On the Use of Size Modifiers When Referring to Visible Objects

Margaret Mitchell (m.mitchell@abdn.ac.uk)
Kees van Deemter (k.vdeemter@abdn.ac.uk)
Ehud Reiter (e.reiter@abdn.ac.uk)
Computing Science Department, University of Aberdeen
Aberdeen, Scotland, U.K.

Abstract

We present a study on how people use size modifiers when referring to visible objects. We find strong evidence that the selection of modifiers like *tall*, *thin*, and *big* is brought about by several interacting factors, including how a target object's physical dimensions differ from another object of the same type, and the relationship between the target object's individual dimensions. Findings from this study are used to inform the design of a referring expression generation algorithm capable of referring to objects naturally, providing a further link between visual cues and corresponding linguistic forms.

Keywords: size adjectives; size modifiers; visual features; referring expression generation

Introduction

Over the past two decades, detailed psycholinguistic models of utterance planning have emerged (Ferreira & Swets, 2002; Griffin & Bock, 2000; Levelt, 1989; Levelt, Roelofs, & Meyer, 1999). These models seek to explain the relationship between thought and language, connecting internal mental processes to the timing and structure of produced expressions. A significant amount of recent work has focused on the relationship between the visual world and the references used to identify items therein (Bock, Irwin, Davidson, & Levelt, 2003; Henderson & Ferreira, 2004), but this research has been underutilized in computational approaches to modeling a language generation process (Dale & Reiter, 1995; Krahmer, van Erk, & Verleg, 2003).

It has been well established that dimensional modifiers, such as those denoting size, play a central role in reference to objects in a visual scene, particularly when objects of the same type are in the scene (Brown-Schmidt & Tanenhaus, 2006; Sedivy, 2003). This property of reference is not only important for work in referring expression generation (REG) that uses size modifiers (Kelleher, Costello, & Genabith, 2005; van Deemter, 2004; Viethen & Dale, 2008), but it offers a clear link between language generation and machine vision techniques that provide detailed information about an object's physical dimensions (Friedland, Jantz, & Rojas, 2005; Zheng, Yuille, & Tu, 2010). Systematically manipulating the visual feature of size to develop an account of how size is used in reference furthers the goal of developing a grounded semantic core for natural language (Gorniak & Roy, 2004), tying visual perception to linguistic reference.

In this study, we seek to better understand the relationship between an object's dimensions and the words used to identify it. We evaluate three hypotheses that explore this relationship. Our results suggest that the selection of size modifiers is governed by several interacting and competing factors, with preferences for overall size modifiers ("big", "small") versus individual-axis size modifiers ("tall", "thin") emerging in different contexts. Additionally, we are able to confirm earlier findings on modifier preferences grounded in physical object properties (Hermann & Deutsch, 1976), and further build on these results. This research will inform a natural language generation (NLG) system that refers to real-world items naturally, and provides a fundamental connection linking natural language generation to a vision-based input.

Background and Motivation

Methods for reasoning about the basic properties common to all visual scenes have isolated the properties of color, location, size, and type as the building blocks for visual reference (Roy & Pentland, 2002; Skočaj et al., 2007). Detailed accounts of several of these factors have been developed, including how to produce natural expressions with appropriate use of color modifiers (Mojsilović, 2005) and spatial descriptions (Gorniak & Roy, 2004; Kelleher et al., 2005).

However, our knowledge of how people use size modification to refer to an object is extremely limited. There has been considerable research on the behavior of size modifiers for other purposes, such as the semantics of dimensional modifiers (Bierwisch & Lang, 1989; Eilers, Oller, & Ellington, 1974; Tucker, 1998; Morzycki, 2009), the acquisition of the meaning of such modifiers (Bartlett, 1976), when dimensional modifiers are used (Brown-Schmidt & Tanenhaus, 2006; Sedivy, Tanenhaus, Chambers, & Carlson, 1999), and how language reflects dimensional properties such as height and width (Landau & Jackendoff, 1993; Landau, 2001). We also know roughly how to choose between different forms of a size adjective ("larger", "largest") (van Deemter, 2004).

A primary open question this research leaves is whether people distinguish objects by focusing on one single dimension or by combining dimensions, and how these are realized as surface forms. Given information about an object's height and width, it is unclear how it will be referred to.

Most REG algorithms presuppose that referents are individuated using "absolute" properties, whose applicability to a referent does not depend on the context in which the referent appears. Size is no exception. Dale and Reiter (1995), for example, let their algorithms start from a Knowledge Base in which some objects are listed explicitly as large, while others are listed as small. Van Deemter (2000, 2004) modifies this procedure by storing actual sizes (e.g., in centimeters) in the Knowledge Base, making the decision of whether something is larger or smaller context dependent. However, neither of these approaches pays attention to the choice between words like "big" and "tall"; presumably, this choice is made by a later module that translates properties into words.

But these words may mean something very different and reflect different properties of a referent. For example, consider an object A that is taller and wider than an object B. It is true that A is *taller* than B; it is true that A is *wider* than B; it is also true that A is *bigger* than B. All three words may be appropriate to refer to A, and we do not know whether there is a preference for one over the other. Landau and Jackendoff (1993) point out that a modifier like "big" selects different dimensions depending on the nature of the object, and tends to be used in cases where an object is large in either two or all three of its dimensions, while modifiers like "thick" and "thin" may be applied when an object extends in a single dimension.

Some information about what to expect in a computational model of size modification is provided by Hermann and Deutsch (1976), who find that subjects are more likely to use words like "fat" rather than "short" when a candle is much fatter but only a little shorter than a comparator. In another vein of computational work, Roy (2002) finds that words like "small" and "large" cluster together, but that "tall" is placed in a separate cluster. A second clustering approach based on visual properties finds that "thin" is associated most strongly with surface area, and only weakly with height-to-width ratio.

These findings suggest that the dimensional properties of a referent may be reasoned about to produce different kinds of expressions. An REG algorithm that generates natural reference to visible objects should be equipped to handle this variation, and building such an algorithm can aid in modeling how people use size modification.

We therefore set out to examine how the words proposed to refer to specific axes, like "tall" and "thick", are used differently than words proposed to refer to overall size, like "large" and "small". The first type we will call *individual-axis size modifiers* and the second *overall size modifiers*.[1] Our hypotheses are designed to formalize aspects of size reference that have been implied by earlier work (e.g., Landau and Jackendoff (1993)), but have not yet been systematically tested. This provides a basis from which to design an REG algorithm that refers to an object's size.

Experiments

We examine what happens when a referent object is different in size from a comparator object (1) along a single axis; (2) along two axes, in the same direction (both axes larger or both smaller); and (3) along two axes, in opposite directions (one axis larger, one smaller). Our hypotheses are listed below.

H_1 When a single dimension differs between a referent object and another object of the same type, an individual-axis size modifier will be produced more often than an overall size modifier.

H_2 When two dimensions differ in the same direction between a referent object and another object of the same type, an overall size modifier will be produced more often than an individual-axis size modifier.

H_3 When two dimensions differ in opposite directions between a referent object and another object of the same type, an individual-axis size modifier will be produced more often than an overall size modifier.

It is relatively straightforward to write a deterministic algorithm capturing what we predict the majority of people will do when there is a difference in at least one dimension between two similar objects, and we sketch such an algorithm in Figure 1. Note that some aspects are still left unspecified, and the algorithm does not address how large a difference must be in order to be salient – clearly, some differences between referent and comparator may be too small to elicit a corresponding modifier. This is an area for future work.

Lines 2–3 and 9–10 represent H_2, returning an overall size modifier depending on the differences between dimensions. Lines 2, 4–7; and 9, 11–14 roughly represent H_3, and call to a second function motivated by Hermann and Deutsch (1976), LARGEST-DIMENSION-DIFF, which returns the dimension with the greater difference. Lines 2, 8; 9, 15; and 16–18 represent H_1. The final size modifier structure is sent to the GENERATE function, requesting an overall size modifier (<over>), or an individual-axis size modifier picking out a specific axis (<ind, width> or <ind, height>), along with whether the modifier should capture a larger (+) or smaller (-) difference. Thus, for example, (<over>, +) could be realized as "large" or "big", while (<ind, height>, -) could be realized as "short".

```
r = referent object, d = object of the same type (comparator)
$r_h$, $r_w$ = referent height, referent width
$d_h$, $d_w$ = comparator height, comparator width
01. GenSizeMod(r, d):
02.   if $r_h > d_h$:
03.       if $r_w > d_w$: generate(<over>, +)
04.       elif $r_w < d_w$:
05.           if largest-dimension-diff($r_h$, $r_w$, $d_h$, $d_w$) == width:
06.               generate(<ind, width>, -)
07.           else: generate(<ind, height>, +)
08.       else: generate(<ind, height>, +)
09.   elif $r_h < d_h$:
10.       if $r_w < d_w$: generate(<over>, -)
11.       elif $r_w > d_w$:
12.           if largest-dimension-diff($r_h$, $r_w$, $d_h$, $d_w$) == width:
13.               generate(<ind, width>, +)
14.           else: generate(<ind, height>, -)
15.       else: generate(<ind, height>, -)
16.   else:
17.       if $r_w > d_w$: generate(<ind, width>, +)
18.       elif $r_w < d_w$: generate(<ind, width>, -)
```

Figure 1: Initial algorithm for generating size modifiers.

However, we expect that this is not the whole story, and return to this issue in the last section.

We consider size differences in two different gradations: A small negative difference (-, 10/11th size) or a small positive difference (+, 11/10th size) between the axis of the referent and the corresponding axis of the comparator; and a large negative difference (- -, 4/5th size) or large positive difference

[1] Note that *individual-axis* size modifiers may occasionally pick out more than one axis, e.g., as in the word "thick".

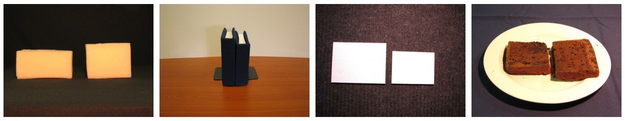

Figure 2: Example stimuli: sponges (++/- -), books (-/0), boards (- -/- -), and brownies (++/0).

(++, 5/4th size) between the two axes. These are operationalizations of what it means for height and width to be different, and serve as a starting point to sample the space of height and width contrasts. Values for these measurements are provided in Table 1.

The stimuli in this study were photographs of real-world objects, physically cut and shaped into different sizes. This follows work in developing computational models that bridge the symbolic realm of language with the physical realm of real-world referents (Herzog & Wazinski, 1994; Roy & Reiter, 2005; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995).

Method

Participants 95 subjects collected using Amazon's Mechanical Turk (Amazon, 2008) were paid for their participation. 87 of these participants labeled themselves as "Native" or "Fluent". From this set, we randomly chose a subset of 60 total participants, spread evenly as groups of 20 in each of our three experiments.

Materials Several different objects were used to elicit size modifiers. These objects were sponges, boards, books, and brownies. All objects were rectilinear solids, varied along their height and width dimensions. The objects were intermixed with fillers, discussed in further detail below.

Each object appeared to the right of a comparator object of the same type (see Figure 2). The target object could appear in 24 different sizes, created by combinations of 5 gradations relative to the comparator object on both the object's horizontal and vertical axes: smaller (- -); a little smaller, (-); no difference (0); a little larger, (+); and larger, (++). The 25th possible size, no difference from the comparator on both the horizontal and vertical axes, was not included. The difference between the height and width of the target object itself was different across the different objects. All target objects had the same relative ratio of difference from the comparator on each axis.

Design We conducted three experiments, addressing each of our hypotheses. The design for each was dimension (2: height, width) x degree of difference (2: small, large) x direction of difference (2: bigger, smaller).

Experiment 1: Differences of degree, single dimension. Responses were elicited for objects with height/width combinations of ++/0, 0/++, +/0, 0/+, -/0, 0/-, - -/0 and 0/- - (8 conditions). Each target item differed from its comparator item in one dimension.

Experiment 2: Differences of degree, matching across dimensions. Responses were elicited for objects with height/width combinations of ++/++, ++/+, +/++, +/+, - -/- -, - -/-, -/- - and -/- (8 conditions). Each target item differed from its comparator item in two dimensions and in the same direction for each; the target item was either bigger overall or smaller overall than the comparator.

Experiment 3: Differences of degree, different polarities across dimensions. Responses were elicited for objects with height/width combinations of ++/- -, - -/++, ++/-, -/++, +/- -, - -/+, +/- and -/+ (8 conditions). Each target item differed from its comparator item in two dimensions and in the opposite direction for each; the target item had one axis bigger and one axis smaller than the comparator.

For each experiment, we followed a Latin square design where all participants saw each of the four object types, with two examples per condition. This yielded 16 experimental stimuli per participant. Each experiment had two subgroups, where one half (10 participants) saw 2 stimuli per condition, and the other half (10 participants) saw the other 2 stimuli per condition.

Stimuli in each experiment were intermixed with the 24 filler pictures, consisting of spatulas, Legos, and shoes. Spatulas appeared in groups of three and Legos and shoes appeared as sets of two. Most objects in filler conditions could be distinguished using part-whole phrases, e.g., "the one with the red Lego" or "the shoe with the laces untied". In total, each subject provided responses for 40 object pictures. Each picture was 400 pixels wide x 300 pixels high, and could be enlarged to 700 x 525 by clicking on it. Pictures were presented in random order, and experimental groups were assigned randomly.

Procedure Instructions informed participants that they had been chosen as "the thrower", tossing objects down a tube to a person below, and their goal was to clearly identify the object on the right so that the person below could pick it up.

Responses were manually corrected for spelling and normalized for punctuation and capitalization. For each expression, we annotate the modifiers as being an individual-axis size modifier (*ind.*), overall size modifier (*over.*), or other. Each single-dimensional modifier was annotated by three postgraduates as being a height modifier or a width modifier. We use the annotations from the annotator who had the highest agreement with the other two, with a Cohen's kappa of 0.90 (95% CI, 0.87–0.94) and 0.71 (0.66–0.76). Table 2 lists the vocabulary and modifier types based on this data. Most base modifiers have corresponding comparative (ending in -er) and superlative (ending in -est) forms.

Results

Results are based on the 320 responses for each experiment. Each response to the test stimuli is counted as either including or not including an individual-axis size modifier (0 or 1) and including or not including an overall size modifier (0 or 1). Note that the two are not exclusive. For each participant, we

Table 1: Measurements for objects along each axis (in cm).

object	height					width				
	++	+	0	-	- -	++	+	0	-	- -
brownies	11.25	9.90	9.00	8.18	7.20	11.25	9.90	9.00	8.18	7.20
sponges	6.25	5.50	5.00	4.54	4.00	12.50	11.00	10.00	9.09	8.00
books	25.00	22.00	20.00	18.18	16.00	6.25	5.50	5.00	4.55	4.00
boards	19.05	16.76	15.24	13.84	12.19	25.4	22.35	20.32	18.47	16.26

Table 2: Size vocabulary.

ind.	height:	high long narrow short skinny slender squat tall thick thin
	width:	fat lengthy long narrow skinny slim thick thin wide
over.		big large small

Table 3: Example responses.

condition	object	expression
h++w++	books	taller fatter book
h+w- -	sponges	taller sponge
h- -w++	boards	the shorter and slightly wider board with a diagonal top side
h0w+	brownies	longer brownie
h- -w- -	boards	smaller board

sum the total number of responses with each type of modifier. This provides two sets for a two-tailed paired t-test in each of our analyses.

Examples of normalized responses are given in Table 3. Table 4 provides the proportions of responses that included an individual-axis size modifier, an overall size modifier, both, or neither for each experiment.

H_1: **When a single dimension differs between a referent object and another object of the same type, an individual-axis size modifier will be produced more often than an overall size modifier.**

We do not find a strong trend to include individual-axis size modifiers, with such modifiers occurring in an average of 8.4 responses per participant, compared to 6.1 responses on average containing an overall size modifier. The difference is not significant ($t = 1.382$, $df = 19$, $p = 0.183$).[2]

H_2: **When two dimensions differ in the same direction between a referent object and another object of the same type, an overall size modifier will be produced more often than an individual-axis size modifier.**

Table 4: Proportion of responses including either 1+ individual-axis size modifiers, 1+ overall size modifiers, both, or neither.

Experiment	ind.	over.	both	neither
1	50.0%	35.6%	2.5%	11.9%
2	29.1%	65.9%	4.7%	0.3%
3	70.6%	8.8%	6.3%	14.4%

We find a strong trend to include overall size modifiers, with such modifiers occurring in an average of 11.3 responses per participant. Individual-axis size modifiers occur in an average of 5.4 responses. The difference in this distribution is significant ($t = -4.914$, $df = 19$, $p < .001$).

H_3: **When two dimensions differ in opposite directions between a referent object and another object of the same type, an individual-axis size modifier will be produced more often than an overall size modifier.**

We find that when two dimensions differ in opposite directions, individual-axis size modifiers are chosen in an average of 12.3 responses per participant, while overall size modifiers are chosen in an average of 2.4 responses. The difference in this distribution is significant ($t = 8.866$, $df = 19$, $p < .001$).

Based on these results, we can confirm Hypotheses 2 and 3. Overall size modifiers tend to be used when both axes are different from a comparator in the same direction, and individual-axis size modifiers tend to be used when both axes are different from a comparator in opposite directions. Results are significant at $\alpha = .01$. We cannot reject a null hypothesis in favor of Hypothesis 1; we do not see a significant difference in the distribution of size modifier types when a single axis is different between a target and a comparator. Further factors that may be affecting participant responses are discussed in the next section.

We have illustrated some basic principles of how people use size in reference. However, these experiments also provide much richer information on how people use size. One immediate question these findings leave is whether it is common to include two individual-axis modifiers, each referring to a separate axis, when the objects have differences of degree, different polarities across dimensions (Experiment 3). We find that this occurs in a minority of responses (mean = 4.8), while it is significantly more common (mean = 11.2) to include just one individual-axis size modifier, an overall size modifier, or neither ($t = -4.292$, $df = 19$, $p < .001$).

We can also confirm the findings in Hermann and Deutsch (1976). Based on responses to Experiment 2 and Experiment 3, in conditions where there is a large difference and a small difference (++/+, +/++, ++/-, -/++, - -/-, -/- -, - -/+, +/- -), if an individual-axis size modifier is chosen, that modifier will refer to the larger difference more often than the smaller difference (mean for large difference = 3.4; small difference = 2.6, $t = 3.629$, $df = 38$, $p < .001$).

[2] Preliminary analysis on a larger dataset suggests that this trend may become significant, and we leave this for future work.

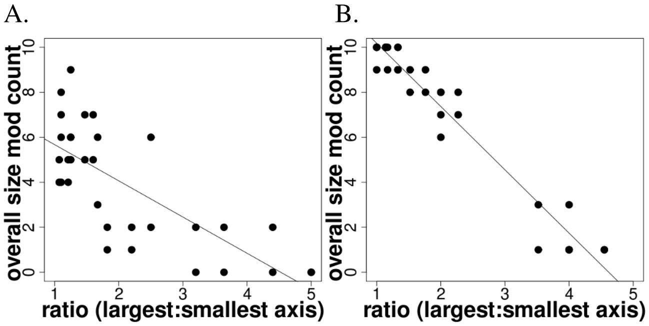

Figure 3: Count of overall size modifiers for different height/width ratios in Experiment 1 (A) and Experiment 2 (B), with linear regression. Ratios shown are for the largest axis divided by the smaller axis.

Further Analysis

This data supports the idea that the selection of size modifier is in a large part determined by dimensional differences between an object and another object of the same type, with a difference along two axes in different directions corresponding to size modifiers like “tall” and “thin”, and a difference along two axes in the same direction corresponding to size modifiers like “small” and “big”.

These experiments have also shed some light on some of the other factors that may affect the selection of size modifier. One trend that emerges in the data is the relationship between the selection of individual-axis or overall size modifier and the ratio between the height and width of the target object itself. Although we did not design the study to test this aspect, our data indicate that the closer the object is to a square shape, e.g., the smaller the difference between height and width, the more likely participants are to use an overall size modifier like *big* or *small*. Figure 3 illustrates this trend, where the x-axis is the ratio between the larger axis (height or width) and the smaller axis (height or width) for each stimulus, and the y-axis is the number of responses to the stimulus that include an overall size modifier. In the data from Experiment 2, this trend is quite strong, $r^2 = 0.95$ ($p < .001$). Across conditions with only height or width differing from the comparator object (Experiment 1) – the conditions where we did not find a tendency to use overall size modifiers – there is also a trend, $r^2 = 0.57$ ($p < .001$). Further testing is necessary to examine this effect.

This suggests that the selection of individual versus overall size modifier may be influenced by the difference in height and width from the comparator object as well as the difference between height and width of the target object itself. Individual-axis size modifiers may be used when only one axis of the target is different from the comparator, however, as the axes of the target itself converge in size, there is a marked increase in preference for overall size modifiers.

We also find a preference to use height modifiers over width modifiers, across the three experiments (mean for height = 6.3, width = 4.7; $t = 4.409$, $df = 59$, $p < .001$). This may reflect that the objects are presented side by side, their heights directly comparable. This brings to light another facet of how the dimensional properties of objects may be reasoned about in a computational model, taking into account a target object’s position with respect to a comparator when selecting a size modifier type.

An obvious area for further analysis concerns the determinism of the size algorithm. The majority of our data comports with the algorithm sketched in Figure 1, however, this data is probabilistic; the algorithm is not. Assigning probabilities to each of the conditional statements may help to better capture how people use size modification.

Implications and Future Research

This study suggests that the selection of size modifier when referring to real-world objects in the presence of another object is influenced by at least two factors:

1. Whether one or both axes differ from a comparator.
2. Which axis is the most different from a comparator.

And may be influenced by two further factors:

1. The location of the target object relative to the comparator.
2. How similar in size the two axes of the target object are.

In future work, we hope to explore our post-hoc findings and refine the algorithm, developing mechanisms for reasoning about the relative size difference between dimensions of the referent object, and including information about where the referent object is placed relative to a comparator. Extending this task to elicit responses from more participants may be used to assign weights to each of the conditions currently in place, and provide the size algorithm with a probability distribution over different possible surface forms. A better understanding of when a difference is small enough not to be salient would help connect this algorithm more closely to a visual input, placing constraints on when the conditional statements outlined above apply.

This research reasons about the interplay between two dimensions, height and width. Scaling up to three dimensions would help further develop a model of how size modifiers are used in the real world. It may be the case that the patterns of individual-axis and overall size modifiers change when there is a third visible dimension available. We also hope to address situations where there are several similar objects, and situations where the target referent is a set of objects. Further work may also examine how this research extends to other kinds of object shapes; this study has focused on rectilinear solids, but whether modifiers pick out the axes for height, width, and depth in less rectangular objects, or objects with irregular shapes, remains an open question.

Acknowledgments

We would like to thank Ellen Gurman Bard for invaluable guidance and the anonymous reviewers for useful suggestions. This research is sponsored by the Scottish Informatics and Computer Science Alliance (SICSA) and the Overseas Research Students Award Scheme (ORSAS).

References

Amazon. (2008). *Amazon mechanical turk: Artificial artificial intelligence.*

Bartlett, E. J. (1976). Sizing things up: the acquisition of the meaning of dimensional adjectives. *Journal of Child Language*, *3*(02), 205–219.

Bierwisch, M., & Lang, E. (Eds.). (1989). *Dimensional adjectives : grammatical structure and conceptual interpretation*. New York: Springer-Verlag.

Bock, J. K., Irwin, D. E., Davidson, D. J., & Levelt, W. J. M. (2003). Minding the clock. *Journal of Memory and Language*, *48*, 653–685.

Brown-Schmidt, S., & Tanenhaus, M. K. (2006). Watching the eyes when talking about size: An investigation of message formulation and utterance planning. *Journal of Memory and Language*, *54*, 592–609.

Dale, R., & Reiter, E. (1995). Computational interpretations of the gricean maxims in the generation of referring expressions. *Cognitive Science*, *19*, 233–263.

Eilers, R. E., Oller, D. K., & Ellington, J. (1974). The acquisition of word-meaning for dimensional adjectives: the long and short of it. *Journal of Child Language*, *1*(02), 195–204.

Ferreira, F., & Swets, B. (2002). How incremental is language production? evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Learning*, *46*, 57–84.

Friedland, G., Jantz, K., & Rojas, R. (2005). SIOX: simple interactive object extraction in still images. *Proceedings of the Seventh IEEE International Symposium on Multimedia*, 253–259.

Gorniak, P., & Roy, D. (2004). Grounded semantic composition for visual scenes. *Journal of Artificial Intelligence Research*, *21*, 429–470.

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, *11*, 274–279.

Henderson, J. M., & Ferreira, F. (2004). *The interface of language, vision, and action: Eye movements and the visual world*. New York, NY: Psychology Press.

Hermann, T., & Deutsch, W. (1976). *Psychologie der objektbenennung*. Bern: Huber Verlag.

Herzog, G., & Wazinski, P. (1994). Visual translator: Linking perceptions and natural language descriptions. *Artificial Intelligence Review*, *8*(2/3), 175–187.

Kelleher, J., Costello, F., & Genabith, J. van. (2005). Dynamically structuring, updating and interrelating representations of visual and linguistic discourse context. *Artificial Intelligence*, *167*, 62–102.

Krahmer, E., van Erk, S., & Verleg, A. (2003). Graph-based generation of referring expressions. *Computational Linguistics*, *29*(1), 53–72.

Landau, B. (2001). Perceptual units and their mapping with language. In T. Shipley & P. Kellman (Eds.), *From fragments to objects: Segmentation and grouping in vision.* The Netherlands: Elsevier.

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

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Levelt, W. J. M., Roelofs, A. P. A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1–37.

Mojsilović, A. (2005). A computational model for color naming and describing color composition of images. *IEEE Transactions of Image Processing*, *14*(5), 690–699.

Morzycki, M. (2009). Degree modification of gradable nouns: size adjectives and adnominal degree morphemes. *Natural Language Semantics*, *17*(2), 175–203.

Roy, D., & Reiter, E. (2005). Connecting language to the world. *Artificial Intelligence*, *167*, 1–12.

Roy, D. K. (2002). Learning visually-grounded words and syntax for a scene description task. *Computer Speech and Language*, *16*, 353–385.

Roy, D. K., & Pentland, A. (2002). Learning words from sights and sounds: A computational model. *Cognitive Science*, *26*, 113–146.

Sedivy, J. C. (2003). Pragmatic versus form-based accounts of referential contrast: Evidence for effects of informativity expectations. *Journal of Psycholinguistic Research*, *32*, 3–23.

Sedivy, J. C., Tanenhaus, M., Chambers, C., & Carlson, G. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, *71*, 109–147.

Skočaj, D., Berginc, G., Ridge, B., Vanek, O., Hutter, M., & Hewes, N. (2007). A system for continuous learning of visual concepts. *Proceedings of the International Conference on Computer Vision Systems.*

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*, 1632–1634.

Tucker, G. (1998). *The lexicogrammar of adjectives: A systemic functional approach to lexis*. London: Cassell.

van Deemter, K. (2000). Generating vague descriptions. *Proceedings of the First Natural Language Generation Conference*, 12–16.

van Deemter, K. (2004). Generating referring expressions that involve gradable properties. *Computational Linguistics*, *32*(2), 195–222.

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

Zheng, S., Yuille, A., & Tu, Z. (2010). Detecting object boundaries using low-, middle-, and high-level information. *Journal of Computer Vision and Image Understanding*, *114*(19), 1055–1067.