

Decision-Theoretic Salience Interactions in Language Generation

T. Patabhraman & Nick Cercone

School of Computing Science & Centre for Systems Science

Simon Fraser University

Burnaby, B.C. Canada V5A 1S6

Abstract

The salience of an entity is interpreted in natural language generation (NLG) as a measure of how well it stands out from other entities and biases the preference of the generator in selecting words and phrases. We delineate two categories, viz., *canonical salience* and *instantial salience*, characterizing the former as a built-in preference in the conceptual- and linguistic knowledge of the speaker, and the latter as the prominence of specific objects in the context of NLG. This paper proposes that the interaction between these categories be formalized in terms of decision-theoretic combinations of probabilities and utilities, and construes *phrase selection* and *basic level/entry level effects* accordingly.

1 Introduction

The salience of an entity, in intuitive terms, refers to its degree of prominence, importance or activation. In natural language generation (NLG), it is interpreted as a measure of how well the entity stands out from other entities. When the generator is faced with the choice of a word or a phrase, a ranking of the available choices is obtained on the basis of the degree of salience of entities in the generation context. The word or phrase is then selected for utterance on the basis of maximum salience. The role of salience in NLG thus involves a causal relation between *prominence* of entities in a ranking, and *preference* of a choice among alternatives.

The first computational study of salience in an NLG system is that of [Conklin and McDonald, 1982], who used it as a heuristic for selecting topics in a scene description task and for ordering the topics in generating paragraphs. In disciplines related to NLG, such as linguistics and psycholinguistics, salience has been invoked in several accounts of language phenomena. For instance, the constituents of sentences are said to be ordered according to their relative salience, and the facility to order the constituents of a sentence is said to be used by speakers to communicate the relative salience of elements in the context of language production [Osgood and Bock, 1977], [Osgood, 1980]. Guided by the premise that the choice of words and constructions on the basis

of salience is pervasive in NLG, our research [Patabhraman, 1992] aims to flesh out the notion of salience in a theoretically-significant form.

A number of factors have been claimed to contribute to the salience of entities. Vividness, pervasiveness, uniqueness, recency of mention and frequency of mention are some of the *determinants of salience*, several of which are at play even in the simplest of NLG settings. A major challenge to the computational study of salience in NLG is in unravelling how the determinants of salience interact in their effects on language generation decisions. As the first step towards meeting this challenge, we delineate an important pair of categories in the salience theory: *canonical salience* and *instantial salience*. The former is characterized as a built-in preference in the general conceptual- and linguistic knowledge of the speaker. An example of such a natural tendency is the preference to express *agent* before *object* in simple action sentences. The latter refers to the salience of specific objects in the context of NLG, and may accrue through such determinants as vividness and recency of mention. These categories are presented briefly in section 2 below, and in more detail, in [Patabhraman, 1992].

We consider the interaction between canonical salience and instantial salience as a crucial subject of study in NLG. We then formalize this interaction in terms of decision-theoretic combinations of probabilities and utilities. The choice of the formal framework is inspired primarily by the conclusions of psycholinguistic experiments most notably, those of [Osgood and Bock, 1977]. In this paper we construe two distinct language generation tasks in terms of decision-theoretic salience interactions: (1) phrase selection (section 3), (2) basic level/entry level preference in noun generation (section 4). A third task, viz., comparison (simile) generation, has been modelled in this framework and presented in [Patabhraman, 1992] and [Patabhraman and Cercone, 1992].

2 Canonical Salience and Instantial Salience

We postulate that there exist built-in preferences or internal biases in the linguistic choice systems of human adult speakers. These biases are regarded as pre-defined relative salience configurations. By this postulate, even

if (for example) sentence *John gave a book to Mary* has the logical, 'transformed' equivalent *John gave Mary a book*, in the presence of a built-in preference in the choice set

$$\{X \text{ gave } Y \text{ to } Z, X \text{ gave } Z \text{ } Y\}$$

the choices will not be equally presented when the generation context involves talking about a(ny) specific *X* giving a(ny) specific *Y* to a(ny) specific *Z*. That is, every time the aforementioned choice set is encountered in generation, the built-in preference will be automatically activated as a determinant of salience, as if by force of habit. If there are other determinants in the instance of generation (like vividness, etc) which define a relative salience configuration among *X*, *Y* and *Z*, their effects will have to be 'gated through' the built-in preference inherent in the choice system.

2.1 Canonical Salience

The inherent, built-in bias by which the generator tends to exercise preferences is termed *canonical salience* in our theory. The *naturalness* hypothesis of [Osgood, 1980] most directly addresses the origin of canonical salience from a psycholinguistic perspective. A major assumption of Osgood's theory is that the cognitive structures underlying the production and comprehension of sentences develop in prelinguistic, perceptual experience; these structures are 'taken over' by the linguistic system as it develops later. The basic structure of simple cognitions is of the form *relation(arg1,arg2)*. Osgood discusses two very basic cognitions which are crucial for the prelinguistic child: stative cognitions of the form *state(figure,ground)*, and action cognitions of the form *action(actor,recipient)*. Osgood then argues that in a prelinguistic child, for stative cognitions, *figure-state-ground* is the natural order, and for action cognitions, *actor-action-recipient* is the natural order.

2.2 Instantial Salience

Salience configurations set up in specific situations or contexts of generation by such determinants as vividness, recency of mention, speaker motivation and so on come under *instantial salience*. For example, in generating *A UFO was spotted by the boy*, the entity *UFO* (the object) is more salient than the entity *boy* (the agent) due to vividness, and consequently gets expressed earlier (as a passive sentence). Instantial salience depends on the presence of other entities in the context: *UFO* is vivid relative to *boy*. It could also be influenced by the presence of speaker-motivation or goals, and on dynamic factors which, as the text is unfolding, may make entities salient for pronominalization and other decisions. The passivization in *My stereo system has been stolen* and in the second clause of *He started arguing loudly, and had to be told to calm down* are licensed by the salience determinants in the respective instances, viz., speaker motivation and recency of mention, respectively.

2.3 Canonical-Instantial Salience Interactions

The choice of a word or phrase is exercised on the basis of *effective salience*, which is the net result of the interaction of canonical- and instantial salience. A distinguish-

ing feature of canonical-instantial salience interactions is that they sometimes agree in their dictates (*synergistic* interaction) and sometimes are opposed in their dictates (*competitive* interaction). For example, in generating the sentence *The Mayor was struck by a car yesterday*, the natural tendency of canonical salience dictates that the instrument 'car' be expressed prior to the patient 'the Mayor'. On the other hand, the prominence (instantial salience) of *the Mayor* presses for it to be positioned earlier in the sentence. Instantial salience 'wins out', and the passive voice sentence is generated.

Psycholinguistic experimental results of [Osgood and Bock, 1977] suggest that the interaction between canonical salience and instantial salience is multiplicative. We capture this interaction in a decision-theoretic formalization which models canonical salience as probabilities, instantial salience as utilities, and the selection criterion based on effective salience as maximization of expected utility (or its variants). *The details of the available evidence and our arguments for the choice of decision theory as the formal framework are presented in [Pattabhiraman, 1992].*

3 Phrase Choice

In this section, we first show how effective salience in the selection between active voice and passive voice is obtained as expected utility. Next, we illustrate the relative rarity of the passive voice by applying the decision criterion to a small set of example sentences. The criterion is then stated in general terms for all binary linguistic choice systems in syntactic linearization. Psycholinguistically-significant characteristics are then presented.

3.1 Effective Salience as Expected Utility

Let us consider the verbalization of a simple *agent-action-object* cognition.

(1) The environment in which the selection decision is carried out is governed by the probabilities that the agent *Ag* and the object *Ob* respectively will be expressed earlier in the utterance. These probabilities correspond to the canonical saliences of *Ag* and *Ob* respectively, and are denoted by $p(Ag)$ and $p(Ob)$ respectively. $p(Ag) + p(Ob) = 1$.

(2) The instantial saliences of *Ag* and *Ob* are given as $u(Ag)$ and $u(Ob)$ respectively. In a specific situation, these are evaluated as functions of such determinants as vividness, recency of mention, etc.

(3) Whatever be the instantial saliences, the agent *Ag* may or may not be expressed sentence-initially, because of the influence of the canonical probabilities. We will have four possible outcomes in this decision, two corresponding to each relative instantial salience configuration. Utilities are associated with those outcomes, expressing their preferability, on the basis of instantial saliences. We expand this point more formally in (4)-(8) below.

(4) We will represent the possible relative instantial salience configurations as a set of states: $S = \{A, O\}$ where $A = u(Ag) \geq u(Ob)$ and $O = u(Ob) > u(Ag)$.

(5) The linguistic choice system is the set of available actions (*decisions*), denoted as $A = \{Ac, Pa\}$ where Ac = "express Ag before Ob ", and Pa = "express Ob before Ag ".

(6) The four possible outcomes are labelled in terms of the states and actions that lead to them, as (*state, action*) pairs: (Ac, A), (Ac, O), (Pa, A), (Pa, O).

(7) Preferences for these outcomes are expressed as utilities. Utility of action $a \in A$ in state $s \in S$ is represented as $U(a, s)$.

(8) We assign values for these utilities on the basis of knowledge of instantial saliences, and the principle of tending to express the more salient element earlier: $U(Ac, A) = u(Ag)$, $U(Pa, A) = 0$, $U(Ac, O) = 0$, and $U(Pa, O) = u(Ob)$.

(9) Given the above information on probabilities and utilities, we can now compute the expected utility of each action as follows:

Expected utility of expressing Ag sentence-initially:

$$\begin{aligned} EU(Ac) &= \\ p(Ag) \cdot U(Ac, A) + p(Ob) \cdot U(Ac, O) &= \\ p(Ag) \cdot u(Ag) + p(Ob) \cdot 0 &= \\ p(Ag) \cdot u(Ag) \end{aligned} \quad (1)$$

Expected utility of expressing Ob sentence-initially:

$$\begin{aligned} EU(Pa) &= \\ p(Ag) \cdot U(Pa, A) + p(Ob) \cdot U(Pa, O) &= \\ p(Ag) \cdot 0 + p(Ob) \cdot u(Ob) &= \\ p(Ob) \cdot u(Ob) \end{aligned} \quad (2)$$

(10) Decision theory recommends choosing the action with maximum expected utility. Saliency principles in NLG recommend selecting the alternative with maximum effective saliency. The above derivation shows how the effective saliency of an entity corresponds to the expected utility of a linguistic action, thus giving saliency a causal role in a linguistic effect.

3.1.1 An Illustration

Consider the set of sentences

1. *Carl Sagan discovered a new comet* last month.
2. *A baffling new galaxy* has been discovered at the centre of the Milky Way.
3. *I bought a pen*.
4. *The astronomers at Palomar* discovered a new asteroid.
5. *Carl Sagan bought a shirt*.

For these sentences, table 1 summarizes the decisions, and the decision-theoretic quantities on which the decisions are based. $p(Ag)$ and $p(Ob)$ are 0.87 and 0.13 respectively (approximates, from the empirical work of [Svartvik, 1966]). The instantial saliences have been subjectively assigned based on the vividness of entities. 3 of the 5 sentences above have a high-salient object. 4 out of 5 have an object with instantial saliency no less than that of the agent. Yet, only 1 out of the 5 cases produces a passive.

S#	$u(Ag)$	$u(Ob)$	$p(Ag)u(Ag)$	$p(Ob)u(Ob)$	Dec
1	100	100	87	13	act
2	10	100	8.7	13	pass
3	10	10	8.7	1.3	act
4	10	50	8.7	6.5	act
5	100	10	87	1.3	act

Table 1: Utilities and Expected Utilities in the Illustration in 3.1.1. (S#=Sentence#; Dec=decision; act=active; pass= passive)

3.2 General Form for Binary Choice Contexts

The derivation in section 3.1 can be expressed easily in a general form, applicable to all binary linguistic choice contexts in linearization. Let x_1 and x_2 be entities in the context of NLG whose saliences are being compared. Denote the canonical saliences of x_1 and x_2 as $p(x_1)$ and $p(x_2)$ respectively ($p(x_1) + p(x_2) = 1$) and the instantial saliences as $u(x_1)$ and $u(x_2)$ respectively. Let $\{X_1, X_2\}$ be a linguistic choice context, with the alternatives X_1 and X_2 corresponding to the earlier expression of x_1 and x_2 respectively.

Then, analogous to the derivation in section 3.1,

$$EU(X_1) = p(x_1) \cdot u(x_1) \quad (3)$$

and

$$EU(X_2) = p(x_2) \cdot u(x_2) \quad (4)$$

The construct selected is the one with greater EU .

3.2.1 Psycholinguistically-Relevant Characteristics

Several important psycholinguistically-relevant aspects are captured straightforwardly in this model:

1. Canonically unbiased linguistic choice context: *If there is no skew in the canonical probabilities, then the decision for earlier expression is governed by instantial saliency.* That is,

$$\begin{aligned} \text{If } p(x_1) = p(x_2) \text{ then} \\ \max\{EU(X_1), EU(X_2)\} = \max\{u(x_1), u(x_2)\} \end{aligned} \quad (5)$$

Choice here is apparently situation-controlled.

2. Equi-salient entities in the generation context: *If there is no skew in the instantial saliences, then the decision for earlier expression is governed by canonical saliency.* That is,

$$\begin{aligned} \text{If } u(x_1) = u(x_2) \text{ then} \\ \max\{EU(X_1), EU(X_2)\} = \max\{p(x_1), p(x_2)\} \end{aligned} \quad (6)$$

Choice here is apparently grammar-controlled.

3. Extremes of lack of availability of choice: *If X_1 is the only available linguistic choice then x_1 will always be expressed earlier.* That is,

$$\text{If } p(x_1) = 1 \text{ then } EU(X_1) > EU(X_2) \quad (7)$$

Similarly,

$$\text{If } p(x_2) = 1 \text{ then } EU(X_2) > EU(X_1) \quad (8)$$

4 Basic Level/Entry Level Preference in Noun Generation

Central to canonical salience is Osgood's conception of *naturalness* as arising in prelinguistic, perceptual experience, and as being cultivated in childhood and persisting in adulthood. We proposed viewing the interactions between canonical salience and instantial salience (in linearization) as decision-theoretic combinations of probabilities and utilities. We next asked ourselves whether we could find other language generation tasks (than syntactic linearization) in which a similar, built-in expressive preference played a crucial role, and whether, in such a task, a possibility exists for overriding the natural preference, and under what conditions such a possibility is admissible. If we can find such a phenomenon, we thought, we can learn more about the nature of canonical salience. Moreover, we can consider applying the decision-theoretic model of salience interactions, suitably instantiated, to model decision criteria in that phenomenon.

Interestingly, we found several dimensions of similarity between the role of *basic levels* of taxonomic concept knowledge in the *free naming* of objects, and the nature of canonical salience discussed in section 3. To our knowledge, this similarity has not been observed or explained in the literature. In day-to-day speech, one refers to objects by labels which are neither too general nor too specific. For example, we generally say *look at that dog!* rather than *look at that collie!* or *look at that mammal!* This effect is described in cognitive psychology by [Rosch and Mervis, 1975] in terms of **basic levels** in human taxonomic knowledge. Later work by [Jolicoeur *et al.*, 1984] recognizes exceptions to the norm of basic level preference, and introduces the notion of **entry level effect**, by which some specific instances of certain categories are named at a level other than the basic level. For example, when pointing a 'generic' bird to a friend, you are more likely to say *look at that bird*; however, when you see an ostrich, you are more likely to say *look at that ostrich*. Basic level preference is one of several constraints on noun generation in [Reiter, 1991]'s system. While Reiter handcodes basic level annotations in his knowledge base, our work deals with the non-linguistic basis for determining basic levels in a concept hierarchy.

The stark similarities between the role of basic levels in object naming and that of canonical salience in linearization have been pointed out in detail in [Pattabhiraman, 1992]. In the same thesis, we also argue, using the conclusions of Jolicoeur *et al.*'s experiments, that entry level effects are an instance of decisions based on *effective salience*. In this section, we present some major aspects of our results in modelling free naming of objects in terms of salience interactions. Using probabilistic concept representations in a taxonomic hierarchy, we formalize entry level preference as maximization of expected utility. Three major components constitute this model (A fully detailed illustration is given in [Pattabhiraman, 1992]): (1) Modelling canonical salience (basic level preference) as a probability distribution. By this model, basic level is the level in the classification hier-

archy of an object at which the object is most likely to be described in a free naming task. The derivation of this distribution depends on *differential category utility*, a measure modified and adapted from *category utility* proposed by [Gluck and Corter, 1985]. (2) Utilities to express the instantial salience of the objects being named. (3) Effective salience as expected utility, determining the choice of names.

4.1 Probabilistic Concepts in Taxonomic Knowledge

Formally, a *concept* C is a labelled set of *properties* P_i , and each property P_i is a pair $A_i : V_i$, where A_i is an *attribute* and V_i is its *probabilized value space*. V_i is a set of pairs $V_{ij} : p_{ij}$, with $p_{ij} \in (0, 1]$, and $\sum_j p_{ij} = 1$. For example, in one of our implementations, the concept *mango* has, among its properties, *colour* = [*yellow:0.25, golden:0.60, green:0.15*]. These representations are a variant of the dimensional approach to probabilistic concept representation, discussed in cognitive psychology literature by [Smith and Medin, 1981]. We can denote the probabilities more perspicuously so as to reflect the associated property and concept, as follows: $p(A_i = V_{ij})$ is the probability that the attribute A_i (of an object) has value V_{ij} . $p(A_i = V_{ij} | C_k)$ is the probability that an object in class C_k has value V_{ij} for attribute A_i .

In the taxonomic knowledge base, probabilistic summaries of value distributions for attributes are stored at several levels. Each category in the taxonomic hierarchy is thus a probabilistic concept. It is to be noted that the probabilities at successive levels of a path in the hierarchy are related to one another: for example, the probability of the red colour of apples summarizes contributions from the probabilities of the red colour of all subtypes of apples.

The root node of the hierarchy is denoted as C_{entity} ; the leaf nodes are the most specific object classes. $C_{sup(k)}$ is the immediate superordinate of C_k . Apart from *property probabilities* $p(A_i = V_{ij} | C_k)$ stored at probabilistic concept nodes, *concept probabilities* are also relevant to our computations. The concept probabilities we consider are: $p(C_{entity}) = 1$ for the root node of the hierarchy, and $p(C_k | C_{sup(k)})$ for all other concepts C_k . The concept taxonomy hierarchy encodes $p(C_k | C_{sup(k)})$ as a value stored at the node C_k .

4.2 Canonical Salience from Differential Category Utility

The scenario we consider is a guessing game in which a participant guesses the values for various attributes of a specific object. The participant is assumed to possess probabilistic concept knowledge, and to adopt what [Gluck and Corter, 1985] call a *probability matching strategy*: for an attribute A_i , (s)he guesses a value V_{ij} with a probability equal to the value of $p(A_i = V_{ij})$ in her own knowledge. The guess is likely to be correct with the same probability.

Assuming that the category information (C_k) is given prior to guessing (as in: "*I have an apple with me; can you guess its colour?*"), for an attribute A_i for which the value is guessed, the score awarded to the participant is

the expected value of $p(A_i = V_{ij} | C_k)$ taken over all possible values of A_i , given by $\sum_{j=1}^{n_{ik}} [p(A_i = V_{ij} | C_k)]^2$ where n_{ik} is the number of possible values of attribute A_i in concept C_k . If guessing proceeds for every one of the l attributes of the object, the score is $\sum_{i=1}^l \sum_{j=1}^{n_{ik}} [p(A_i = V_{ij} | C_k)]^2$. This is the expected number of values of attributes that can be correctly guessed for an object given that it belongs to the class C_k .

At higher (more abstract) levels of taxonomic knowledge, probability distributions for attribute values tend to be more dispersed than at lower (more specific) levels. (By dispersion we mean the number of possible values, as well as the flatness of the distribution envelope). That is, when I am told, "I have an entity on mind; can you guess its colour?", all possible values on the colour dimension are candidate guesses. When more specific category information is given, as in, "I have a type of apple on mind; can you guess its colour?", the possible values are more restricted, and the probability distribution for attribute values generally tends to show more skew or prominence.

This tendency will be reflected accordingly in the values of $\sum_j [p(A_i = V_{ij})]^2$: for an attribute dimension A_i having n_i possible values, $\sum_{j=1}^{n_i} [p(A_i = V_{ij})]^2 \in [\frac{1}{n_i}, 1]$. It is minimum ($= \frac{1}{n_i}$) when the attribute values are equiprobable, and maximum ($= 1$) when there is only one possible value for A_i . Therefore, if more specific categorial information is given, the score in the guessing game is higher. That is, for any C_k , $\sum_j [p(A_i = V_{ij}) | C_{sup(k)}]^2$ will tend to be less than $\sum_j [p(A_i = V_{ij} | C_k)]^2$.

[Mervis and Rosch, 1981] note that basic levels are where a number of inference-related abilities are maximized in humans. [Rosch and Mervis, 1975] state that the categories subordinate to the basic level are also bundles of predictable attributes and functions, but contain *little more information than* the basic level to which they are subordinate.

Consider a traversal of the hierarchy starting at the root (C_{entity}): as we go down, not only does the score in the guessing game increase, but the rate at which it increases, increases. Soon, there comes a point where *the rate of increase decreases*, though the score itself continues to increase. We consider the expected *increase* in score, made possible by knowledge of a category C_k over the score with prior knowledge of its superordinate $C_{sup(k)}$. That is, compare *I have an apple with me; can you guess its colour?* versus *I have a fruit with me; can you guess its colour?* Basic level is the level at which the increase of score over the score at the immediate superordinate is maximum.

4.2.1 Differential Category Utility

The differential category utility (DCU) of a concept C_k , written as $DCU(C_k)$, is:

$$p(C_k | C_{sup(k)}) \cdot \left\{ \sum_{i=1}^l \sum_{j=1}^{n_{ik}} [p(A_i = V_{ij} | C_k)]^2 - \right.$$

$$\left. \sum_{i=1}^l \sum_{j=1}^{n_{ik}} [p(A_i = V_{ij} | C_{sup(k)})]^2 \right\} \quad (9)$$

That is, DCU is defined as the expected increase in score made possible by knowledge of C_k over the score with knowledge only of $C_{sup(k)}$. As a boundary case, $DCU(C_{entity}) = 0$. DCU differs from **category utility** [Gluck and Corter, 1985], defined as the expected increase in score made possible with knowledge of C_k over the score with *no prior category information*—that is, over the score with knowledge of C_{entity} . We argue in [Pattabhiraman, 1992] that category utility will always predict the most specific level, which need not necessarily be the basic level.

4.2.2 Probabilistic Definition of Canonical Saliency

Given a set of concepts $\{C_1, C_2, \dots, C_n\}$ constituting a path in taxonomic concept knowledge such that $C_k = C_{sup(k+1)}$ for $k = 1 \dots (n-1)$, define:

$$\mathcal{P}_B(C_k) = \frac{DCU(C_k)}{\sum_{i=1}^n DCU(C_k)} \quad (10)$$

so that $\mathcal{P}_B(C_k) \in [0, 1]$ and $\sum_{i=1}^n \mathcal{P}_B(C_k) = 1$. The canonical saliency of concept C_k is $\mathcal{P}_B(C_k)$. The basic level is the 'most probable' concept in the $\mathcal{P}_B$ distribution.

4.3 Instantial Saliencies and Utilities

According to the experimental results of [Jolicoeur *et al.*, 1984], in the presence of atypicality, preference tends to shift down from the basic level to the subordinate level. That is, when we see an object which is in some way (instantially) salient, we tend to use a more specific name. Atypicality of concepts is assessed in the context of their immediate superordinate. *More than one subconcept of a concept can be atypical*: Consider a set of concepts $\{C_1, C_2, \dots, C_n\}$ grouped under C . One or more of the C_k 's could be salient in C . For example, in

bird = {robin, raven, sparrow, peacock, ostrich, ...}

both *ostrich* and *peacock* are salient, the former because of its size, flightlessness and shape, and the latter, for its vivid feathers.

We propose that the saliencies of the concepts C_k be evaluated by a function $S : C \rightarrow [0, 1]$, with $S(C_k)$ being determined as a single number representing a combination of scores from determinants like atypicality and vividness. In the *bird* example, we would expect $S(\text{peacock})$ and $S(\text{ostrich})$ to be high (> 0.5). The choice between C_k and C in a free-naming task is governed in part by a function u which apportions the number 1 between C_k and C as follows:

$$\begin{aligned} \text{if } S(C_k) > 0.5 \text{ then } u(C_k) &= S(C_k) \\ \text{else } u(C_k) &= 0.5 \end{aligned} \quad (11)$$

and

$$u(C) = 1 - u(C_k) \quad (12)$$

Moreover, since the context in which the saliency of C_k is evaluated is provided by its superordinate C , $u(C') =$

0 for all concepts C' in the path, other than C_k and C . For example, the salience of *crabapple* is evaluated with respect to *apple*. We take $u(C) = 0$ for all other higher concepts C (*fruit*, *plant_product* and *entity*) in the *crabapple* path. When a specific concept is salient, the u score is 'taken away' from the superordinate, to the extent that the specific concept is salient. Thus, for the choice context $\{bird, robin\}$, owing to the low salience of *robin*, we will have $u(robin) = 0.5 = u(bird)$. For the choice context $\{bird, ostrich\}$, owing to the high salience of *ostrich*, we will have (say), $u(ostrich) = 0.8$ and $u(bird) = 0.2$.

4.3.1 Probabilistic Measure of Instantial Salience

We propose that the function S evaluate the instantial saliences of the concepts C_k as a number in $[0, 1]$. In general, S will be evaluated as a composition of several subfunctions, one corresponding to each determinant of salience. We are interested in determining the extent to which an object classified under C_k possesses an unusual property for its category C . The unusualness we consider is one by which we judge, for example, a crabapple, to be too sour or too small for an apple.

The index of this unusualness is computable from the probabilistic concept representations in taxonomic knowledge, as follows: given a concept C_k , a superconcept C and an attribute A_i , if the most probable value $A_i = V_{ij}$ is atypical in C_k , the probabilistic degree of salience of C_k because of V_{ij} in the context of concepts constituting C is denoted by $S_p(C_k, A_i)$, and is computed as:

$$p(A_i = V_{ij}|C_k) - [p(A_i = V_{ij}|C) - p(A_i = V_{ij}|C_k) \cdot p(C_k|C)] \quad (13)$$

The subtraction of the square-bracketed term prevents double-counting of the contribution of C_k to $p(A_i = V_{ij}|C)$.

4.3.2 Using a Threshold to Recognize High Salience

The attributes *taste*, *size* and *colour* are highly salient in crabapples in the context of apples. There could, however be attributes which are not salient in the probabilistic sense. For instance, the shape of crabapples does not particularly stand out when the shape of apples is considered. In fact, the distributions of *shape* in crabapple and apple could be identical. Such low scores do not contribute to the instantial salience of objects, and are ignored if they are below a threshold θ_s . We take θ_s to be 0.5 in our computations.

4.3.3 Prioritizing Property Saliences with Intrinsicness Weights

We then compute the net (instantial) salience of concepts C_k as a weighted sum of S_p 's:

$$S(C_k) = \sum_i w(A_i) \cdot S_p(C_k, A_i) \quad (14)$$

where $\sum_i w(A_i) = 1$. The weights $w(A_i)$ serve the purpose of prioritizing the contribution of the unusualness

of attribute A_i to the overall salience of C_k . For example, in judging crabapples as salient subspecies of apples, it may matter more that crabapples are very sour than that they are yellow. This could be because taste is in some sense more essential to the identity of an object, and 'less alienable' from the object than colour. This conception of attribute salience tallies with the notion of *intrinsicness*, discussed in cognitive linguistics by [Langacker, 1987].

4.3.4 Evaluating the Functions S and u

The degree of atypicality of C_k with respect to $C_{sup(k)}$, is computed as

$$S(C_k) = \sum_i w(A_i) \cdot S_p(C_k, A_i) \quad (15)$$

$u(C_k)$ and $u(C_{sup(k)})$, which represent the instantial saliences of C_k and $C_{sup(k)}$ respectively, are now calculated from $S(\text{crabapple})$ according to expressions (11) and (12). These utilities will interact with the probabilities $\mathcal{P}_B$ (expression 10) in choosing the label as a referring expression in a free-naming task.

4.4 Entry Level Effect as Expected Utility Maximization

The derivation of the expected utility model for selecting a concept label as a referring expression for an object closely parallels the derivation of the criterion for voice selection illustrated in section 3. We are given:

- an object classified under a set of concepts $\{C_1, C_2, \dots, C_n\}$ representing a path in taxonomic concept knowledge such that $C_k = C_{sup(k-1)}$ for $k = 1 \dots (n-1)$.
- the probabilities $\mathcal{P}_B(C_k) \in [0, 1]$ for $k = 1 \dots n$, representing the canonical salience of the concepts, and
- $u(C_k) \in [0, 1]$ for $k = 1 \dots n$, standing for the instantial salience of the concepts.

In decision theoretic terms, we have:

- a set of actions $A = \{A_{C_1}, A_{C_2}, \dots, A_{C_n}\}$, where A_{C_i} stands for the selection of the label corresponding to concept C_i
- a set of states representing the maximal statuses of the instantial saliences of concepts: $S = \{S_1, S_2, \dots, S_n\}$ where S_i is the state in which $u(C_i) = \max(\{u(C_1), u(C_2), \dots, u(C_n)\})$
- a set of possible outcomes in the form of (state, action) pairs. Here we have n^2 of them: (A_{C_i}, S_j) for $i = 1 \dots n, j = 1 \dots n$.
- Utilities $\mathcal{U}(A_{C_i}, S_j)$ associated with the outcomes. $\mathcal{U}(A_{C_i}, S_j)$ is the utility of selecting the label corresponding to C_i when C_j has maximum instantial salience among all the concepts. The values of $\mathcal{U}(A_{C_i}, S_j)$ are derived from the instantial saliences $u(C_i)$ and are shown in matrix form in table 2.
- An environment of decision making, characterized by the probability distribution $\mathcal{P}_B$ with n values, one corresponding to each state. We can transfer the distribution on to the set of states and write: $p(S_j) = \mathcal{P}_B(C_j)$.

	S_1	S_2	...	S_n
A_{C_1}	$u(C_1)$	0	0	0
A_{C_2}	0	$u(C_2)$	0	0
...	0	0	...	0
A_{C_n}	0	0	0	$u(C_n)$

(16)

Table 2: Matrix of $U(A_{C_i}, S_j)$

The expected utilities of action A_{C_i} , denoted by $EU(A_{C_i})$, represents the effective salience of C_i , and stands for the strength of the motivational factor for selecting the label corresponding to C_i . Given the utilities tabulated in the matrix 2, we get:

$$EU(A_{C_i}) = \sum_{j=1}^n p(S_j) \cdot U(C_i, S_j) = P_B(C_i) \cdot u(C_i) \quad (17)$$

When C_j has maximum instantial salience among all the concepts, the utility of selecting the label corresponding to C_i for any $i \neq j$ is 0. Thus, the only nonzero contribution to the expected utility of A_{C_i} comes from the pair (A_{C_i}, S_i) , and is given by the product $P_B(C_i) \cdot u(C_i)$.

5 Concluding Remarks

Decision-theoretic formalization of salience-induced phrase choice helps capture the relative frequencies of linguistic constructs, as illustrated in section 3 for the passive voice construction. Our view of salience-based phrase choice complements as well as reinforces probabilistic themes in systemic linguistics, such as those of [Halliday, 1991]: we discuss the overlaps and complementarity in [Pattabhiraman, 1992]. With the canonical-instantial interactions and decision theory, we have been able to 'travel' to another phenomenon, viz., basic/entry-level effect in object naming, and have fleshed it out computationally.

Our research plans for the future include work on structuring the decisions on the basis of the probabilities and utilities that enter into them. This will yield new NLG architectures in the long term. The multitude of numbers that one has to deal with in a quantitative approach can be a deterrent to ventures of large-scale implementations in existing AI languages and environments. However, the profusion of interacting numbers is more easily manageable in more 'linguistic' tasks like syntactic linearization than, say, in comparison generation, where heavy reliance on conceptual knowledge entails many numbers coming into play. Probabilistic contexts in linguistic choice are mostly binary, while those in conceptual choice typically involve several possibilities.

Acknowledgments: The authors thank David McDonald, Christian Matthiessen, Norman Foo, Ed Hovy and G.Sivakumar for comments, criticism and discussions. Thanks also to the three anonymous referees for their comments.

References

- [Conklin and McDonald, 1982] E.J. Conklin and D.D. McDonald. Salience: The key to the selection problem in natural language generation. In *Proceedings of the 20th ACL Conference, Toronto*, pages 129–135, 1982.
- [Gluck and Corter, 1985] M.A. Gluck and J.E. Corter. Information, uncertainty, and the utility of categories. In *Proceedings of the Seventh Annual Conference of the Cognitive Science Society*, pages 283–287, 1985.
- [Halliday, 1991] M.A.K. Halliday. Towards probabilistic interpretations. In *Functional and Systemic Linguistics: Approaches and Uses*, Ventola, E., Editor, pages 39–63. Mouton, 1991.
- [Jolicoeur et al., 1984] P. Jolicoeur, M. Gluck, and S. Kosslyn. Pictures and names: Making the connection. *Cognitive Psychology*, 16:243–275, 1984.
- [Langacker, 1987] R.W. Langacker. *Foundations of Cognitive Grammar, Volume 1: Theoretical Prerequisites*. Stanford University Press, 1987.
- [Mervis and Rosch, 1981] C.B. Mervis and E. Rosch. Categorization of natural objects. In *Annual Review of Psychology, Volume 32*, Rosenzweig, M.R. and Porter, L.W., Editors, pages 89–116. Annual Reviews, Inc., 1981.
- [Osgood and Bock, 1977] C.E. Osgood and J.K. Bock. Salience and sentencng: Some production principles. In *Sentence Production: Developments in Research and Theory*, Rosenberg, S., Editor, pages 89–140. Erlbaum, Hildale, N.J., 1977.
- [Osgood, 1980] C.E. Osgood. Salience dynamics and unnaturalness in sentence production. In *Lectures on Language Performance*, Osgood, C.E., pages 185–210. Springer-Verlag, 1980.
- [Pattabhiraman and Cercone, 1992] T. Pattabhiraman and N. Cercone. Communicating properties using salience-induced comparisons. In *Proceedings of the 14th Annual Conference of the Cognitive Science Society*, pages 1032–1037, 1992.
- [Pattabhiraman, 1992] T. Pattabhiraman. *Aspects of Salience in Natural Language Generation*. PhD thesis, School of Computing Science, Simon Fraser University, 1992.
- [Reiter, 1991] E. Reiter. A new model of lexical choice for nouns. *Computational Intelligence*, 7(4):240–251, 1991.
- [Rosch and Mervis, 1975] E. Rosch and C.B. Mervis. Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7(4):573–604, 1975.
- [Smith and Medin, 1981] E.E. Smith and D.L. Medin. *Categories and Concepts*. Harvard University Press, 1981.
- [Svartvik, 1966] J. Svartvik. *On Voice in the English Verb*. Mouton, The Hague, 1966.