Vagueness in referring expressions of quantity: effects on the audience

Matt Green
School of Psychology, University of Aberdeen

Kees van Deemter
Department of Computing Science, University of Aberdeen

Author note: This SIPI 2012 article is adapted from Green and van Deemter (2011)

NLG systems that generate natural language text from numerical input data must decide between alternative surface linguistic forms for the natural language output. When using referring expressions to identify numerical quantities, the system must decide between vague and crisp surface forms of the referring expression. Ideally, the system would be equipped with heuristics that it could use to make these decisions in the way that best suits the audience: however there is currrently little empirical data to draw on concerning the differential audience benefits of vague and crisp surface forms. In this paper we describe a series of experiments that investigate whether different surface forms affect the audience's cognitive load in different ways. We estimate cognitive load by measuring the response latencies in a forced choice referent identification task in which we vary the surface form of the referring expression that constitutes the instruction in the task. We find that the pattern of audience responses across the series of experiments provides little support for the *cost reduction* hypothesis that vague surface forms should place fewer cogntive demands on the audience than crisp surface forms: instead the results support the view that referring expressions that contain numerals are more taxing for the audience than referring expressions that use natural language quantifiers, at least in the context of a forced choice referent identification task. We offer this work as an initial foray into the provision of heuristics to augment NLG systems with audience-sensitivity.

Introduction

When quantities need to be referred to by an NLG system, it should consider which manner of presentation will benefit the readers of its output. Data from experiments with readers have the potential to guide systems in making such decisions, as for example in Power and Williams (2012).

Keefe and Smith (1996) give what can be regarded as a widely accepted definition of vagueness: "vague predicates have borderline cases, have fuzzy boundaries, and are susceptible to sorites paradoxes" (p. 4). A canonical example is the concept *tall*.

Lipman (2011) argues that a limitation of standard game-theoretical models of communication is that they are unable to explain why so much of everyday language is vague. Lipman suggests one explanation for the prevalence of vagueness, that is cast in terms of audience effort: "For the listener, information which is too specific may require more effort to analyze". Van Deemter (2009) extends this line of argument to cover the speaker's costs too, suggesting that benefits of vagueness can accrue when vagueness reduces costs for speaker or hearer. We use the term *cost reduction hypothesis* to refer to this cost-based line of argument, and we test this hypothesis empirically in the experiments in this paper.

NLG systems that generate text from numerical input must decide between alternative linguistic forms of the same numerical content, including how vague to be. Such systems include (Goldberg, Driedger, & Kittredge, 1994) in the domain of weather forecasting; (Hunter et al., 2008) in the domain of medical decision support.

Currently there is little empirical data to support NLG decisions about vagueness. Mishra, Mishra, and Shiv (2011) demonstrated that vague (but not precise) feedback can be distorted by the hearer in such a way as to help the hearer achieve his own goals better. However, their *vague* condition took the form of an exact range of values (e.g., 30–35, compared with a single value e.g, 30 in the *precise* condition) and thus lacked borderline cases. (Peters et al., 2009) assessed the impact of adding vague evaluative categories to numerical quality of care information in the rating of hospitals, but it is unclear whether the effects were due to the mere addition of information, or to adding *vague* information.

Experiments

Introduction

We designed a series of three experiments to test the cost reduction hypothesis in a laboratory setting: this required us to define *cost* in a narrow sense susceptible to measurement; and to set up a task that would place the participants in a situation of cognitive load; as well as to operationalise vagueness as a manipulable independent variable.

We used the forced choice paradigm to provide an experimental setting where we placed participants under cognitive load by requiring them to identify the referent of a referring expression of quantity. Participants were presented with a small number of dot arrays together with an instruction to identify one of them: the instruction referred to the target dot array by indicating the number of dots it contained, using either a vague or a crisp referring expression of quantity. We measured the response latency between the appearance of

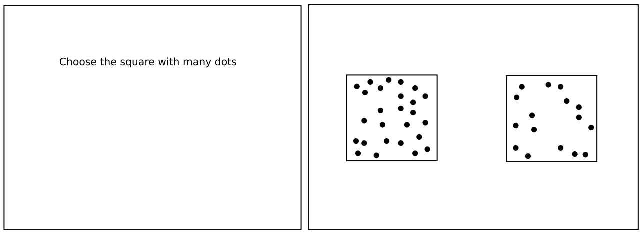

Figure 1. Example stimulus from experiment 2: left panel shows the instruction; right panel shows the dot arrays

the instruction and the selection of a dot array as an indicator of cognitive load: participants under large load should respond slower than participants under small load. With other variables held constant, it was plausible to attribute the load to the vagueness or crispness of the instruction. The cost reduction hypothesis predicts that vague instructions would engender faster responses than crisp instructions: in contrast we consider that standard game theory predicts either no difference, or an advantage for crisp instructions.

Experiments one and two

In the first two experiments we used as the vague instructions “Choose the square with {many, few} dots.”; and as the crisp instructions “Choose the square with {n} dots” (where *n* was the number of dots in the target array). In both these experiments participants were required to choose one from two arrays.

A pilot experiment that used small numbers of dots (less than 10 in each of two squares) found that when both numbers of dots were above the subitizable range (Trick & Pylyshyn, 1994), vague quantifiers (e.g., *many* dots) attracted faster accurate response times than precise quantifiers (e.g., *nine* dots), but only when they identified the larger of the two numbers.

A second experiment used larger numbers of dots and two squares in a similar experiment. Pairs of numbers were: (5,25) (10,25) (15,25) (20,25) (30,25) (35,25) (40,25). Figure (1) illustrates the task. The results indicated that vague quantifiers attracted faster accurate response times and fewer errors than did the precise quantifiers, for all the combinations of numbers that we used.

These two experiments used *few* and *many* as the vague quantifiers. These quantifiers have the potential for vagueness in the sense of Keefe and Smith (1996). However, this potential for vagueness was not realised because the quantifier uniquely identified a single square and there were no borderline cases. Also, the vague instructions used quantifiers that did not contain a number (few, many), whereas the precise instructions did contain numbers (e.g., five in experiment one, 30 in experiments two and three), and so there was a potential confound of vagueness with the absence of a number in the instruction.

Experiment three

In a third experiment we set out to address these concerns. Firstly, we used three squares, and indefinite articles in the instructions in the vague conditions. For example, an instruction like *Please choose a square with many dots* was used in the context of three squares (e.g., (6,15,24)), where two plausibly contained many dots, but one contained more dots and was therefore considered to be a better match for the instruction, leaving the other as a borderline case. Secondly we used both vague and precise versions of instructions that specified numbers, and of instructions that did not specify numbers.

We found faster responses in the no-number conditions than the number conditions; but no main effect of the vague / precise manipulation; and no interaction, suggesting that when the instructions allowed the potential for vagueness to be realised, vagueness did not exert a beneficial influence on response times.

The response time patterns for experiments 2 and 3 show that while the results from experiment 2 suggest an advantage for vagueness, the results from experiment 3 suggest an alternative explanation of the same pattern such that what appeared to be a vagueness advantage in experiment 2 is revealed to be an advantage for instructions that avoid numerals.

Conclusions

Vagueness in a referring expression is a combination of the referring expressions potential for borderline cases; and the specific situation in which the referring expression is used. We found that expressions with the potential for vagueness attracted faster response times than expressions without; but only when the referent set did not allow the possibility of borderline cases. When the referent set did allow borderline cases, we found differences between expressions that used numbers and those that did not, but no differences between vague and precise expressions in either case. Although our experiments were limited in focussing on vagueness in descriptive noun phrases only, and although they did show up advantages for certain vague expressions, they do more to cast doubt on the cost reduction hypothesis than to confirm it.

References

Goldberg, E., Driedger, N., & Kittredge, R. (1994). Using natural-language processing to produce weather forecasts. *IEEE Expert*, *9*(2), 45–53.

Green, M., & van Deemter, K. (2011). Vagueness as Cost Reduction: An Empirical Test. In *Proceedings of ‘Production of Referring Expressions’ workshop at 33rd Annual Meeting of the Cognitive Science Society Production of Referring Expressions.* Boston, MA.

Hunter, J., Freer, Y., Gatt, A., Logie, R., McIntosh, N., Van Der Meulen, M., … Sykes, C. (2008). Summarising complex icu data in natural language. In *Amia annual symposium proceedings* (Vol. 2008, p. 323).

Keefe, R., & Smith, P. (1996). *Vagueness: a reader. a bradford book.* The MIT Press.

Lipman, B. L. (2011). *Why is language vague.*

Mishra, H., Mishra, A., & Shiv, B. (2011). In praise of vagueness. *Psychological Science*, *22*(6), 733.

Peters, E., Dieckmann, N., Västfjäll, D., Mertz, C., Slovic, P., & Hibbard, J. (2009). Bringing meaning to numbers: The impact of evaluative categories on decisions. *Journal of Experimental Psychology: Applied; Journal of Experimental Psychology: Applied*, *15*(3), 213.

Power, R., & Williams, S. (2012). Generating numerical approximations. *Computational Linguistics*, *38*(1), 113–134.

Trick, L., & Pylyshyn, Z. (1994). Why are small and large numbers enumerated differently? a limited-capacity preattentive stage in vision. *Psychological review*, *101*(1), 80.

Van Deemter, K. (2009). Utility and language generation: the case of vagueness. *Journal of philosophical logic*, *38*(6), 607–632.