Attributes in Visual 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

Introduction

An emerging body of research has begun to look at how people refer to visible referents (Gatt, Belz, & Kow, 2009; Viethen, Goudbeek, & Krahmer, 2012; Mitchell, Deemter, & Reiter, 2013). However, the selection of which visual attributes are most important to study has been somewhat ad-hoc. In this work, we extensively annotate and report on trends from two corpora of visible object references. We seek to make an initial step towards understanding what kinds of attributes people use when referring in real-world visual domains, how frequent these attributes are, and what this suggests for research in referring expression generation (REG).

Below, we examine the attributes used to refer to visible objects in two corpora, the Craft Corpus available from (Mitchell, van Deemter, & Reiter, 2010), and the Typicality Corpus available from (Mitchell, Reiter, & van Deemter, 2013).[1] We refer interested readers to the relevant papers for further details on data collection.

In these domains, in contrast to semantically transparent domains such as those available in the TUNA Corpus (van Deemter, van der Sluis, & Gatt, 2006) or the GRE3D3 Corpus (Viethen & Dale, 2008), the objects are visually complex. It is therefore not always clear what all the different attributes relevant for a referent are. We annotate those properties used by participants. Although the attributes in any one corpus are not representative of the attributes that people use *in general*, by examining the attributes used for real, everyday objects, we can begin to understand how traditional assumptions made in REG fare in real-world scenes.

We find that several assumptions may need to be revised to better fit to generating reference in visual domains. For example, previous research has assumed that domains can be adequately modeled using a set of <attribute:value> pairs, where one of the attributes is TYPE; but we find that the notion of a TYPE is not always clear-cut. Other types of reference, including reference to parts (*part-whole modularity*) are also extremely common, although current REG models are not built to refer in this way. We also see utterances that suggest a more complex knowledge representation than previously made available in REG corpora.

[1] Both are available at http://m-mitchell.com/corpora.html

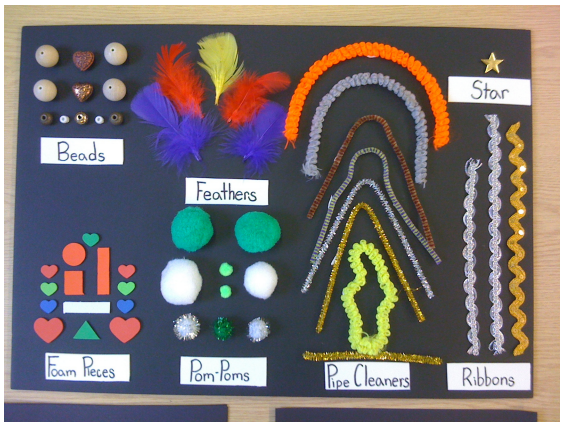

Figure 1: Objects used in (Mitchell et al., 2010).

Corpus 1: Crafts

The first corpus we use is available from (Mitchell et al., 2010), which provides references and descriptions of craft objects. Craft items are bundles of visual properties by design, and so serve as a rich source of information about how different visual attributes are realized linguistically. Objects were chosen from various craft sets, and included pipe-cleaners, beads, feathers, etc., with different colors, textures, shapes, patterns, and were made of different materials. Subjects recorded directions for how to create craft "faces", using specific objects on a board (see Figure 1).

Annotation. Initial single object references were annotated as <attribute:value> pairs and extracted. For example, the reference *the small heart red foam-piece* was annotated as <size:small, color:red, type/material:foam-piece, shape:heart>.[2]

A difficulty that arose during annotation for this study was that an object TYPE was often not distinct from SHAPE or MATERIAL. As shown in Table 1, references may include a SHAPE or MATERIAL word as a head noun to refer to the object; but referential head nouns in REG are commonly considered to be realized from a distinct TYPE attribute. For example, several participants used the referential noun phrase "a heart foam-piece", where a MATERIAL attribute was realized as the head noun, and "a heart", where the SHAPE attribute

[2] Additional annotators to measure inter-annotator agreement were not available for this study.

Attribute	Count	Example Values
COLOR	594	red, green, silver, yellow
SIZE	192	big, medium, short, thick
SHAPE	120	heart, circle, ball, square
TYPE/MATERIAL	94	foam piece
TYPE/SHAPE	89	heart, square, rectangle
MATERIAL	73	foam, wooden, plastic
FORM	36	bent, twisty, straight
SHEEN	22	sparkly, glitter, shiny
TEXTURE	16	fluffy, fuzzy, furry
ANALOGIES	13	"like a fuzzy caterpillar"
ORIENTATION	12	upside-down, horizontal
PART-WHOLE	11	"with tinsel on the outside"
PATTERN	3	striped, (with a) pattern
LOCATION	1	"at the bottom"

Table 1: Attributes in Craft Corpus.

was realized as the head noun. This mirrors psycholinguistic research on object naming that suggests that different objects may be given the same name if they are the same shape (Landau & Jackendoff, 1993) or made of the same material (Markman, 1989), and suggests that if an explicit TYPE attribute is not available for reference, a SHAPE or MATERIAL attribute may be used to generate a head noun for the referent.

The Attributes. A list of the attributes with counts and example values is shown in Table 1. In line with recent REG work, we find that COLOR and SIZE are dominant visual attributes. However, these are followed by SHAPE and MATERIAL, which have not been examined in detail by the NLG community. LOCATION, which has received significant attention in computer-generated graphic visual domains (Viethen & Dale, 2008; Kelleher & Kruijff, 2006), is all but absent from the references produced in this domain.

We also observe that one attribute may sometimes be inferred from another. For example, although COLOR was a preferred attribute, many people referred to *the wooden bead* rather than, e.g., *the brown bead*; presumably, the fact that something is wooden implies typical color information, in addition to typical texture, opacity, etc.

The distribution of number of modifiers used for a referent is shown in Table 2. We find that participants tended to use one or two modifiers, including those realized prenominally (e.g., adjectives) and postnominally (e.g., relative clauses). Example modifying phrases are shown in Table 5. To the best of our knowledge, research in REG does not yet provide information about when to realize a property prenominally or postnominally, nor the possible syntactic forms it may take in each position.

Corpus 2: Everyday Objects

A notable problem with the previous corpus is that the objects are not common; the language annotated in this corpus may be different than the language used to describe more "everyday" objects. We address this issue by constructing a corpus of common objects released with this work.

Count	Frequency	Count	Frequency
0	14 (2.7%)	3	41 (7.9%)
1	248 (47.5%)	4	7 (1.3%)
2	211 (40.4%)	6	1 (0.2%)

Table 2: Distribution of Modifier Counts in Craft Corpus. Some attributes were repeated.

	Prenominal modifiers	Head noun	Postnominal modifiers
the	slightly darker small green	heart	
<SIZE:*small,* COLOR:*dark-green,* TYPE/SHAPE:*heart*>			
the	yellow	pipe-cleaner	that's in a circle
<COLOR:*yellow,* SHAPE:*circle,* TYPE:*pipe-cleaner*>			
a		pipe-cleaner,	which is straight and it's golden
<COLOR:*gold,* TYPE:*pipe-cleaner,* FORM:*straight*>			
the	orange	pipe-cleaner,	just like a kind of cotton material
<COLOR:*orange,* TYPE:*pipe-cleaner,* LIKE:<MATERIAL:*cotton*>>			

Table 3: Example prenominal and postnominal modifiers in Craft Corpus, with annotations in italics.

.

The Corpus. Thirty-five objects ("Everyday Objects") were selected randomly from an office and a home. Subjects instructed a hearer to place these objects in different spatial configurations (see Table 4). 14 additional objects used in this study were carefully controlled for shape/material typicality, and we do not include a discussion of these objects here.

Annotation. Inter-annotator agreement from two annotators on the study items from this corpus was "good" (Cohen's $\kappa > 0.75$). A difficulty that arose during annotation was the difference between subtype and type. For example, the broad label *clip* may be applied to the *paperclip* as well as *hair clip* and *binder clip*. Compound nouns of this sort may define a subtype label, basic category label, or a label with modifiers.

The Attributes. A list of the attributes with counts and example values is shown in Table 6. COLOR, SIZE, MATERIAL and SHAPE again emerge as the most common attributes.[3] We also find frequent examples of *part-whole* modularity, whereby speakers refer to parts of an object to identify the object as a whole. For example, "black-handled" and "with

[3]However, it should be noted that because the test items in this elicitation experiment were controlled for MATERIAL and SHAPE, they may have affected participant responses to the Everyday 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.

Table 4: Objects reported in this study. See Figure 2 for corresponding images of the objects.

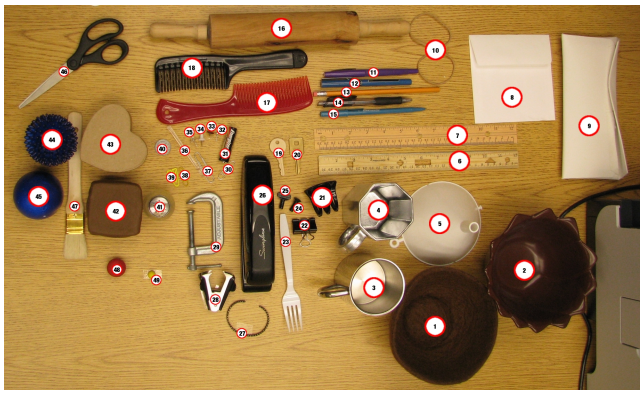

Figure 2: Objects used in study, keyed to descriptions in Table 4.

a wooden handle" pick out the handle part of an object and refers to attributes of the part (COLOR and MATERIAL). "With the lid at the top" picks out both a part of the object ("lid") and its location relative to the whole ("at the top"). To our knowledge, a model for generating part-whole relations has not yet been developed.

Approximately 2% of the references in this corpus additionally include analogies. For example, "shaped like a P" is an analogy to define SHAPE. "That opens like a purse" is an analogy to describe a PROCESS that the object may undergo. People appear to compare current visual object properties against typical properties of objects stored in memory when referring. This was particularly apparent when they did not know the name or category of the object.

The distribution of number of modifiers used for a referent is shown in Table 7. As before, we find that participants rarely used more than three modifiers, with the largest preference for just one modifier.

Discussion

We have provided an overview of some of the phenomena that emerge from reference to real-world, visual objects. We find that many of the traditional assumptions in REG may need to be revised or refined in order to generate natural reference in visual domains. We also discover evidence for much richer knowledge representations than have previously been provided in REG corpora.

	Prenominal modifiers	**Head noun**	**Postnominal modifiers**
the	blue	ball	that's not smooth
<COLOR:*blue,* TEXTURE:*not-smooth,* TYPE:*ball*>			
the	metal	tool	that has like a handle on it
<MATERIAL:*metal,* TYPE:*tool,* PART-WHOLE:*handle*>			
a	triangular black	object	with two ears
<COLOR:*black,* SHAPE:*triangular,* TYPE:*object,* PART-WHOLE:<COUNT:*two,* TYPE:*ears*>>			

Table 5: Prenominal and postnominal modifiers in Craft Corpus, with annotations in italics.

In line with previous work in referring expression generation, the current corpora support the idea that COLOR and SIZE are useful attributes to research when trying to understand visual properties in natural reference. However, beyond this, we find that several assumptions may need to be further examined. Object TYPE is not always clear, as the distinction between TYPE and SUBTYPE may not be obvious (e.g., is *paperclip* a subtype of *clip*? When do we choose one over the other?). There may not always be a TYPE value distinct from other properties, such as SHAPE and MATERIAL, which appear to be particularly salient in visual domains.

Further, properties may not be completely distinct from one another, with something like MATERIAL:*wood* implying related properties (e.g., COLOR:*brown* and TEXTURE:*wood-grain*); and we find a preference for mentioning these kinds of natural materials. This suggests that when selecting which properties to refer to, people may take into account the amount of information each property conveys, selecting properties that entail or imply several others.

We see frequent occurrences of part-whole modularity, where speakers refer to parts of an object within reference to the whole, which has received little attention in REG, and may require richer knowledge representations. Traditionally, objects are represented as a series of attribute:value pairs with a single type; but the description of parts that we observe within description of a whole object suggests that a recursive structure may be a more suitable representation in a domain of real-world objects.

We also find analogies, which require comparing the properties of an unknown object against properties stored in a knowledge base in order to determine what an object is sim-

Attribute	Count	Example Values
COLOR	165	brown, red, yellow
SIZE	75	big, small, skinnier
MATERIAL	74	metal, plastic, wooden
SHAPE	58	triangular, heart-shaped
PART-WHOLE	50	with a handle, with the cap
PROCESS	22	adjustable, unsharpened
LOCATION	13	here, there
USE	10	for removing staples
ANALOGY	10	sea urchin, kind of claw
TEXTURE	4	smooth, fuzzy, hairy
OTHER	4	other, another
INTENSITY	3	lighter, light
COUNT	2	two
WEIGHT	1	heavy
SHEEN	1	shiny
PATTERN	1	pattern
OPACITY	1	clear
LUMINESCENCE	1	bright
HARDNESS	1	hard
FILL	1	empty

Table 6: Everyday Object attributes in Typicality Corpus.

Count	Frequency	Count	Frequency
0	219 (32.5%)	3	45 (6.7%)
1	251 (37.2%)	4	5 (0.7%)
2	153 (22.7%)	5	1 (0.0%)

Table 7: Distribution of modifier counts for Everyday Objects in Typicality Corpus. Some attributes were repeated.

ilar to when its category is unknown. In future work, it may be possible to refer to an unknown object by analyzing its properties and finding the object with the closest match in the knowledge base; this may then be used to say an object is *kind of like* something else.

It is clear from this work that current models for REG may be further expanded by considering the rich complexities of real world objects, including further details of how properties are represented and analyzed, which properties are important to focus on, how object TYPE is considered, and how target referents are compared against stored objects. We hope that these findings suggest several new avenues to explore in generating reference to visible 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 anonymous reviewers for useful suggestions.

References

Gatt, A., Belz, A., & Kow, E. (2009). The TUNA REG challenge 2009: Overview and evaluation results. *Proceedings of the 12th European Workshop on Natural Language Generation (ENLG 2009)*, 174–182.

Kelleher, J., & Kruijff, G.-J. M. (2006). Incremental generation of spatial referring expressions in situation dialog. *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the ACL (COLING-ACL 2006)*, 1041–1048.

Landau, B., & Jackendoff, R. (1993). "what" and "where" in spatial language and spatial cognition. *Behavioral and Brain Sciences*, *16*, 217–265.

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

Mitchell, M., Deemter, K. van, & Reiter, E. (2013). Generating descriptions of visible objects. *Proceedings of NAACL*.

Mitchell, M., Reiter, E., & van Deemter, K. (2013). Typicality and object reference. *Proceedings of CogSci 2013*.

Mitchell, M., van Deemter, K., & Reiter, E. (2010). Natural reference to objects in a visual domain. *Proceedings of the 6th International Natural Language Generation Conference (INLG 2010)*.

van Deemter, K., van der Sluis, I., & Gatt, A. (2006). Building a semantically transparent corpus for the generation of referring expressions. *Proceedings of the 4th International Conference on Natural Language Generation (INLG 2006)*.

Viethen, J., & Dale, R. (2008). The use of spatial relations in referring expression generation. *Proceedings of the 5th International Natural Language Generation Conference (INLG 2008)*, 59–67.

Viethen, J., Goudbeek, M., & Krahmer, E. (2012). The impact of colour difference and colour codability on reference production. *Proceedings of the 34th Annual Meeting of the Cognitive Science Society (CogSci 2012)*.