3D3) or *small fan* (TUNA).

We examine the role of typicality in reference to real world, everyday objects, focusing on material and shape properties. Objects are presented so that one of these two properties will distinguish the object. We test whether there is a significant difference between groups when participants in one may choose between *atypical shape* or *typical material* to describe target objects, while participants in a second group may choose between *typical shape* or *atypical material* to describe target objects. Our findings suggest that there is a tendency to select the *atypical* over the *typical*.

Although this study focuses on shape and material typicality, we release the full corpus from our experiments, annotated with a variety of visual properties, in hopes of helping further work in constructing models of reference to real objects. Current available corpora for reference to visible objects, such as the GRE3D3 Corpus (Viethen & Dale, 2008) or the TUNA Corpus (van Deemter, Gatt, van der Sluis, & Power, 2012), were built from reference elicited to graphics of simple objects presented on a computer screen (see Figure 1). In this work, we seek to better understand the rich details of reference in real world visual domains, where a multitude of different visual properties interact. This opens up several aspects of reference that have not been researched before and gives rise to further questions about the factors influencing initial reference and visual object descriptions. We discuss some of these issues, and their implications for a computa-

[1] http://m-mitchell.com/corpora/typicality_corpus/

tional model of reference.

To establish what properties are typical for an object, we use semantic feature production norms. Semantic feature production norms provide a set of common properties for basic-level concepts, and are collected to explore conceptual representations such as typicality (Rosch & Mervis, 1975) and semantics (Wu & Barsalou, 2009). We use McRae's norms (McRae, Cree, Seidenberg, & McNorgan, 2005; McRae, 2011), which to our knowledge is the largest source of production norms to date. McRae's norms were collected by providing participants with 10 blank lines for each basic category and asking them to list features for each, such as physical (perceptual) properties (how it looks, sounds, smells, feels, and tastes), functional properties (what it is used for and where and when it is used), and other information, such as encyclopedic facts (e.g., where it is from).

For this study, we are interested in perceptual properties, specifically shape and material properties, which are available from the norms.[2] For example, objects belonging to the *bowl* category are listed as typically having a "curved" or "round" shape, and made of a "ceramic" or "plastic" material. We consider *atypical* properties to be properties (1) mentioned by no participants for the object, and (2) difficult to find during our object collection period.

Our initial list of possible objects included all inanimate objects from McRae's norms that could fit on an experiment table, and this set was narrowed down by availability and our abilities to control the visual properties of the objects. The final set of test objects are listed in Table 1, along with their typical and atypical shapes and materials. The full set of objects used in this study are shown and labeled in Figure 3.

Using Real World Objects

A notable complication in this study is that we seek to use a variety of everyday objects, while controlling the typicality of particular properties of those objects. This means that the objects must look relatively commonplace, matching as closely as possible on every visual property except for shape/material; and for these properties, one must be clearly atypical while the other must be clearly typical. Finding everyday objects that fit within these rigid constraints is difficult. In some cases (bowl, mug, screw), we colored the objects to match one another, while in other cases (atypical envelope, key, ruler), we physically created the objects in order for them to have all the desired properties. Real world objects bring with them a set of complications for any model of reference, and we discuss some of these briefly. Although we cannot address all of the issues we list, we hope to provide evidence for a preliminary model of typicality in reference while bringing to light areas for further research.

Cultural and Individual Differences

What is typical for an object varies person to person, culture to culture. This study was conducted with a range of students and professionals in two countries (the U.S. and the U.K.), but ideally in testing and modeling the production of reference, the set of typical properties would be defined with respect to a culture or group of people, or tailored to a specific person. For our study, we use one set of objects, without changing typical/atypical properties.

Interconnection

It is clear that there is an *interconnection* between different visual properties. For example, material often entails color and texture. An object made of wool is fuzzy or rough (texture values), while an object made of wood is usually tan or brown, and for everyday objects, tends to be smooth (color and e.g., smoothness values). Ideally, participants would refer only to those properties that we vary; but they may instead refer to interconnected properties, calling a woolen bowl "coarse" or "flexible", or a mug made of ceramic but painted silver a "metal mug".

Lexicalization

Another competing factor in this study is how easy a property is to lexicalize. Some shapes (e.g., "square") are common and may be quick to access, while other shapes (e.g., "octagonal") may take longer to produce, affecting the object description. Further complications may arise when there is competition over whether to use a prenominal modifier ("the flower-shaped bowl") or a postnominal modifier ("the bowl that looks a bit like a flower").

Shapes, Parts, and Object Categories

Figure 2: Bowl, Sugar Bowl, Creamer, Teacup, Mug, Pitcher: Similar objects with different shapes tend to have different names.

An object's shape is often indicated by its name (Markman, 1989; Landau & Jackendoff, 1993), and therefore an object designed to have an *atypical* shape may instead appear to belong in an entirely different object category (see Figure 2). For some objects, we found that changing its full shape made it unclear what the object was, or else created a subtype of our target basic-level object category; in a few cases, we therefore manipulated a *part* of the object's shape. Rather than a round head of a key, the head was square; the straight rectangular center of a ruler was cut out with geometric shapes; and the circular head of a screw was made atypically oval.

The Objects

The objects used in the study, as they were presented to participants (without the superimposed identifiers), are shown in Figure 3. Test objects are listed in Table 1 along with their shapes and materials. Those values in italics could not be found in McRae's norms and were added based on intuition and object availability. Filler objects are listed in Table 2.

[2] We use McRae et al.'s "external_surface_property"/"external_component" labels for shape and "made_of" labels for material.

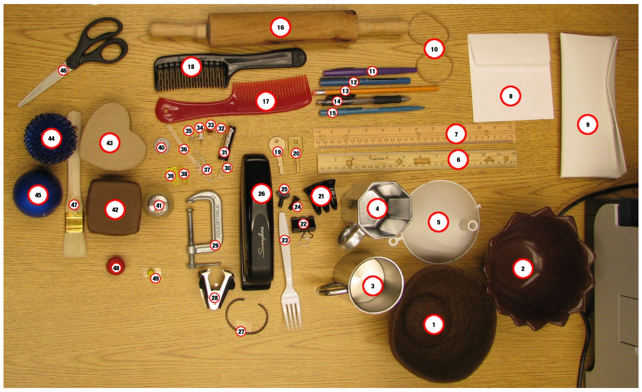


Figure 3: Objects used in study, keyed to descriptions in Tables 1 and 2 below.

Table 1: Test objects with shapes and materials. Values listed in italics were provided by the authors because they were not listed in the McRae et al. (2005) norms.

	Group 1			Group 2		
	Atypical Shape			Atypical Material		
Object	ID	Shape	Mater.	ID	Shape	Mater.
bowl	2	flower	ceramic	1	round	wool
box	43	heart	cardboard	42	square	clay
envelope	8	square	paper	9	*rectangle*	foam
key	20	square head	metal	19	*rounded* head	wood
mug	4	octagonal	ceramic	3	*round*	metal
ruler	6	with holes	wood	7	rectangle	paper
screw	25	oval head	metal	24	flat *circular* head	plastic

Table 2: Filler objects. See Figure 3 for corresponding images of the objects.

44	ball	40	coin	33	pushpin
45	ball	17	comb	34	pushpin
31	battery	18	comb	35	pushpin
27	bracelet	48/49	cube*	16	rolling-pin
29	c-clamp	23	fork	10	rubber-band
21	clip	5	funnel	41	salt-shaker
22	clip	11	pen	46	scissors
30	clip	12	pen	48/49	sphere*
37	clip	14	pen	28	staple-remover
38	clip	15	pen	26	stapler
39	clip	13	pencil	36	toothpick
		32	pushpin		

* These objects were varied by color/size/type as part of a separate pilot experiment.

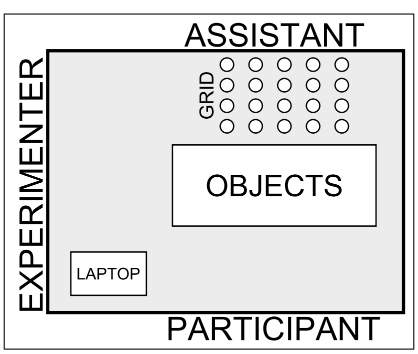


Figure 4: Subjects sat across from an assistant, with objects between them. An experimenter sat at the head of the table, moving objects to their original positions between trials.

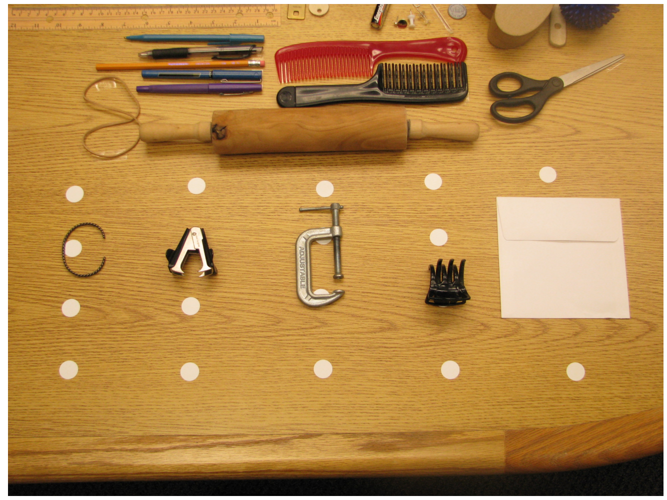


Figure 5: Example stimuli, Atypical Shape group. Here, the test object is the square envelope.

Experiment

Participants sat at a table across from an assistant (Figure 4), with a variety of objects on the table between them (Figure 3). Subjects were asked to explain to the person sitting across how to recreate images of the objects grouped in different patterns (Figure 5). There were two objects for each test object category on the table, matched for color and size. One object had atypical material and typical shape; the other had atypical shape and typical material. Subjects could therefore not distinguish a test object to the hearer by its type alone, but could distinguish it by mentioning its shape or material.

Atypical feature is a grouping variable both by subjects and by materials. For the Atypical Shape participants, shape properties of the test objects were atypical, while material properties were typical. For the Atypical Material participants, material properties of the test objects were atypical, while shape properties were typical.

Method

Participants Thirty native English speakers with normal or corrected vision in the United States and the United Kingdom were paid for their participation ($5 or £5). Subjects were recruited through word of mouth and online ads, 17 males and 13 females, aged 20–55, and randomly assigned to one of the two experimental groups, (1) Atypical Shape or (2) Atypical Material. Four male subjects and one female subject were randomly removed to a held-out set to balance gender, leaving 6 female and 6 male subjects in each group (24 subjects total).

To check for possible outliers in each group, we calculate the average number of references with shape, and the average number of references with material. Participants whose total number of references with shape or material were more than two standard deviations from the mean for that property were identified as possible outliers. We found no outliers in the Atypical Material group, and two possible outliers in the Atypical Shape group. The data for these two subjects (one male, one female) were removed and replaced with gender-matched data from the held-out set.

Materials Participants sat in front of a large set of everyday objects (rulers, envelopes, pins, etc., as shown in Figures 3 and 4), with test objects and fillers mixed. Test objects for the two experimental groups with their corresponding shape and material properties are listed in Table 1.

Procedure & Design All participants consented to participate in the study. The experiment followed a director-matcher paradigm, where the director (the participant) instructed the matcher (the assistant). Participant and assistant sat opposite one another while the experimenter sat on a third side of the table. Participants alone viewed pictures on a laptop (positioned so screen was not visible to the assistant). Each participant saw 8 pictures in randomized order including a different atypical stimulus each time (see example in Figure 5). On each trial, the participant viewed the picture and explained to the assistant where each pictured object should go on the grid laid out on the table between them. At the end of each trial, the experimenter returned objects to their original positions. Participants' instructions were recorded onto the laptop.

Results

Annotation Annotations are provided by the first author. To test for adequacy of the annotation system, a second annotator annotated a random subset of 20 references to the test objects. The annotator was given mark-up instructions and examples of a variety of visual properties (shown in Table 3), and told to mark which words referred to which properties *as best they could* following their understanding of the examples. Treating shape and material as binary categorical variables, Cohen's κ is very good for shape (κ=.894) and good for material (κ=.798) between annotators. Disagreements were over whether "metallic" in "the non-ribbed metallic cup" was a material or a texture, whether "heart" in "a heart-shaped box" was a shape or a type, and whether "silver" in "a silver round cup" was a color or a material. The total number of expressions produced for test items with shape and material modifiers in each experimental group is shown in Figure 6.

Analysis We see a tendency to choose the atypical over the typical in both groups. In the ATYPICAL SHAPE group, 54 expressions contain a shape modifier while 36 contain a material modifier; in the ATYPICAL MATERIAL group, 28 expressions contain a shape modifier while 41 contain a material modifier (see Figure 6). There is a slight preference for shape over material across the groups; 82 expressions contain a shape modifier while 77 contain a material modifier.

We are interested in understanding whether there is a significant difference in the selection of modifiers between groups. For each participant, we subtract the number of test object expressions containing a modifier for the object material from the number of test object expressions containing a modifier for the object shape. In other words, for each participant p, given the number of expressions with material modifiers M_p and the number of expressions with shape modifiers S_p, we calculate $V_p = S_p - M_p$. This provides a vector for each group with differences in the number of modifiers for shape and material. For participants in the ATYPICAL SHAPE group, where shape is atypical, we expect positive values for V_p. For participants in the ATYPICAL MATERIAL group, where material is atypical, we expect negative values for V_p.

An independent samples t-test was conducted to compare the effect of property typicality on the production of modifiers between groups. There is a significant difference at $\alpha = .01$ between the ATYPICAL SHAPE group (n=12, mean=1.50, sd=1.62) and the ATYPICAL MATERIAL group (n=12, mean=-1.08, sd=2.07); t(21)=3.406, p=0.003.

Table 3: Attributes annotated and example surface forms.

ANALOGY	"like a cat"	SHAPE	"square"
COLOR	"blue"	SHEEN	"shiny"
FILL	"empty"	SIZE	"little"
FLEX	"flexible"	SUBJECTIVE	"weird"
FORM	"open"	SUBTYPE	"ball point"
HARDNESS	"hard"	TEXTURE	"rough"
INTENSITY	"dark"	TYPE	"box"
LOCATION	"close to me"	USE	"for oil"
LUMINESCENCE	"bright"	WEIGHT	"light"
MATERIAL	"copper"	**Examples of phrases conveying more than one attribute**	
OPACITY	"clear"	ANALOGY:SHAPE	"shaped like a P"
OTHER	"other", "another"	TYPE:SHAPE	"diamond"
PART-WHOLE	"with a slot on top"	PROCESS:ANALOGY	"that opens like a purse"
PROCESS	"adjustable"		

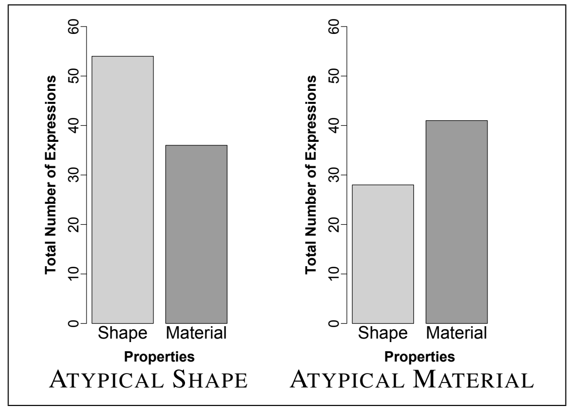


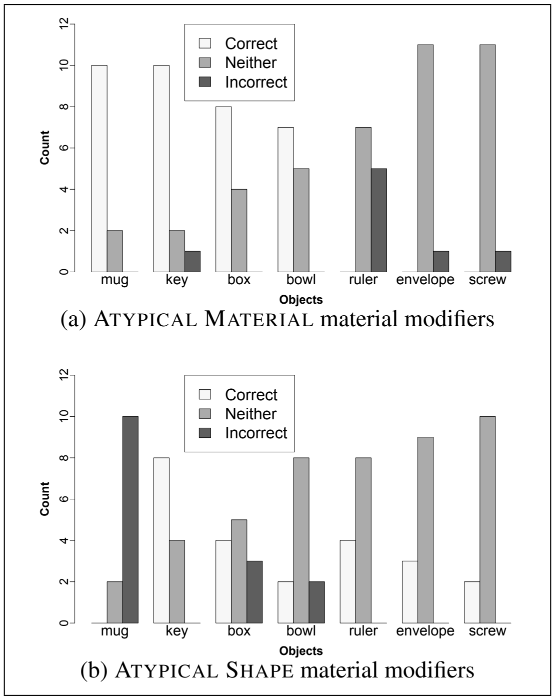
Figure 6: Number of expressions with shape and material modifiers in each experimental group.

Discussion

Current Study

These results suggest that atypicality affects object reference. We find a tendency to select the atypical property over the typical one, with participants in the ATYPICAL SHAPE group preferring shape modifiers, and participants in the ATYPICAL MATERIAL group preferring material modifiers. This difference is significant between groups.

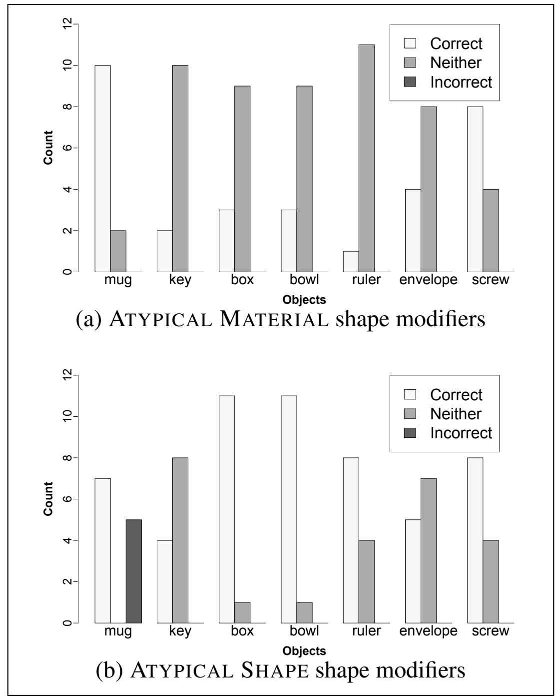
Shedding further light on these findings, when material was included in a reference in the ATYPICAL MATERIAL group, it was often incorrect. Figure 7 illustrates how frequently subjects were correct and incorrect in the description of an object's material. In the ATYPICAL MATERIAL group (Figure 7a), the plastic screw was called "metal", the paper ruler was called "wooden". The ruler in particular gave rise to incorrect material modifiers – it was printed on paper with a wood print,



(a) ATYPICAL MATERIAL material modifiers

(b) ATYPICAL SHAPE material modifiers

Figure 7: Number of participants who included material modifiers that were correct, incorrect, or did not use a material modifier at all (neither). In the ATYPICAL MATERIAL group (a), ruler tends to evoke incorrect material modifiers. The ruler, envelope, and screw had no correct material modifiers. In the ATYPICAL SHAPE group (b), the mug, box, and bowl tended to evoke incorrect material modifiers. The mug (painted silver) had no correct material modifiers.



(a) ATYPICAL MATERIAL shape modifiers

(b) ATYPICAL SHAPE shape modifiers

Figure 8: Number of participants who included shape modifiers that were correct, incorrect, or did not use a shape modifier at all (neither). In the ATYPICAL MATERIAL group (a), there were no incorrect shape modifiers. In the ATYPICAL SHAPE group (b), only the mug received incorrect shape modifiers (e.g., "hexagonal" rather than "octagonal").

so it was called "wooden". Most participants in the ATYPICAL MATERIAL group did not use material modifiers for the envelope and the screw, which may be partially due to the fact that the screw was painted black and so was not clearly plastic; and the envelope was made of foam, which may have not been clear without physically touching the object. Some examples of expressions in the ATYPICAL MATERIAL group that do not include material modifiers are given in Table 4.

A similar tendency to refer to incorrect material emerged in the ATYPICAL SHAPE group (Figure 7b), where the ceramic mug painted silver was called "metal" or "steel", the ceramic bowl was called "plastic", and the cardboard box was called "wooden". In contrast, subjects were rarely incorrect about shape (Figure 8). The only exception to this is the atypically shaped mug (Figure 8b), which was called "octagonal" (correct), "hexagonal" (incorrect) and "septuplet" (incorrect).

These trends suggest that material is not purely visual, but may also be guided by our tactile sensations of the objects; without tactile input, our expectation of the typical material for the object may be used in our reference rather than its actual material, or we may disprefer material altogether. It is not enough to judge whether a visual property is atypical or not; it must also be judged whether that value is visually clear, and whether other properties suggest another interconnected property (which may not actually be true of the referent, as when subjects use incorrect modifiers).

Future work could shed more light on these issues with a variant of the experiment in which we look at how reference here compares to reference towards objects of the same category that have typical shape and typical material, and likewise, atypical shape and atypical material. This should also provide useful data on how to separate general preference for shape or material versus a preference for atypical properties.

Towards a Model of Reference

As discussed in the introduction, current research on reference production has focused on very simple, constrained domains. In this work, we propose taking the object's typicality into consideration when deciding which properties to add to an initial description. In a computational model, typical information may be made available in a knowledge base queried at runtime. As part of the reference production process, the target object category could then be compared against a stored object category. Property selection in such a model could be a function of the typicality of the property for the object, as well as, e.g., its contrastive value against the other objects. This offers an extension to current models of referring expression generation (e.g., Dale and Reiter (1995)), and may help to further explain the process of reference generation.

Ruler	Envelope	Screw
“the ruler”	“the white envelope”	“the screw”
“ruler that’s flatter”	“the uh weird padded looking envelope thing”	“black flat head screw”
“the darker tan ruler”	“long rectangular envelope”	“the screw with the flat head”

Table 4: Examples of references without MATERIAL in the Atypical MATERIAL group. We see underspecified references and references describing the object’s size.

Conclusions and Future Work

This study has sampled a handful of real world objects to understand the role that typicality plays on reference. We have focused on two visual attributes, shape and material, in a visual scenario where either may be used to identify an object. We see a tendency to select the atypical property over the typical one, and find a significant difference between the selection of atypical shape over typical material versus typical shape over atypical material between groups.

Our study has focused on relatively crisp properties of objects. It would be interesting to explore whether our findings extrapolate to properties that come in degrees, such as height, weight, age, and so on. Based on our study, we hypothesize, for example, that the length of a screw is more likely to be mentioned if it is unusual (e.g., unusually long or unusually short) than if it is not. In line with this, gradations of atypicality for an attribute may also affect reference; some values may be more atypical than others, and thus more likely to be included in a final description.

This study leaves many further open questions. To fully model reference production to real world, visible objects, we must better understand how the production of visual modifiers is affected by interconnection and lexicalization issues, and how notions of typicality are changed culture to culture.

In future work, we aim to focus directly on interconnection, understanding how the degree of correlation between properties affects description. We would like to extend our set of objects in order to examine reference when both properties are typical, or both atypical. There may be a tendency to select the atypical over the typical, but we have only begun to scratch the surface of the factors at play when we refer to real world objects.

Acknowledgments

The first author of this paper was supported by funding from the Scottish Informatics and Computer Science Alliance (SICSA), the Overseas Research Students Awards Scheme (ORSAS) and Oregon Health and Science University (OHSU). Thanks to Ellen Bard at the University of Edinburgh for help with design and analysis. Thanks to students at OHSU for volunteering their time as assistants and thanks to SignWorksOfOregon – particularly Shannon Zuiderweg, for the construction of the wooden key.

References

Arnold, J. E. (2008). Reference production: Production-internal and addressee-oriented processes. *Language and Cognitive Processes*, *23*.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*.

Clark, H. H., & Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language*, *50*.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*.

Dale, R., & Reiter, E. (1995). Computational interpretations of the gricean maxims in the generation of referring expressions. *Cognitive Science*, *19*.

Krahmer, E., & van Deemter, K. (2012). Computational generation of referring expressions: A survey. *Comp. Ling.*, *38*.

Krahmer, E., van Erk, S., & Verleg, A. (2003). Graph-based generation of referring expressions. *Comp. Ling.*, *29*.

Landau, B., & Jackendoff, R. (1993). “what” and “where” in spatial language and spatial cognition. *Behavioral and Brain Sciences*, *16*.

Markman, E. (1989). *Categorization and naming in children: Problems of induction*. Cambridge, MA: MIT Press.

McRae, K. (2011). *McRae’s norms.* Available from `http://amdrae.ssc.uwo.ca/McRaeLab/norms.php.` `Accessed 5.Oct.2011`

McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods, Instruments, & Computers*, *37*(4).

Rosch, E., C. Mervis, C., W. Gray, W., Johnson, D., & Braem, P. B. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*.

Treisman, A. M., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology*, *12*.

van Deemter, K., Gatt, A., van der Sluis, I., & Power, R. (2012). Generation of referring expressions: Assessing the incremental algorithm. *Cognitive Science*, *36*(5).

Viethen, J., & Dale, R. (2008). The use of spatial relations in referring expression generation. *Proc. INLG 2008*.

Wolfe, J. M. (2006). Guided search 4.0. In W. D. Gray (Ed.), *Integrated models of cognitive systems.* New York, NY: Oxford University Press.

Wu, L. L., & Barsalou, L. W. (2009). Perceptual simulation in conceptual combination: Evidence from property generation. *Acta Psychologica*, *132*.

Yeh, W., & Barsalou, L. W. (2006). The situated nature of concepts. *American Journal of Psychology*, *119*.