Utility and Language Generation: the case of vagueness

Kees van Deemter
University of Aberdeen
k.vdeemter@abdn.ac.uk

Abstract

This paper[1] asks why information should ever be expressed vaguely, reassessing some previously proposed answers to this question and suggesting some new ones. Particular attention is paid to the benefits that vague expressions can have in situations where agreement over the meaning of an expression cannot be taken for granted. A distinction between two different versions of the above-mentioned question is advocated. The first asks why human *languages* contain vague expressions, the second question asks when and why a speaker should choose a vague expression when communicating with a hearer. While the former question is purely theoretical, the latter has practical implications for the computational generation of utterances in Natural Language Generation (NLG).

1 Introduction

In recent years, researchers such as Barton Lipman have asked why vagueness is such a prominent feature of language. On the face of it, vagueness is just sloppiness, so why have we tolerated "a world-wide several-thousand-year efficiency loss" (Lipman 2006) by speaking vaguely? The present paper reviews the answers to this question that have so far been suggested, and adds another answer that has, to the best of my knowledge, not been discussed before. Even though some of these answers raise questions of their own, it appears to me that, collectively, they start to amount to a satisfactory (though probably still incomplete) answer. The question of the *rationale* for vagueness, in other words, does not represent a baffling puzzle any longer, but as an area of plausible conjectures.

I will start by explaining one of my reasons for being interested in this question, which lies in its relevance for Natural Language Generation (NLG). At the end of the paper, we shall return to NLG, to see how our findings are relevant for this area of applied research.

[1]A much-enhanced version of Van Deemter (2009), the written form of a keynote speech at the 12th European Workshop on Natural Language Generation, Athens, March 2009; it is related to chapter 11 of Van Deemter (in preparation).

2 Language generation as a choice problem

Natural Language Generation (NLG) is the area of computational linguistics that is concerned with the mapping from non-linguistic to linguistic expressions (e.g. Reiter and Dale 2000, Bateman and Zock 2009). Prominent application areas of NLG include weather forecasting, where computers convert meteorological data into text (e.g. Goldberg et al. 1994), and medical decision support, where NLG systems produce medical reports on babies in intensive care, for example, to help doctors and nurses make clinical decisions (Portet et al. 2009). The systems in question detect salient trends and events (e.g. ususual heart rhythm), they conceptualise, aggregate and order them, and they express the resulting information in coherent sequences of grammatical sentences. The challenges in NLG are huge of course, and success to date is only partial, although there exist application areas where systems are starting to outperform human authors in terms of the quality of the texts that are produced, as rated by subjects in controlled evaluation studies (e.g. Reiter et al. 2005).

It might be thought that NLG is best viewed as a kind of *translation* problem, where the key challenge is to find a way to convert a formal expression into (for example) an English one. In its early years, this may have been a fruitful way to think about NLG but, these days, a better perspective is of NLG as a *choice* problem. For after the advances of recent years, the problem is no longer so much "How on Earth can this information be expressed in English?", but rather "From all the possible ways to express this information in English, which one is the most effective choice?" If later modules of the system express the text through speech and human-like characters on a computer screen then many more choices need to be made, varying from the phonological and phonetic domain to that of gestures, gaze and body movement (e.g. Van Deemter et al. 2008), but in this paper we focus on textual choices, which vary from decisions concerning the overall structure of the text to the manner in which objects are referred to (e.g., using pronouns or full noun phrases), and from decisions about syntactic structure (e.g., the choice between active or passive form) to the choice of lexical items.

It is often fair to assume that the formal expressions from which NLG takes its departure are themselves clear and unambiguous. Let us call the inputs to the generator *Contents*. Now suppose we have a grammar that tells us how each Content can be expressed in a language such as English. The task for NLG is to choose, for each of these Contents, which of all the different linguistic Forms that can express it according to the grammar is the *best* expression of this particular Content. Ultimately, this choice is likely to depend on a number of other parameters, such as the identity of the hearer, and the words that have been used elsewhere in the text. In the present paper, these "contextual" issues will largely be ignored, allowing us to simplify by thinking in terms of a mapping from Contents to Forms. The perspective that views NLG as a choice problem is far from new (see e.g. McDonald 1987, where it takes a central position); in fact, it forms the methodological spine of Systemic-Functional Grammar, whose AND/OR graphs formalise the dependencies that exist between such choices

(Bateman 1997).

An important type of information in many situations – including the ones that feature in NLG's most central applications – arises when quantitative information is expressed in ordinary language. Vague language is often a natural choice in such situations. When a thermometer, for example, measures the temperature in your living room as 24.82 Celcius, this can be expressed, among other possibilities, by saying that your temperature is '24.8 degrees', or by saying that the temperature is 'approximately 25 degrees', but in many situations we would be more likely to say that it is 'warm'. Which of these linguistic Forms is preferable in a given situation, and why?

3 Utility

At the most abstract level, the answer to such questions is likely to have something to do with the "utility" of the different Forms that can be generated. To utter a sentence is to perform an action, after all, and the choice between different actions is naturally thought of as governed by utility, understood in the broadest possible sense. Utilities can be positive or negative, in which they might be thought of as costs. The idea, in other words, is to look at NLG as an area of applied *decision theory* (e.g., Jeffrey 1983).

The analysis of language production/generation as driven by the utility of utterances has rarely been pursued yet it feels natural to people familiar with practical applications, where texts are generated for a concrete purpose. This type of analysis should definitely suit anyone who is interested in the effects of an utterance on an audience (e.g., Mellish and Van der Sluis 2009). To see the relevance of a decision-theoretical analysis, in which the expected pay-offs associated with different texts are compared, consider a concrete NLG system that informs roadgritters' decisions about the condition of the roads in Scotland, to help them decide which ones are icy enough to require treatment with salt to avoid traffic accidents (e.g. Turner et al. 2008).

Computerised weather forecasts can tell road gritters which roads are likely to be icy, and hence dangerous. Let us assume that generally-accepted criteria exist for determining when a road is considered dangerous on a given night, so the input to the system is essentially a set of roads. There can be thousands of dangerous roads on a given night, and it is often impossible to say in a few words *exactly* which roads are expected to be dangerous (Turner et al. 2008). The generator might approximate the data by saying either (a) or (b):

(a) Roads *in the Highlands* are icy
(b) Roads *above 500 metres* are icy.

Both descriptions are probably best considered as crisp (i.e., not vague), with the second covering a slightly larger area. It matters substantially which of these summaries is generated, because each summary will lead to a different set of roads being treated with salt (i.e., gritted). Suppose summary (a) has 10 false positives (i.e., roads gritted unnecessarily) and 10 false negatives (i.e., dangerous roads not gritted); summary (b) might have 100 false positives and only 2 false negatives. In a situation such as this one, which involves a tradeoff between safety on the one hand, and money and environmental damage (from salt) on the other, decision theory would be a natural framework in which to compare the utility of the two summaries. If a false positive has a negative utility of -0.1 and a false negative one of -0.5, for example, then the first summary wins the day. Needless to say, the choice of these constants is crucial, and tricky to justify.[2]

More specifically, many NLG systems invite a *game-theoretical* analysis – or an Optimality-Theoretic analysis, which can come down to the same thing, as was demonstrated by Dekker and Van Rooij (2000). Suppose I want to state that all old people are entitled to certain benefits (cf. Khan et al. 2008):

a. Old men and old women are entitled to benefits.
b. Old men and women are entitled to benefits.

Which of these two linguistic Forms should the speaker choose? This depends on the strategy of the hearer. If the hearer interprets (b) as concerning *all* women (rather only the old ones) then my utterance will have misfired to an extent. The success (for speaker and/or hearer!) of the speaker's generation strategy, in other words, depends on the hearer's interpretation strategy.[3]

The interaction between speakers' and hearers' strategies means that decision theory is not the best tool for analysing the situation, for where different agents' strategies interact, Decision Theory gives way to *Game Theory*. The mathematical theory of games has roots in the works of Pascal and Huygens and took shape in the nineteen fourties (Von Neumann and Morgenstern 1944); it has come to be used extensively by economists, sociologists, biologists, and others. Far from being limited to games in a narrow sense of the word, Game Theory is essentially the study of rational social interaction and, as such, it can shed light on language use and logic as well. (For a recent overview of the issues, see Van Benthem 2008). Perhaps more than anything, Game Theory promises to have the potential to explain *why* communication works the way it does. For if we could show that people's linguistic behaviour conforms with what it would be rational for them to do, then this would have substantial explanatory value (cf., Bermúdez 2009 for discussion).

David Lewis' work on *coordination games* helped to make Game Theory relevant for situations where the players cooperate (Lewis 1969). A classic example involves two generals who are both intent on attacking a joint enemy. But while each general individually is weaker than the enemy, they can beat him if they attack at the same time. Communication ("I am going to attack now!") can help the generals to cooperate and win the battle. Much the same thing happens when you try to meet a friend: neither of you may care much

[2] In fact, the roadgritting system of Turner et al. (2008) does not allow *any* false negatives and could therefore be described as assigning a value of ∞ to these (even though an explicit decision-theoretic perspective is not discussed by the authors).

[3] For a game-theoretical perspective on the generation of referring expressions, where success depends on alignment between hearer and speaker strategies, see Kibble 2003.

where and when to meet, as long as the two of you end up in the same place at the same time; communication can help you achieve this goal.

Vagueness and ambiguity pose a direct challenge to utility-based approaches to language. For, on the face of it, it would be rational to be clear and precise all the time. From this perspective, vagueness and ambiguity are just manifestations of sloppiness, comparable to typos: undesirable occurrences which we prune away from a text once we recognise them for what they are. Although this might be the right perspective in some cases, it seems doubtful that sloppiness is the only explanation. Some facts may serve to reinforce this doubt. Vague words, for example, are prevalent among the words first learned by a typical infant (Peccei 1994) and many of their subtleties are understood by children of only 24 months old (Ebeling and Gelman 1994).[4] And it is not as if we "unlearn" them when we grow up: as many as 8 out of the 10 most frequent adjectives in the British National Corpus are vague (Van Deemter 2006). Based on observations of this nature, most researchers suspect that vagueness and ambiguity are irradicably central to human communication. The challenge is to explain why. In recent years this challenge has been clearly recognised, and a number of tentative answers have been proposed (Lipman 2000, 2006; De Jaegher 2003, Board and Blume 2009). To get a first glimpse of the way in which game theory has been applied to ambiguity and vagueness, let us look at the work of Aragones and Neeman (Aragonès and Neeman 2000).

4 Vagueness in situations of conflict

First, let us focus on a type of situations where it is relatively easy to understand what the differential benefits of vagueness can be. We start by examining a game-theoretic study of a different but related phenomenon: ambiguity.

4.1 The utility of ambiguity: Aragonès and Neeman

Like most other researchers, we take *vagueness* to arise if an expression allows borderline cases.[5] The word 'tall', for example, is vague because in a typical context there can be people who are difficult to categorise as either tall or not tall: they are somewhat in between, one is tempted to say. *Ambiguity* is something else. It arises when an expression can be meant in a limited number of different ways. The word 'letter', for example, is ambiguous because it can refer to one individual character or to an entire epistle. In 1994, two game theorists asked whether a Game Theoretical explanation might be given for strategic use of ambiguity, and they came up with the following answer (Aragonès and Neeman 1994).

[4] The authors show that children of this age have little difficulty understanding utterances like 'the dress is large for the doll', and 'pick up the large doll', in a context where there are dolls and dresses of different sizes.

[5] See section 6.2 for a brief discussion of higher-order vagueness.

Suppose two unscrupulous politicians position themselves for an election. Not burdened by stong convictions, they are free to choose between three different ideologies (left, right, center), depending on what gives them the highest utility; additionally, they can choose between two *commitment*[6] levels, c_{high} and c_{low}, both representable as real numbers with $c_{high} > c_{low}$. Aragonès and Neeman do not say what a commitment level is, but one might think of a more and a less extreme version of the politicians' chosen ideology.

What combination of an ideology and an commitment level should each politician choose? This depends on the electorate, of course. Suppose there are three blocs of voters: V(left), V(right) and V(center). A leftist voter prefers a leftist politician, and preferably one with a high commitment level. Confronted with a choice between two rightwing politicians, our leftist voter will prefer one with a low commitment. A rightwing voter behaves as the mirror-image of the leftist voter, while the neutral voter is neutral between the two ideologies but, weary of ideology, she prefers low commitment over high commitment. Commitment, in other words, is only relevant for a choice between politicians of the same ideology.

If this is the whole story then politicians will choose an ideology and commitment level based on their estimates of the numbers of voters in each bloc, trying to maximise their expected payoff, formulated solely in terms of the likelihood of winning the election. The task for Game Theory is to work out what *combination* of strategies might give both politicians the highest possible payoff, for example in the sense that a policy change by just one of the two politicans can never improve his expected payoff.

But Aragonès and Neeman's model allows politicians to look beyond the election, towards their anticipated time in government. Surely, a low commitment is easier to fulfil than a high commitment, particularly in view of unforeseen contingencies, so it is nicer to be elected on a low-commitment platform that does not tie one's hands too much. To model this, Aragonès and Neeman formulate utility in a way that multiplies the probability of a politician's winning the elections with a constant that is negatively correlated to his commitment. Let $U_i(I_1, c_1; I_2, c_2)$ be the utility for politician i given that politician 1 chooses ideology I_1 with commitment level c_1, while politician 2 chooses I_2 with level c_2. Furthermore, $P_i(I_1, c_1; I_2, c_2)$ represents the probability of i winning the elections given this same constellation of choices. Let $k \geq c_{high}$.

Utility Formula: $U_i(I_1, c_1; I_2, c_2) = P_i(I_1, c_1; I_2, c_2) * (k - c_i)$

Under these assumptions one can show that a low commitment level (i.e., c_{low}) can sometimes give a politician a higher overall utility than a high one (i.e., c_{high}). Suppose politican 1 adopts c_{low}. Depending on how many leftist voters there are, this may cause the probability $P_1(left, c_{low}; I_2, c_2)$ to be lower than $P_1(left, c_{high}; I_2, c_2)$, because leftist voters will be less inclined to vote for politician 1 than if he had chosen a higher commitment level (although the reverse is

[6] Aragonès and Neeman call these *ambiguity* levels, but since the relation with ambiguity is debatable we opt for a more neutral term. Low commitment equals high ambiguity and conversely.

true for centrist and rightist voters). On the other hand, $k - c_{low} > k - c_{high}$, reflecting the fact that a politician's time in office will be easier if he gets there on a low commitment. The effect is that, depending on the size of $P_1(left, c_{high}; I_2, c_2) - P_1(left, c_{low}; I_2, c_2)$, the overall utility for politician 1, as determined by the Utility Formula, may be *higher* than if he had chosen a high commitment level. For details see Aragonès and Neeman (1994).

4.2 The utility of vagueness

It is sometimes thought that Aragonès and Neeman's model demonstrates how ambiguity can be used strategically, but that it fails to shed light on *vagueness* (e.g. De Jaegher 2003). It appears, however, that an interpretation of their work is possible that involves vagueness rather than ambiguity.

Suppose the ideology in question – a leftist, or perhaps a populist one – is to take away money from the 10% of richest people and give it to the 10% poorest. Commitment level, in this case, could be a way of making explicit *what percentage* of the income of the top 10% richest people to give away. An ambiguous commitment, presumably, would be a disjunctive one:

> The ambiguous politicians' game:
> – I: take money from the 10% richest people and divide it equally over the 10% poorest.
> – c_{50}: perform I with 50% of the money of each of the richest people.
> – $c_{ambiguous}$: perform I with either 5% *or* 50% of the incomes of each of the richest people.

Analogously, a *vague* version of Aragonès and Neeman's game would arise if a politician left a range of options open, as when he said "I'll redistribute a *large* portion of the 10% richest people's income":

> The vague politicians' game:
> – I: As above
> – c_{50}: As above
> – c_{vague}: perform I with a *large* portion of the money of each of the richest people

The vague politicians' game is isomorphic to the ambiguous politicians' game: fierce advocates of redistribution would favour c_{50} over c_{vague}, for example, because the latter leaves them uncertain over the amount of redistribution. It is also plausible that politicians would prefer to avoid a commitment as clear as c_{50}, because future contingencies might make it difficult for them to honour this promise. In fact, one could extend the game with a second election, where the electorate could give its verdict on a politician's time in office, in which case one could adapt the utility formula with a third term to represent the probability of winning that second election. The breaking of promises doesn't do much for a politician's changes of being re-elected, and a precise promise is easier to break than a vague one.

It seems that, with help from Aragonès and Neeman, we have found a situation in which vagueness has a higher utility than precision. Other situations have been discussed by De Jaegher (2003)[7] and more recently by Board and Blume (2009). All these models, however, hinge on the fact that the interests of the speaker and the hearer differ: what's good for the politician, in Aragonès and Neeman's paper, may be bad for his voters for example. NLG systems can be faced with similar asymmetries, for example when an artificial doctor decides to keep its predictions vague to avoid being contradicted by the facts; a doctor who says "These symptoms will disappear fairly soon" is less likely to get complaints, and to be sued, than one who says "These symptoms will have disappeared by midnight". Something similar holds for a roadgritting system (e.g., Turner et al. 2008), which could face lawsuits if it gets things wrong. Advertisements also come to mind, because the interests of the advertiser may not coincide with those of the customer. – Cases where vagueness can save money or face are plentiful, yet one wonders whether vagueness can also be advantageous in situations where it is one's aim to inform an audience as well as one can.

5 Vagueness in the absence of conflict

The question why vagueness is used in situations where the interests of speakers and hearers are essentially aligned was asked perhaps most forcefully by the economist Barton Lipman. Lipman assumed a game-theoretic model in the tradition of Crawford and Sobel (1982). In such a model, speakers express Contents using Forms, and hearers use Forms to decide upon Actions. In the simplest types of situations, choices are deterministic, hence both strategies are functions:

(Speaker strategy) S : Contents → Forms
(Hearer strategy) H : Forms → Actions

Deterministic strategies are known as *pure* strategies; an example of a pure speaker strategy would be to say 'tall' to describe any person whose height is above, say, 180cm, and 'short' in all other cases. Strategies can also be probabilistic, in which case they are called *mixed*; an example of a mixed speaker strategy would be to say 'short' to describe any person whose height is below 160cm, 'tall' to describe any person whose height is above 180cm, and to choose by flipping a coin in all other cases (i.e., when observing a person between 160cm and 180cm). Each action is associated with a utility value, which is represented as a number between 0 and 1. Speaking in general, the speaker and hearer may associate different utilities with a given action, which may, without loss of generality, be assumed to lie between 0 and 1:

U_S Actions $\rightarrow [0, 1]$
U_H Actions $\rightarrow [0, 1]$.

[7] De Jaegher's is a more complex version of the game of the *two generals* (section 2). It lets one general tell the other about the *preparedness* of the enemy. The utility of vagueness hinges on a subtle asymmetry between the generals, only one of whom will suffer if the enemy turns out to be *prepared*. Instructive though his paper is, I find it difficult to see how De Jaegher's game is relevant to everyday communication or NLG.

To model the absence of conflict, Lipman focusses on situations where these two functions are equal: each action is valued in the same way by the hearer and the speaker.

This formal model allows Lipman to make his question precise, but he does pay a price, because there may be cases where an intuitively reasonable answer does not "fit" into the mold of Crawford and Sobel. For example, their model does not easily allow one to associate a cost (i.e., a negative utility) with a Form, for instance on the basis of its length, as one might be tempted to do (see e.g., the third answer in section 5 below).

The core of Lipman's paper is a theorem which says, in essence, that in the absence of conflict, the best speaker strategy is always pure. If we apply this idea to the description of people's heights again then this means that, in the absence of conflict, it must be optimal for both hearer and speaker to always (i.e., deterministically) describe each given height in the same way. Borderline cases, and hence vagueness, cannot be advantageous under these assumptions. To make the lessons from his theorem concrete, Lipman used an *airport scenario* involving two players. Here is the gist of the scenario in its simplest form.

> *Player 1 asks player 2 to go to the airport to pick up an acquaintance of player 1. Player 1 knows the referent's height with full precision, while player 2 carries a perfect measuring device. There are two other people at the airport (whose heights are not known in advance), and it is assumed that heights are distributed uniformly between a maximum denoted by* 1 *and a minimum denoted by* 0*; this means that every height occurs equally often. The payoff for both players is* 1 *if player 2 successfully picks the referent, while it is* 0 *if the first person she taps on the shoulder turns out to be someone else.*

Lipman argues that, under these assumptions, player 1 would be foolish to use vagueness: why say 'He is tall' if you can say 'He is 183.721cm'? By stating his acquaintance's exact height, player 1 will allow player 2 to identify this person with almost complete certainty, given that the chance of two people having the exact same height is almost nil. Lipman realises that this strategy would imply that the language is able to distinguish between all possible heights, so he also asks what would happen if only two words were available to player 1, regardless of the person he needs to identify. He shows that, under these assumptions, optimal communication arises if these words are used in accordance with the following rule:

> Say 'tall' if height(person) $> 1/2$, else say 'short'.[8]

This concept does not involve vagueness, of course, because the rule does not allow any borderline cases: everyone is either tall or short. In other words, no

[8] Under the assumption of uniformity, the value 1/2 represents both the average and the median height of all the people at the airport. Note that the crisp expressions considered by Lipman can be seen as instances of estimation ('183.721 cm') and categorisation ('tall', 'short') respectively. This distinction will be taken up in section 6.2.

rationale for vagueness has been found so far.

Based on these insights, Lipman asks what explains people's abundant use of vagueness. For, on the strength of his theorem, vagueness might easily seem "a world-wide, several-thousand-year efficiency loss". Note that Lipman is not simply asking whether vague utterances can be useful; he is asking whether vague expressions can be *more* useful than any non-vague expression – or at least equally useful as these. Like us, he is interested in situations where there is no conflict between speaker and hearer. He discusses reasons why vagueness might sometimes be optimal nevertheless, even though this appears to fly in the face of standard game theory. This can only happen, of course, in situations where one or more of the assumptions on which his theorem rests are not met. Such situations cry out for standard game theory to be modified.

5.1 Answering Lipman

In what follows, we consider a number of possible answers to Lipman's puzzle, some of which are also mentioned by Lipman. We choose to express these possible answers in an informal, nontechnical style that matches the exploratory nature of the ideas. Unlike Lipman, we shall not assume any particular game-theoretical model, asking more generally how vagueness can benefit the speaker and the hearer. In other words, our focus will be on sketching and evaluating possible answers to Lipman's question. The question how best to "implement" these answers in a game-theoretical formalism will be left aside here.

First answer: Necessary vagueness. Sometimes I speak vaguely because that is the best I can do. This happens, for example, when all I have is someone else's spoken report: if you tell me there are 'many' students on a given university course and I trust your judgment then if someone asks me, the best I can do may be to pass on your assessment *verbatim*. Practically important though this may be, second-hand vagueness does not give us much insight, because it begs the question why I was given vague information in the first place.

Various authors have claimed plausibly that vagueness can also stem from the limitations of human *memory* (Dow 1991, Rubinstein 1998).[9] Suppose I told you that 324,542 people perished in some cataclismic earthquake. For a while, you might remember the exact death toll. In the longer term, however, details are likely to be corrupted (if you remember the wrong number) or lost: the next day, you may only recall that the victims numbered in their hundreds of thousands; a year later, you may only remember that there were many. Similarly, I may remember the position of a shooting star with considerable precision after a few seconds, but a minute later I can only point vaguely at some vast section of

[9]Baddeley 2007 discusses a number of mechanisms that might play a role in memories becoming increasingly vague over time, including *semantic coding* (whereby only the meaning of an expression is retained, while the expression itself is forgotten) and *chunking* (whereby a handful of items are clustered together as one unit). There exists a rich literature on memory and forgetting, but I do not know of any psychological research directly addressing the fading (in the sense of becoming vague) of memories over time.

the sky. Human memory appears to economise on memory space, while only retaining the gist of the information presented to it – on a good day, that is.[10]

Lipman also discusses the question whether the imprecision of perception itself might be responsible for vagueness. Suppose you are looking at a person, trying to estimate his height. Lipman doubts that the imprecision associated with your observation implies that vagueness is unavoidable, because uncertainty can be expressed precisely, for example through a probability distribution. But suppose the person whose height you are estimating is sitting in a chair. In this situation you may still be able to tell that he is tall, but height estimation becomes even harder than before. If your assessment of his height can be captured though a probability distribution then it seems thoroughly unclear where these probabilities come from. – Considering complications of this kind, we leave the imprecision of perception here as an unresolved question.

In search of other situations where vagueness is unavoidable, let us return to the original airport scenario, assuming that the speaker describes the person in question as 'taller than 175cm'. Lipman saw this as a crisp description of the person, because he assumed it to be based on absolute measurement. One could question this position because, on closer inspection, even 'taller than 175cm' is actually *vague*, because height measurement, like all distance measurements, is affected by imprecision (Van Deemter, in preparation). Still, the concept 'taller than 175cm' is far *less* vague than our colloquial word 'tall': surely almost every person is either taller than 175cm or not; only people whose height is such that it might equally be measured as above or below 175cm are borderline cases. So although the imprecision of measurement and perception force us to be vague, they do not force us to be *as vague as we actually are*. In this sense, Lipman is right: these limitations do not explain the extent to which vagueness affects our communication. (At the end of this section, when discussing the seventh answer to Lipman, we shall return to the issue of perceptual limitations.)

Second answer: Apparent vagueness. Let us modify Lipman's scenario. In a new *modified airport scenario* the speaker knows the heights of all three people at the airport. Suddenly it becomes easy to understand why vagueness can be useful. For suppose your acquaintance happens to be the tallest of the three. You can then identify him as *'the tall guy'*. Arguably, this is safer than citing the person's height in centimeters, because 'the tall guy' does not require the players to make any measurements: *comparisons* between heights can often be made in an instant, and with more confidence than absolute measurements.

To see what I mean, suppose two similarly shaped old mansions are standing next to each other in a street where all other buildings are much lower: one of them is infested by wood-destroying insects and needs to be demolished as soon as possible, while the other will be lovingly restored to its former glory. The

[10] It would be possible to go one step further and ask *why* memories are vague. Instead of storing the information about the earthquake vaguely, you could have reduced information by remembering that there were between 10^5 and 10^6 casualties, which does not involve borderline cases and is therefore not vague. Issues of a similar nature will be discussed in connection with the third answer.

first house is 20 metres tall, the other one 19. Arguably, it is safer to tell the demolition company to destroy “the tall house” than to destroy ‘the house that is 20m tall’, since the latter requires some laborious measurement, whereas the former does not. It is far easier to see that one house is taller than another than which of the two is 20 metres tall.

In cases like this, one can argue that vagueness is only local. It is well known that *ambiguity* can be merely local, for example when the sentence as a whole allows one to disambiguate an ambiguous word in it, for example when a pronoun gets resolved or a lexical item disambiguated. This happens, for example, when we say ‘the letter was conceived and written within an hour’, the ambiguous word ‘letter’ is clarified by the following words. Similarly, in the modified airport scenario, the noun phrase as a whole (e.g., ‘the tall guy’) allows no borderline cases, so there is no *global* vagueness here even though no exact threshold is implied.

The precisifying effect of referring expressions on the vague terms in them is well understood (e.g., Kyburg and Morreau 2000 for a semantic analysis; DeVault and Stone 2004 and Van Deemter 2006 for computational accounts), but vague words can be precisified in many ways. When I say of a gymnastic exercise that it is ‘good for young and old’, for example, there is nothing vague about my description: I am using vague words to say that this exercise is good for people, regardless of their age. Or consider the slogan ‘bad for bacteria, good for gums’, on a toothpaste tube. It would be beside the point to ask ‘How bad?’ or ‘How good?’, since the aim of the description is to identify the mechanism through which the toothpaste claims to work: bacteria are killed, and this will benefit your gums. The removal of vagueness works differently from one context to another. For example, we say ‘His feet were large but his hands were small’, we essentially say that his hands were smaller than one might expect given the size of his feet; the vagueness in the words ‘large’ and ‘small’ is only apparent. Yet another type of apparent vagueness arises in law-like statements such as ‘when there is much water in the bathtub, a lot of water will flow through the drain’, where we are making a crisp statement about the connection between two variables: the amount of water in the tub and the amount of water flowing out in any given unit of time. Once again, vague terms are employed to make a crisp statement. – Far from saying that vague expressions are *always* completely resolved by context, it is worth realising that they often are.

It has, of course, long been understood that when a vague adjective is conjoined with a noun, the interpretation of the adjective is influenced by the noun: to be counted as a tall baseball player involves different standards of tallness than to be counted as a tall jockey, for example (e.g., Klein 1980, Bartsch and Vennemann 1983). Even though this can be analysed as letting the noun sharpen up the meaning of the adjective, the cases discussed above involve a much more radical sharpening: the domain of the two houses, for example, gets divided into two crisp sets: the house that is tall and the house that is not tall. Although this can leave uncertainty of the categorisation of houses that are not in the domain (e.g., when they are 19.5m tall), such houses are irrelevant in the situation at hand: the referent of the expression ‘the tall house’ is denoted with complete crispness, so the communication does not involve any borderline cases.

Third answer: Vagueness as cost reduction. It has been suggested that strategic vagueness can arise from a desire to reduce the "cost" of the utterances involved, as long as this does not cause too large a reduction in their benefits (Van Rooij 2003, Lipman 2006, Jäger 2008). Broadly speaking, the idea is that vague expressions are easier to produce and interpret than precise ones. In principle, this sounds highly plausible. The problem is, however, that whenever vague expressions reduce cost, other factors tend to be implicated. We have encountered one situation where this was the case: cost is a factor underlying the success of the referential expressions discussed above: 'the tall house' is easier to produce and interpret than 'the house that is 20 metres tall'; we have seen, however, that the vagueness of 'tall' in this context is only apparent, so it is unclear whether such cases are best viewed as using vagueness for cost reduction. Other cases of cost reduction are complicated in other ways. Let us look at an example.

Consider a doctor who measures your body temperature as 38.82 Celcius. By stating that you have 'a high fever' (instead of 'thirty eight point eighty two degrees') the doctor is pruning away details that are of questionable relevance in the situation at hand. Perhaps this is vagueness as cost reduction. But a Lipman-style argument suggests that cost reduction is no reason for using vague language, as defined by the existence of borderline cases. The doctor could have achieved a similar economy without vagueness by saying that your temperature is 39 degrees, using the normal rounding conventions; or she could have said that your temperature is above 38. Either way, she would have reduced information without being vague. Conversely, she could have been vague without achieving any cost reduction, for example by saying that the temperature was '(somewhat close to) 40 degrees'.[11] In other words, cost reduction does not imply vagueness, nor does vagueness imply cost reduction.

The situation is similar if one argues that vagueness leads to cost reduction in *learning* the concepts involved (see also Lipman 2006), because vague words are part of a general mechanism that makes their meaning dependent on the context in which they are used. The basic observation here is that the size constraints on 'a small elephant', for example, are very different from those on 'a small mouse', but they follow from what it means to be an elephant and what it means to be a mouse. For example,

(1) tall X = those X that are taller than the X are on average; or
(2) tall X = those X that are taller than most X.

The *context-dependence* of vague words like 'small' makes them much easier to learn than they would otherwise be, and this is, no doubt, one of the mechanisms that help make vagueness such a useful device. But once again, it is not clear why this mechanism means that the words in question must be vague (i.e.,

[11] Round numbers tend to be interpreted more vaguely than other numbers; also, they tend to be brief and hence easy to produce and process. See Krifka (2002) for details.

allow borderline cases). Crisp concepts can be context dependent as well. In fact, definitions such as (1) and (2) result in a *crisp* interpretation for expressions of the form 'a small so-and-so': if we took these definitions seriously then 'tall X' would not contain borderline cases and hence would not vague at all.

Fourth answer: The benefits of bias. Frank Veltman and others have stressed that vagueness does not only give us *less* of something (e.g., less precision), it can also give us *more* of something, namely bias or evaluation. Perhaps the clearest evidence comes from the fact that they are often used in situations where other functions are difficult to detect. If you tell me you paid 200 pounds for your shoes, I might respond "That's expensive!". My utterance cannot be designed to tell you what the price is, because you know this. My point is merely to tell you what I think about it. My utterance is of the form $p > n$, where p is the price of these shoes, and n the normal price (or the highest acceptable price) for shoes of this kind. Since the value of p is a given, the news for you can only lie in what you learn about n.

Bias is particularly useful if measurement is problematic (cf., the fifth answer). Most of us, who measure coffee in cups or spoonfulls, are none the wiser if a medical information leaflet informs us that a pill against headache contains 25mg of caffeine. (Is that enough to keep me awake?) The same is true if we hear that our blood pressure is 140/90, for example, until we are told what these figures mean in terms of our health. Expressions like 'high blood pressure' finesse this problem handsomely, by telling us that there is cause for concern. There is, of course, a fine line between helping and patronising; in many situations, the expression of bias is only justified if full, unbiased information is provided as well, but that's another matter.

There is, however, a puzzle to be solved here, along the same Lipmanian lines as before. If bias needs to be expressed, then why not simply say it? There are a number of options here: The doctor could have said explicitly that this reading should be considered worrisome. For why should bias necessarily be coupled with vagueness only, given that it is as easy to think of a crisp expression that contains bias as it is to think of a vague expression that does not (e.g., in the case of an adjective like 'tall')? There is no need to assume that this should necessarily lead to very lengthy descriptions (cf., the third answer). A good example of *crispness* + *bias*, expressed succinctly, is the word 'obese', in the presently-accepted sense of having a Body Mass Index of over 30. For the reason why obesity was defined in this way is precisely that this degree of overweight is considered medically worrisome (van Deemter, in preparation.

Fifth answer: The lack of a good metric. For simplicity, we have pretended that when information is expressed, the starting point is always a well-defined chunk of information. Similarly, we have assumed that our words must always correspond with well-defined metrics in well-understood ways. In many situations, however, this is far from true. Suppose someone is happy, or beautiful, or pleasant to be with. (Lipman discusses the word 'nice' in this connection.) It is difficult to quantify such things. All we have in such cases is vagueness, it seems, and how this vague information is represented in our heads is a mystery: "the throbbing centres of our lives appear to be describable only in vague terms" (Sainsbury 1990).

It might be thought that this problem affects only highly complex social domains, but this seems incorrect to me. If I call a house small, for example, it may well be unclear to you how I combine its square footage with its height. Perhaps that attic, which extends across the entire footprint of the house but whose height is barely 1 metre, is of no use to me at all, and this may have affected my qualification. In situations like this, even comparatives can fail to apply, because whether house x or house y is considered the larger of the two may depend on the way in which the different dimensions of the house are weighed. It is, consequently, difficult to see how even such an apparently simple thing as the size of a house can be measured objectively.

Sixth answer: Future contingencies. When a doctor tells you that your symptoms will disappear "in a matter of days", the vagueness of his prediction makes it more difficult for you to complain if, in fact, they linger for a full week. The doctor is uncertain how the future will turn out, which is why he, sensibly, wraps his predictions in vagueness. Some of the same issues are relevant here as in the case of the first answer, where the imprecision of perception was discussed, and we asked whether it would have been feasible to express uncertainty in a way that does not involve vagueness. An interesting class of cases, where vagueness seems difficult to avoid, arises from the area of laws and contracts.

Laws and contracts notoriously contain expressions which can only be applied to a concrete situation after certain unclarities in them have been resolved (Hart 1994, Mcleod 2007, Lipman 2006). These unclarities can sometimes be an attempt to avoid a conflict between people who disagree over some issue. A good example is the timeframe within which abortions should be allowed, which is sometimes denoted vaguely as 'a reasonable time period' for this reason. But vagueness can also arise for reasons that have little to do with conflicts between lawmakers. In such cases, vagueness can be seen as a way of *parameterising* the law and thereby making it applicable to a larger variety of situations, whose details are impossible to foresee. Lipman, who discusses these cases as well, cites phrases like "appropriate care" (which is often used in clinical contexts, among others) as examples.

In areas where social norms or technology play a role, the usefulness of this kind of vagueness is particularly obvious. In 1981, for example, the British Parliament passed a law forbidding the public display of *any indecent matter* (except, notably, in art galeries) but it did not specify what was meant by indecent. Another often-invoked example is an injunction against using vehicles in a park. Does this apply to scooters? Bicyles? And how about the newfangled skates that allow you to go very fast? The word 'vehicle' is not well-enough defined to tell (Mcleod 2007). The vagueness in these laws means that they stand a good chance of surviving for a long time: the law against indecency, for example, forbids us from doing what is indecent *at a given moment*. The downside is that there are bound to be differences of opinion over what indecency is and how indecent something has to be to be covered by this Act.

Strictly speaking, the mere fact that expressions like ‘indecent’ and ‘vehicle’ depend on context for their interpretation does not make them vague. (Compare the word ‘I’, which refers to a different person depending on who is speaking, but always only to just that person, so no vagueness is involved.) I take it that they are vague because context affects the interpretation of these words in ways that are impossible to foresee: their precisification depends on who it is that does the precisifying, for example – and what they have had for lunch.

Seventh answer: Vagueness facilitates search. The interpretation of vague predicates is subject to substantial variation between different people (and probably within a given person too). The evidence for this claim is ample, and perhaps best understood physiologically in the case of colour. The density of the pigment layers on the lens and retina differs even among people who are able to distinguish between the same number of colour shades, and the same is true for the sensitivity of the photo-receptors in their eyes (e.g., Hilbert 1987). These differences mean that some colour shades that can be separated by one person cannot be separated by another. Consequently, when different people call a book blue, for example, they do not necessarily mean the same part of the colour spectrum. The existence of such mismatches has considerable consequences. Rohit Parikh showed that not all is necessarily lost when they arise (Parikh 1994). Unlike the present paper, his point was not to show any advantages that might stem from vagueness, but merely that when conceptual mismatches occur, communication can still be useful (and to draw certain conclusions concerning the nature of propositions and the methododology of formal semantics). To explain his point, he tells a story which is slightly reworded here:

Ann and Bob work at a college where Ann teaches Maths and Bob History. One day around noon, Ann is in her office while Bob is working from home. Ann calls Bob asking him to bring her book on topology, when he comes in after lunch. Ann has only one topology book, and Bob will know the book when he sees it. Bob asks *What does it look like?*, and Ann says *It is blue.* ‘Blue’ is understood differently by different people, so when Bob starts checking all the books that he calls blue, he *may* be lucky and find the topology book among them. But if he is unlucky, the book is one of those Ann calls blue but he does not. If this happens, he may have to check *all* Ann’s books. Parikh assumes that Ann and Bob understand ‘blue’ as a crisp concept, and he demonstrates by means of a concrete numerical example that, as long as the overlap between the two concepts, blue_A and blue_B is large enough compared to their symmetric difference, then Ann’s instruction (‘It is blue’) helps to reduce Bob’s expected search time.

Our seventh answer to Lipman is of a subtly different nature than the other ones: it does not explain primarily why the speaker opts for a vague expression, but why it is advantageous (for both speaker and hearer) if the hearer understands an expression as vague. To understand the idea, it might help to realise that, in Parikh’s story, it would be advantageous for Bob to rise above thinking in terms of one crisp dychotomy between blue and non-blue. Bob might argue, for example, that if the target book is not found among the ones he considers blue, then it's most likely to be one of those he considers *almost* blue, or somewhat blue; so after inspecting the books he considers blue, he would be wise to inspect these other books, which he considers only somewhat blue.[12] A few variants of this idea will be elaborated in the following section. For simplicity, we focus on the one-dimensional word 'tall' (rather than the notoriously complex colour terms), which will have the effect that, of any two extensions that the word may be assigned in a given situation, one must always be a subset of the other.

6 Vagueness facilitates search

The story of the diamond is set in Beijing's Forbidden City, long ago.

> *A diamond has been stolen from the Emperor and, security being tight in the Forbidden Palace, the thief must have been one of the Emperor's 1000 eunuchs. A witness see a suspicious character sneaking away. He tries to catch him but fails, getting fatally injured in the process, while the scoundrel escapes. With his last breath, the witness reports "The thief is tall!", then he gives up the ghost. – How can the Emperor capitalise on these momentous last words?*

Suppose the Emperor thinks of 'tall' as a *crisp* concept, meaning 'taller than average', for instance. In this case, his men will gather all those eunuchs who are taller than average, perhaps about 500 of them. In the absence of any further clues, he should expect to search an average of as many as 250 tall people (i.e., half of the total number).

Matters get even worse if we allow the possibility that the witness may have used a more relaxed notion of 'tall' than the Emperor. If this Parikh-style mismatch arises, it is possible that the Emperor does not consider the thief to be tall, so the perpetrator will not be among the eunuchs whom the Emperor considers to be tall. Once again, since classical logic does not make height distinctions between non-tall eunuchs, all of them are essentially alike, so the Emperor's men can only search them in arbitrary order. In other words, he first searches 500 eunuchs in vain, then an expected $0.5 * 500 = 250$, totalling 750. In this unlucky situation, the witness' statement ('the thief is tall') was actually detrimental to the Emperor's search effort, because it caused him to search 500 people in vain and loose precious time as a result; the Emperor would have been better off without any description of the thief's height, in which case he should have expected to search $0.5 * 1000 = 500$ eunuchs. The Emperor could have diminished the likelihood of a false start (i.e., a search strategy based on a

[12] Parikh does hint briefly at a closely related possibility, but without discussing its implications: "It may be worth pointing out that probably Bob does have another larger set of Bluish(Bob) books which includes both Blue(Ann) and Blue(Bob). After looking through Blue(Bob), he will most likely look only through the remaining Bluish(Bob) books. (Parikh 1994, p.533)."

notion of 'tall' that excludes the thief) by counting more eunuchs as tall. But in doing so, he would have increased the search times that are necessary to inspect all the eunuchs he considers tall. The only way to avoid the possibility of a false start altogether is by counting *all* eunuchs as tall, which would rob the witness' statement of its usefulness.

6.1 The benefits of borderline cases

If the Emperor thinks of 'tall' as *vague*, however, then he might separate the eunuchs into three rather than two groups: the ones who are definitely tall, the ones who are definitely not tall, and the borderline cases that are characteristic of vague concepts. For concreteness, assume 100 eunuchs are definitely tall, 500 are definitely not tall, and 400 are doubtful. Surely, the eunuchs in the "definitely tall" category are *more likely to be called tall* than the ones in the "doubtfull" category, while noone in the "definitely not tall" category could be called tall. To put some figures to it, let the chance of finding the thief in the group of 100 be 50% and the chance of finding him in the doubtful group of 400 likewise. Under this scenario, it pays off to search the "definitely tall" eunuchs first, as one may easily verify.[13] In other words, the Emperor benefits from regarding 'tall' as containing borderline cases (i.e., being vague).

This thought experiment suggests that borderline cases, and hence vagueness, can facilitate search, because borderline cases allow us to distinguish more finely than would be possible if all our concepts were crisp. If your language contains only crisp concepts then separating the eunuchs into three different groups does not make sense to you: there are tall eunuchs, not-tall ones, and that's it. Only if you understand 'tall' to have borderline cases can you distinguish between the different people whom you do not consider tall, as well as between the ones you consider tall and all the others. But if distinguishing between three different categories (tall, not-tall, and borderline tall) is better than distinguishing between just two, then surely it must be even better to distinguish even more finely. The logical extreme of this idea is explored in the following section.

6.2 The benefits of degrees

As is customary, we have taken vagueness to be the phenomenon where a linguistic expression has borderline cases: a man who is borderline tall, for example. But if the word 'tall' was modelled accurately by a Partial Logic that distinguishes crisply between tall, not-tall, and borderline-tall then this word would not pose much of a puzzle to semantic theories, as various philosophers have pointed out (e.g. Wright 1976): the sorites paradox, for example, could easily be avoided. What is problematic, from this perspective, is the existence of borderline cases of borderline cases, and so on: a phenomenon known as *higher-order* vagueness. Assuming that we understand why languages contain words

[13] The expected search effort will amount to $0.5 * 50 + 0.5 * (100 + 200) = 175$ in this case, compared with $0.5 * 200 + 0.5 * (400 + 50) = 325$ if the wrong order is chosen.

that allow borderline cases, one could go on to ask why languages contain words that allow higher-order vagueness. Some of the factors discussed in section 5 do not apply to higher-order vagueness; it seems, for example, hard to see how higher-order vagueness causes more cost reduction than first-order vagueness. Answer 5, by contrast, offers a much better explanation of higher-order vagueness because, in the absence of a good metric, a partial model (involving a crisp separaton between tall, not-tall, and borderline-tall) would be as difficult to enforce as a classical one. In what follows, let's see whether the same might be true for answer 7, which focusses on benefits for search.

Our starting observation is that, in the story of the diamond, the Emperor can do even better than was suggested above, if he uses what I shall call a *ranking* strategy. Suppose he has the eunuchs arranged according to their heights. First the tallest eunuch is searched, then the tallest but one, and so on, until the diamond is found. This strategy is faster than each of the other ones if we assume that *the taller a person is, the more likely the witness is to have described him as tall.* Under this assumption, the same type of advantage obtains as in the previous case (where only borderline cases were acknowledged), but at a larger scale.

Note that we are ascribing a ranking strategy to the Emperor (i.e., the hearer) only. For all we know, the witness (i.e., the speaker) may be ignorant of the eunuchs' heights while only possessing a rough impression of that of the thief. The Emperor and his men, by contrast, can rank the eunuchs at their ease.

What conclusions are we to draw? First, it is useful for the Emperor to understand that 'tall' is a *gradable* concept. Without this awareness, it makes no sense for him to rank the eunuchs. So not only are there benefits to giving words like 'tall' and 'blue' borderline areas; there are even greater benefits to understanding them as applying to different degrees.

But we can learn even more from examples of this kind, because our observations appear to have important implications for the semantics of vagueness. To see why, one should realise that viewing 'tall' as a crisp concept is what happens if the meaning of the word is modelled using *Classical Logic*; viewing 'tall' as inviting a separation between three sets of people is consistent with *Partial Logic*; ranking them in order of their height requires a richer semantic model, where degrees are acknowledged in one way or other, as is the case, for example, in Chris Kennedy's work (e.g., Kennedy 2001) and in theories that use multiple truth values (e.g., Fuzzy Logic or the probabilistic proposal of Edgington 1996; see also van Deemter (in preparation)). So firstly, Partial Logic is better placed to understand our use of words like 'tall' and 'blue' than Classical Logic. Furthermore, our reasoning suggests that degree theories are in an even better position to do this. In other words, the story of the diamond emerges as something of a tie-breaker in the contest between different theories of vagueness.

A few objections to this line of argument are worth mentioning briefly. In theory, the ranking strategy is consistent with Classical Logic, but only under some highly nontrivial assumptions. One might, for example, grant a classical logician the knowledge that other agents use other models than his own; furthermore, he might be allowed to argue that if x is larger than y, then x is more likely to be in an agent's denotation of 'tall' than y. Under these assumptions, even a classical logician could use a ranking strategy. Something similar would hold if we granted the classical logician knowledge of a large number of qualifiers, such as '(large)-ish', 'somewhat (tall)', and so on, in which case he could start by considering all the eunuchs he considers 'tall', then (if the perpetrator is not found among them) all remaining eunuchs that he considers tall-ish, and so on. It appears, however, that these alternative explanations smuggle in degrees through the backdoor: a classical logician who knows what models corresponds with each possible threshold for 'tall' would have an essentially graded understanding of what it means to be tall.

Our analysis of the story of the diamond has contrasted crisp and vague interpretations of *categories* (i.e., adjectives) with each other. One might ask why the witness did not use a numerical *estimate* instead, estimating the height of the thief as accurately as he can (cf. section 5, footnote 7). But it is difficult to compare a category with an estimate. To compare like with like, one would have to compare a *precise* estimate with a *vague* one. A precise estimate might say, "The thief is 190cm tall, +/− 5cm"; a vague one "The thief is roughly 190cm tall", leaving the margin of error unspecified. When vague and precise estimates are compared in this way, similar issues come up as in our discussion of vague and precise categories: a precise estimate is, strictly speaking, unable to allow borderline cases, for example, and it cannot take into account that values become less likely as they get further removed from the estimated number. Crisp estimates, in other words, appear to have the same drawbacks as crisp categories.

7 Conclusion

Two strands can be discerned in the body of this paper. First, we have argued that *utility* should play a crucial role in our thinking about language generation and production. Thanks to authors such as Van Benthem, this claim might now be taken for granted by many students of language and communication (e.g., Parikh 1994, Merin 1999, Rubinstein 2000, Van Benthem 2000, 2008, Parikh 2000, Van Rooij 2003, Benz et al. 2009), but it is still something of a novelty for computational linguists. Secondly, this paper has discussed some specific questions concerning the *rationale for vagueness* in language and communication. Having reviewed and commented on a number of answers that have come up in the literature, we concluded with the discussion of an answer that has, to the best of my knowledge, not been suggested before.

7.1 Why language is vague

We set out to answer the question why language is vague. We found that it is fairly clear why vagueness can have differential benefits in communication between agents whose interests differ substantially[14], such as a politician and his potential voters (Aragonès and Neeman 2000), or like a professional who does not wish to be sued by his clients. In situations of this kind it can be beneficial to obfuscate, exploiting the borderline cases inherent in vague expressions.

A harder question, in the footsteps of Lipman (2000, 2006), is whether vagueness might be useful when speakers' and hearers' interests are closely aligned. In attempting to answer this question, we started off by examining various situations in which it might be argued that vagueness is unavoidable. This led us to consider one undeniable but relatively uninteresting trigger for vagueness (namely *second-hand* vagueness), one plausible but so-far unsubstantiated one (vagueness stemming from the *fading of memories*), and a more doubtful one (vague perception). Lipman's arguments led us to question *cost reduction* and *bias/evaluation* as independent explanations of the vagueness of language (but see below, where these factors are regarded from a less theoretical perspective). A number of other triggers were more difficult to dismiss. This was true for what we called *apparent* or *local* vagueness, and for the types of vagueness that stem from the *absence of a generally-agreed metric*, from the anticipation of *future contingencies* and, particularly, from the *facilitation of search.*

7.2 When to be vague?

Since these explorations were partly motivated by a desire to shed light on the decisions that are made by an Natural Language Generation (NLG) system, let's see how our findings can inform such choices, and the choices of a human speaker as well. We address the seven different factors one by one, this time focussing on language as *parole* (i.e., language use) rather than *langue* (i.e., the language itself) (de Saussure 1916).

1. Necessary vagueness. Second-hand vagueness occurs in NLG when the knowledge base from which the generator takes its departure makes use of vague concepts. Far from being unusual, this is probably the norm. Knowledge bases routinely incorporate vaguely defined concepts, for example when a table lists a risk as 'high', or a day as 'cloudy'. The decision to allow vague concepts was made by the designer of the database scheme in such cases, and the decision to populate the database in a particular way (e.g., by classifying yesterday as cloudy and today as sunny) was made by the person who entered the different items into the knowledge base.

2. Apparent vagueness. Reference is one of the tasks that NLG systems routinely perform. When NLG systems produce referring expressions that contain vague expressions, apparent vagueness plays precisely the "disambiguating" role that was discussed in the second answer, section 5.1. (See Van Deemter 2006 for details.)

3 and 4. Cost reduction and bias. These factors were tentatively dismissed in the theoretical discussions above, but they do play a considerable

[14]Compare Horton and Keysar 1996 for experimental evidence of speaker's laziness in situations where their interests are approximately aligned. Work in this psycholinguistic tradition casts doubt on the the idea that speakers' and hearer's interests can ever be perfectly aligned.

role in the decisions of a speaker, whether the speaker is a human or a machine. The reason for the difference is as follows. Lipman's original question was why *language* is vague. To answer this theoretical question, it was not sufficient to consider all the different ways in which something can be said, for example in English, and demonstrate that the best of these is vague: one would have to demonstrate that it could not have been any other way. Our discussion of bias demonstrates, for example, that English *might* have contained a crisp expression that means 'your shoes cost more than 100 pounds and I disapprove of that', but if no such expression is actually available in the language then an English speaker (or an NLG system), may be forced to use a vague expression. Lipman asked why language is what it is, not why language is used (or should be used) in a particular way.

5 and 6. Lack of a good metric, and future contingencies. These factors are relevant to language production (by people), and they affect language generation (by machines) in principle. For instance, NLG applications such as weather forecasting are clearly subject to future contingencies, and these can trigger vagueness as discussed in section 5. Admittedly, present-day NLG systems tend to focus on highly concrete domains, which are based on data that is fairly easy to measure. Present NLG systems therefore seldom use vagueness because of the absence of a good metric, but this may well change in future.

7. Vagueness facilitates search. Suppose a speaker had a choice between a crisp concept tall_1 and a vague concept tall_2. If the hearer understands what meaning the speaker attaches to tall_1 and tall_2 then it is preferable that the speaker use the crisp concept tall_1. As we have seen, the same may not be true if the hearer has to guess how these concepts are intended. For a practical NLG system, it is often advisable to avoid semantic mismatches altogether, by defining its terms explicitly and sticking rigidly with these definitions.[15]

7.3 When *not* to be vague?

The flipside of Lipman's question is *When is it* bad *to be vague?* Evidently, vagueness is bad in situations where more precision would have given the speaker and hearer a higher utility than other expressions. Lipman's airport scenario is a good example of a situation where this is the case. But additionally, it is bad to be vague when vagueness is likely to be confused with precision. This is perhaps clearest when vague expressions are combined with numbers, as when we are told that $x\%$ of people feel unsafe in public spaces, that y commonly available cleaning products are poisonous, or that $z\%$ of children at a school have difficulty reading. In all such cases, the numbers suggest a clarity that does not exist, because they are combined with concepts (unsafe, commonly

[15] This was done very deliberately in the weather forecasting system of Reiter et al. (2005), for example. The authors show that words like 'evening' are used in surprisingly different ways by different professional weather reporters, and construct a generation system in which evenings (afternoons, etc.) start and end at well-defined hours. The consistency that results from this approach is probably the main reason why their system-generated texts are appreciated more highly than texts written by human authors.

available, poisonous, difficulty) that allow borderline cases. I take it that it is undesirable uses like this which give vagueness a bad name.[16]

Acknowledgments

Thanks are due to Albert Gatt, Chris Mellish, Ehud Reiter, Graeme Ritchie, Hans van Ditmarsch, Robert van Rooij, and Frank Veltman for discussions on the theme of this paper. I also thank two anonymous reviewers for very constructive comments. Funding from the UK's Engineering and Physical Sciences Research Council (EPSRC) under the Platform Grant "Affecting People with Natural Language" (EP/E011764/1) is gratefully acknowledged.

8 References

Aragonès and Neeman 2000. Enriqueta Aragonès and Zvika Neeman. Strategic ambiguity in electoral competition. *Journal of Theoretical Politics* **12**, pp.183-204.

Baddeley 2007. Alan Baddeley. *Working memory, thought, and action.* Oxford University Press. Oxford.

Bartsch and Vennemann 1983. Renate Bartsch and Theo Vennemann. *Grundzüge der Sprachtheorie: eine Linguistische Einfuerung.* (Principles of language theory: a linguistic introduction.) Tübingen: Max Niemeyer.

Bateman 1997. John Bateman. Sentence generation and systemic grammar: an introduction. Iwanami Lecture Series: Language Sciences. Iwanami Shoten Publishers, Tokyo.

Bateman and Zock 2009. John Bateman and Michael Zock's list of Natural Language Generation Systems. Downloadable from `http://www.fb10.-uni-bremen.de/anglistik/langpro/NLG-table/NLG-table-root.htm`. (Consulted 23 April 2009.)

Benz et al. 2009. Anton Benz, Gerhard Jäger, and Robert van Rooij (Eds.) Game theory and pragmatics. Palgrave Macmillan. Houndsmills, Basingstoke, UK.

Bermúdez 2009. José Luis Bermúdez. *Decision Theory and Rationality.* Oxford University Press. Oxford.

Blastland and Dilnot 2008. *The tiger that isn't: seeing through a world of numbers.* Second, expanded edition. Profile Books. London.

Board and Blume 2009. *Intentional Vagueness.* Working Papers 381, University of Pittsburgh, Dept. of Economics. Version of May 2009.

Crawford and Sobel 1982. V.Crawford and J.Sobel. Strategic information trans-

[16] For an informal discussion of many cases of "false clarity", see Blastland and Dilnot 2008 (*passim*), who appear to have coined this expression. See also Van Deemter (in preparation), Chapter 1.

mission. *Econometrica* **50**, pp. 1431-1451.

de Jaegher 2003. Kris de Jaegher. A game-theoretical rationale for vagueness. *Linguistics and Philosophy* **26**: pp.637-659.

de Saussure 1916. Ferdinand de Saussure. Course in general linguistics. (trans. Roy Harris.) London. Duckworth.

Dekker and Van Rooij 2000. Paul Dekker and Robert van Rooij. Bi-directional Optimality Theory: an application of Game Theory. *Journal of Semantics* **17**: 217-242.

DeVault and Stone 2004. David DeVault and Matthew Stone. Interpreting vague utterances in context. In Proceedings of the 20th International Conference on Computational Linguistics (COLING-2004), Geneva, Switzerland.

Dow 1991. James Dow. Search decisions with limited memory. *Review of economic studies* **58**, pp.1-14.

Ebeling and Gelman 1994. K.S.Ebeling and S.A.Gelman. Children's use of context in interpreting "big" and "little". *Child Development* **65** (4): 1178-1192.

Edgington 1996. Dorothy Edgington. Vagueness by Degrees. In Keefe and Smith 1997.

Goldberg et al. 1994. Eli Goldberg, Norbert Driedger and Richard Kitteridge. Using Natural-Language Processing to Produce Weather Forecasts. *IEEE Expert* **9** (2), pp.45-53.

Hart 1994. Herbert Lionel Adolphus Hart *The Concept of Law.* Oxford. Clarendon Press.

Hilbert 1987. David R.Hilbert. Color and color perception: a study in anthropocentric realism. CSLI Lecture Notes 9, Center for the Study of Logic and Information. Stanford, Ca.

Horton and Keysar 1996. William S. Horton and Boaz Keysar. When do speakers take into account common ground? *Cognition* **59**, pp.91-117.

Jäger 2008. Gerhard Jäger. Applications of Game Theory in Linguistics. *Language and Linguistics Compass* **2/3**.

Jeffrey 1983. Richard Jeffrey. *The Logic of Decision.* University of Chicago Press. Chicago.

Keefe and Smith 1997. Rosanna Keefe and Peter Smith (Eds.) *Vagueness: a Reader.* MIT Press, Cambridge (Massachusetts) and London.

Kennedy 2001. Chris Kennedy. Polar opposition and ontology of 'degrees'. *Linguistics and Philosophy* **24**, pp.33-70.

Khan et al 2008. Generation of referring expressions: managing structural ambiguities. 22th Int. Conf. on computational linguistics (COLING-2008), pp. 433-440.

Kibble 2003. Rodger Kibble. Both sides now: predictive reference resolution in generation and resolution. Proc. of Fifth International Workshop on Computational Semantics (IWCS-2003). Tilburg, The Netherlands.

Klabunde 2009. Ralph Klabunde. Towards a game-theoretic approach to content determination. Proc. of 12th European Workshop on Natural Language Generation (ENLG-2009).

Klein 1980. Ewan Klein. A semantics for positive and comparative adjectives. *Linguistics and Philosophy* **4**, pp.1-45.

Krifka 2002. Manfred Krifka. Be brief and be vague! In David Restle & Dietmar Zaefferer (eds), Sounds and Systems. Studies in Structure and Change. A Festschrift for Theo Vennemann, Mouton de Gruyter (Trends in Linguistics 141), Berlin. 439-458.

Kyburg and Morreau 2000. Alice Kyburg and Michael Morreau. Fitting Words: Vague Language in Context. *Linguistics and Philosophy* **23**, pp.577-597.

Lewis 1969. David Lewis. *Convention – A Philosophical Study.* Harvard University Press.

Lipman 2000. Barton L.Lipman. Economics and Language. "Comments" section, Rubinstein (2000).

Lipman 2006. Barton L.Lipman. Why is language vague? Working paper, December 2006, Department of Economics, Boston University.

Mcleod 2007. Ian McLeod. *Legal Theory.* Fourth Edition. Palgrave MacMillan Law Masters. Basingstoke and New York.

McDonald 1987. Natural Language Generation. In S.Shapiro *Encyclopaedia of Artificial Intelligence*, Volume 1. John Wiley, New York.

Mellish and Van der Sluis 2009. Chris Mellish and Ielka van der Sluis. Towards empirical evaluation of affective tactical NLG. Proc. of 12th European Workshop on Natural Language Generation (ENLG-2009)

Merin 1999. Information, relevance, and social decisionmaking. In L.Moss, J.Ginzburg, M.de Rijke (Eds.) *Logic, Language and Computation* II. Stanford.

Von Neumann and Morgenstern 1944. John von Neumann and Oskar Morgenstern. *Theory of games and economic behavior.* Wiley & Sons, Princeton, NJ.

Parikh 1994. Rohit Parikh. Vagueness and utility: the semantics of common nouns. *Linguistics and Philosophy* **17**: 521-535.

Parikh 2000. Prashant Parikh. Communication, meaning, and interpretation. *Linguistics and Philosophy* **23** pp. 185-212.

Peccei 1994. Jean Stilwell Peccei. *Child Language.* Routledge.

Portet et al. 2009. F.Portet, E.Reiter, A.Gatt, J.Hunter, S.Sripada, Y.Freer, and C.Sykes. Automatic generation of textual summaries for neonatal intensive care data. Artificial Intelligence, DOI: 10.1016/j.artint.2008.12.002.

Rasmussen 2001. Eric Rasmussen. *Games & Information: an introduction to* *game theory.* Third Edition. Blackwell Publishing.

Reiter and Dale 2000. Ehud Reiter and Robert Dale. *Building natural language generation systems.* Cambridge University Press. Cambridge.

Reiter et al. 2005. E.Reiter, S.Sripada, J.Hunter, J.Yu, and I.Davy. Choosing words in computer-generated weather forecasts. *Artificial Intelligence* **167**: 137-169.

Reiter 2007. Ehud Reiter. An architecture for data-to-text systems. In Procs. of 11th European Workshop on Natural Language Generation (ENLG-2007): pp.97-104.

Rubinstein 1998. Ariel Rubinstein. *Modeling Bounded Rationality.* MIT Press, Cambridge Mass.

Rubinstein 2000. Ariel Rubinstein. *Economics and Language: Five Essays.* Cambridge University Press. Cambridge.

Sainsbury 1990. Mark Sainsbury. Concepts without boundaries. London: King's College London. Reprinted in Keefe and Smith (Eds.) 1997.

Theune et al. 2001. M.Theune, E.Klabbers, J.R. de Pijper and E.Krahmer. From data to speech a general approach. *Natural Language Engineering* **7** (1): 47-86.

Turner et al. 2008. R.Turner, S.Sripada, E.Reiter and I.P.Davy. Using spatial reference frames to generate grounded textual summaries of georeferenced data. In Proceedings of INLG-2008. Salt Fork, Ohio, USA.

Van Benthem 2000. Johan van Benthem. Economics and Language. In Rubinstein (2000), Part (III), Comments.

Van Benthem 2008. Johan van Benthem. Games that makes sense: logic, language, and multi-agent interaction. In K.Apt and R.van Rooij (Eds.) Proceedings of KNAW Colloquium on games and interactive logic. Texts in logic and games, Amsterdam University Press.

Van Deemter 2006. Kees van Deemter. Generating referring expressions that involve gradable properties. *Computational Linguistics* **32** (2).

van Deemter et al. 2008. Kees van Deemter, Brigitte Krenn, Paul Piwek, Marc Schroeder, Martin Kleesen, Stephan Baumann. Fully Generated Scripted Dialogue. *Artificial Intelligence* 172 (10), pp. 1219-1244.

Van Deemter 2009. Kees van Deemter. What Game Theory can do for NLG: the case of vague language. Proc. of 12th European Workshop on Natural Language Generation (ENLG-2009).

Van Deemter (in preparation). Kees van Deemter. *Not Exactly: in Praise of Vagueness.* To appear with Oxford University Press in 2010.

Van Rooij 2003. Robert van Rooij. Being polite is a handicap: towards a game theoretical analysis of polite linguistic behavior. In Procs. of Theoretical Aspects of Rationality and Knowledge (TARK-9), Bloomington, Indiana.

Veltman 2002. Frank Veltman. Het verschil tussen ‘vaag’ en ‘niet precies’. (The difference between ‘vague’ and ‘not precise’.) Inaugural lecture. Vossiuspers, University of Amsterdam.

Wright 1976. Crispin Wright. Language-mastery and the sorites paradox. In *Truth and Meaning: Essays in Semantics*, G.Evans and J. McDowell (Eds.), p.223-247. Oxford: Clarendon Press. Reprinted in Keefe and Smith 1997.