Communicating Properties Using Salience-Induced Comparisons

T. Pattabhiraman and Nick Cercone
Centre for Systems Science & School of Computing Science
Simon Fraser University
Burnaby, B.C., CANADA V5A 1S6
patta@cs.sfu.ca and nick@cs.sfu.ca

Abstract

A method for generating simple comparison sentences of the form *A is like B* is proposed. The postulated input to the generator consists of the name of an entity *A*, and a set of descriptors about *A* in the form of attribute:value pairs. The main source of knowledge that controls decision making is a probabilistic conception of salience of empirically observable properties among concrete objects. We also use a salience heuristic based on the notion of property intrinsicness. The information-theoretic concept of redundancy is used to quantify salience in probabilistic contexts. Salience factors influencing selection decisions are modelled as utilities and costs, and the decision for selecting the best object of comparison is based on the maximization of net expected utility. The method proposed has been implemented in a generation system written in CProlog.

Introduction

In this paper we present a method for generating simple comparison sentences with the aid of the notion of salience. The method has been implemented in a generation system written in CProlog. The postulated input to the generator consists of a set of descriptors about an entity (say, *Mary's cheeks*) in the form of attribute:value pairs (like *colour:red, texture:smooth*, etc). The task we are focussing on consists of describing the entity through a comparison sentence of the form *A is/are like B* (e.g., *Mary's cheeks are like apples*), by means of which the hearer can infer the intended descriptors of *A*. Based on the format *A is like B*, we will refer to the entity being described (*Mary's cheeks*) as the *A-term* and the chosen example (*apples*) as the *B-term*.

Modelling Property Salience

The approach taken in this work for selecting a good example of an object (dually, concept) possessing a certain property relies mainly upon probabilistic knowledge of the distribution of values for certain essential, empirically observable attributes among concrete (physical) objects considered as objects of comparison. In addition, we employ a salience heuristic based on property intrinsicness.

Concept Representations

The knowledge base includes representations of concrete objects with probabilized value spaces representing the distribution of possible values for attributes. Such representations are assumed to be integral to the *general world knowledge* of the speaker about the concepts.

Formally, a *concept* C is a labelled set of *properties* P_i, and each property P_i is a pair $A_i : \mathcal{V}_i$, where A_i is an *attribute* and $\mathcal{V}_i$ is its *probabilized value space*. $\mathcal{V}_i$ is a set of pairs $V_{ij} : p_{ij}$, with $p_{ij} \in (0, 1]$, and $\sum_j p_{ij} = 1$. In our implementation, for example, we have *colour = [yellow:0.25, golden:0.60, green:0.15]* as a property of the *mango* concept. Noting that the probabilities of values in these representations are *conditional* upon the concept possessing them for the respective attributes, we can write, for example: $p(colour = green|mango) = 0.15$.

These representations are a variant of the dimensional approach to probabilistic concept representation, discussed in cognitive psychology literature by [Smith & Medin, 1981]. In our implementation, these representations are embedded in a semantic network. Nominalized versions of property values (like *redness* for *red*) are represented in the network as abstract concepts, with links directed towards their extensions (like *apples* and so on).

Salience Based on Probabilities

When generating comparisons, a concept (say, *apple*) is considered as a candidate example of an input property (say, *colour:red*) if its **most probable** value for the corresponding attribute (*colour*) matches the value (*red*) in the input property. Several objects (all of which are mostly *red*) may present themselves as candidate examples, and the preference is based on a measure of salience of the property. However, salience is not simply *equal to* the probability, since the number of other possible values (say, *green, yellow*) as well as their respective probabilities have the effect of either enhancing or suppressing the prominence of the most probable value in a distribution.

If the most probable value (*red*) has a high probability (*0.85*), and there are very few other possible values (say, only *green* with probability *0.15*), then *red* has high salience in the context of the *colour* distribution. If, on the other hand, even if red was the most probable colour (with probability *0.25*), if other colours were also equally probable (say *green, brown* and *yellow*, each with probability *0.25*), then *red* would not *stand out* in the distribution, and its salience would therefore be very low. Information theory helps us capture these notions of salience precisely, through the concept of **redundancy**, which we use in our work to quantify property salience.

Moreover, when looking for the best example of a red object, not only should the candidate example, say, apple, be mostly red, and highly salient among apples of all colours, but the redness of apples should be more salient than the redness of, say, *strawberries* and other *fruit* which are also mostly red. To model this aspect of salience, for a given property value (say *red*), we consider the redundancy of a concept (say *apples*) in the context of the redundancies of other candidates (say *strawberries*, etc) through the definition of **normalized redundancy**.

Similar information-theoretic measures have been used by [Iwayama, Tokunaga & Tanaka, 1990] in their computational modelling of metaphor comprehension. We adopt their work as a good point of departure to examine the modelling of salience and its role in generating comparisons for describing object properties, and use their example sentence *Mary's cheeks are like apples* to convey our algorithm in this paper. In interpreting the above sentence, their system (called *AMUSE*) calculates the salience of the properties of *apples*, matches the high-salient properties of *apples* with the properties of *cheeks* and infers the properties of *Mary's cheeks* intended by the speaker.

To compute the redundancy of the property of an object, we proceed through the following information-theoretic concepts. Given a discrete probability distribution with n probabilities $p_i \in (0,1]$, with $\sum_{i=1}^{n} p_i = 1$, the **entropy** of the distribution H is given by

$$H = \sum_{i=1}^{n} p_i log_c \frac{1}{p_i} \tag{1}$$

We take c (the base of the logarithm) to be 2 throughout this paper.

The entropy H of a distribution is zero if there is only one possible value (with unit probability). For a given set of n possible values, H is maximum if the values are equiprobable ($p_i = \frac{1}{n}$ for all i), and equals $log_2 n$. H quantifies the 'flatness' or 'dispersion' of a probability distribution, and can be interpreted as measuring the extent to which the prominence of the most probable value in the distribution is *suppressed* in the context of possible values. For instance, for the *colour* property of *mango* represented by *colour = [yellow:0.25, golden:0.60, green:0.15]*, H turns out to be 1.35272. If the colours were equiprobable, H would have been maximum, at 1.58496 (i.e., $log_2 3$). If all mangoes were golden, H would have been 0.

Relative entropy H_{rel} expresses the entropy of a distribution in the unit interval. By normalizing H with respect to maximum entropy for a given set of samples, H_{rel} expresses entropy independently of the sample size. It is defined as

$$H_{rel} = \left(if\ n = 1\ then\ 0\ else\ \frac{H}{log_2 n} \right) \tag{2}$$

For any $n > 1$, $H_{rel} = 1$ if the values are equiprobable, and less if the most probable value is higher in probability, and the number and magnitude of other values, smaller. If only one value is possible (probability = 1), $H_{rel} = 0$. H_{rel} quantifies the extent of suppression, or lack of salience, of the most probable value in a probability distribution. The quantitative complement of H_{rel}, viz., redundancy, therefore measures the degree of salience of the most probable value in a probability distribution:

Redundancy of a distribution is computed by

$$R = 1 - H_{rel} \tag{3}$$

In our system, the *colour* space of *apple* is *[red:0.75, green:0.15, yellow:0.10]*. The redundancy (R) of this distribution is 0.33499. By comparison, the colour of *orange*, with a distribution of *[orange:0.80, yellow:0.10, green:0.10]*, has a (greater) R of 0.41833. As a final example, the colour of *grapefruit* with *[yellow:0.75, pink:0.25]*, has $R = 0.18872$. Even though *red* in *apple* and *yellow* in *grapefruit* occur with the same probability, the former has higher salience (R) in its context of possible values.

Finally, in a set of concepts $C = \{C_1, C_2, \ldots C_n\}$ in which all C_i possess the same most probable value (say *red*) for a given property P (*colour*), the salience of a concept C_k in the context of C due to property P is measured by its **normalized redundancy**, computed as

$$NR(C_k, P) = \frac{R(C_k, P)}{\sum_{i=1}^{n} R(C_i, P)} \tag{4}$$

where $R(C_i, P)$ is the redundancy of the property P of C_i. In a context of fruits which are all mostly red, $NR(apple, colour)$ measures the relative salience of apples, and governs the candidacy of apples as good examples of red fruits. Note also that by the above definition, in a given context C, all the NRs add up to 1, and $NR(C_i, P)$ can be interpreted as the conditional probability $p(C_i|P)$. $p(apple|colour : red)$ is the likelihood of choosing *apple* when looking for a good example of *colour* : *red* among fruits.

Decision Making in Comparison Generation

In formulating the choice of the B-term in the comparison as a decision making problem, we first derive the formal equivalence between information-theoretic redundancy and expected utility.

Redundancy as Expected Utility

Substituting (1) in (2) and (2) in (3), the redundancy R associated with a probability distribution over n possible values ($n > 1$) is:

$$\begin{aligned} & 1 - \frac{1}{log_2 n} \cdot \sum_{i=1}^{n} p_i log_2 \frac{1}{p_i} \\ &= \frac{log_2 n + \sum_{i=1}^{n} p_i log_2 p_i}{log_2 n} \\ &= \frac{\sum_{i=1}^{n} p_i log_2 n + \sum_{i=1}^{n} p_i log_2 p_i}{log_2 n} \\ &= \sum_{i=1}^{n} p_i \cdot \left(\frac{log_2(np_i)}{log_2 n} \right) \end{aligned} \tag{5}$$

The summation (5) is the familiar form of expected utility, viz., $\sum_{i=1}^{n} p_i u_i$, with $u_i = log_2(np_i)/log_2 n$. The utility u_i is derived from the probability p_i and the size of the value space, n. It is interpreted as the reward associated with the selection of the value with probability p_i. For the special case of $n = 1$, we have $u_i = 1$ and $p_i = 1$. As is evident from the above expression, the reward u_i is maximum for the most probable value, and redundancy measures the expected reward in the context of all possible values. Choosing a concept as an example on the basis of maximum R among competing concepts, and equivalently, on the basis of maximum NR, can hence be modelled as a decision problem of maximizing expected utility.

Generating comparisons describing one property: The algorithm for choosing the best comparison when there is one input property P can now be stated as follows: P in the generator input serves as a source of activation in the knowledge base, which spreads towards concepts (forming a set C) in which P occurs redundantly. For each $C_i \in C$, compute the expected utility $EU(C_i, P) = EU_i = NR(C_i, P)$. Select the concept with maximum EU_i as the best (available) example. In our generator implementation, for the input entity *Mary's cheeks* and the descriptor *[shape:round]*, the following EUs are computed (note that they add up to 1):

C_i	$EU(C_i, [shape : round])$
plum	0.2555
apple	0.3434
grape	0.0908
peach	0.0908
lemon	0.0317
grapefruit	0.1878

The generator outputs *Mary's cheeks are like apples.*

Comparisons Describing Two or More Properties

When two or more properties are intended to be communicated about the input entity (for example, that *Mary's cheeks are red, smooth and round*), each property initiates a search (ideally, in parallel) in the knowledge base for a good example in which the respective property occurs redundantly. The decision criterion for choosing the best example is now one of maximizing *total expected utility* (TEU), the total being the sum of individual EUs from each property.

Table 1 shows a portion of the matrix of EUs computed in our implementation for the input entity *Mary's cheeks* and the descriptors *[colour:red]*, *[texture:smooth]* and *[shape:round]*. The generator outputs *Mary's cheeks are like apples* based on maximum TEU.

The Problem of Zero Credit

The above method is effective as a straightforward extension of the one-property case, and works well when concepts receive non-zero EU for each property to be communicated, as is the case for *apple*

C_i	$EU(C_i, red)$	$EU(C_i, smooth)$	$EU(C_i, round)$	$TEU(C_i)$
apple	0.3333	0.3075	0.3434	0.9842
grape	--	0.4133	0.0908	0.5041
strawberry	0.5284	--	--	0.5284
plum	0.1383	0.0819	0.2555	0.4757
...	...	...	...	...

Table 1: A portion of a matrix of *EU*s

and ***plum*** in Table 1. Note however in the same matrix, that while ***plum*** receives an *EU* for each property, ***strawberry***, though having no *EU* for ***round*** and ***smooth***, scores a higher *TEU* than ***plum*** due to sheer intensity of high *EU* for ***red***. In this case, ***strawberry*** isn't exactly a good example of something ***red*** *and smooth and round*!

Similarly, when the input contains *[colour:yellow, texture:smooth, shape:round]*, our generator, on the basis of *TEU* alone, would still say ***Mary's cheeks are like apples***. Miscommunication results in this case, as the hearer, who also uses knowledge of salience in comprehension, ends up inferring that Mary's cheeks are ***red***. If ***apple*** was not present in the knowledge base, and if the input property values were *red, round* and ***smooth***, the generator would have said: ***Mary's cheeks are like strawberries***. In this case, there may be either non-communication (nothing at all will be inferred about texture and shape of Mary's cheeks) or anomalous communication (the inferred shape and texture of Mary's cheeks will not be consistent with the general knowledge of the hearer about the shape and texture of cheeks.) In the parlance of connectionism, we may say then that it is not sufficient to have high convergent (***total***) activation: there should also be sufficient activation from ***each*** source.

While it is tempting to get a quick mathematical fix by defining arbitrary minimum thresholds on the individual *EU*s in the *TEU* criterion, we motivate the solution by considering the comprehensibility of the generated comparisons, and derive *cost* measures (antipodal to *utility*) to use in decision making. Two different cost measures are proposed, corresponding respectively to the problems of zero credit (described above) and unintended properties (described later).

Exception Clauses

It is fairly common in day-to-day speech to come across comparisons in which miscommunication due to zero credit (in the sense discussed above) is averted by generating additional clauses that make explicit the inexactitude of the match: as in ***Mary's cheeks are like apples, except...(that) they are yellow***. We call the latter clause an ***exception clause***, the property (attribute) described in it (here, ***colour***) the ***exception property (attribute)***, and the value communicated in it (here, ***yellow***) the ***exception value***. In simple settings like the ones under consideration, the utility of communicating through comparison drops off rapidly with the number of exceptions that may be sought. One exception is fairly common, as not always do we find ***one*** best example of all properties we want to communicate about an entity.

We quantify the cost of zero credit by focussing on the generation and acceptability of comparisons with exception clauses. We found it helpful to visualize that the speaker's evaluation of comparisons with exceptions is mediated by an imaginal process in which the B-term object without the exception property is ***mentally distorted*** into the B-term object with the exception property. For example, in ***Mary's cheeks are like apples, except.. they are yellow***, a ***red*** (salient colour) apple is 'repainted' into a ***yellow*** one, and offered as an object of comparison describing the intended properties of ***Mary's cheeks***. A cost is added to the *TEU* as a negative number, and the less the cost is the better. The cost will be less if it is in some way ***easier*** to 'distort' a red apple into a yellow apple. We propose the following as a probabilistic measure of cost for the running example:

$$Cost(apple, colour) = \\ p(colour = red|apple) \cdot (1 - p(colour = yellow|apple))$$

The more probable red apples are, and the less probable yellow apples are (in the general knowledge of the speaker), the more difficult will it be to 'mentally distort' a red apple into a yellow one, and the costlier will be the exception clause. In general, given a concept C and an exception attribute P, if the exception value is EV and the most probable value is MPV, then

$$Cost(C, P) = \\ p(P = MPV|C) \cdot (1 - p(P = EV|C))$$

Property Intrinsicness

There seems to be more to the cost of zero credit than probabilistic knowledge as modelled here. For instance, given the input property values of *yellow, round* and *smooth*, compare the acceptability of *Mary's cheeks are like apples.. except, they are yellow*, with what our generator said charmingly oddly in an earlier version: *Mary's cheeks are like bananas, except.. they are round*! Distorting the colour of objects seems easier than distorting the shape of objects as shape is in some sense a more salient attribute than colour. This conception of salience is discussed in cognitive linguistics [Langacker, 1987] under *intrinsicness*. We annotate the properties in the knowledge base with this heuristic measure, giving a higher score to shape than to colour. This is added to the cost of distortion when the distortion entails conception of an impossible object. This cost is zero for *yellow apples* since they do exist in the speaker's knowledge; it is less for *purple apples* than for *round bananas*. Even among impossible objects, some seem more impossible than others!

For every zero credit entry in the TEU matrix we compute such costs. When there is positive EU for an entry, cost is zero, since the property to be communicated is salient in the entry. Finally, when n properties are intended to be communicated, and the TEU matrix is formed in the manner discussed earlier, the probability p_{zc} that any one of the entries is zero is given by $\frac{2^{n-1}-1}{2^n-1}$, analogous to the probability that any one of the inputs of an n-input OR-gate is 0, given that the gate output is 1. The *expected cost* (EC) of an entry in the TEU matrix for the property P of concept C is therefore

$$p_{zc} \cdot Cost(C,P) + (1 - p_{zc}) \cdot 0$$
$$= \left(\frac{2^{n-1}-1}{2^n-1}\right) \cdot Cost(C,P)$$

We can now compute the *total expected cost* (TEC) of a candidate example as the sum of individual ECs for each property. The decision criterion is now one of maximizing

$$TEU - \alpha \cdot TEC$$

where α is a non-negative number. $0 \leq \alpha < 1$ corresponds to *speaker-oriented* generation models, and $\alpha \geq 1$ corresponds to *considerate* or *listener-oriented* generation models. For $\alpha = 1$ and for the values *yellow, smooth and round*, we generated *Mary's cheeks are like grapefruit*, which, though unlikely to be uttered by humans, was the best that our system could do with its modest knowledge base, and without the use of the 'unintended properties' cost, to be described next.

Cost of Unintended Properties

Another problem arises when the list of descriptors in the generator input does not include certain properties in the *generic* representation of the input entity in the knowledge base. For instance, for the input values *yellow* and *smooth*, the generator would say *Mary's cheeks are like bananas*. This is because the input which encodes the speaker's communicative intentions does not include *round* as a descriptor of *Mary's cheeks*. Since both *yellow* and *smooth* receive positive EUs for banana and make its TEU score the highest, the above comparison is generated. The problem arises because the inferred (salient) shape of bananas (B-term) is in conflict with the salient value of the unintended (= not included in the generator input) property in the *general knowledge* about cheeks, viz., *round*. To counter the problem of miscommunication or anomaly due to such interference, we use another cost measure very similar to the one discussed earlier, using both probabilistic knowledge and property intrinsicness. For every property in the general knowledge of the A-term not included in the input, the cost of mismatch with the corresponding salient property of the B-term is computed, and summed up as the total expected cost TEC_{B-A}. The former TEC (for the zero credit problem) can be now relabelled as TEC_{A-B}.

Decision Criterion for Comparison Generation

The final decision criterion we have developed is termed *net expected utility*, computed as

$$NEU = TEU - \alpha_1 \cdot TEC_{A-B} - \alpha_2 \cdot TEC_{B-A}$$

where α_1 and α_2 are non-negative. Although our generation algorithm does not explicitly advert to the concept of *similarity* in its design, its decision criterion NEU has striking formal resemblance to the similarity metrics proposed in the cognitive psychology literature, such as those of [Tversky, 1977] and [Ortony et al., 1985]. NEU is more like Ortony et al's measure, however, since TEU is evaluated with respect to the B-term. With NEU, for the input properties *yellow* and *smooth*, we generate *Mary's cheeks are like lemons* for $\alpha_1 = \alpha_2 = 1$.

General Discussion

The research reported in this paper is integral to our investigation of salience and its role in natural language generation decisions [Pattabhiraman, 1992]. The measure of salience used in this paper is intuitively appealing, while being at the same time mathematically well-grounded. Our method of generating comparisons, while aiming to satisfy cognitive concerns, also relies upon decision theory and information theory for its formal foundations, and thereby lends itself to computational usability in NLG systems, as our implementation has demonstrated. A fresh perspective on the problem of similes emerges when we view it from the speaker's angle. However, several additional factors will have to be treated before presenting our work as a full-fledged descriptive model of human comparison generation, or as a model for computer generators with very large knowledge bases.

First, other determinants of salience like *vividness* and *imageability* ([Osgood & Bock, 1977]) should be explored for their influence on comparison generation. Their influence can be modelled quantitatively in terms of utility functions and incorporated into our decision criterion. A red object may be selected not necessarily because it is redundant in the information-theoretic sense, but because its redness is very vivid. Secondly, preferences for basic-level terms and for concrete objects, which are implicit in our method, should be treated explicitly in the theory. Preference for concrete objects can help generate *indirectly grounded* comparisons like *proud as a peacock*: though peacocks are not quite proud (cf. general discussion in [Ortony et al., 1985]), we suggest that a link exists through *ostentation*: ostentation is a possible symptom of pride, and peacocks when they unfurl their feathers appear ostentatious. The collective noun *an ostentation of peacocks* is another telltale indicator! Such indirect associations may get short-circuited as direct conceptual- or lexical-collocational associations through the process of entrenchment. Thirdly, the initial stage of processing the generator input to form the set of potential objects of comparison can be refined to control selection between literal and metaphorical comparisons.

Finally, while we have assumed tacitly that the property values are independent, covariances among property values and among relations between property values are common in natural concepts [Goldstone, Gentner & Medin, 1989]. The accuracy of decision making increases when knowledge of correlations between properties is represented and used. However, even in small knowledge bases, an immense number of covariances will have to be identified and represented; a decrease in computational efficiency is inevitable with their use. The precise nature of this trade-off and its implications for descriptive models of comparison generation merits further research.

Acknowledgments. Our thanks are due to the anonymous reviewers for their helpful comments and suggestions, and to the CSS and the School of Computing Science at SFU for the use of their facilities.

References

[Goldstone, Gentner & Medin, 1989] Goldstone, R., Gentner, D., and Medin, D. (1989). Relations Relating Relations. In *Proceedings of the 11th Annual Conference of the Cognitive Science Society*, pages 131–138.

[Iwayama, Tokunaga & Tanaka, 1990] Iwayama, M., Tokunaga, T., and Tanaka, H. (1990). A Method for Calculating the Measure of Salience in Understanding Metaphors. In *Proceedings of AAAI'90, Boston*, pages 298–303.

[Langacker, 1987] Langacker, R. (1987). *Foundations of Cognitive Grammar, Volume 1: Theoretical Prerequisites.* Stanford University Press.

[Ortony et al., 1985] Ortony, A., Vondruska, R., Foss, M., and Jones, L. (1985). Salience, Similes and the Asymmetry of Similarity. *Journal of Memory and Language*, 24:569–594.

[Osgood & Bock, 1977] Osgood, C. and Bock, J. (1977). *Salience and Sentencing: Some Production Principles*, In *Sentence Production: Developments in Research and Theory*, Rosenberg, S., Editor, pages 89–140. Erlbaum, Hilldale, N.J.

[Pattabhiraman, 1992] Pattabhiraman, T. (1992). *Aspects of Salience in Natural Language Generation.* PhD Thesis, School of Computing Science, Simon Fraser University. In preparation.

[Smith & Medin, 1981] Smith, E. and Medin, D. (1981). *Categories and Concepts.* Harvard University Press.

[Tversky, 1977] Tversky, A. (1977). Features of similarity. *Psychological Review*, 84(4):327–352.