Lib (ext) copy

ASPECTS OF SALIENCE IN NATURAL LANGUAGE GENERATION

by

T. Pattabhiraman

B.Tech., Indian Institute of Technology, Madras, 1982

M.S., University of Kentucky, Lexington, 1984

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
in the School
of
Computing Science

SIMON FRASER UNIVERSITY
August 1992

APPROVAL

Name: T. Pattabhiraman

Degree: Doctor of Philosophy

Title of Thesis: Aspects of Salience in Natural Language Generation

Examining Committee: Dr. Bill Havens, Chairman

Dr. Nicholas J. Cercone, Senior Supervisor

Dr. Robert F. Hadley, Supervisor

Dr. Raymond E. Jennings, Supervisor

Dr. Frederick P. Popowich, Examiner

Dr. David D. McDonald, Professor,
Computer Science, Brandeis University,
External Examiner

Date Approved: 7 August '92

Title of Thesis/Project/Extended Essay

Aspects of Salience in Natural Language Generation.

Author:

(signature)

Thiyagarajasarma Pattabhiraman

(name)

August 11, 1992

(date)

Abstract

This dissertation examines the role of *salience* in natural language generation (NLG). The salience of an entity, in intuitive terms, refers to its prominence, and is interpreted as a measure of how well an entity stands out from other entities and biases the preference of the generator in selecting words and complex constructs. Through an analysis of previous work in diverse disciplines, we show the variety of salience effects in NLG. Next, we classify several important *determinants* of salience, corresponding to different factors contributing to salience.

We then delineate two theoretically-significant categories: *canonical salience* and *instantial salience*. The former is characterized as a built-in preference in the general conceptual- and linguistic knowledge of the speaker. 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. Psycholinguistic results of Osgood and Bock are highlighted to suggest the multiplicative interaction between canonical salience and instantial salience. This interaction is captured in a decision-theoretic formalization which models canonical salience as probabilities, instantial salience as utilities, and the selection criterion as maximization of expected utility.

Next we consider the phenomena of *basic level-* and *entry level* preference in naming objects. We argue that basic level preference in taxonomic concept knowledge is a form of canonical salience. By establishing the stark similarity of the conclusions of the psychological experiments of Jolicoeur *et al.* to those of Osgood and Bock, the dynamics of preferring names for objects is conceptualized as an interaction between canonical salience and instantial salience. We demonstrate the suitability of decision theory to model this interaction.

Our model of salience interactions is further reinforced in a study of simile generation. We model property salience as information-theoretic *redundancy* and equivalently, as expected utility. To avert miscommunication and anomaly, we propose two different *cost* measures (antipodal to utility) using probabilistic knowledge and intrinsicness, and develop *net expected utility* as a decision criterion for selecting objects of comparison in similes.

We conclude in this thesis that the multi-aspect notion of salience provides subtle, quantitative control in language generation decisions, and captures interesting and significant effects such as graded judgments and relative frequencies of linguistic constructs.

To Amma and Appa

Acknowledgments

I acknowledge with gratitude the support, help and encouragement I received from many individuals in conducting my graduate studies at SFU and completing this dissertation.

First and foremost, my thanks are due to my Senior Supervisor Dr. Nick Cercone for providing individualized and patient supervision, constant encouragement and opportunities for all-round development, and most of all, for creating an intellectually liberal and open-minded environment for conducting research. I thank Dr. Ray Jennings for the many fruitful discussions, and for inspiring creativity ('*Never do a safe thesis*'), and Dr. Bob Hadley for encouraging me to seek psychological clues. I also thank Dr. Veronica Dahl for introducing me to computational linguistics and to natural language generation (NLG), and Dr. Romas Aleliunas for the many fascinating discussions we had at the incipient stages of this research. Drs. Fred Popowich, Dan Fass and Jim Delgrande helped with literature search.

Drs. Binay Bhattacharya and Art Liestman provided moral support and good counsel. Kersti Jaager helped in numerous ways and showed a personal interest in my progress. Tim, Stephen, Honman, Frank, Peter, Steve and Ed helped with the use of the computing facilities at the Centre for Systems Science and the School of Computing Science. SFU library was very responsive to my suggestions for book acquisition.

Generous financial support for sustaining my studies and travelling to conferences was provided by Simon Fraser University (several graduate research fellowships), the Natural Sciences and Engineering Research Council of Canada and Centre for Systems Science (thanks to Dr. Nick Cercone), the British Columbia Advanced Systems Foundation (graduate scholarship) and IJCAII (conference travel award). Valuable initial experience for NLG research was gained through participation in the project supported by IBM Canada and conducted by Dr. Veronica Dahl.

My special thanks are due to the members of the international NLG research community (the list is too long to enumerate here!) who have helped me immensely by obliging literature requests, discussing research issues over e-mail, and providing feedback and encouragement at conferences.

Sanjay and Sumathi, Mahesh and Pathmini, Ram, Ramesh, Rao, Ekam and their families, Param and Kalpagam and innumerable other friends and colleagues have made my stay at SFU pleasant, and have helped me in countless ways. I thank them all. Special thanks to Bob and Audrey who provided a home away from home.

I am greatly indebted to my parents, grandparents, brother, sisters, teachers, friends and relatives for their role in shaping and supporting my interests and instilling in me a zest for life.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

This thesis examines the role of *salience* in natural language generation (NLG). The salience of an entity, in intuitive terms, refers to its degree of prominence or level of activation. In NLG, it is interpreted as a measure of how well an 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 effective role of salience for prioritizing alternatives in NLG was first noted by Conklin and McDonald [4], who used it as a heuristic for selecting topics in a scene description task and ordering the topics in producing a paragraph. Guided by the premise that the choice of words and constructions on the basis of salience is pervasive in NLG, this thesis aims to flesh out the notion of salience in a theoretically-significant form.

NLG is most naturally viewed as a process of decision making under constraints (McDonald [38]). Throughout the process of producing an utterance, decisions are made on a number of different types of tasks, such as:

- *selecting utterance content*: For example, in giving route directions, what kind of information should be given to someone familiar with the area? to someone new to the area? How is a car described in a junior encyclopaedia? in a car manual?

- *selecting words*: Should I say *boy, youth, teenager* or *adolescent*? Should the object in question be called a *dog*? Or should it be called a *terrier*, or a *mammal?*

- *ordering a set of alternatives for sequential expression:* Should the *agent* of a proposition be expressed sentence-initially, in an active voice sentence, as in *The program generates a random sentence*, or, sentence-finally, in a passive voice sentence, as in *A random sentence is generated by the program*? We call this decision *linearization* in this thesis.

- choosing an appropriate reference for an entity:

  - Should a nominal entity be expressed as a pronoun? Can it be omitted, and the meaning still be recovered? Which of the following sounds more natural?

    * *John wanted to meet Mary yesterday, but John could not see Mary.*
    * *John wanted to meet Mary yesterday, but could not see her.*

  - Should an object be realized as a definite expression? Consider: *Harry fell several times. The snow was cold and wet*: How can we say '*the* snow', without having mentioned 'the snow' in prior discourse?

Salience provides rationale for choices in tasks like the above. For example, If words are ranked on the basis of salience in a lexical selection task, the most salient word is chosen. If the entities to be expressed as sentence constituents are ranked in terms of salience, then the more salient entity is expressed as the earlier sentence-constituent. The role of salience in NLG thus involves a relation between *prominence* of entities in a ranking, and *preference* of a choice among alternatives.

In this thesis, we explore the pervasiveness of salience as a principle of preference in NLG, by examining several facets of salience, such as the following:

- What factors determine the salience of an entity?
- What effects does the salience of an entity have in NLG?
- How do different factors determining salience of entities interact in NLG decisions?

We aim to shape a hitherto-non-existing theory of salience in NLG by asking a broad set of questions and seeking to formalize the most relevant aspects. We depend extensively in this thesis on discussions about salience in related disciplines of psycholinguistics and cognitive psychology.

1.1 Overview of the Thesis

Since we are seeking to mould a theory of a phenomenon which has not been explored for its theoretical significance and pervasiveness, we simultaneously aim towards breadth and depth. If there is some generality to salience, then we should explore a number of its facets, from the viewpoint of cause (factors that make entities salient) as well as effect (various decisions in NLG which are affected by salience). Once we see some general patterns, we can begin to formulate categories and distinctions in the evolving theory, explore them in depth and examine some of their consequences.

We begin our enterprise in chapter 2 by describing a number of types of decisions or phenomena in NLG in which salience plays a role. The decisions described are: content selection, content ordering, lexical selection, linearization and reference generation. We draw from works in NLG using salience as a heuristic, as well as works in other disciplines using salience in their explanations. The chapter suggests that these phenomena are sufficiently similar so as to be studied together, yet sufficiently different so as to form different aspects of a coherent study.

In the same spirit as chapter 2, chapter 3 covers a very broad ground by examining a number of factors which have been claimed to contribute to the salience of entities. Entities can be salient by being very vivid, by being pervasive, by being unique, or by being spoken about most recently, and so on. We attempt to explicate such factors as **determinants of salience.** This chapter takes as raw material discussions from a very diverse set of disciplines, each of which has used salience as a device to explain some facet of language behaviour or cognition. All the determinants assembled in this chapter have been invoked in explaining language generation decisions; yet, most of them are unformalized or not defined precisely. This chapter also testifies to the scattered nature of available knowledge about salience.

Chapter 4 then delineates an important pair of categories in the evolving theory of salience: *canonical salience* and *instantial salience.* The former is characterized as a natural, built-in preference which is inherent in the general conceptual knowledge and linguistic knowledge of the speaker/generator. An example of such a natural tendency is the preference to express *agent* before *object* in simple action sentences. Psycholinguistic evidence that we seek in characterizing canonical salience reveals that there *are* such principles of preference, and that they have their roots in prelinguistic perceptual experience and are cultivated during childhood. We also bring in clues from other disciplines like linguistics. We arrive at a set of characteristics of canonical salience which are relevant for a computational study of its role in NLG. While canonical salience is inherent in the generator, *instantial salience* arises in the generation context, in particular instances of language generation, because of a number of factors like *vividness*, *speaker motivation* and so on: factors which are discussed in chapter 3 as determinants of salience. We consider the interaction between canonical salience and instantial salience (that is, between built-in factors and situationally-arising factors) as a very crucial subject of study in NLG, and explain our intuitions on why this interaction is relevant in a theory of salience in NLG.

Next, in chapter 5, we explore the canonical-instantial interactions, and seek detailed psycholinguistic evidence that sheds light on the nature of their interactions. A distinguishing feature of these interactions is that canonical salience and instantial salience sometimes agree in their dictates, and sometimes are opposed in their dictates. 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. We then propose a formalization with several aspects:

- Canonical salience biases are probability distributions.
- Instantial saliences are utilities, expressing degree of preferability of choices.
- Determinants of salience are independent variables that compute utility functions.
- The interaction between canonical salience and instantial salience is computation and maximization of expected utility.

We thus put salience-induced decision making in NLG on a formal footing by correlating aspects of salience in NLG to elements of decision theory. We illustrate the formalization of interactions with an example of choosing between active voice and passive voice.

In chapter 6, we present a completely different aspect of language generation wherein salience has not been explicitly invoked in the existing discussions: **basic levels** in taxonomic concept knowledge. Basic level is a psychologically privileged level in a path in human taxonomic concept knowledge. One effect of basic levels on NLG is that when we name objects in day-to-day speech, we are most likely to choose a name which is neither too general, nor too specific. For example, we are most likely to say *give me an apple* instead of *give me an edible plant product* or *give me a red delicious apple.* We provide a set of reasons for considering basic level preference as canonical salience. Then, we bring into focus **entry level** effects, by which exceptions to the tendency of basic level preference are exercised. For example, while we are most likely to say *look at that bird!* when we point to most birds, we prefer to say *look at that ostrich!* when we point to an ostrich. We bring in experimental evidence from cognitive psychology to argue that the dynamics of preferring labels from a concept hierarchy in naming tasks can be modelled in terms of interaction between canonical salience and instantial salience. We demonstrate this argument by formalizing the interaction, and illustrate the effects with a detailed example.

Chapter 7 presents a method for generating comparison sentences of the form *A is like B*. The choice of the B in the comparison is investigated in terms of the salience of properties of candidate B's. We use elements of information theory to model property salience. We show how modelling the choice of the object of comparison in information-theoretic terms has an equivalent decision-theoretic formulation. We then develop the decision-theoretic criterion of *total expected utility* further, and augment it with salience heuristics based on *property intrinsicness* and thereby suppress generation of anomalous comparisons and avert miscommunication. The final decision criterion, termed *net expected utility*, has striking formal resemblance to some metrics of similarity. We have implemented the method described in this chapter in a C-Prolog generator that produces sentences like *Mary's cheeks are like apples*. We conclude this thesis document with chapter 8 which summarizes our findings and suggests directions for further work.

Chapter 2

Salience Effects in NLG

The first computational use of salience in an NLG system is that of Conklin and McDonald ([40], [3]). Since then, salience has been quite familiar to the NLG research community, and is fairly widely understood as a quantitative heuristic for prioritizing alternatives in a choice context and selecting from among them. The salience 'score' of a potential choice would indicate its degree of importance or prominence, mostly according to intuitive criteria, and govern its retrievability or likelihood of selection.

In disciplines related to NLG, such as linguistics and psycholinguistics, salience has been used 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.

While such explanations have been assimilated as part of common knowledge in NLG and computational linguistics, most of what is known about salience is available in works which are disparate, which are in diverse disciplines, and many of which themselves do not discuss salience as a theoretically-significant concept in its own right. Perhaps because of lack of availability of knowledge about salience in systematic form, the exploration of salience has not been seriously followed up in NLG since the early efforts of Conklin and

McDonald, and is at present regarded as little more than a convenient numerical prioritizing heuristic.

Nor are there any established methodological principles available for guiding a systematic computational study of salience. Such a study, if sustained, could improve our understanding of human language production, and perhaps of human language faculty itself, and its relation to cognition.

Given this state of affairs, to embark on a serious quest of salience to capture its perceived or alleged ubiquity would seem unmanageable and arduous, and much worse, risky, since the ubiquity of salience might after all turn out to be illusory! However, if one is sufficiently curious, one forges ahead with the endeavour nevertheless, and in spite of the ponderous initial steps, is bound to be rewarded with interesting results.

2.1 Studying Salience in NLG

The following very general questions can be regarded as constituting a comprehensive study of salience in NLG:

1. What *can* be salient?

2. What are the effects of salience in NLG?

3. How do entities gain or lose salience?

4. How do the factors contributing to the salience of entities interact in their effects in NLG?

5. Can these interactions serve as a basis for deriving architectures for language generators? If so how?

Detailed answer to all of these questions entails nothing short of a substantial long-term research programme. We present these questions here to suggest how an exploration of salience could be planned, and to use them also as a basis for organizing the exposition in some of the following chapters. Our first step is to get a feel for the pervasiveness of salience in language generation phenomena through a presentation and analysis of previous work pertaining to salience, done in the NLG research community as well as in other related disciplines.

2.2 Salience and NLG Decisions

In this chapter, we will present in brief a number of different NLG phenomena which have been explained or rationalized in terms of salience (question 2 above). The next chapter is devoted to different conceptions of salience, corresponding to different factors contributing to the salience of entities (question 3 above). We do not claim these discussions to be exhaustive of all previous work that has a bearing on our research. However, through an analysis of some of the major works which we found relevant, our aim, at the end of these chapters, is to improve our understanding of the facets of salience in NLG, in terms of effects (this chapter) as well as causes (next chapter), and as well, to entrench a better overall gestalt of salience as a phenomenon.

In the remainder of this chapter, we discuss, in turn, NLG decisions on content selection, content ordering, lexical selection, linearization and reference generation, from the viewpoint of salience effects. This portion of the chapter is an expanded and refined version of our paper [61]. We conclude this chapter with a summary discussion.

2.2.1 Content Selection

The salience of an entity in the knowledge base can determine whether the entity is selected for description in language. The higher the salience of an entity is, the greater is its likelihood of selection during content planning. The role of salience as a major influence on content selection decisions is particularly tangible in settings where natural scenes (which may involve motion) or their representations (e.g., photographs, movie segments) are described. The scene or the picture that forms the knowledge-base may be recalled from long-term memory, or concurrently perceived and understood as description generation proceeds.

We take the work of Conklin and McDonald as exemplars of how the salience of entities in the generation context can be used to plan and generate an entire discourse. In the case of Conklin and McDonald, the context is primarily a perceptual one, consisting of suburban scenes about which multisentential descriptions are generated. Conklin's text planner GENARO [3] prepares its output in the form of realization specifications which are taken up as input by McDonald's MUMBLE generator for producing surface text.

GENARO uses salience for determining topics in text planning. The knowledge base in GENARO is a network representation of a winter house scene, consisting of such objects as *house, fence, door, driveway, gate, mailbox* and so on. These objects, as well as their properties (e.g., colour of the house) and certain relations between the objects (spatial relations like *in-front-of*), are represented in GENARO as KL-ONE *concepts*. Numerical salience values are attached to objects, properties and relations, and are calculated from a perceptual experiment involving human subjects describing a photograph of the house scene [4].

The salience of the objects in the picture is determined by such aspects as centrality (in the picture), size, degree of unexpectedness and certain other aspects of world knowledge that determine the *prominence* of the object. The saliences of properties and relations are also determined by unexpectedness (for example, the colour of a *red* house in New England would be quite salient [3]) and general knowledge of prominence (for example, *red* is in general an attention-getting colour). Relation salience depends also on the salience of the objects related.

In GENARO, objects in the knowledge base are chosen as discourse topics solely on the basis of their salience. In addition, decisions on selecting subtopics to elaborate the main topics are influenced widely, if not exclusively, by the salience of properties of objects and spatial relations between objects. In short, GENARO embodies the principle that salience information in the knowledge base at least partially determines which aspects of the knowledge base get expressed in descriptions.

The following is an example of a paragraph generated by the GENARO-MUMBLE combination, reported by Conklin and McDonald in [4]:

> "*This is a picture of a white house with a fence around it. The door of the house is red, and so is the gate of the fence. There is a mailbox across the street in front of the fence, and a large tree obscures part of the driveway on the right. It is a cloudy day in winter.* "

2.2.2 Content Ordering

The role of salience in content ordering is correlated with the role in content selection (in human NLG) in settings that involve incremental language production. In these settings, once a topic is selected, it is immediately expressed in discourse. Content ordering decisions implicitly emerge out of the combination of content selection decisions and incremental generation.

This has been implemented in the GENARO text planner [3] using a stack mechanism to store the unmentioned salient objects. The top of the stack is the most salient of the currently unmentioned objects. Items are popped off the stack in turn, and expressed in discourse.

2.2.3 Lexical Selection

The role of salience in lexical selection is evident in several distinct aspects. Here we present three instances wherein salience has been invoked in lexical selection tasks:

1. salience of lexical feature values contributing to the meaning of words. Equivalently, salience of attribute values contributing to the definition or identity of concepts.

2. prominence of words related to their capacity to evoke intense affective responses.

3. factors contributing to the salience of objects being explicitly verbalized as adjectives.

Feature Salience

In the lexical representation of a concept, salience associated with an attribute signifies how essential the attribute is to the definition or identity of the concept. For example, *majesty* of lions and *cuddliness* of koalas may be recorded as salient attributes. In the lexical selection system of Nirenburg and Nirenburg [50], *importance values* are attached to the attributes in the dictionary slots. In the representation of the word/concept *boy*, the *sex* attribute carries a higher salience (= importance value) than the *age* attribute. This annotation enables selection of *youth* rather than *girl* as a synonym for *boy*.

A similar idea has been used by McKeown in the TEXT system [43]. When queried about the difference between two entities, the answering system should convey (what McKeown calls) the *most salient distinction* between the entities. What constitutes the most salient distinction between the entities depends, in the TEXT system, on the relative positions of the compared entities in the generalization hierarchy. Thus for the question

> What is the difference between a part-time and a full-time student?

the following response is appropriate:

> A part-time student takes 2 or 3 courses per semester while a full-time student takes 3 or 4.

On the other hand, for the question

> What is the difference between a raven and a writing desk?

the following response is odd:

A writing desk has 4 legs while a raven has only 2.

A better answer would have perhaps pointed out that ravens are animate while writing desks are inanimate objects. Although McKeown does not rely on an explicit treatment of salience in her TEXT system, her use of the term *salience* in her discussion about response generation is very much like the use of importance values in the dictionary slots of Nirenburg and Nirenburg.

We discuss property salience in chapter 3 as well, where we present some of the conceptions of salience in cognitive psychology. We present our own formalizations of property salience in chapter 6, where we model the influence of atypicality of objects in object-naming tasks, and in chapter 7, in comparison generation.

Lexemic Vividness

Salience values may be associated with lexemes so as to reflect their intrinsic *vividness*. For example, here is a set of near-synonymous alternatives in increasing order of their vividness:

kill, murder, assassinate

In this example, lexemic prominence does not by itself guarantee preference. The word *assassinate* is inherently prominent, but not inherently preferable.

Which of the above words is chosen in particular cases *might depend* on who the victim of the killing was: the more famous or prominent the victim is, the more vivid could the chosen verb be. Thus, the position of the chosen word on its prominence scale matches that of victim of the killing (or murder, or assassination!) on another prominence scale.

What we have here is an interaction between salience configurations in the input to the lexical selection process and those in the lexicon, associated with the lexemes. The word is selected through a process of 'resonance' of salience levels. Vividness of lexemes is defined by their ability to evoke affective responses. It is discussed in chapter 3 as a determinant of salience.

Verbalizing Salience Determinants as Adjectives

Words are sometimes chosen so as to indicate explicitly some of the factors that contribute to the salience of objects. This is particularly common when the objects are highly salient. For example, in

> A *baffling, new* galaxy has been discovered at the centre of the Milky Way.

the high salience of the galaxy is evident not only in its early positioning in the sentence (cf. linearization, discussed below), but also in the use of adjectives *baffling* and *new* that indicate the roots of the galaxy's salience.

2.2.4 Linearization

By linearization, we refer to the class of ordering decisions involved in expressing fragments of meaning in the form of clausal or sub-clausal units. In general, linearization encompasses ordering decisions occurring at several levels of discourse structure, and as such, could subsume content ordering decisions described above. In this thesis, however, we will use the term *linearization* to refer to ordering at the sentence level.

Relative positions of meaning fragments in sentences could be stated in several ways, viz., in terms of:

- semantic roles: *agent* $\prec$ *patient* ('$\prec$' = precedes)
- grammatical relations: *direct object* $\prec$ *indirect object*
- syntactic categorial labels, typically naming nodes of parse trees: $NP \prec VP$
- abstract syntactic categories: *head* $\prec$ *complement*

and so on. In NLG, the eventual order of words in an utterance arises out of a complex interaction of language-specific syntactic constraints with a broad class of functional constraints imposed in the context in which utterances are produced.

The general principle of the role of salience in linearization in NLG may be stated as follows: the higher the relative salience of an element is, the earlier is it likely to be expressed in the utterance. In psycholinguistics, the effects of salience on linearization have been studied with the aid of detailed experiments, by Osgood and Bock [58] and Sridhar [69].

Salience influences linearization to produce sentences exhibiting a wide variety of syntactic phenomena, some English examples of which are listed below:

- 'movement' phenomena like passives:

  A baffling new galaxy has been discovered at the centre of the Milky Way.

  The big black ball is hit by the smaller black ball. (from the experimental data of [69]).

- Inversion phenomena, like dative shift:

  I wrote *the President* a letter.

- Adverb movement:

  The big boy *greedily* ate the pizza. (a variation of an example in [58]).

Objects could be rendered salient by a variety of factors, such as vividness, speaker's intentional emphasis, and so on. Salience effects in linearization could also emerge implicitly, as a derivative effect of verb choice. For example, with reference to

John sold a book to the President

the higher salience of *the President* could trigger

The President bought a book from John

rather than the passive

The President was sold a book by John

Salience effects in linearization have been found to occur in a wide variety of languages. However, while speakers may tend to express anything they find or consider salient earlier in the sentence, the extent to which sentences could be linearized in response to salience depends also on language-specific morphosyntactic constraints. Sridhar [69] notes that SOV languages, when compared to SVO languages, tend to rely less on word order to express salience. Salience effects on linearization are considered further in chapter 5.

Computational Mechanisms for Salience-Induced Linearization and Lexical Selection in NLG systems

The recent work of Wanner and Bateman [79] is a good example of a mechanism in an NLG system to allow salience of particular aspects of semantic input to the generator to control lexical selection. Wanner and Bateman represent the input semantic structure as an *abstract situation* in the Meaning-Text Theory (MTT) of Mel'čuk [44], consisting of a key term and a set of participants. The choice set of potential configurations of salience (prominence) is represented as a network consisting of MTT *lexical functions* (LFs) at several levels. These LFs capture co-occurrence dependencies.

For instance, the LF *Oper1* corresponds to the salience of the first participant in the situation. *Oper2* corresponds to the salience of the second participant. Thus, *Oper1(influence)* = *exert*, and could generate

A exerts influence on B

and *Oper2(influence)* = *be under*, corresponding to

B is under the influence of A

In the former instance, the abstract situation representation is accompanied by information that the agentive participant of the situation is more salient than the patient. In the latter instance, the patient is more salient than the agent. In both cases, the more salient participant will be expressed earlier in the sentence. It is to be noted that the choice of the verb could depend on the relative saliences of its arguments, which are determined in several possible ways earlier during generation. Choosing the verb prior to processing salience information on the arguments could force the generator to backtrack and reconsider its verb choice.

2.2.5 Generating Referring Expressions

In this subsection we identify two more phenomena in NLG which have been explained on the basis of salience: metonymic reference and anaphoric reference. Metonymies are determined by the salience of *association* between objects. Anaphoric reference is governed by a *dynamic* notion of salience, by which the salience of entities introduced into the discourse varies as the discourse progresses.

Metonymy

Metonymic (or indirect) reference involves referring to an object in terms of an associated attribute or another object. For example, in the sentence (adapted from Lakoff and Johnson [31])

The ham sandwich paid his bill and left

the ham sandwich refers to the person who ordered the ham sandwich.

Metonymic references are pervasive in language, and are usually explained in terms of salience (Herskovits [25]). An object B (say, ham sandwich), in order to function as a metonymic reference to an object A (the person who ordered the ham sandwich) is required to be a *typically salient* attribute of A, or saliently related to A in the context of generation.

The knowledge of what constitutes a typically salient association is an aspect of world knowledge. The attribute that licenses metonymic reference (object B) can be associated with object A subordinately (as a part or a subclass), superordinately (as a whole or a superclass) or through other associations. In each of the following pairs of examples, (a) involves direct reference, while (b) involves indirect reference.

(a) buy a carton of milk
(b) buy milk

In (b) above, what is contained in the carton is substituted for the carton. Contents are typically salient attributes of containers.

In the following pair, (b) involves metonymic reference by a shift from the occupants of the stadium to the stadium itself:

(a) The spectators in the stadium rose in applause.
(b) The stadium rose in applause.

In (b) below, the author of a book is used as a label in referring to the book.

(a) The book by Plato is on the shelf.
(b) Plato is on the shelf. (from Fauconnier [8])

Pronominal and Definite Reference

The level of activation of entities in the short-term memory or the discourse model of speakers and hearers is an important variable in generating connected multisentential text, An entity just recently mentioned stays at a high level of activation, while an entity which has not been mentioned for a long time is quite likely to have dissipated its once-high activation level, and faded out of the discourse model. Thus, the processing history of the generator as it produces text governs the activation levels of entities in the discourse model, which, in turn, determine aspects of subsequent text produced by the generator.

The notion of activation level of entities in the context of NLG, which varies with time as the discourse progresses, is synonymous with an important dimension of salience which we call *dynamic salience.* Dynamic salience is crucial to several phenomena in multisentential discourse generation. The level of dynamic salience of an entity at any time during discourse production has a high impact on whether the entity can be pronominalized or referred to by means of a definite noun phrase. Factors determining dynamic salience are discussed in chapter 3 as *processual* factors, since they pertain to the processing history of the generator.

The influence of dynamic salience on processes in language comprehension, such as pronoun resolution and definite reference resolution, have been studied in the tradition of Czech functional linguistics by Hajičova and her colleagues ([18], [19] and [17]).

Reference generation phenomena have also been explained in terms of several alternative notions related to dynamic salience, such as *focus, context activation* and *concept activatedness.* We refer the reader to Robert Dale's Ph.D. thesis [6] for informative discussions of the literature pertaining to these notions. A context activation mechanism for use in NL interpretation processes such as reference resolution and ambiguity resolution has been proposed and implemented by Alshawi [1]. Most recently, Alshawi's proposal has been applied in reference generation in a multimodal environment by Claassen [2].

Several aspects of the influence of dynamic salience on reference generation are illustrated by the following short discourse, adapted from [19]:

(1) The school garden was full of children
(2) They talked noisily
(3) But the teacher did not restrain them

Pronominalization of entities, introduced overtly into the discourse earlier, is allowed on the basis of their (dynamic) salience. As long as the salience of an entity stays above a threshold, it stands a good chance of being pronominalized in subsequent reference. Thus, in (2) above, *they* refers to *children* mentioned in (1). A recently-mentioned entity retains a high level of salience in the speaker's discourse model.

When an entity is mentioned, and given high salience thereby, certain other entities that are closely associated with the recently-mentioned entity also gain salience by virtue of their association with the original salient entity, through the process of *spreading activation*. The degree of salience gained by associated entities is related to the proximity of association, and decreases as the distance of association increases. There is thus a ripple-effect in propagating salience to other entities through spreading activation.

When an entity is overtly mentioned for the first time in discourse, it is normally not referred to by means of definite expressions. However, in the example discourse given above, the use of the definite expression *the teacher* in an initial reference in (3) is permitted under the influence of the salience gained by the concept *teacher* because of its association with the *school garden*.

The (dynamic) salience of an entity not mentioned in the discourse tends to decrease. If it is not mentioned for long enough, (i.e., if the salience of the entity drops below a certain threshold level), a pronominal reference to the entity is not permissible, as the reference is not likely to be resolved correctly. If the object needs to be mentioned again, it has to be re-introduced overtly in the text, and thereby, re-assigned high salience.

In the above example, *them* in (3) refers to *children*. Though it is a bit distant from the initial reference in (1), it is allowed, as the salience of *children* is still above the threshold for pronominalization. The pronominalization in

Floyd wanted Roscoe to get *his* fishing pole (from [42])

can be similarly analyzed.

2.3 Synthesis

The variety of NLG phenomena which have been independently observed and explained in terms of salience in different research communities testifies to the pervasiveness of salience. The term *salience* has been applied in different senses, depending on the source of salience, as well as on the context in which salience has been examined for its effect: perceptual prominence (for example, by Conklin and McDonald), vividness (Osgood) and a dynamic sense of activation level in discourse model (Hajičova). Moreover, salience has been referred to by such different labels as *importance value*, *prominence*, *degree of foregrounding* and *activation level*. However, these cases appear sufficiently comparable for lending themselves to an eventual definition as a coherent concept with multiple facets.

All of the NLG effects discussed in this chapter involve a relation between *salience* and *preference*:

- An entity becomes salient in a certain context.
- A choice among alternatives needs to be exercised in the generation process.
- By interpreting the degree of salience as having motivational significance, by which it is transferred on to words or phrases as their degree of preferability in a context of choice among alternative words and phrases, a cause-effect relation is brought about between the context of salience and the NLG process.

The degree of salience is evaluated relative to, or in the context of, other entities vying for salience. Correspondingly, the degree of preferablity of a word or a construction is evaluated in a context of alternative words or constructions. In this manner, salience plays a role in decision making in NLG.

When a cause-effect relation is brought about between a context of salience and a context of preference of words or phrases, salience is said to be expressed in language, and language is said to serve as a vehicle for communicating salience. Salience is expressed in language in a variety of ways as seen in this chapter:

- through uncommon or special orderings (as in passives)

- through the use of unusual words, or the use of words in unusual senses (as in metonymy)
- through explicit verbalization of factors contributing to salience of entities (as in a baffling, new galaxy)
- through the use of vivid words evoking intense affective responses (as in *assassinate*)

The more unusual or vivid the chosen word or construction is, the more *tangible* salience is in the utterance generated.

The sources of salience may be found in a number of factors such as perceptual centrality, vividness, speaker motivation, recency of mention and so on. We devote the next chapter (chapter 3) to discussing these factors as *determinants of salience*. While each of the effects and causes could be conveniently studied in isolation, a major challenge for NLG research is to understand and explain how they work in combination: how the different determinants interact in their effects, and what the interdependencies are between the effects themselves. We will begin to address the problem of interactions after discussing the determinants of salience in the next chapter.

Chapter 3

Determinants of Salience

As suggested in the previous chapter, a study of salience in NLG ought to approach the problem from several angles, viz., from the viewpoints of the entities that become salient, the effects of salient entities on NLG decisions, factors determining the salience of entities, the nature of interactions between the determinants of salience, and finally, architectural implications of the interactions on the generator.

We envision architectural consequences as an advanced stage of the theory, since a lot will have to be understood about salience interactions before we can use them as a principled basis for designing generators. Much of this thesis is devoted, therefore, to issues of modelling rather than to implementational ones. We do have a non-trivial implementation accompanying chapter 7. However, the implementation was carried out more in the nature of a useful ancillary to our modelling interests rather than for immediate adoption by the NLG community as a blueprint for implementing large generators.

Usage: In this thesis, by *dynamics of salience*, we will informally refer to any one or more of the processes by which:

- entities gain or lose salience
- salience is propagated from one entity to another
- saliences of entities interact in influencing NLG decisions

We will refer to salient entities as *bearers of salience*, and to the factors contributing to the salience of entities as *determinants of salience*. Entities, the saliences of which are compared in context, are said to constitute, or participate in, a *relative salience configuration*.

3.1 Bearers of Salience

Various entities can possess salience in NLG, as indicated by our presentation in the previous chapter. Salience values may be ascribed to all ontological categories such as objects, properties, relations, states, events, actions and so on. Conklin [3] classified salience determinants into the types *object salience*, *property salience*, *relation salience* and *gestalt salience*, corresponding to the kinds of entities that became potentially salient in the picture-description domain.

Relative salience configurations could exist within an ontological category, as, for example, in:

- objects: *house* $\succ$ ('is more salient than') *tree*, say, in a picture.
- properties: *colour* of a house $\succ$ *price* of the house, say, in attracting visual attention.
- relations:

  $ordered(customer, ham\_sandwich) \succ did\_not\_like(customer, ham\_sandwich)$

  in referring to the *customer* metonymically as *ham sandwich*.

They could be defined *across* categories as well. For example, Sridhar [69] notes that in natural descriptions of scenes, the following relative salience configuration expresses speaker's preference in content selection:

$$actions \succ changes\ of\ state \succ constant\ states$$

Abstract entities which get reified during the process of discourse production can become salient; for example, complex entities such as

> the fact that the President fell off a horse when he was vacationing in his ranch

can be salient.

As a first cut at classifying bearers of salience, we may classify them into three broad groups:

1. salient entities in the resources, like words, items in general knowledge, specific items which are input to the generator, etc.

2. salient entities in the evolving text: e.g., fragments of semantic representations, noun phrases, etc.

3. salient items installed in the discourse model or short-term memory as the discourse progresses. This category includes the types (1) and (2) in the dynamic sense of time-varying aspects of salience of entities.

For example, in generating the short discourse

> A *baffling new galaxy* has been discovered at the centre of the Milky Way. Astronomers are fascinated with its unusual characteristics.

1. The entity *galaxy* is salient (because of its baffling, new nature).

2. The noun phrase *A baffling new galaxy* is salient.

3. As the discourse proceeds, the baffling new galaxy is still salient enough in the discourse model towards the end of the second sentence that it licenses pronominalization in *its unusual characteristics*.

3.1.1 Propagation

Our intuition, at this stage, is that depending on the type of bearers of salience, different aspects of the generation process could serve as natural mechanisms for propagating salience. For example, if syntactic structural knowledge is explicitly used as a mechanism for surface generation, then salience could be propagated through phrase structure. If generation is modelled as a mapping between different levels of representation, mapping rules could be designed so as to preserve relative salience configurations. The links in structures in the knowledge base could be used as vehicles for transmitting salience, say from a property to an object which possesses it, or from one object to another through association (by spreading activation). Managing dynamic salience over time for tasks like reference generation will require special update mechanisms.

We will give little more detail about salience propagation in the rest of this thesis, since the specific technical details of propagation require architectural commitments as prerequisites. However, accommodating considerations about salience propagation in the formative stage of the theory on the basis of the *feel* we have for the process is bound to prove itself useful for future developments.

The remainder of this chapter is devoted to determinants of salience. We rely for our exposition in this chapter upon discussions in a number of disciplines. We will also attempt to clarify the relation between different researchers' conceptions of salience.

3.2 Salience Determinants

Following the usage of Sridhar [69], we designate factors contributing to salience of entities as *determinants of salience*. Determinants contribute to the salience scores of entities, and also serve (in the theory) as explanations of why or how something is salient. Several determinants may combine to contribute to the salience score of an entity. One determinant may define scores for several bearers of salience, which could all constitute a relative salience configuration.

It may not be possible to catalogue the possible determinants of salience in NLG exhaustively, nor is it (we feel) a particularly pressing issue now. Characterizing them in terms of primitives is important. However, it appears to be a daunting task, and has not been done successfully so far in any of the ventures that dealt with salience. Nevertheless, different researchers have characterized the salience determinants that arose in their respective studies. We will understand the existence of interdependences among salience determinants, and try to unlock some of them.

Moreover, the criteria that characterize the determinants are often intuitive, and hence they do not reveal all the relevant details, nor do they admit watertight formal definitions. We will include them in our framework as long as they are intuitively appealing, and their role in NLG decisions is incontrovertible. Besides, we are inclined to view the intuitive and diffuse nature not as a problem, but as an opportunity for exploration and eventual precise characterization. We will now discuss the determinants in turn, grouping similar ones together 'locally', but otherwise following a somewhat arbitrary sequence. We will mark several intermediate checkpoints at which we will offer discussions on the determinants just presented.

3.2.1 Vividness and Imageability

Vividness of an entity is described as its capacity to evoke emotional or affective representations. For example, *vampire* is vivid; *man* is relatively *pallid*. In psycholinguistics, Osgood and Bock [58] and Sridhar [69] have investigated its role in linearization. In social cognition, Nisbett and Ross [51] and Fiske and Taylor [11] discuss its effects on human social- and attitude judgments. In consumer research, it has been examined for its persuasive effects, by Taylor and Wood [72], among others.

Haagen [16] characterizes vividness of connotation as the clarity or graphicness of the impressions which words arouse. Vivid words evoke attitudes and feelings quite like those created by the actual experiences, events or objects they represent. Pallid words do not stir emotion, nor do they bring to mind graphic imagery- they are little more than mere identifiers.

The ability to evoke clear internal visual representations is termed **imageability**. All the researchers cited here have noted the close connection between imageability and vividness. However, imagebility and vividness are not identical. *Triangle* is easily imageable, but not vivid; *scandalous* is very vivid, but not easily imageable.

Osgood and Bock measure vividness as the intensity of an intersection of both denotative semantic features (e.g., +relative size: a very large ball is more vivid than an average-size ball), and (especially) affective features which are highly polarized (too high or too low). **Novelty** is also noted as contributing to vividness.

Both psycholinguists and social psychologists regard vividness as an inherent property of words (or stimuli). However, for vividness to have its impact, a contrast is necessary: in our parlance, an entity should be highly vivid in a relative salience configuration in which other entities are pallid, or not as vivid. (For example, a bright star is vivid in a constellation of dim stars). Here are some more examples of vivid/pallid word pairs, mostly from Osgood and Bock: *fairy tale/story, orphan/child, bomb/letter, tornado/rain, castle/house, avalanche/snow.*

The vividness of words has been measured in psycholinguistics using the *semantic differential technique* [59]. Haagen [16] reports experimental evaluation of vividness of 400 pairs of adjectives.

Entities rendered salient by vividness tend to get expressed earlier in the sentence, as in the passive

> A vampire was seen by the maid

In spoken language production, vividness leads to extra stress.

3.2.2 Uniqueness

Uniqueness is the term we use to refer to salience in the *one-in-a-million* effect of how entities stand out in a group by being exceptional or rare in some way: the 'million' creating a background for the 'one' to stand out in the foreground. Related determinants discussed in the literature are: *unusualness* and *atypicality*. While uniqueness signifies an extreme, *atypicality* is perhaps better in indicating the graded nature of the determinant.

By this determinant, a young person is highly salient in a roomful of octogenarians [11], while probably not so in a roomful of young persons. The perceived uniqueness of an entity changes when it appears in different contexts. Fiske and Taylor call this determinant *salience*, and distinguish *salience* from *vividness*.

Other related determinants that fall naturally in this group are: (degree of) *unexpectedness, novelty, surprise* and *expectation failure*. Uniqueness can be perceived in a perceptual context when an object stands out in a group, or when an unprecedented event, or an event contrary to expectations occurs.

3.2.3 Pervasiveness

Entities can become highly salient by being the most pervasive, abundant, frequent or probable in a context. Related notions include *typicality* and *familiarity*.

An entity will receive complementary (opposite) ratings when evaluated for its uniqueness and pervasiveness. However, it is possible for an object to gain salience by being pervasive (abundant) in one context and unique in another simultaneously. For example, given a context of fruits of different colours, *apples* will stand out as good examples of red fruit if apples are mostly red in colour (pervasiveness of *red* among *apples*) and if there are very few other fruits in which red colour stands out to any high degree (uniqueness of *red apples* among *fruits*).

Discussion

At this stage of our enumeration of salience determinants, two chief characteristics of salience emerge:

1. It is a graded notion with potential values on a continuous scale, although it is frequently discussed in terms of qualitative high/low contrasts.

2. It is relative, or context-dependent: a foreground needs a background in order to stand out.

Fiske and Taylor note that while uniqueness (their *salience*) is a property of objects *in context*, vividness is an *inherent* property of objects. Osgood also characterizes vividness of words as *inherent* in, or *intrinsic* to, their individual semantic encodings.

However, Fiske and Taylor as well as Osgood recognize that a differential (in the form of a contrast against pallid items) is necessary for the vividness of an object to have effects, whether persuasive effects (the former) or NLG effects (the latter).

Since saliences of entities have to compete in context to stand a chance of causing an impact, [Fiske and Taylor]'s emphasis on uniqueness being contextual and (in contrast) vividness being intrinsic would seem a bit misleading.

Yet, there does exist a difference between vividness and uniqueness: while the evaluation of the vividness of a word (say) can be done in isolation, the evaluation of the uniqueness of an entity requires a context. It is on the basis of whether the measurement of salience requires explicit reference to external objects, that vividness and uniqueness differ. This distinction is also captured clearly in Tversky's characterization of salience in cognitive psychology (described below).

3.2.4 Property Salience in Cognitive Psychology

Some of the most interesting discussions available on salience are in cognitive psychology, specifically, those of Ortony and Tversky. Their problem of interest is similarity judgments in humans. The degree of similarity between objects compared in a similarity judgment is conceived as a function of salience of properties (attributes) of the objects. The early work of Tversky [73] on literal similarity judgments is furthered and modified by Ortony [53] and Ortony *et al.* [54], and applied especially to metaphorical comparisons.

Tversky's Salience Factors

Tversky [73] specified two chief determinants of property salience: (1) *intensity* and (2) *diagnosticity*. Intensity is characterized by factors that increase signal-to-noise ratio, such as brightness of a light, clarity of a picture, amplitude of a tone, vividness of an image, and so forth. According to Tversky, intensity is independent of the object.

Diagnosticity pertains to the discriminability of an object from other objects with which it is classified. It is object-dependent, in the sense of presupposing a set of alternatives. Ortony *et al.* [54] call this determinant *relevance*.

We can easily see that Tversky's *intensity* is the same as *vividness* of Osgood and others except for a difference in emphasis (in Osgood, of the intensity of affective features and of emotional interest in determining vividness). Tversky regards intensity to be *object-independent* in the same way Osgood and [Fiske and Taylor] regard vividness to be *intrinsic* or *inherent*. Tversky defines diagnosticity to be *object-dependent* in the same way Fiske and Taylor characterize uniqueness (their *salience*) as a property of stimuli in context.

Ortony's Salience Measures

Ortony's critique [53] of Tversky's measures of salience is largely directed at the definition of salience as an object-independent measure. Specifically, Ortony criticizes Tversky's position for its implication that intensity and diagnosticity together exhaust all there is to attribute salience (as applied to similarity judgments), and when diagnosticity does not come into play in similarity judgments, the salience of an attribute is independent of the object of which it is an attribute.

Conceptual Centrality Ortony then proposes what we may regard as a third dimension of attribute salience. In this dimension, regardless of diagnosticity, an attribute can be more important or prominent with respect to one object than it is with respect to another, in a person's representation of an entity or a category. For example, *being red* is a more important attribute of a fire truck than it is of a brick; in particular, neither need be necessarily red. In a later article ([54]), this third determinant is termed *conceptual centrality*.

The difference between diagnosticity (which Ortony calls *relevance*) and conceptual centrality is illustrated in [54] by considering the category of general anaesthetics and sleep-inducing drugs. Within this category, the attribute *induces sleep* has high salience (conceptual centrality) for all its members. However, for the same reason, it has little *relevance* for similarity judgments between members, which cannot be distinguished on the basis of *induces sleep*. Thus, within the context of the given category, the relevance (diagnosticity) of a high-salient (conceptually central) attribute is low.

Salience and Similarity Ortony then uses his determinants of salience in experiments with his *imbalance model* of similarity, which supersedes Tversky's *contrast model*. With the imbalance model, Ortony and his colleagues are able to explain certain asymmetries common in human similarity judgments (for example, North Korea is judged to be more similar to China than China is to North Korea. Also compare *A smile is like a magnet* with *A magnet is like a smile*. They also examine the distinction between literal, metaphorical and anomalous comparisons, for the details of which we refer the reader to [53] and [54].

Determinants of Ortony et al. The determinants of salience used by Ortony *et al.* in their experiments are:

1. *conceptual centrality*: given above

2. *applicability*: applicability to all or most instances of a concept, for example, the attribute *has a windshield* for automobiles. This determinant is the specialization of what we discussed earlier under *pervasiveness* to the distribution of properties among concepts.

3. *characteristicness*: this refers to salient attributes of objects which need not be conceptually central- e.g., the attribute *being hairy* of gorillas. According to Ortony *et al.*, while *being hairy* is a salient attribute of gorillas, it is not *conceptually central* to gorillas, since a shaven gorilla is no less a gorilla.

Discussion

To understand the subtle dimensions of property salience is of crucial importance to NLG in a number of tasks in which conceptual knowledge plays a salient(!) role: tasks such as generating referring expressions and object names, and the vast unexplored territories of generating comparisons, similes and (certain kinds of) metaphors. Ortony's theoretical discussion of salience is not detailed in computational terms. An attempt at fleshing out some of these salience determinants computationally should prove interesting. In particular, the conception of property salience in terms of pervasiveness within category and uniqueness across categories, coupled with the potential for the degree of salience to lie anywhere in the continuum between pervasiveness and uniqueness, suggests a probabilistic treatment. Inspired in part by Ortony's discussions, we have developed quantitative models of some aspects of property salience. We present them in chapter 6 (atypicality in object naming tasks) and chapter 7 (comparison generation).

Intermediate Summary In sum, objects can become salient due to their properties in a variety of ways, through one or more of the following determinants:

1. vividness (= intensity, Tversky)
2. uniqueness (= diagnosticity, Tversky)
3. pervasiveness (= applicability, Ortony)
4. conceptual centrality (Ortony)

5. characteristicness (Ortony)

3.2.5 Perspectives from Cognitive Linguistics

We now turn to the discussions of Langacker [32] which amplify and clarify our earlier discussion on property salience. Langacker relates the degree of centrality of a property of an object to the traditional distinction between *essential* properties and *contingent* properties. The former have to be adduced in explaining to someone what the concept is (e.g., shape of cats), while the latter involve knowledge of its association with other concepts, or very specific events involving specific instances of the concept (e.g., cats' association with Halloween, what my neighbour's cat ate this morning, etc). Such properties may enrich our understanding of a cat, but are peripheral to its meaning.

Langacker notes that all the numerous specifications in the human encyclopaedic conception of an entity form a cline with respect to their centrality. No specific point along the cline is a logical choice for *a priori* demarcation of linguistically-relevant properties on one side and linguistically-irrelevant properties on the other.

Langacker's Determinants

Langacker gives the following as determinants of centrality:

1. *Conventionality*: (estimated) extent of sharing by the speech community.

2. *Genericity*: as opposed to specificity. For example, compare

   I am allergic to my neighbour's cat (very specific)

   with

   Many people are allergic to cats (generic)

3. *Intrinsicness*: A property is intrinsic to the extent that its characterization makes no essential reference to external entities. For example,

taste of edible fruits is extremely intrinsic. *Shape* is highly intrinsic. *Size* is not as intrinsic as shape. Behavioral properties involving relation to other entities (like cats chasing mice) are more extrinsic (to cats). The symbolic and cultural roles of entities are highly extrinsic (for example, associating doves with peace, or peacocks with pride), since they are largely a matter of how others regard the entities. We apply a heuristic based on intrinsicness in chapter 7 to avert anomaly and miscommunication in comparisons.

4. *Characteristicness*: Uniqueness to a class of entities, sufficient to identify a class member. For example, shape of cats. The shape of California is more characteristic than the shape of Colorado.

Synthesis

Langacker's *characteristicness* coincides with a combination of Ortony's characteristicness and applicability. While Langacker gives it as a component of centrality, Ortony distinguishes characteristicness from centrality (a characteristic attribute need not be conceptually central, according to Ortony) and gives them both as components of salience. All characteristic properties are applicable to all or most instances of a category, but not all widely-applicable properties are characteristic of a category (e.g., the attribute *has two eyes* for a cat).

We then adhere to Ortony's convention of labelling all the factors as determinants of salience, and arrive at the following list of salience determinants as an outcome of unioning and intersecting the lists of Ortony and Langacker:

1. conceptual centrality (Ortony)

2. pervasiveness (= applicability: Ortony)

3. characteristicness (Ortony; Langacker, sans the applicability connotation)

4. intrinsicness (Langacker)

5. conventionality (Langacker)

6. genericity (Langacker)

This list completes the determinants of salience we have seen so far, along with:

7. vividness/intensity (Osgood, Tversky and others)
8. uniqueness (= diagnosticity: Tversky; relevance: Ortony)

3.2.6 Association, Conventionality and Entrenchment

Through association, entities can gain salience from other entities. Association is implicated in explaining metonymic shifts in reference generation (chapter 2). Association between entities forms a structural basis for organizing knowledge (memory) in associative networks. Schema, frames and scripts are other structured representations that capture associations.

Chains of association can be 'short-circuited' into direct associations gradually. The direct associations get more *entrenched* through repeated use over the long term. Such associations can be found in similes such as *proud as a peacock*, and in symbolisms such as *doves* standing for *peace* (pointed out by Lakoff and Johnson [31]). This determinant is related to Langacker's conventionality. It is also shaped by another determinant, viz., *frequency of mention* in the long term, which contributes to the entrenchment of the association.

Finally, we can count *entrenchment* itself as a determinant of salience: the degree of salience of an entity in the knowledge base or a phrase in a linguistic repertoire is determined in part by its relative degree of entrenchment. Novel constructions and words are less well-entrenched than familiar and common constructions or words.

In NLG models dealing with adult language, as well as in computer generators which are not linked to a learning component, entrenchment does not have to be dealt with explicitly, nor will it figure in on-line computations, since every item in the resources (conceptual or linguistic) can be assumed to be well-entrenched. However, entrenchment is a crucial factor in language acquisition in children, as well as in gradual long-term change of a language as a resource.

3.2.7 Concreteness

Concreteness as a salience determinant reflects the observation that objects in highly abstract domains are in general less salient than those which are related more directly to sensory experience [32]. The simile *proud as a peacock*, alluded to under *association*, reflects a tendency to prefer concrete entities as objects of comparison in similes. Concreteness thus influences the direction in which association tend to develop. Most discussions of similes in the literature implicitly incorporate concreteness as a preference principle.

3.2.8 Speaker Motivation

In this subsection we will present three closely related determinants which capture the personal significance that a choice has for the speaker: *ego salience, speaker-motivation* and *relevance*.

Ego Salience

Ego salience does not admit precise characterization, but its role as a determinant is undisputed in psycholinguistics. Ego salience is related to the '*me first*' principle of Cooper and Ross [5] who studied word orders frozen into idiomatic conjunctions like *now and then* and *here and there* in which the conjunct expressed earlier is closer to the speaker's coordinates. Ego salience is also related to the *personal salience* of Sarason and Sarason [66] which directs attention to the particular elements of a situation which have a personal significance to the speaker.

Speaker Motivation

Closely related to ego salience is the *speaker motivation* of Osgood and Bock [58] which is the salience attributed to entities by speakers in order to entertain their personal concerns, goals and motivation. Speaker motivation triggers preference of

My car was stolen

over

Someone stole my car

Relevance and Goals

The preference of words and constructs as a process carried out deliberately in accordance with the speaker/generator's communicative goals is most naturally placed in this group. The link between choices in NLG and speaker's goals has been defined through the determinant *relevance* by the linguist Leech [33], and following Leech, by Haslett [23] in the field of communication. Note that this sense of relevance is entirely different from Ortony's sense which was discussed earlier as identical to diagnosticity.

Salience and Relevance

Leech [33] defines relevance as follows:

> An utterance is relevant to a speech situation if it can be interpreted as contributing to the conversational goals of the speaker or hearer.

Haslett [23] modifies this definition to account for the communicative actions and knowledge. Her definition of relevance is stated as follows:

> An utterance, action or unit of knowledge is relevant to a speech situation if it can be interpreted as contributing to the communicative goals of the speaker or listener.

We pointed out the distinction between salience and relevance in one of our early publications, [60]. Several researchers have sought to distinguish relevance from salience on the basis of whether the choice involved speaker's deliberation or volition, in service of fulfilling personal (communicative) goals.

Herskovits [25] notes that salience involves foregrounding of objects arising in our interactions with (and perception of) our environment, while relevance pertains to what the speaker *wishes* to express or imply in the current context. Herskovits notes that relevance 'becomes entangled' with the question of planning in NLG.

The same distinction is indirectly articulated by David Waltz in his work on understanding and generating scene descriptions [78]. Waltz notes that in generating scene descriptions, what the speaker notices (selects for description) is a function of *external factors* and *internal factors*. External factors are perceptual factors that attract attention, while internal factors include goals and desires.

Philosophical Issues in Brief

We are treading on complex terrain when we deal with issues of deliberation in NLG. We are not certain about all the mechanistic implications of the distinction between *intentional* (involving goals, deliberations, etc) and *non-intentional* (in the sense of not involving volition or deliberation to explain choice) factors influencing selection.

The significance of language as a vehicle for goal satisfaction is undisputed. However, we are drawn towards appreciating accounts (like [Osgood and Bock]'s) which include goals/deliberation as *one of several* potential determinants of choice of words and constructions. Such a view will be in harmony with NLG as a process of decision making under constraints, goals being one source of constraints. *Decision making* is used here in a sense which is neutral between volitional and non-volitional selection. In a theory which takes goals as forming the prime overarching input to NLG, one could end up positing an endless number of degenerate goals to explain or implement choices.

Protracted philosophical disquisitions, 'beyond a point', are considered unproductive in attempts to arrive at computational models. Yet, to some extent, we should be aware of existing philosophical differences and know our affiliation.

3.2.9 Humanness and Animacy

Fillmore [9] proposed humanness as a salience determinant. For example, consider

> (a) I hit Harry with the stick
> (b) I hit the stick against Harry

In (a) which is more preferable according to Fillmore, *Harry*, because of the determinant *humanness*, is expressed prior to *the stick* which is demoted to an oblique position as an indirect object.

Similarly,

> A Burnaby resident was struck by lightning last night

is more probable than

> Lightning struck a Burnaby resident last night

Extending out from out from *humanness* towards inanimate objects, *animacy* is considered a determinant of salience (for example, by Osgood and Bock [58]), and is implicated in preferences over inanimate objects. Accordingly,

> A deer on Burnaby mountain was hit by a car.

is more probable than

> A car hit a deer on Burnaby mountain.

3.2.10 Processual Determinants

We now turn to the last major group of determinants. They pertain to a dynamic conception of salience. The *processual* determinants refer to the increased availability of potential choices due to aspects of the generator's processing history, which include linguistic as well as perceptuo-cognitive processing.

The processual determinants we include here are:

- recency of mention
- frequency of mention

In linguistics, these determinants are recognized by Hajičova *et al.* in [18] and their later works. As discussed in chapter 2, the processual determinants play a role in pronominal- and definite reference generation.

In some contexts, the language generated will exhibit processual effects even when there is no explicit prior mention. Osgood [56] gives an example of two people walking along a street, and when they see "a particular female entity named Mary, wearing a mini-mini skirt", one tells the other: '*she also dyes her hair*'. Here, the pronoun *she* is licensed by presuming the high salience of *Mary* after cognizing the 'equivalent' of the sentence *Mary is wearing a mini-mini skirt*.

Sridhar [69] generalizes recency of mention to *recency of prior cognizing*, and calls the determinant *topicality*. Topicality is related to the *given/new* distinction of Halliday [20], but is distinct from it. Given/new is a discourse-related notion that explicitly brings in assumptions made by the speaker regarding what is present in the listener's consciousness, and as such, is a major salience determinant along the *interpersonal* dimension.

Discussion

Topicality, according to Sridhar, is a speaker-based notion, and what is topical for the speaker need not have been mentioned in prior discourse. Recall the spread of salience to *teacher* through association to *school garden* in the discourse (given in chapter 2):

(1) The school garden was full of children
(2) They talked noisily
(3) But the teacher did not restrain them

It is quite possible that in natural speech, some of the NLG effects based on spread of salience through association could cause problems for the hearer when knowledge is not shared, or when activation spreading patterns are different for speakers and hearers. Again, explicit assumption on shared knowledge along the interpersonal dimension will be required in listener-oriented models, and in all computer generators engaging in dialogue. Functional linguistic theories (Czech as well as systemic) treat such assumptions explicitly. Although we do not discuss interpersonal determinants of salience in any detail in this thesis, our strategy of identifying salience determinants as clearly as possible provides a natural place for extensions.

Most of the determinants we discussed here are perceptuo-conceptual, and can be grouped under the *ideational* domain in the systemic-functional enterprise (the other two systemic-functional domains being *interpersonal* and *textual*). The processual determinants just discussed are naturally classified under the *textual* domain.

The *interpersonal domain* will require explicit treatment of speaker-hearer interactions. Of particular significance to the modelling of interpersonal salience is the sophisticated pragmatic context in Ed Hovy's PAULINE system ([26]). We envision extension of salience theory along the now-neglected dimensions as not just worth while- but necessary developments in the future. Part of the reason for not including all the interpersonal aspects of communication at this stage is a methodological safeguard against having 'too many irons in the fire' at the start.

Moreover, our work is fuelled tremendously by psycholinguistic and cognitive questions. As could be discerned from this chapter, psycholinguists and cognitive psychologists have contributed the most to the existing (though scattered) discussions of salience. We intuit that an effort based on these discussions will blossom into fairly major developments for NLG as a computational problem.

Given the (now apparent!) pervasiveness of salience in NLG, extending existing linguistic theories with room for salience is a very vital task. However, while it is by no means premature to embark on the task now, to consign salience completely as an application area of *any* linguistic theory at this stage is shortsighted, and will mutilate psycholinguistic and cognitive interests in their infancy.

Chapter 4

Canonical Salience

Of the multitude of salience determinants presented in the last chapter, need not all be present in every instance of language generation. Scene descriptions could be generated, influenced primarily by perceptual salience; persuasive rhetoric, on the other hand, will involve a high degree of speaker-motivation. Depending on the determinants at play, the generation process could vary in its degree of automaticity or weight of deliberation. Even in the simplest of settings, however, several determinants could be at play.

Two of the major issues in bringing about salience effects in NLG are:

1. the computation of saliences of bearers given the determinants

2. combining the contributions of several (possible) determinants in arriving at the aggregate salience score (what we will call *effective salience*) of a bearer.

Once we arrive at the combined effect of all the salience determinants in the form of a single number (or a qualitative assignment of position on a discrete scale- like *high* or *low*), we can rank such salience assignments, correlate the ranking with available linguistic alternatives, and base the decision on simple optimality criteria that select the alternative corresponding to the maximum (strictly, optimum) effective salience score.

Before we address the nature of interactions, we should pay attention to the nature of the interactants. To be more precise, clues on how the scores combine to report an effective salience score could depend on where the scores come from. For an incipient theory of salience, the implications are that categories should be developed in the theory that accommodate the distinctions necessary for capturing the observed, alleged or intuited interactions. The categories, combined with psycholinguistic support, could help us channel our speculations towards models of interactions which are more subtle and farther-reaching than simple numerical models in which the contributions of the determinants are simply added up.

Assuming that the manner of combination of scores could depend on the nature of the determinants involved, the question we confronted next was: is any one of the determinants *privileged* in any sense? Given that any of a number of determinants could be at play during an instance of language generation, will we be successful in isolating a *privileged* determinant and confer special theoretical significance on it?

Note that the privilege we are seeking is one of theoretical status, not one of maximum priority in an NLG decision, for if the latter were the case, whenever the privileged (maximum priority) determinant is present, the decision will be the same. Moreover, among all the determinants we have seen in the previous chapter, despite the possibility of (say) speaker-motivation or vividness having maximum priority in specific decisions in specific settings, we have, at this stage, no convincing grounds for setting any one of them apart from the rest as privileged. If, on the other hand, we manage to be even half-correct in isolating a privileged determinant, the implications will be interesting.

At this stage, we directed our speculation as follows: if none of the determinants working in the context of a specific instance of generation is a-priori privileged, then how are choices exercised when *all* the salience determinants are *absent* from the context? That is, imagine an instance where the objects being talked about are all pallid, not particularly unique or pervasive, not mentioned recently, not having motivational significance for the speaker, and so on! In such a situation, are the choices which are usually ranked by salience equally available to the speaker/generator? The possibility we pursue in this chapter is that they may not be: there *could* exist preferences *in vacuo*.

4.1 Built-in Preference

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 the sentence

John gave a book to Mary

has a logical, 'transformed' equivalent

John gave Mary a book

in the presence of a built-in preference in the choice set

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

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.

4.2 Canonical Salience and Instantial Salience

The two major categories which, we propose, are most relevant in our study of salience in NLG are:

- canonical salience
- instantial salience

The inherent, built-in bias by which the generator tends to exercise preferences, considered as a determinant of salience, is termed *canonical salience* in our theory. We regard canonical salience configurations as most naturally associated with conceptual and linguistic knowledge, if not as their integral component.

4.2.1 Instantial Salience

Salience configurations which are set up in a specific instance or context 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 relative positions of *agent* and *object* need to be determined. In this instance, the entity *UFO* is more salient than the entity *boy* due to vividness. Equivalently, the *object* is more salient than the *subject* in the sentence, and triggers generation of the passive.

Instantial salience depends on the presence of other entities in the context: *UFO* is vivid relative to *boy.* It also depends on the presence of speaker-motivation or goals in context, and on dynamic (processual) determinants which, as the text is unfolding, may make entities salient for pronominalization and other decisions. Thus, the passivization in

My stereo system has been stolen

and in the second sentence of

He started arguing loudly, and had to be told to calm down

are licensed by the salience determinants in the instances, viz., speaker motivation and recency of mention, respectively.

4.2.2 Remarks

We are not in a position at this stage to state formal definitions of the categories *canonical salience* and *instantial salience*. Our major objective, rather, is to seek support and evidence which will help us arrive at such definitions. The better part of this chapter is devoted to shaping the identity of canonical salience with the aid of psycholinguistic, cognitive and analytical support.

The support is quite encouraging, as several works point to the existence of canonical salience. The strongest support we get is from Osgood's *naturalness hypothesis* discussed in [58] and [55]. We also bring to bear on our discussion in this chapter support from cognitive linguistics and formal linguistics. Considerations of both cause (the origin of canonical salience) and effect (expressive preferences due to canonical salience, recorded in linguistic literature) are brought in.

We do not *prove* the existence of canonical salience in the sense of formally deducing it as a logical consequence of the support gleaned. Rather, we seek rationale for the postulation of canonical salience as a useful category in the theory now for exploration later. The authors whose works we present and discuss in the remainder of this chapter are:

- from psycholinguistics: Osgood, [Osgood and Bock], Sridhar
- from AI/NLP: Herskovits
- from cognitive linguistics: Langacker
- from formal syntax: Falk

We conclude the chapter with a discussion.

4.3 Naturalness and the Perceptual Connection

The naturalness hypothesis of Osgood ([55], also presented in [58]) most directly addresses the question of origin of canonical salience. A major assumption of Osgood's theory is that the cognitive structures underlying the production and comprehension of sentences develop in prelinguistic, perceptual experience, and these structures are 'taken over' by the linguistic system as it develops later.

The basic structure of simple cognitions are of the form

$$relation(arg1, arg2)$$

In Osgood's work, such structures are called *tripartite* cognitions, since they have three parts: a relation and two arguments. Osgood then discusses two very basic cognitions which are crucial for the prelinguistic child: stative cognitions (relations) 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 relations,

$$figure - state - ground$$

is the natural order, and for action relations,

$$actor - action - recipient$$

is the most natural order.

4.3.1 Evidence for Naturalness

Osgood's Introspective Evidence: One usually perceived *actors* before *recipients*, with *actions* associated more with actors. Actors are mostly *human* and generally *animate*) and *mobile*, whereas recipient of actions are mostly inanimate and immobile, passive entities. The prelinguistic child learns to distinguish between agents and recipients or objects of actions, and finds agents more salient.

For stative relations, Osgood invokes gestalt-type evidence to argue that *figures* stand out from *grounds*, and are typically more *animate* and more (potentially) *mobile* than grounds. Thus,

The newspaper is on the sofa

is natural, while

The sofa is under the newspaper

is not natural.

Osgood defines naturalness as the ordering of constituents within clauses corresponding to the inherent salience of components of simple cognitive structures, as shaped by experience through perceptuomotor comprehension and behaviour during the prelinguistic stage. Natural preferences thus cultivated tend to get reinforced through perception and language use in the succeeding adult years.

The term *naturalness* is a bit misleading, since all types of salience (including those discussed in the last chapter), whether inherent or not, are equally natural in the sense of reflecting the underlying principle of NLG corresponding to each type. However, Osgood points out, naturalness is natural in the sense of reflecting structures developed during prelinguistic perceptual experience, and are, in that sense, special. Other salience determinants in a generation context *could* bring about effects which reflect disorderings of natural biases.

Cross-Cultural Studies: Osgood and Bock [58] report evidence from cross-cultural studies involving some 30 cultures around the world, corroborating that figural concepts are universally more salient than ground concepts, and that actors and instruments are more salient than objects and recipients.

Experimental Evidence: Using perceptual stimuli in a simple description task, Osgood [56] has experimentally verified that figures tend to get expressed earlier than grounds, as in

The ball is on the table

and sources of actions tend to get expressed earlier than recipients, as in

A rolling black ball hits a blue ball.

Other Evidence: According to Osgood's naturalness analysis, in bitransitive action sentences involving a dative recipient, the direct object is naturally expressed earlier than the dative object. For example,

The boy tossed the ball to the dog

is more natural than

The boy tossed the dog the ball

because the entity being transferred is part of the action relation, and is more tightly bound to the verb, so that we have

$$tossed\_the\_ball(boy, dog)$$

which is an instance of the tripartite action cognition

$$action(actor, recipient)$$

familiar to the prelinguistic child.

Osgood [55] reports cross-language linguistic data, evidence from language development in children as well as experimental evidence that support his analysis of natural ordering on bitransitive action sentences like the above.

4.3.2 Discussion

Naturalness is the most important determinant of ordering, since it is based on regularities in perceptual experience. The high theoretical status of canonical salience, viewed as incorporating naturalness in its nucleus, is in consonance with the intimate relation between language and perception, observed in detail by Miller and Johnson-Laird [48].

The strength of the naturalness principle could vary among the instances where it is found. For example, in

(a) The newspaper is on the sofa
(b) The sofa is under the newspaper

the asymmetry between the alternatives is total: (a) is very natural, while (b) is unacceptable, and is hardly ever considered as a possible choice by speakers (except in linguistic illustrations!).

On the other hand, in

(a) Dan Quayle gave a potatoe to the student
(b) Dan Quayle gave the student a potatoe

though (a) is more natural than (b), (b) is not unacceptable in the same way (b) was in the previous pair. The asymmetry of canonical salience is less stark in this pair, compared to the one above. Both alternatives are probable, with the more natural alternative (a) being more probable. (And both may be equally unacceptable for a different reason altogether.)

4.4 Sridhar's Perceptual Hypotheses

Sridhar [69] incorporates Osgood's naturalness as a source of his perceptual hypotheses on linear order in sentences. The hypotheses are reproduced below, from [69]. Note that Sridhar relates *perspective taking* ([9]) to linear order.

Hypothesis 1: (adopted from Osgood): In describing stative and action perceptions, people will tend to take the perspectives of the state and the sources of action, respectively. They will also tend to express figures and sources earlier in the sentence than, respectively, grounds and recipients.

Hypothesis 2: Natural psychological salience yields a continuum of expressive preference:

$$actions \succ changes\ of\ state \succ constant\ states$$

According to this, when a constant state changes, or when a state change involves an agent, speakers tend to attend to the more salient aspects of the situation. To quote an example from Sridhar,

A boy threw a rock through the window

is more likely (*ceteris paribus*)

A rock came through the window

is more likely than

The window broke

when the speaker is someone who witnessed a boy throw a rock through the window.

Hypothesis 3: Given that the primary perceptual receptors of the human organism are located atop and in front of the body, people will tend to describe (a) vertical arrays from the top down (that is, by locating objects at the top with reference to those below), and (b) horizontal arrays from the front to the back (that is, by locating objects near to the perceivers with reference to those away from them).

By this hypothesis, in the hypothetical situation of someone viewing a car parked 5 feet away, and a van 10 feet further away, the viewer is more likely to describe the scene as

There is a car with a van behind

than as

There is a van with a car in front

This hypothesis also underlies the assignment of high salience to the *in-front-of* relation in Jeff Conklin's GENARO system [3].

Hypothesis 4: The natural order for describing temporally related events will be one that retains the order in which the events are ordinarily perceived in non-linguistic experience.

This hypothesis predicts that

(a) John made breakfast and Mary went to work
(b) After John made breakfast, Mary went to work

are more probable/frequent than

(c) Mary went to work after John made breakfast
(d) Before Mary went to work, John made breakfast.

Notice than none of (a)-(d) are forbidden on grounds of grammaticality.

4.5 Herskovits on Spatial Prepositions

Herskovits [25], who investigated the semantics of spatial prepositions, affirms the centrality of the figure/ground asymmetry to perceptual experience, and links it to tendencies of choice in language generation, specifically, generating locative expressions involving prepositional phrases. Assuming that figures tend to get expressed prior to grounds, Herskovits focusses on factors that condition the choice of figures and grounds.

4.5.1 Figure/Ground Asymmetry

Talmy's definition of figure and ground: According to Talmy [71], the *figure* object is a moving, or a conceptually moving point whose path or site is conceived as a variable, the particular value of which is the issue. The *ground* object is a reference point, having a stationary setting within a reference frame, with respect to which the figure path or site receives characterization.

In a free description,

> The charter van is near the restaurant

is more natural than

> The restaurant is near the charter van.

If the location of the restaurant needs to be communicated intentionally, and if the van is the only suitable object available as the ground (say on the highway, at a rest area) then the need to have a fixed ground is fulfilled by the use of periphrases as in

> The restaurant is near where the charter van is.

Herskovits notes the figure-ground relation as a central example of an asymmetric perceptual relation. When two objects are involved in a figure-ground relation, one of the two possible role assignments is natural, while the converse, though logically possible, is often unacceptable. The degree of acceptability varies idiosyncratically.

In the pair

> (a) The house is near the church
> (b) The church is near the house

both (a) and (b) are equally acceptable (in a free description).

In the pair,

> (a) The van is near the restaurant
> (b) The restaurant is near the van

(a) is much more acceptable than (b).

Finally, in the pair

> (a) The cap is on the bottle
> (b) * The bottle is under the cap

(b) is completely unacceptable.

4.5.2 Remarks

We interpret Herskovits' discussion as strong support for the existence of canonical salience. with the more natural of the choices being more acceptable or more frequent. We are reminded here of the view of [38] of grammaticality as an *inescapable* consequence of the control structure that produces surface sentences. We propose a parallel view in this context, according to which we can regard the absence or infrequency of odd and unacceptable (though logical and syntactically well-formed) sentences in uncontrived, natural speech as an inescapable consequence of the strength of preference embodied in canonical salience.

4.6 Langacker's Analysis

Langacker [32] asserts that the figure/ground asymmetry is fundamental to relational predications. Noting that *the choice of figure and ground is not automatically determined for a given scene*, Langacker discusses factors which condition choice of figure and ground. Of specific interest to us is his observation that a particular figure/ground relation could be occasionally overridden (i.e., the assignments reversed). He notes that the notions of figure, foreground and focus of attention are naturally associated, but not always coincident.

With regard to bitransitive actions sentences with a dative object, Langacker affirms Osgood's position of naturalness. He analyzes

(a) He sent a letter to Susan

as more basic, and

(b) He sent Susan a letter

as an acceptable, but derived form. However, rather than analyze it from the production perspective (cf. Osgood's line of reasoning involving perceptual factors), Langacker focusses on the saliences of the images created by the sentences in the *comprehension* process. The notions of *path* and *state* are evident in both (a) and (b) above. In sentence (a), the path taken by the letter with Susan as goal is more salient, while in (b), the resulting state in which Susan possesses the letter is more salient.

In the pair

(a) The shortstop threw a ball to the fence

(b) * The shortstop threw the fence a ball

(a) is much more acceptable than (b), since a fence is more easily interpreted as the endpoint of a path than as the possessor of a ball.

Langacker thus ranks the preferability of alternatives on the basis of what aspects of the situation conveyed in the sentence emerge as salient ones in comprehending the sentences, and notes that a salience-based analysis permits and entails graded judgments of acceptability.

Langacker's discussions, on the whole, point to the existence of natural asymmetries of preference between linguistic alternatives, and the relation of this asymmetry to perceptuo-cognitive factors. These preferences need not be exclusive, and can be occasionally overridden.

4.7 Clues from Formal Linguistics

Falk [7] is among the earliest of formal linguists who considered factoring out ordering rules from constituency rules (that is, linear precedence from immediate dominance) in phrase structure based grammars. Falk, a student of Jane Grimshaw, worked within the X-bar convention.

He distinguishes between *absolutely essential order* rules and *preferred order* rules, the latter of which can be overridden. His discussion of preferred order is most relevant for our categories of canonical salience and instantial salience.

Considering examples of bitransitive sentences, the object NP normally precedes the PP, as in (a) below:

(a) Max donated the book to the library.
(b) * Max donated to the library the book.

(Note, by the way, that we are now referring to the ordered elements in the sentences in terms of syntactic categorial labels rather than perceptual or semantic relations- we are at a formal syntactician's homeground!)

However, the precedence pattern can be overridden, as for example, in the following example of a heavy-NP shift:

(a) ? Max donated the big green book with the blue pages and an orange bookmark to the library.
(b) Max donated to the library the big green book with blue pages and an orange bookmark.

The same consideration licenses extraposition in this statement of Sherlock Holmes from *The Adventure of the Musgrave Ritual*:

> A collection of my trifling achievements would no doubt be incomplete which contained no account of this very singular business.

Falk states:

> *"Preferred order is interpreted as being obligatory unless there is a good reason to violate it, where good reason will (at least in most cases) be extra-linguistic. Thus, preferred word order is a meeting point between grammar and other fields of cognition."*

From the NLG perspective, we interpret Falk as follows: Canonical salience naturally conditions preference among potential alternatives. This preference often shows a high bias. Instantial salience, arising in the NLG context, can occasionally override the preference voted by canonical salience.

Moreover, while *preferred order*, which allows only occasional overriding effects reflects canonical salience with a high bias towards the preferred alternative, ***absolutely necessary order*** corresponds to the extreme of canonical salience which 'blanks out' the other alternative, with no possibility of overriding. In the absence of an ordering constraint, both alternatives (assuming a binary choice system for expository convenience) are equally available for choice. Canonical salience can thus be *graded* in its bias.

4.8 Discussion

Canonical salience is privileged by virtue of being cultivated through experience during prelinguistic development (aided by innateness- [55]), and being wired into the adult linguistic system as a built-in preference. It is unchanging, except on long term basis when the language itself changes. It is closely linked to a number of notions current in linguistics and computational linguistics: *basic, preferred, default, natural* and *unmarked.*

4.8.1 Forms of Statement of Canonical Salience

The bias in canonical salience can be expressed in several forms, viz., in terms of:

- general perceptuo-semantic relations:

$$figure \succ ground$$

- semantic relations or roles:

$$agent \succ object$$

- syntactic constituents:

$$object\ NP \succ PP$$

and so on, where '$\succ$' = 'more salient than'.

It can also be expressed equivalently in terms of its *effects* in NLG, as, e.g.,

$$figure \prec ground, \quad agent \prec object$$

and so on, where '$\prec$' = 'precedes'.

4.8.2 Computationally Relevant Characteristics of Canonical Salience

In this subsection we summarize the characteristics of canonical salience which are relevant for the modelling stage, as well as for the later implementational stage of the salience theory.

- Canonicality is a graded notion; the strength of preference is rooted to various degrees. Correspondingly, canonical salience can be overridden to various degrees.
- The degree of canonical salience captures both degrees of grammaticality/acceptability and degrees of relative frequency. The less preferred could be perfectly grammatical, but rarer.
- The preference is highly automatized, corresponding to the built-in, 'force-of-habit' nature of the preference.
- Canonical salience itself is not under volitional control, though it interacts in context with other preferences which could be exercised voluntarily.
- It is inescapable, in the sense that every time a choice system is activated, its canonical salience 'casts its vote'. Factors in the generation context (instantial salience), whether they override canonical salience or not, interact with canonical salience, rather than bypass it.

4.8.3 Difficult Issues

Canonical salience compels us to confront the issue of the naturalness or basicness of linguistic constructions. Though the judgment is clearcut in many instances, ascertaining which of two given alternative constructions is more basic is a very difficult problem. Introspective evidence alone is not sufficient; interaction between linguists and psycholinguists seems necessary.

Moreover, Osgood and Bock [58] point out that the analyses of psycholinguists and linguists on questions of basicness have disagreed in several cases. Even within linguistics, issues of basicness are far from settled. For example, in the dative inversion pair (from [58])

(a) The boy tossed the frisbee to the St.Barnard

(b) The boy tossed the St.Barnard the frisbee

Osgood's naturalness analysis contends that *frisbee*, the direct object, is more tightly bound to the verb, and is part of the action relation:

$$threw\_frisbee(boy, St.Barnard)$$

which is an instance of the tripartite action cognition familiar to the prelinguistic child. Therefore (a) is more basic. Osgood and Bock say that the generative linguists have not agreed on which of these is more basic. Their paper was published in 1977, and the indications I have are that modern linguistic analysis agrees with Osgood's analysis, as I could discern from the examples of Pollard and Sag [63] and Falk [7].

On the other hand, with regard to active- vs passive voice, both linguists and psycholinguists concur that the active voice is the basic construction. Here, the natural salience of *agent* over *object* is clearly available for introspection.

In the genitive pair (from [58])

(a) The powwow was held at the wigwam of the Sitting Bull.

(b) The powwow was held at the Sitting Bull's wigwam.

linguistic analysis asserts that (a) is more basic (head precedes possessive complement in the noun phrase) while the naturalness analysis of Osgood contends that the converse holds. In a natural, perception-based cognition of *possessor-has-possession*, possessors are mostly human (and usually animate), while possessions are mostly inanimate (and usually not human). Hence *possessor* has greater salience than *possession* and (*ceteris paribus*) would linearly precede it in the utterance.

It seems, therefore, that when linguistic (introspective) analysis is not settled on the basicness issue, psycholinguistic and cognitive evidence can be brought to bear on the matter (as in the dative inversion case). When linguistic and psycholinguistic analyses disagree, further research involving interactions between linguists and psycholinguists is necessary.

With regard to the study of salience in NLG, we do not need to take a stand on basicness issues at this stage when many questions remain unsettled. We can be aware of existing problems, assume the existence of canonical salience as described so far, and proceed with examining its interactions with instantial salience.

Moreover, while canonical salience is widely present in linguistic choice systems, not every linguistic choice system necessarily displays an inherent bias. Choices *could* be equally available, and be conditioned solely by instantial factors in the generation context. This, however, is not a problem for modelling the interactions. The presence or absence of a bias in canonical salience can be modelled uniformly in all linguistic systems, as we will see in the next chapter.

Chapter 5

Salience Interactions in Phrase Choice

The determinants of salience can be of different origins, and can be varied independently. However, they follow the same general principles that link salience to NLG, viz., that higher salience implies higher likelihood of selection and earlier expression. We assume that the net result of interactions among the determinants at play during an instance of NLG would be evaluated for each entity as a single number which represents its instantial salience. The instantial saliences interact with canonical salience which represents the inherent preference of the generator. A decision is then reached on the basis of this interaction. It is to the nature of this interaction that we turn to in this chapter.

5.1 Interactions between Canonical Salience and Instantial Salience

As we noted earlier, in chapter 4, any one of the determinants can be present or absent in a particular instance. However, owing to its inherent nature, the canonical salience associated with a linguistic choice system will *always* be activated, every time the choice system is considered.

1. Therefore, we attach greater importance, at this stage, to an examination of the canonical-instantial interactions, rather than interactions among the instantial determinants.

2. We assume that all instantial determinants combine among themselves to evaluate a single number.

3. Any clues we get on the nature of interactions of any one specific instantial determinant with canonical salience are assumed to be representative of the interaction of instantial salience with canonical salience *in general*, whatever the constituent determinants are. (We will use the shorthand 'canonical-instantial interactions' from now on).

4. The characterization of canonical salience is applicable, in principle, to all types of NLG decisions (be they lexical selection decisions or linearization decisions) in which a built-in bias, of the kind we discussed in the last chapter, could exist in the respective choice system. Therefore, we conjecture that whatever we may learn, of the nature of canonical-instantial interactions in any specific NLG decision, is suggestive of canonical-instantial interactions in *all* NLG decisions in which canonical salience and instantial salience can indeed be observed.

5. Finally, if we succeed in obtaining a formal model of canonical-instantial interactions in any one specific NLG decision, it is potentially applicable to several other NLG decisions.

5.1.1 Two Major Types of Interactions

We presented several examples, in the last chapter, of instances where instantial salience 'overcomes' canonical salience. To recall one, in generating the sentence

A UFO was seen by the boy

1. Canonical salience dictates that the agent (the boy) be expressed prior to the object (UFO), i.e., that an active voice sentence be generated.
2. *A UFO*, because of its vividness, has higher instantial salience than *the boy*, and is therefore preferred for prior expression in a passive voice sentence.
3. Vividness overrides canonical salience, and causes generation of the passive sentence.

In contrast, in generating the sentence

President Bush watched a movie

both canonical salience and instantial salience agree in their tendency to express the agent prior to the object.

We will give an informal, preliminary definition of these types of interactions, as follows:

Synergistic interaction: interactions between canonical salience and instantial salience are said to be *synergistic* when they dictate the same effect.

Competitive interaction: interactions between canonical salience and instantial salience are said to be *competitive* when they dictate opposing effects.

These interactions have been investigated in psycholinguistics as *congruence* and *incongruence* respectively by Osgood and Bock [58], and later called *mutual reinforcement* and *mutual competition* respectively, by Osgood [57].

This chapter is devoted first to the psycholinguistic evidences we have on the interactions, and next, to the derivation of implications for modelling the interactions. We conclude the chapter with a proposal for a formal framework to model canonical salience, instantial salience and their interactions.

5.2 Evidence on the Canonical-Instantial Interactions

In this section we present some psycholinguistic evidence on canonical-instantial interactions. The first work we discuss, viz., that of Osgood [56], is an experiment conducted with perceptual materials, while the next one, viz., that of Osgood and Bock [58] is a sentence reconstruction experiment performed with linguistic materials. We present shorter discussions of several others, including some from linguistics.

5.2.1 Experiments in the 'Simply Describing' Domain

Osgood [56] reports a sentence production experiment, conducted with a series of perceptual demonstrations. The objective of the experiment was not particularly to observe salience interactions, but to investigate, in general, how non-linguistic (especially, perceptual) aspects of a situation influenced the form and content of sentences produced.

The demonstrations consisted of the manipulation of a variety of objects (balls of different sizes and colours, plates, tubes, etc) by the experimenter, standing at one end of a large table. The demonstrations included simple events like rolling balls on the table, causing them to collide, placing spoons and plates in the middle of the table, and so on. The adult subjects of the experiment were told to describe the perceptual event they saw, in a single sentence, *'that a hypothetical six-year old child, just outside the door and not able to see what is happening would comprehend'*. This was the *Simply Describing* domain, a 'naturalistic' approach, designed to minimize all the complicated variables that usually determine sentence production, except simple perception, and to encourage the use of simple ordinary language with no jargon. The sentences elicited were thus meant to be 'linguistic paraphrases' of perceptual events.

Osgood [56] reports a number of findings on different aspects of sentence production, like the use of determiners, modifiers, pronouns and so on, the details of which need not concern us here. In fact, Osgood, in [56], says nothing about salience. However, in a later paper pertaining to salience [58], Osgood re-analyzes the findings of his *Simply Describing* experiments from the viewpoint of *naturalness* (cf. chapter 4) and its interactions with vividness, and reports how *improbable* unnatural sentences are (unnatural = instantial salience overcoming high bias in canonical salience). Vividness is estimated from novelty, size, colour of entities, etc.

Of relevance to us are the following points:

1. In the absence of instantial salience (Osgood: perceptual vividness), the sentences produced followed canonical salience (Osgood: naturalness) overwhelmingly, for perceived relations which were stative (*the ball is on the table*) as well as action (*one squash balls rolls and hits another*).

2. When instantial salience and canonical salience were synergistic (Osgood: congruent), the sentences were ordered overwhelmingly in their canonical order.

3. When instantial salience and canonical salience were competitive (Osgood: incongruent), the results favoured canonical order. The 'competitive' situations involved novel or vivid entities in recipient or ground roles, which are canonically less salient than actor and figure roles respectively. For example, in an experiment involving the experimenter holding a big, bright blue ball, only 4 of the 26 subjects produced passive sentences. Even in demonstrations deliberately calculated to elicit passive sentences (*a very big orange ball being hit by a small black ball*), only 3 of 26 speakers produced passives.

4. Canonical salience had higher weight over instantial salience, in the sense that in spite of several deliberate attempts to create 'unnatural' orderings, the sentences were generated in canonical order.

5.2.2 Experiments of Osgood and Bock

The experiments of Osgood and Bock [58], involving a sentence reconstruction task, were designed to examine the role of vividness in sentence production. This extremely complex experiment is described in detail in [58]. We will not describe all its intricacies here. We will, however, sketch the essential elements of its design, to give an indication of how intricate psycholinguistic experiments can be. We then summarize its main findings.

The experiments used pre-designed sentences belonging to several types of optional transformations: dative, equative, genitive, passive, etc. Vividness of the NPs in clauses was manipulated by designing sentences with all possible permutations of relative salience (vividness) configurations for both 'basic' and 'transformed' sentences for each group: namely, H(igh)/H(igh), H(igh)/L(ow), L/H and L/L for first/second constituent, totalling four. A set of four such for the U(ntransformed) sentences and four for T(ransformed) sentences made up a total of 8. Here is the clause set used for the dative group (from Appendix B of [58]):

U-H/H: The boy tossed the frisbee to the St.Barnard
U-H/L: The boy tossed the frisbee to the dog
U-L/H: The boy tossed the ball to the St.Barnard
U-L/L: The boy tossed the ball to the dog
T-H/H: The boy tossed the St.Barnard the frisbee
T-H/L: The boy tossed the St.Barnard the ball
T-L/H: The boy tossed the dog the frisbee
T-L/L: The boy tossed the dog the ball

Eight such sets were written for each of the 8 sentence (transformation) types.

Next, a 'couple set' was composed, consisting of 'half sets' of sentences from the original sets, with each half set belonging to a *different* clause type. An example of such a couple set of half sets is:

- Half set #1 (dative inversion):

The maharajah gave a rug to the queen
The maharajah gave a rug to the visitor
The maharajah gave the queen a rug
The maharajah gave the visitor a rug

- Half set #2 (phrase conjunct reversal):

Tornados and high winds devastated many areas
Storms and high winds devastated many areas
High winds and tornados devastated many areas
High winds and storms devastated many areas

Each half set was composed by selecting at random two untransformed clauses from the original set of 8 clauses, and then their corresponding transforms.

The half sets put together in a couple set were such that it was possible to make up a non-bizarre complex sentence out of them, consisting of one clause from each half set and a conjunction. A list of conjunctions was put together for the experiment.

Next, 'clause pairs' were composed out of each couple set, with the constraints that: (1) each clause in the pair was from a different half-set in the couple and (2) it would be possible to generate a complex sentence using them and a chosen conjunction such that the clauses could appear *in either order* as subclauses of the complex sentence.

The clause pairs were typed on IBM cards (those days!). An appropriate conjunction for each pair was also typed on an IBM card, and placed in a deck immediately after the card with the clause pair.

Sentence reconstruction experiment: Subjects were shown the IBM cards with the clause pair and the conjunction, and told to make up a sentence using the two clauses in the pair and the conjunction, with the conjunction free to appear either between the two clauses or at the sentence beginning (for all except *and* and *but*). Next, the subjects were told to create a plausible 'story' in the form of a short description, using their complex sentence as the last sentence.

The last sentences of the stories were then analyzed, mainly to test the prediction that the more salient (vivid) constituent is expressed earlier in the sentence. In particular, the ordering of constituents in the subclauses of the complex sentences in the story was compared with the ordering as it appeared in the respective IBM card, shown to the subject at the start, to determine if there was any tendency *to shift* the order in the reconstructed sentence.

For example, if the clause on the IBM card had an *active* voice sentence with a vivid *object* and a pallid *agent*, which goes against the 'vivid constituent first' hypothesis, the experimenters were interested in checking whether the story generator showed a tendency to produce it as a passive *sentence* in the reconstruction, with the vivid entity in the earlier position.

Ignoring errors and omissions, the reconstructed clauses were categorized as *corrects* and *shifts*, the former preserving the constituent order on the IBM card, and the latter shifting the order and producing the alternative syntactic form.

Subsidiary experiments: Two additional experiments were run, involving other subjects giving *preference ratings* for the clauses and *vividness ratings* for the constituents. In the former, the subjects were shown pairs of clauses, consisting of an 'untransformed' clause and its transformed equivalent, and were asked to indicate which of the two 'sounded better', and which one they would use in a situation in which either was appropriate. In the vividness ratings experiment, the subjects were shown the clauses used in the sentence reconstruction experiment, and were asked to rate the vividness of constituents on a seven-point scale. The authors used these ratings mainly to confirm their intuitions on their own a priori ratings of vividnesses as H and L, and (regrettably) do not report examples of the numerical scores in the paper.

Results: The experiments were beset with several complications, such as verbatim-recalls of original sentences by the subjects, despite the authors' efforts to minimize them. The statistical variance in the results was somewhat high. Yet, several the experiments permitted drawing several sound conclusions. Here is a short summary of the results:

- The percentage number of shifts for L/H (second constituent more vivid) was in general higher than for H/L, indicating a general tendency to place the more vivid constituent earlier.
- Clauses in the canonical order (natural form) were recalled *correctly* (no shifts) more frequently than the 'non-basic' transformed clauses.
- Overall conclusions: Both naturalness and vividness had independent effects on ordering in clauses. They combined to produce then general trends predicted by the salience principles (viz., more salient element in earlier position). Naturalness had greater weight. Moreover, vividness effects were not restricted to the familiar active-passive case, but appeared general to all the linguistic choice systems (clause types) used in the experiment. This particular conclusion was a clear advance in psycholinguistic research when it was published, since most previous experimental research was restricted to the study of passives.

5.2.3 Other Evidence

Sridhar's Cross-Linguistic Experiments

Sridhar [69], a student of Osgood, conducted elaborate cross-linguistic experiments using a filmed version of the perceptual demonstrations of Osgood [56]. Sridhar's experiments were not designed to test the nature of canonical-instantial interactions. However, the conclusions of the experiments supported the cross-linguistic validity of Sridhar's salience hypothesis that entities rendered salient by virtue of vividness, speaker motivation and topicality tend to be displaced from their canonical positions for prior (earlier) expression in the sentence.

Note that Sridhar's interest was restricted to the possibility of the displacement of a canonically low-salient entity towards a sentence-initial position, as may be caused by instantial salience (for example, the generation of passives because of instantial salience of the *object* entity). In other words, Sridhar does not address synergistic canonical-instantial interactions, but rather those cases of competitive interactions in which instantial salience tends to 'win out'.

Sridhar also tested the cross-linguistic validity of his perceptual hypotheses (cf. chapter 4) and pragmatic hypotheses (not discussed here). The film was used in experiments with 10 language-culture communities around the world, and shown to native-speaking undergraduate students who produced sentences describing the demonstrations in the film.

Studies of Passives

We cite two experimental findings as relevant to our study:

1. Flores d'Arcais [13] provided a set of pictures to his subjects, and asked them to describe the pictures using an active sentence or a passive sentence, giving either the name of the agent or the object as the cue word, thus increasing its *topicality*. He found that the latency of passive production decreased when the object of the action was given as the cue, suggesting that increasing the topicality (salience due to recency of mention) of the object facilitates use of passives.

2. Johnson-Laird [28] performed an experiment in which subjects drew diagrams to illustrate pairs of sentences such as

   (a) Red follows blue
   (b) Blue is followed by red

   He found that the area assigned to the earlier colour in the sentence was typically larger than that for the later colour.

Linguistics

Research on linear precedence in linguistics, most notably those of Uszkoreit ([74], [75]) and Pollard and Sag [63], recognizes the presence of competing determinants of linear precedence. The competition is discussed in terms of conflicts among ordering rules in the grammar. The elements ordered in these rules are syntactic constituents like NP, PP, etc. However, the factors determining order among these constituents can refer to thematic roles, and (what they call) discourse phenomena like (what they don't define) focus.

We cite here one instance of an ordering phenomenon discussed in [63], which sheds light on the presence of competitive canonical-instantial interactions. The sentences

(a) * Kim put on the table the book
(b) ? Sandy gave to Kim the book

are unacceptable ((a) more so than (b)), because they violate their 'obliqueness' linear precedence constraint. However, the sentences

(a) Kim put on the table the book he bought in Vienna
(b) Sandy gave to Kim the book she bought in Vienna

are acceptable.

The explanation of Pollard and Sag here is that the element *the book (s)he bought in Vienna* is *focussed*, and is expressed last in the sentence. (Our commentary: This sense of *focussing* corresponds to the intentional use of *recency of mention* by the speaker to achieve salience in the listener's mind.) Pollard and Sag then posit a *focus rule* (designated $LP3$), stating that a focussed element which fails to precede some more oblique sister constituent should not count as a violation.

We interpret this concession in terms of salience as follows: The canonical bias in expressing the object of an action prior to other non-agentive semantic roles is mirrored in English syntax in the core of Pollard and Sag's obliqueness hierarchy, by which NP complements precede PP complements (their $LP2$ rule). This bias is quite strong. However, it is possible for instantial salience to overcome the canonical bias in competition, so that the speaker expresses a peripheral non-agentive (in this example, locative) role prior to the object.

The effect of this is mirrored in the environment of syntactic constituents of the sentence by a PP (unusually) preceding an NP complement. Sentences so generated can be considered 'acceptable'.

Pollard and Sag are interested in grammaticizing linear precedence regularities in English by positing an obliqueness hierarchy which captures as wide a set of observed regularities (in linear order) as possible. The obliqueness hierarchy starts out (in the course of its theoretical development), by incorporating syntactic, language-specific precedence regularities like head-initialness in English, for which perceptuo-cognitive motivations are not needed, as well as unproblematic regularities discernible in canonical biases for which perceptuo-cognitive explanations are available. As more exceptions to the rules of obliqueness hierarchy are encountered in the data, the immutable rules of obliqueness hierarchy flex themselves by accommodating new rules in the hierarchy at a lower level, thereby preserving their own higher strength, while conceding a positive grammaticality judgment even when they themselves are violated. We may say that the obliqueness hierarchy *partially* mirrors canonical salience in syntactic environments, and the part corresponding to canonical salience is obscured when several other disjunctively-applicable conditions arising for diverse reasons are added to the hierarchy, along with enumerations of the conditions under which the lower (weaker) rules can bypass the immutability of higher rules.

5.3 Prelude to Formalization

In this section, we summarize the available psycholinguistic support, verbalize our extrapolations and suggestions on our road to a formal model of interactions, and conclude the section with a proposal for the same.

5.3.1 Psycholinguistic Support

In this subsection, we summarize the available clues on interactions, classifying them as *strong* and *weak*, and gradually derive their implications for our (as yet not existing) model. The citations in the brackets are authors whose findings imply or suggest the corresponding point.

(1) (*Strong*): In the absence of a bias in instantial salience, the linguistic choice is determined by canonical salience. [56] has evidence for this when instantial saliences are all low (all pallid entities). We extrapolate this to be valid even when the entities in the relative instantial salience configuration are *equally high* in vividness. It is well-established that vividness needs a contrast to be able to have an impact (cf. chapter 3).

(2) (*Strong*): Instantial salience and canonical salience interact with each other, rather than bypass each other (when they win in competition). The combined effect of the interactions determines which linguistic construct is chosen, whether they interact synergistically or competitively. ([56], [58]).

With the aid of this point, we visualize the interaction as a process by which canonical salience and instantial salience *modulate* each other. We call the outcome of this process an *effective salience configuration*, on the basis of which decisions are taken. *Effective salience* reflects the combination of canonical salience and instantial salience. Moreover, in competitive interaction, the effective salience configuration will be a shuffle of either one of the interacting configurations. (When there is competition, there is only one winner). This manner of combination can be easily modelled using the tools of arithmetic. Therefore, we visualize the degrees of canonical and instantial saliences as numbers, and effective salience as a numerical function of canonical salience and instantial salience. In other words, *the interactions are quantitative*.

(3) *Strong*: When the canonical-instantial interaction is synergistic, the outcome of the decision coincides with the dictates of canonical salience [56]. This is a somewhat trivial observation.

(4) *Strong*: When the interaction is competitive, a higher relative instantial salience does not necessarily imply higher relative effective salience. (For instance: higher vividness does not guarantee earlier positioning) [58].

We relate this aspect to the *degree of bias* in canonical salience. We visualize instantial salience as being forced to surmount a certain critical threshold in order to keep the effective salience configuration in its favour, unshuffled (i.e., to win) Since canonical and instantial salience are the only interactants, the 'critical threshold' for instantial salience is set up by canonical salience!

(5) *Weak*: Suppose that canonical salience has a high bias towards A in the context $\{A, B\}$. Then, in a competitive interaction, increasing the instantial salience of B *facilitates* the process of B gaining more effective salience, and thereby facilitates the process of B winning [13]. The evidence for this point is not as strong as for the other points listed earlier.

Osgood and Bock [58] note that the *magnitude* of vividness difference between the constituents of a clause, as rated by their subjects in a subsidiary experiment, was *not* systematically related to a higher number of correct recalls for H/L sentences or shifts for L/H sentences (H/L and L/H indicate vividness of constituents appearing in that order in the clause. H = high; L=low).

We suggest that this may have been due to the lack of a quantitative model of interactions which the state of art in experimental psycholinguistics did not (and perhaps does not) permit. Moreover, Osgood and Bock do not discuss or point out the possibility that *the bias in naturalness can be of different degrees in different clause types*- an aspect which can be captured in a quantitative model easily.

Finally, the relation between vividness (as perhaps measured in an experiment) and the instantial salience induced by vividness may not, and need not, be a linear one. If vividness is tacitly taken to be synonymous with, or identical to, instantial salience, a linearly-proportional increase in salience is anticipated when vividness is increased. However, if instantial salience is taken as a (monotonically increasing) *function* of vividness, then the possibility of a non-linear relation is opened up. In particular, we may be able to model a *law* *of diminishing returns* (as in $y = \sqrt{x}$), in which at higher levels of vividness, the increase in vividness is less perceptible, or a *law of increasing returns* (as in $y = x^2$) in which the increase in vividness is more perceptible at higher levels, or some combination of the two. In all cases, the function can still be kept monotonically increasing, reflecting the correlation between higher vividness and higher salience. In short, we suggest a quantitative model of the dependence of salience on its determinants!

(6) *Strong*: In competitive interactions, instantial salience, with a sufficiently high bias, *can* manage to overcome the bias of canonical salience, assuming that canonical salience does permit competition and allows room for its bias to be overridden ([58] and several others).

5.3.2 Empirical Support

Osgood and Bock note that the effects of vividness was very noticeable for passives in their experiments. They conjecture that a possible reason for this effect may be the relative infrequency of the passive voice. Jan Svartvik's extensive study of passives, which involved a frequency count in a large corpus, reported that the passive voice was only one-seventh as frequent as the active, and over 70 percent of the passives were agentless. Osgood and Bock take the abundance of the agentless form among passives as testifying to the uninteresting (we: considerably less salient) nature of the logical subject when passive forms are employed.

Coupled with the observations of naturalness having higher weight over vividness, we interpret the remarks of Osgood and Bock on passives as implying the following:

In the $\{active, passive\}$ choice system, canonical salience has a high bias towards choosing *active*. The canonically-less-salient *object* cannot get expressed sentence-initially simply by having higher instantial salience than *agent*. Otherwise, passives would be more frequent than they are. However, a very high relative instantial salience of *object can* succeed in triggering a passive, by enabling *object* to surmount the threshold set up by the canonical bias towards *agent*.

Osgood and Bock suggest a tantalizing connection between built-in preferences having perceptuo-cognitive significance (canonical salience) and *relative frequency* of constructs in text, offering a hint on a possible way of estimating or *measuring* the degree of canonical salience!

A Digression on Salience and Ellipsis Interestingly, we also notice a connection between deletion/ellipsis phenomena in English and salience. There is a co-dependency between the uninterestingness of agents and the opportunity that objects have to gain a very high salience over agents. On the one hand, if agents were any more interesting, objects would not stand a chance of surmounting the canonical privilege of agents. On the other, as a result of objects gaining a very high salience because of vividness or other determinants, agents are rendered uninteresting, and slip to a later position in the sentence, or more frequently, get omitted.

There may be a number of situations in which the uninterestingness of agents can be a priori agreed upon as a conventional judgment, justifying a heuristic to select an agentless passive form, as pointed out by McDonald in a very recent paper: in the example (from [39])

> Don E.Miller was named senior vice president and general counsel

McDonald points out that we do not need to be told who did the naming of Mr.Miller: this is a conventional default judgment, licensing generation of the (agentless) passive. This example also admits a parallel, salience-based analysis, for if President Bush had nominated Mr.Miller, we might have obtained a sentence such as

> The President named Don E.Miller senior vice president and general counsel

unless the speaker intentionally uses recency of mention as a device to achieve salience of the President in the listener's mind, as in

> Don E.Miller was named senior vice president and general counsel
> by none other than President Bush himself.

A similar effect is noticeable in *pleonastic* verbs in which a specific entity that can function as some argument of a verb is so much a part of the meaning of the verb that to mention it would be redundant, as in

> I *drove* to the university *in a car*

in which *in a car* is redundant. However,

> I *drove* to the university *in a stolen antique car*

is quite a bit different. It will be interesting to explore the connection between salience and ellipsis/deletion in greater detail.

Another Digression on Usage Observe that while instantial salience is most naturally thought of as a property of specific entities, canonical salience, being inherently associated with linguistic and conceptual knowledge of the speaker, is available in a more 'parametric' form over the possible specific entities that are spoken about (*agent, object*, etc), or over the constructions in the choice system to which the salience ranking is transferred (*active, passive*, etc).

During the interactions when effective salience is evaluated, there is also a process of diffusion of identities, which equates or correlates the specific entities which are bearers of instantial salience, and the abstract, parametric entities which are bearers of canonical salience. This is an extremely important mechanism for implementing the interactions. Perhaps regrettably, we will say no more about this mechanism in this thesis, since our current focus is on modelling the interactions, not on the details of lower-level technicalia. We assume the necessary mechanism to be unproblematic, and refer to the bearers of salience interchangeably with their various identities as, for example, in:

- "instantial salience of Mr.Miller" (a specific entity)
- "effective salience of *agent*" (a semantic role)
- "canonical salience of active voice" (a linguistic construction), and so on.

Indeed, the mechanism by which items of salience information stated at different levels of abstraction come together in canonical-salience interactions is part of the generation process proper.

5.3.3 The Penultimate Steps

We recapitulate our suggestions pertaining to the quantitative nature of the canonical-instantial interactions:

1. Canonical salience and instantial salience are expressed as numbers.
2. Effective salience of an entity is a numerical function of its canonical salience and instantial salience.
3. If canonical salience has a bias, instantial salience has to surmount the threshold set up by canonical salience to win in competition.

Following (1) above, let us consider two entities x_1 and x_2, with associated canonical saliences p_1 and p_2 respectively and instantial saliences u_1 and u_2 respectively.

Following point (2) above, designate the effective saliences of x_1 and x_2 as e_1 and e_2 respectively. We can write:

$$e_1 = f(p_1, u_1) \; and \; e_2 = f(p_2, u_2) \tag{5.1}$$

Point (3) above can be modelled by taking f to be a product function. Here is a simple proposal:

$$e_1 = p_1 \cdot u_1 \; and \; e_2 = p_2 \cdot u_2 \tag{5.2}$$

In competitive interaction, if $p_1 > p_2$, then u_2/u_1 should exceed p_1/p_2 in order to bring about $e_2 > e_1$.

5.3.4 Quest for a Formal Framework

We are seeking to place the canonical-instantial interactions and their effects on word/phrase selection in NLG in a formal framework of decision making. Notice that until now, we have made no commitment to any established or standard formal theory in vogue, for the research was fuelled primarily by curiosity about a hitherto-unformalized phenomenon, clues about which lay scattered in diverse disciplines.

Formalization is however a desirable 'ultimate' goal, and can reveal to us a tremendous lot more about the phenomenon that interests us. At this stage, we feel that time is ripe for us to suggest a formal framework as a vehicle to convey salience and its interactions in NLG decisions. When we began our research, we had no idea that we would arrive at this juncture!

5.3.5 A Proposal

Our proposal consists of several interrelated parts, all of which fit together in one coherent framework. We present the proposal below:

1. Canonical salience is interpreted as a built-in preference which is not under volitional control. It is always exercised every time the choice context is activated. In interacting with canonical salience, the preference dictated by instantial salience may or may not be satisfied in the linguistic decision. Therefore, *canonical salience can be modelled as a probability distribution constituting an uncertain environment.*

2. Instantial salience is interpreted as a strength of preference arising in the context or situation of NLG decision making. Therefore, *instantial saliences can be modelled as utilities.*

3. *The dependence of instantial salience on its determinants can be modelled as a utility function.*

4. *The process of a speaker attempting to satisfy, or responding to, motivational factors encapsulated in instantial salience in the presence of canonical salience can be modelled as a process of decision making in an uncertain environment.* ("attempting to satisfy or responding to" allows room for different degrees of deliberation in the instantial factors.)

5. (As a consequence of (1)-(4),) *Choosing a word or a phrase on the basis of maximum effective salience can be modelled as a process of maximizing expected utility.*

After a brief overview of decision theory, these proposals are fleshed out further.

5.4 An Overview of Decision Theory

Decision theory deals with three essential aspects of decision making:

1. Uncertainties associated with the outcomes of a decision

2. Preferences for the possible outcomes of a decision

3. A rational criterion for decision making which balances considerations of uncertainties of outcomes with preferences for possible outcomes.

5.4.1 Decision Making under Certainty

Let us first consider a decision making scenario in which uncertainties are absent: actions lead to *certain* outcomes. Associated with each outcome is a *utility*, which represents the degree of preferability of the outcome: the higher this ranking is, the better.

The best decision is the one which leads to the outcome with *maximum utility*. It is not necessary for the utilities to be quantitative: decisions can be reached on the basis of preference ranking among the outcomes expressed qualitatively, in terms of 'ordinal' utilities: e.g., *high, low*, etc.

5.4.2 Decision Making under Uncertainty

In most real world situations, however, actions are carried out (= decisions are taken) in environments in which their outcomes are uncertain. The decision maker, however, may know the possible outcomes of an action, and may be able to estimate the probabilities of the outcomes (*action* and *decision* are used synonymously).

In this scenario of decision making under uncertainty, the decision maker takes into account both the *probabilities* of the outcomes of an action, and the *utilities* of the outcomes which reflect their preferability. Figure 5.1 shows the basic structure of decision-theoretic choice under uncertainty.

In figure 5.1, the square node is the *decision node* at which alternative actions $a_1, a_2, \ldots$ are available. Outcomes of actions are governed by uncertainties at the circular nodes called *chance nodes*. An action a_i leads to an outcome C_{ij} with probability p_{ij}. Utility u_{ij} represents preferability of outcome C_{ij}. What is the decision criterion in this scenario?

On the basis of a number of axioms on *rational* decision making, a theorem is derived in decision theory, stating that the most rational action is the one which *maximizes expected utility*. Expected utility of an action a_i is defined as:

$$EU(a_i) = \sum_{j=1}^{m} p_{ij} u_{ij} \tag{5.3}$$

Accordingly, in figure 5.1, action a_1 is preferred over action a_2 if

$$\sum_{j=1}^{m} p_{1j} u_{1j} > \sum_{j=1}^{n} p_{2j} u_{2j}$$

It is to be noted that figure 5.1 is not a state transition diagram for a process unfolding in time; it merely shows the relation between decisions and factors influencing the decisions. In many complex real-world problems (e.g., a city government's decision to curb pollution) in which decision analysis (an offshoot of decision theory) is applied, the probabilities and utilities are ascertained with extensive analyses. Once the decision inputs are all available, the decision recommended by the theory is carried out over time, perhaps

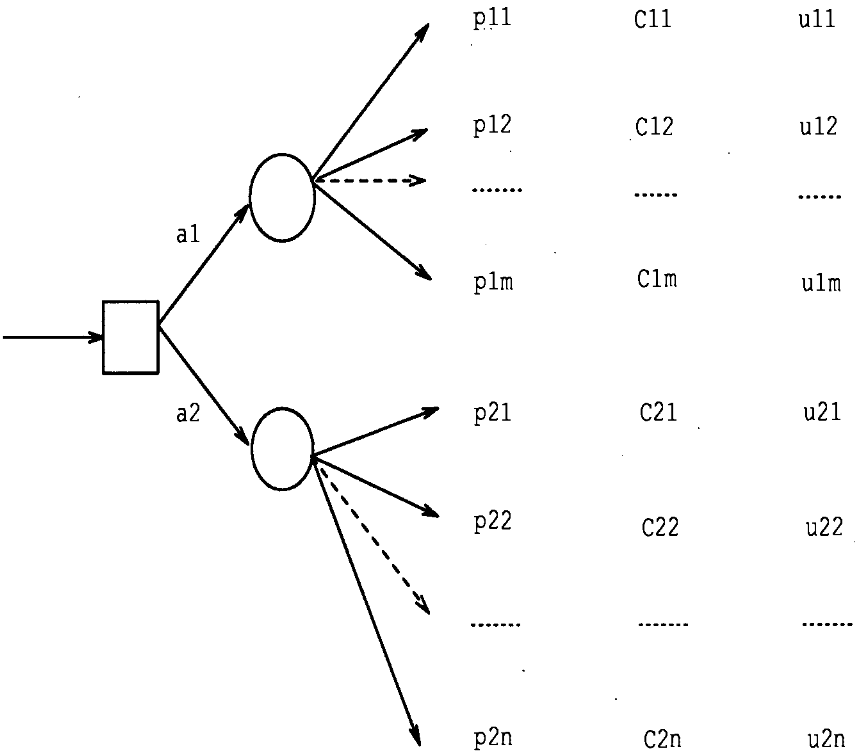

Figure 5.1: Decision Making with Probabilities and Utilities

over a long duration, detached from the initial steps of problem structuring, analyses of decision inputs and the application of decision theory to reach the decision.

We believe that the usefulness of decision theory extends beyond situations where human decision makers have the time to carry out statistical and preference analyses and thus *know* all the numbers involved before reaching a decision consciously for subsequent execution. We feel that as a very general theory of rational decision making, decision theory is applicable in principle in all probabilistic situations governing outcomes for which preferences are assigned, as depicted in figure 5.1, regardless of whether the decisions are carried out 'on-line', or spanning a long duration.

Utility Functions

The utilities u_{ij} are seldom available a priori as a ready number. They are computed by *utility functions* that relate the strength of preference for an outcome with the determinants of preference (called *attributes* in decision analysis). Multiattribute utility theory deals with the combination of preferences stated along several dimensions into a single utility. A lucid overview of several mathematical aspects of decision theory can be found in Luce and Suppes [34].

5.4.3 A Historical Note

The origins of decision theory can be traced back to the 16th century ideas on probability and value (utility), when they were applied to explain human economic behaviour, and was thus the province of economists and psychologists. Since then, decision theory has diverged from psychological questions, and has been developed extensively in mathematics as a *normative* theory of rational decision making. An axiomatization of decision theory may be found in Savage [67]. Decision theory is closely related to game theory (von Neumann and Morgenstern [76]).

von Neumann (a mathematician) and Morgenstern (an economist) contributed significantly to the modern (post-World War II) developments in decision theory. They underscored the subjective nature of probabilities and utilities, and replaced ordinal (qualitative) utility with *cardinal* (quantitative) utility. According to von Neumann and Morgenstern, while qualitative utilities are adequate in *certain* (deterministic) environments, quantitative utilities are necessary for rational behaviour in uncertain (probabilistic) environments. [Note: Let us recall our rationale for representing canonical salience and instantial salience as numbers, and our arguments for a quantitative approach to salience interactions! In total absence of canonical biases in linguistic choice, a qualitative ordering of instantial saliences will suffice. With the presence of canonical biases, however, comes the possibility of interactions we just discussed, pointing to a quantitative approach. The line of thought that led me to a quantitative approach to salience in NLG having such an eerie resemblance to [von Neumann and Morgenstern]'s arguments favouring cardinal utility has been one of the most thrilling discoveries for me in this research.]

5.4.4 Decision Analysis

Decision theory has given birth to the field of decision analysis which applies decision theory to help human decision makers in real-world situations. Decision analysts recognize the discomfort that many people have in working with numbers, and offer lengthy admonishments against using the numbers frivolously. They have developed sophisticated techniques for *constructing* utility functions, on the basis of human judgments on preferences for consequences and tradeoffs among objectives. Some of the good texts available on decision theory and decision analysis are: Keeney and Raiffa [30], Yu [80] and von Winterfeldt and Edwards [77].

Decision Analysis and AI

Decision analysis is being domesticated in AI by the '*uncertainty*' community. One of their chief interests is in network representations for structuring probabilistic and decision-theoretic reasoning: influence diagrams, Bayesian networks and so on. They are also interested in applying elements of decision analysis/theory to AI problems, most notably, expert systems. Henrion *et al.* [24], which appeared in the Winter 1991 AI Magazine, is a good overview article containing an extensive bibliogaphy.

The *uncertainty* community is neutral between the *subjectivist* (Bayesian) and *frequentist* (empirical) interpretations of probability, allowing designers to express their intuitions and to conduct experiments with human subjects, at the same time permitting collection of appropriate data in service of estimating the probabilities.

5.5 Salience Interactions in Decision-Theoretic Terms

In this section we first show how effective salience in salience-induced selection between active voice and passive voice is obtained as expected utility in decision theory. Next, we illustrate the decision criterion for a set of example sentences. The criterion is then stated in general terms for all binary linguistic choice systems in linearization. Important, psycholinguistically-relevant characteristics are then discussed.

5.5.1 Effective Salience as Expected Utility

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

(1). The uncertain 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 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. We have the condition:

$$p(Ag) + p(Ob) = 1$$

(2). The instantial saliences of Ag and Ob are given as $u(Ag)$ and $u(Ob)$ respectively. In an instance of NLG, 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 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, given knowledge 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:

$$S = \{\mathcal{A}, \mathcal{O}\}$$

where

$$\mathcal{A} = u(Ag) \geq u(Ob) \; and \; \mathcal{O} = u(Ob) > u(Ag)$$

(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: $(Ac, \mathcal{A}), (Ac, \mathcal{O}), (Pa, \mathcal{A}), (Pa, \mathcal{O})$.

(7). Preferences for these outcomes are expressed as utilities. Utility of action $a \in A$ in state $s \in S$ is represented as $U(a, s)$. In the case under consideration, we have:

- $U(Ac, \mathcal{A})$ = utility of expressing the agent before the object when $u(Agent) \geq u(Object)$, i.e., when the agent is no less salient than the object in the NLG context.

- $U(Ac, \mathcal{O})$ = utility of expressing the agent before the object when $u(Object) > u(Agent)$, i.e., when the object is more salient than the agent.

- $U(Pa, \mathcal{A})$ = utility of expressing the object before the agent when $u(Agent) \geq u(Object)$, i.e., when the agent is no less salient than the object.

- $U(Pa, \mathcal{O})$ = utility of expressing the object before the agent when $u(Object) > u(Agent)$, i.e., when the object is more salient than the agent.

We may regard these as a form of *conditional* utilities. However, decision theorists have not developed a theory of conditioning for utilities (in a manner parallel to conditional probability).

(8). We assign values for these utilities on the basis of knowledge of instantial saliences, and the principle of tending to express more salient element earlier. These utilities are independent of the probabilities in canonical salience.

- $U(Ac, \mathcal{A}) = u(Ag)$

- $U(Pa, \mathcal{A}) = 0$

- $U(Ac, \mathcal{O}) = 0$

- $U(Pa, \mathcal{O}) = u(Ob)$

Another common way of conceiving of utilities is as *rewards* associated with actions. When the agent is more salient than the object, the reward associated with expressing the object sentence-initially is 0, since the action goes against the principle of expressing the more salient entity earlier. That is, $U(Pa, \mathcal{A}) = 0$ (above). The reward associated with expressing the agent earlier is positive, and is given by $u(Ag)$, the instantial salience of the agent. A similar line of reasoning justifies the third and fourth utilities in the assignment above.

(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:

$$
\begin{aligned}
&EU(Ac) = \\
&\quad p(Ag) \cdot U(Ac, \mathcal{A}) + p(Ob) \cdot U(Ac, \mathcal{O}) = \\
&\quad p(Ag) \cdot u(Ag) + p(Ob) \cdot 0 = \\
&\quad p(Ag) \cdot u(Ag)
\end{aligned}
\tag{5.4}
$$

  This is the effective salience of *Ag* (expression 5.2)!

- Expected utility of expressing *Ob* sentence-initially:

$$
\begin{aligned}
&EU(Pa) = \\
&\quad p(Ag) \cdot U(Pa, \mathcal{A}) + p(Ob) \cdot U(Pa, \mathcal{O}) = \\
&\quad p(Ag) \cdot 0 + p(Ob) \cdot u(Ob) = \\
&\quad p(Ob) \cdot u(Ob)
\end{aligned}
\tag{5.5}
$$

  This is in fact the effective salience of *Ob* (expression 5.2).

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

5.5.2 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.*

Sent#	$u(Ag)$	$u(Ob)$	$p(Ag)u(Ag)$	$p(Ob)u(Ob)$	*decision*
1	100	100	87	13	*active*
2	10	100	8.7	13	*passive*
3	10	10	8.7	1.3	*active*
4	10	50	8.7	6.5	*active*
5	100	10	87	1.3	*active*

Table 5.1: Utilities and Expected Utilities in Illustration

Sent#	$u(Ag)$	$u(Ob)$	*precedence*	*decision*
1	*high*	*high*	$Ag \prec Ob$	*active*
2	*low*	*high*	$Ob \prec Ag$	*passive*
3	*low*	*low*	$Ag \prec Ob$	*active*
4	*low*	*high*	$Ag \prec Ob$	*active*
5	*high*	*low*	$Ag \prec Ob$	*active*

Table 5.2: Qualitative Utilities and Decisions in Illustration

5. *Carl Sagan* bought *a shirt*.

As before, let $u(Ag), u(Ob)$ denote the instantial saliences of agent and object respectively. Let the corresponding canonical saliences be $p(Ag), p(Ob)$. By the derivation in the previous subsection, the expected utility of generating an active voice sentence is the effective salience of Ag, given by $p(Ag)u(Ag)$. Similarly, for a passive voice sentence we have $p(Ob)u(Ob)$ as the expected utility.

For the five sentences in the illustration, tables 5.1 and 5.2 summarize 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. These probabilities are approximates of the results of the empirical work of Svartvik [70] on the relative frequency of passives. The instantial saliences have been subjectively assigned as functions of the vividness of entities.

Remarks:

- In sentence (1), though the object has high instantial salience in absolute terms, it does not stand out in contrast with the agent's. Hence canonical salience has its way in deciding the choice.
- In (2), the instantial salience of the object is high enough to overcome the canonical bias. Hence a passive sentence is obtained. (Note the omission of the agent.)
- In (3), both entities are equally pallid, leaving the choice to canonical salience.
- In (4), though the object has a fairly high instantial salience compared to the agent, it is not high enough to overcome the canonical salience in competition. An active voice sentence therefore results.
- In (5), the agent has a very high instantial salience, which is accentuated by the canonical bias towards the agent in a synergistic interaction. An active voice sentence is (strongly!) recommended.
- 3 of the 5 sentences above have a high-salient object. 4 out of 5 have an object with instantial salience no less salient than that of the agent. Yet, only 1 out of the 5 cases produces a passive.

Note also that all possible interactions are captured numerically in this simple model, without making explicit symbolic references to synergistic or competitive interaction, or to the vividness or pallidness of entities.

5.5.3 General Form for Binary Choice Contexts

The derivation of the expected utility of actions as effective salience of entities in a binary salience configuration can be expressed in a general form, applicable to binary linguistic choice contexts in linearization, similar to the active-passive derivation given earlier.

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. $p(x_1) + p(x_2) = 1$.

Denote the instantial saliences of x_1 and x_2 as $u(x_1)$ and $u(x_2)$ respectively.

Let $\{\mathcal{X}_1, \mathcal{X}_2\}$ be a linguistic choice context, with the alternatives $\mathcal{X}_1$ and $\mathcal{X}_2$ corresponding to the earlier expression of x_1 and x_2 respectively.

Then, analogous to the derivation for the active-passive instance,

$$EU(\mathcal{X}_1) = p(x_1) \cdot u(x_1) \tag{5.6}$$

and

$$EU(\mathcal{X}_2) = p(x_2) \cdot u(x_2) \tag{5.7}$$

The construct selected is the one with greater EU.

Synergistic and Competitive Interactions Defined

We can give formal definitions of synergistic and competitive interactions between canonical salience and instantial salience in binary choice contexts involving entities x_1 and x_2, as follows:

1. *Synergistic Interaction*: Canonical salience and instantial salience interact synergistically if

$$p(x_i) > p(x_j) \Leftrightarrow u(x_i) > u(x_j) \tag{5.8}$$

for $i \neq j$ and $i, j \in \{1, 2\}$.

2. *Competitive Interaction*: Canonical salience and instantial salience interact competitively if

$$p(x_i) > p(x_j) \Leftrightarrow u(x_j) > u(x_i) \tag{5.9}$$

for $i \neq j$ and $i, j \in \{1, 2\}$.

5.5.4 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 salience.* That is,

$$\begin{aligned} &If\ p(x_1) = p(x_2)\ then \\ &\quad max\{EU(\mathcal{X}_1), EU(\mathcal{X}_2)\} = max\{u(x_1), u(x_2)\} \end{aligned} \tag{5.10}$$

   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 salience.* That is,

$$\begin{aligned} &If\ u(x_1) = u(x_2)\ then \\ &\quad max\{EU(\mathcal{X}_1), EU(\mathcal{X}_2)\} = max\{p(x_1), p(x_2)\} \end{aligned} \tag{5.11}$$

   Choice here is apparently grammar-controlled.

3. Extremes of lack of availability of choice: *If $\mathcal{X}_1$ is the only available linguistic choice then x_1 will always be expressed earlier.* That is,

$$\begin{aligned} &If\ p(x_1) = 1\ (and\ hence\ p(x_2) = 0)\ then \\ &EU(\mathcal{X}_1) > EU(\mathcal{X}_2) \end{aligned} \tag{5.12}$$

4. Similarly, *If $\mathcal{X}_2$ is the only available linguistic choice then x_2 will always be expressed earlier.*

$$\begin{aligned} &If\ p(x_2) = 1\ (and\ hence\ p(x_1) = 0)\ then \\ &EU(\mathcal{X}_2) > EU(\mathcal{X}_1) \end{aligned} \tag{5.13}$$

Range of Utility Values

Utilities, unlike probabilities, are not necessarily in [0, 1], nor do they have to add up to 1. However, assuming them to be bounded, we can express them in [0, 1] as *relative utilities* as follows:

$$u'(x_1) = \frac{u(x_1)}{u(x_1) + u(x_2)} \tag{5.14}$$

and

$$u'(x_2) = \frac{u(x_2)}{u(x_1) + u(x_2)} \tag{5.15}$$

so that

$$u'(x_1) + u'(x_2) = 1 \tag{5.16}$$

5.5.5 Graphic Illustration of Degrees of Canonicality

Consider the choice context $\{\mathcal{X}_1, \mathcal{X}_2\}$, and entities x_1 and x_2 in the generation context. Assume that saliences are represented as probabilities and utilities as discussed above.

We assume further that bounded utilities are normalized, so that $u(x_1) + u(x_2) = 1$. Now, for a given, fixed canonical distribution, say $p(x_1)$ and $p(x_2)$ are 0.75 and 0.25 respectively, we can evaluate the expected utilities of $\mathcal{X}_1$ and $\mathcal{X}_2$ for each possible $u(x_1), u(x_2)$ combination, and display them graphically.

We can obtain one such plot for each possible canonical probability distribution. These graphs differ in the probability distributions, and can be taken as visual indicators of degrees of canonicality of inherent bias in linguistic choice systems.

The diagrams 5.2 - 5.6 help us visualize linguistic choice in binary contexts, as is applicable to salience-induced linearization. Each of these diagrams corresponds to a specific canonical salience (probability) distribution, and shows the variation of expected utilities of $\mathcal{X}_1$ and $\mathcal{X}_2$ with different relative values of instantial saliences $u(x_1)$ and $u(x_2)$.

The instantial saliences vary along the horizontal axis CB, with $u(x_1)$ decreasing from 1 at C to 0 at B, and correspondingly, $u(x_2)$ increasing from 0 at C to 1 at B. The line AB is a plot of the expected utility of $\mathcal{X}_1$, computed as $p(x_1)u(x_1)$. CD is a plot of the expected utility of $\mathcal{X}_2$ ($= p(x_2)u(x_2)$).

For any given value of $u(x_1)$ (and $u(x_2)$), the linguistic construct chosen corresponds to the "higher" of the lines AB and CD.

For instance, consider the diagram 5.3 corresponding to $p(x_1) = 0.75$. As long as $u(x_1)$ is above 0.25, the construct $\mathcal{X}_1$ is preferred. $EU(\mathcal{X}_2)$ equals $EU(\mathcal{X}_1)$ at the point O, where $u(x_1) = 0.25$ and $u(x_2) = 0.75$. From this point onwards, till B where $u(x_1) = 0$, the construct $\mathcal{X}_2$ is preferred. The lengths of the altitudes of the triangles OAC and OBD respectively from the vertex O are indicators of the frequency or rarity of the constructs $\mathcal{X}_1$ and $\mathcal{X}_2$ respectively. Indeed, they equal $p(x_1)$ and $p(x_2)$ respectively.

In diagram 5.4, the choices $\mathcal{X}_1$ and $\mathcal{X}_2$ are both equally available, and equally frequent in a large corpus of text.

The extremes: Diagrams 5.2 and 5.6 depict the extremes in the probability distributions. In 5.2 where $p(x_1) = 1$, the construct $\mathcal{X}_1$ is always chosen, whatever the relative instantial saliences be. The built-in preference for the construct $\mathcal{X}_1$ is very strong, and no other alternative is available. In such a case, to convey the higher (instantial) salience of entity x_2 relative to x_1, the speaker may have to resort to other available syntactic means, or more probably, to other modes of expression such as prosody or phonology. In 5.6, $p(x_1) = 0$, and the construct $\mathcal{X}_2$ is always preferred. The point O coincides with B in figure 5.2 and with C in figure 5.6.

5.6 Discussion

The model of interactions between canonical salience and instantial salience discussed in this section captures all of the psycholinguistically-relevant dimensions of the interactions. Modelling the interaction as a decision-theoretic combination of probabilities and utilities appears to be a natural choice. Our work was not motivated by a need to find application for decision theory; the

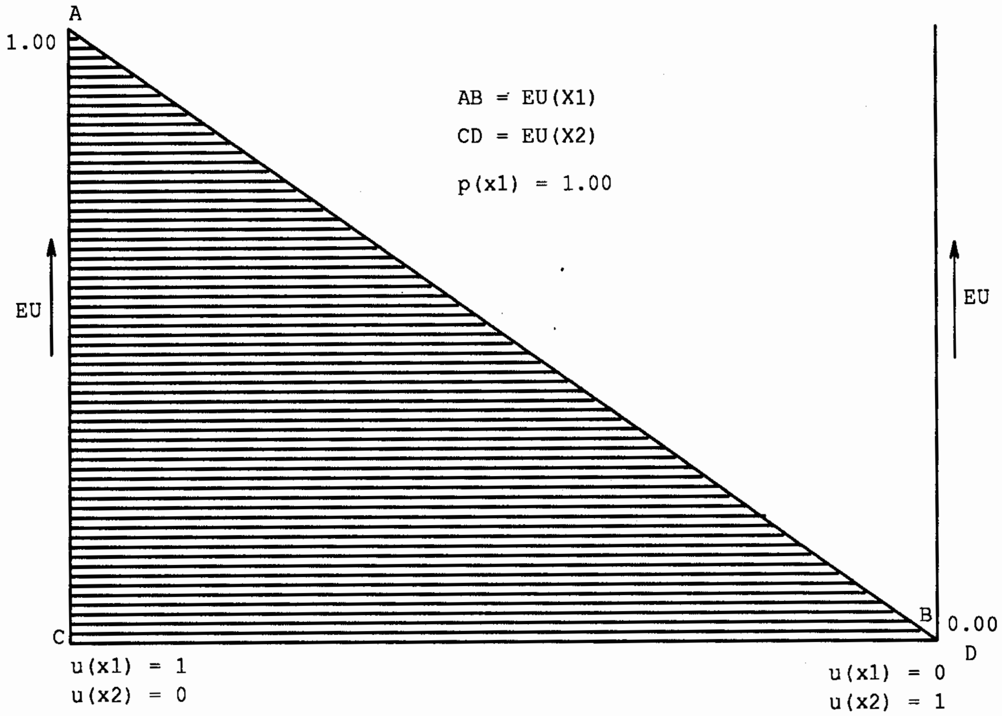

Figure 5.2: Expected utilities of $\mathcal{X}_1$ (AB) and $\mathcal{X}_2$ (CD) for $p(x_1) = 1.00$

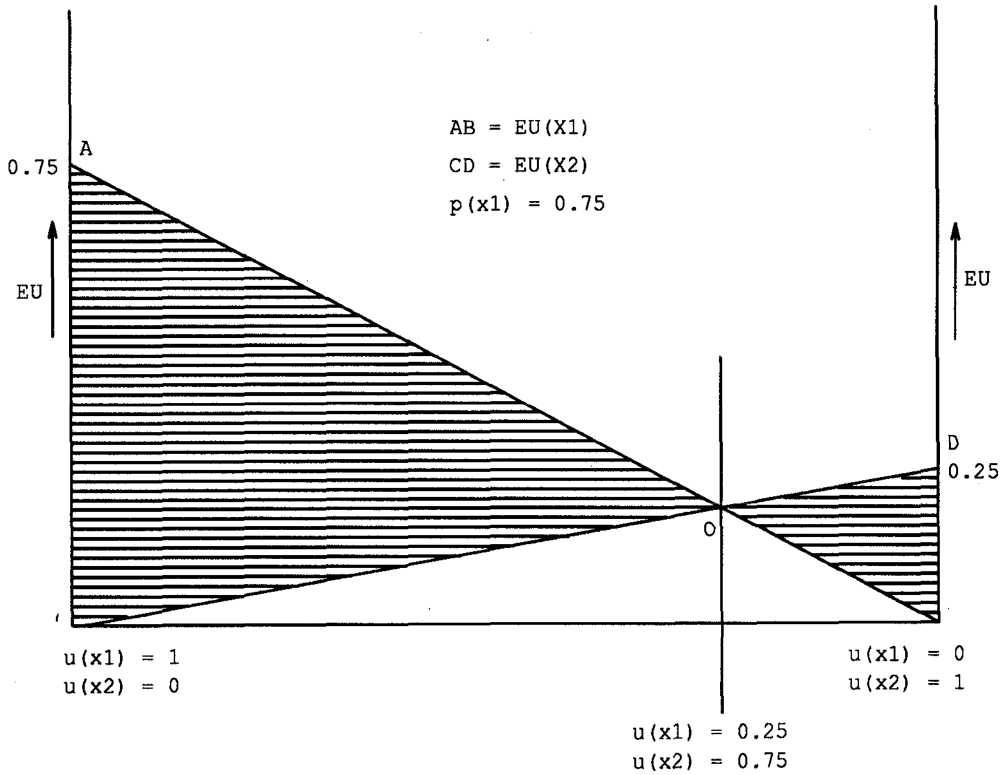

Figure 5.3: Expected utilities of $\mathcal{X}_1$ (AB) and $\mathcal{X}_2$ (CD) for $p(x_1) = 0.75$

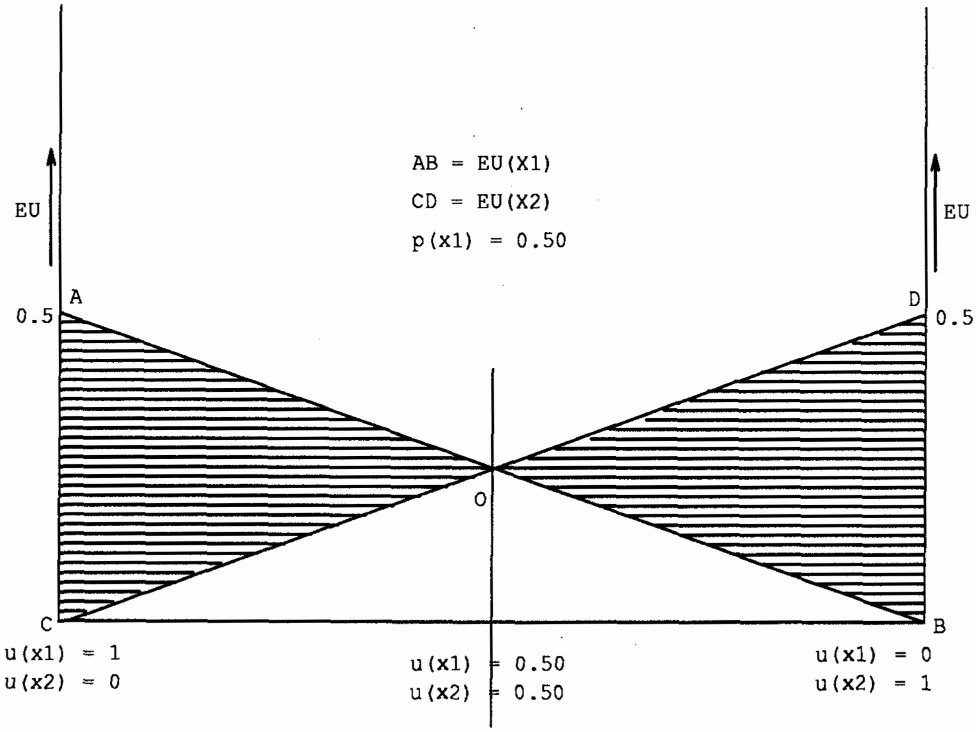

Figure 5.4: Expected utilities of $\mathcal{X}_1$ (AB) and $\mathcal{X}_2$ (CD) for $p(x_1) = 0.50$

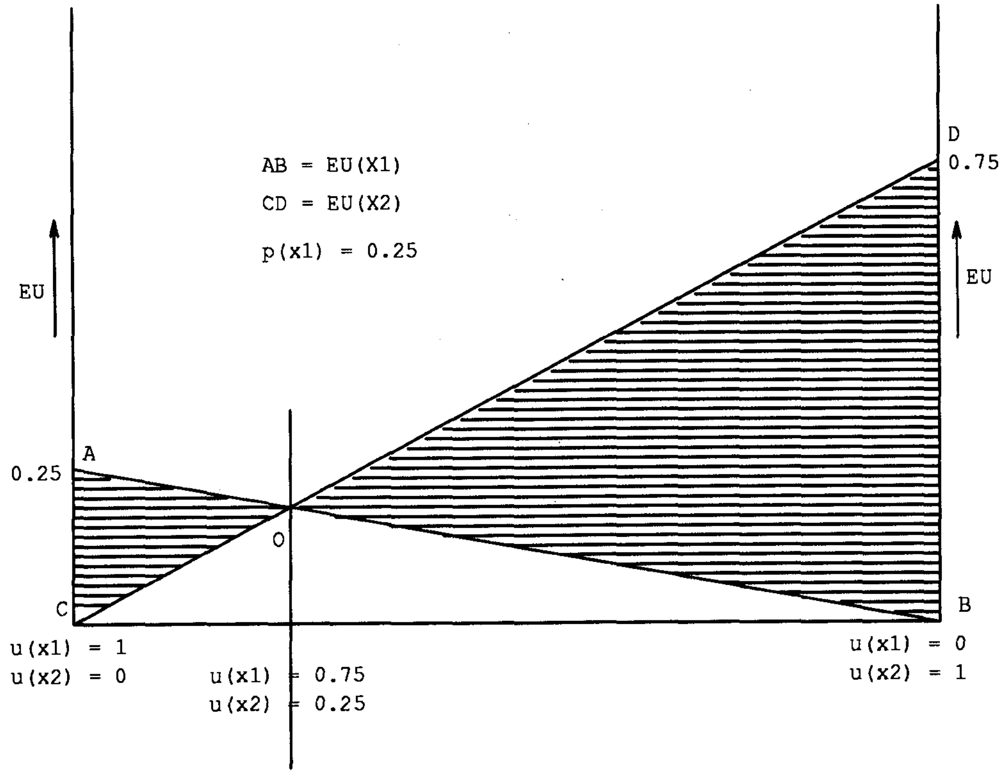

Figure 5.5: Expected utilities of $\mathcal{X}_1$ (AB) and $\mathcal{X}_2$ (CD) for $p(x_1) = 0.25$

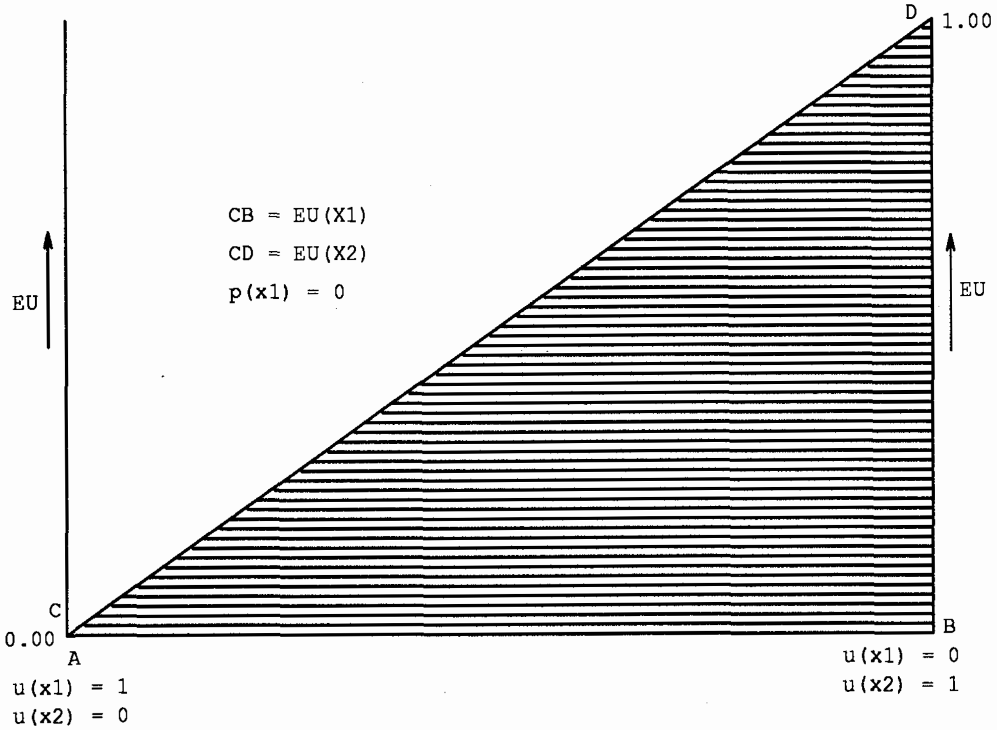

Figure 5.6: Expected utilities of $\mathcal{X}_1$ (AB) and $\mathcal{X}_2$ (CD) for $p(x_1) = 0.00$

manner in which decision theory presented itself as a candidate vehicle for conveying salience interactions was an interesting development in this research.

5.6.1 Psychological Plausibility and Validity

I think this model is psychologically *plausible*. However, psychological *validity*, which I think is different from *plausibility*, is a matter to be established scientifically through a process of *model validation*, by projecting the proposed model back to human behaviour and observing the effects, through a combination of experimental and empirical work. Our work can perhaps inspire some new experiments on sentence production, as well as some corpus work for estimating canonical probabilities of several syntactic constructs. The latter will also be useful for incorporation into computer generators. I am not fully aware of the state of art of experimental psycholinguistics to hazard a guess on what the experiments will be like. At any rate, since the principal aim of this work was to arrive at interesting hypotheses rather than to validate established hypotheses, we suspend the line of thought short of the validation stage, and seek to 'travel' further with the proposed model of interactions to other NLG phenomena...

5.6.2 Using this Model in Generators

A generator which uses this model of interactions for generating simple sentences will not overgenerate unnatural or rarer constructs. The numbers will not be brittle either: very small changes in the numbers will not cause a drastic difference in the decision. 4 numbers are needed (2 canonical- and 2 instantial saliences) for each decision. The former can be naturally associated with the sentence generator's linguistic knowledge, and the latter comes from the instantial saliences in context, and can be left to be evaluated by the text planner.

Depending on how the generator slices the surface generation tasks from planning tasks, specific generators will differ in how these interactions are handled. For example, in the grammar component of Penman, called Nigel, the canonical probabilities are naturally attached to the systems of choice. Before making a choice, these numbers can interact with the instantial saliences input to the generator, and on the basis of the interactions, the choice can be committed. This choice, clearly, will not always coincide with the a priori probabilities attached to the systems.

In Mumble-86 however, the choices are assumed to be already made, as Mumble-86 is a realization component. Therefore, the realization specification input to Mumble-86 should encode linearization directives based on *effective* salience. In either case (Penman/Nigel or Mumble-86), we see salience-induced linearization as an interesting line of further development. We do not discuss the details of these generators here, as we consider them (at this state of development of NLG as a field) part of common knowledge in the field, and as familiar to the researcher as, say, MYCIN or ID3. We point the reader to Meteer *et al.* [46] for a clear documentation of Mumble-86, and to Mann [36] and Mann and Matthiessen [35] for details on Penman/Nigel.

As noted earlier, the use of four numbers (2 canonical, 2 instantial) avoids overgenerating unnatural constructs. However, the performance of computer generators with regard to their ability to capture frequencies of constructs cannot be strictly evaluated on the basis of the frequencies observed in a small number of sentences generated. This is because language is an ergodic system, and the canonical probabilities approximate frequencies in the long run.

5.6.3 Probabilities in Systemic Grammar

There is a small but interesting body of literature in systemic functional linguistics (SFL) associating probabilities with systems of choice in grammar Nesbitt and Plum [49], Halliday [21]. The essence of Halliday's idea is that probabilities, as weights attached to options in the SFL choice systems, are integral to the grammatical description of a language for capturing variation observed in text. SFL thus clearly associates probabilities in the systems with frequencies in text, measurable through corpus work, as was done by Nesbitt and Plum [49].

How Halliday Complements Our View-as-it-stands-now

Halliday's views overlap with our conception of canonical salience. They *complement* our development, in the sense that Halliday sees *register* as a major factor conditioning the probabilities. His view opens up the possibility of enriching canonical salience with sociolinguistic determinants, in addition to perceptual ones.

Mother-Child Interaction Halliday views the probabilities attached to linguistic choice systems as arising in the linguistic interactions between mother and child, a social dimension which complements the prelinguistic perceptuo-cognitive dimension elucidated in psycholinguistics. Halliday cites Hasan [22] to suggest that the linguistic choices made by the mother affect the child's way of reasoning, learning and talking. This is because every use of a (linguistic) pattern by the mother foregrounds the pattern quantitatively in such a manner as to affect the probabilities of the child's own system. This process corresponds to *entrenchment* and *frequency of mention*, given in chapter 3 as determinants of salience. Halliday/Hasan's thesis can be paraphrased as: frequency of mention by the mother contributes to entrenchment of pattern in child, and influences crystallization of canonical salience. Through a gradual process, the child's probabilities become compatible with the mother's.

Moreover, the effect of sociolinguistic factors allows the probabilities to continue to change very gradually, during adulthood. This change can capture certain aspects of the gradual variation of language itself, over time.

How Halliday's View Coincides with Ours

Halliday's probabilities coincide with our canonical distributions, in that they are associated with linguistic choice systems and related to frequencies of constructs observed empirically. Nesbitt and Plum [49] claim that probabilities are in fact *necessary* to bring out the patterns of choice in text observed empirically in frequencies.

Halliday also clearly observes the relation between probabilities associated with the SFL choice systems, and the degree of linguistic *markedness* of constructs. A very low probability indicates a high degree of markedness.

How Halliday's View may be Incompatible with Others'

While probabilities in linguistic systems are related to frequencies in text, one makes an *a priori* commitment to a sample space over which probabilities are defined, before setting out to measure the probabilities. These sample spaces are the linguistic choice systems. The cases we have seen so far, like active-passive, dative inversion and so on didn't seem questionable as sample spaces, as we had psycholinguistic and linguistic arguments that naturally saw the alternatives as constituting a choice system. However, not all theories, or generators, "bundle" linguistic knowledge in the same way: a point clearly observed by McDonald and Pustejovsky [41]. For example, Mumble-86 and Penman/Nigel have different bases for organizing linguistic knowledge. Therefore, it is quite likely that a candidate canonical system may coincide with the presentation of alternatives in one, and be spread across the presentations of choices in another. Nevertheless, the researcher can always define *any* choice system by putting together a reasonable set of alternatives, based on intuitions, and experiment with it by measuring frequencies through corpus work. Interesting patterns of distribution may emerge, and shed light on the basicness of constructs.

How Osgood's View Complements Halliday's

Osgood's naturalness, as a bias shaped by prelinguistic, perceptual experience (and possibly, innateness) and persisting in adulthood reflects a dimension of language acquisition which is entirely different from Hasan-Halliday's social-interactional factors. Both, we believe, are necessary to explain human language acquisition.

How Our View Complements Halliday's

Halliday's view is incomplete in the following sense: While he sees probabilities as mechanisms essential for bringing out patterns of choice as reflected by frequencies in text, he has no account of how, or under what conditions, the *less probable* constructs are generated. To give such an account, we think that SFL has to consider many contrasts or biases in ideational, interpersonal and textual contexts as contributing to choice-motivating factors in the form of *instantial salience* which interacts sometimes synergistically and sometimes competitively with the a priori weights in the SFL choice systems.

A generator with probabilistically-weighted SFL choice systems cannot depend on the probabilities alone, since, in that case, they will never generate the less likely or marked alternatives. Nor is instantial salience by itself sufficient, because without canonical salience, the less-likely constructs will be overgenerated: we need a canonical bias to set up a frequency threshold. Our model of interaction has a very clear place for both generating less likely constructs occasionally, as well as common constructs often, while at the same time satisfying the salience principles.

How Our View Complements Halliday's and Extends Osgood's

Our delineation of canonical salience and instantial salience as very general categories in the salience theory, capturing built-in and instantially-induced preferences respectively, dramatically extends both Osgood's naturalness and Halliday's conceptions of probabilities in the next chapter.

If canonical salience and instantial salience are general, they should be observable in several other NLG tasks- not just phrase selection. In the next chapter, we show how a simple word selection task in naming objects depends on synergistic/competitive interactions between factors identifiable as canonical salience and instantial salience, wherein canonical salience arises in the perceptual experience with the objects one comes across in life, shaping one's taxonomic conceptual knowledge.

Chapter 6

Entry Level Effects in Naming Objects

We have presented canonical salience as a built-in bias in linguistic choice systems, which is applied automatically, every time the choice systems are activated. Central to the characterization of canonical salience is Osgood's notion of *naturalness* as arising in prelinguistic, perceptual experience, and as being cultivated in childhood and persisting in adulthood. We have explored the problem of modelling the interactions between canonical salience and instantial salience as they take place in linearization phenomena. We proposed viewing the interactions 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 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 naming objects, and the role of canonical salience discussed in the previous two chapters. To our knowledge, this similarity has not been observed or explained in the literature. In this chapter, we demonstrate the similarities we found between the role of basic levels in object naming and that of canonical salience in linearization. We present our results in modelling free naming of objects in terms of salience dynamics.

Basic levels: The notion of basic levels is applied to taxonomic concept knowledge. It was characterized by Rosch and Mervis [65] in cognitive psychology as the level of abstraction at which the concrete objects of the world are *most naturally* divided into categories. Accordingly, in day-to-day speech, one refers to objects by labels which are neither too general nor too specific. For example, when you request a friend for a particular instance of an apple, you are more likely to say:

> Can I have that *apple* please?

rather than:

> Can I have that *fruit* please?
>
> Can I have that *edible plant product* please?
>
> Can I have that *red delicious apple* please?

although the particular object you are requesting is classified under all the categories in the sentences above, and correspondingly, all the labels are applicable to the object. *apple* is the label corresponding to the basic level in the path of that object in taxonomic knowledge.

Basic levels are a cognitively-significant aspect of human conceptual categorization, and are shaped by the distribution of properties among the objects that we come across in our experience, according to which we classify objects as instances or non-instances of various classes. If the concrete objects of the world are grouped into categories in a certain optimal way, and if the labels associated with the categories correspond to nouns in natural language, then the selection of nouns (when a choice is available) in naming tasks is governed by the same structural properties of knowledge which is built through experience.

In the remainder of this chapter, we first discuss basic levels as studied in cognitive psychology. and present the similarities we found between basic levels in conceptual knowledge and Osgood's naturalness. We then discuss the findings of the experiments of Jolicoeur *et al.* [29], which demonstrate 'exceptions to the rule' of basic levels. Jolicoeur *et al.* introduce the notion of **entry level**, by which some specific instances of certain categories are referred to at a level other than the basic level.

For example, when pointing a particular *robin*, or some other 'generic' bird to a friend, you are more likely to day:

Look at that *bird*

However, when you see an ostrich (a bird), you are more likely to say:

Look at that *ostrich*

We point out the similarities of the conclusions of Jolicoeur *et al.*'s experiments (in cognitive psychology) to the conclusions on salience dynamics of Osgood, [Osgood and Bock] and others (in psycholinguistics). We discuss the parallels between the dynamics of entry level effects and that of salience, and propose considering entry level preference as induced by *effective* salience.

Basic level is the psychologically-optimal level of preference in a path in taxonomic knowledge. Entry level is the level at which one describes concrete objects. Entry level sometimes coincides with the basic level, and sometimes differs– giving us an excellent opportunity to explore the phenomenon in terms of synergistic and competitive interaction between canonical salience and instantial salience!

In the later part of the chapter, we model entry level preference in object naming, in decision-theoretic terms, and demonstrate it in detail with an example.

6.1 Basic Levels in Human Categorization

We present an overview of basic levels in taxonomic knowledge. The discussions in this section are based on the characterization of basic levels in [65], [45] and [29]. Among all the levels of abstraction at which concrete objects are classified, there exists a basic level which is psychologically optimal, or natural. The basis for categorizing objects is the distribution of properties among objects. These properties can be perceptual as well as functional (having to do with the way humans use the objects). Basic levels (e.g., table, car) are the most general level of abstraction at which many properties are common to all or most members of the categories. In the path *furniture-chair-easy chair*, chair is the basic level.

At levels which are superordinate to the basic levels (e.g. furniture, vehicle), objects share only a few properties, many of which are abstract. Categories below the basic level are subordinates (e.g., dining table, sports car). Subordinates are also aggregations of predictable properties, but contain little more information than the basic level to which they are subordinate.

Objects classified under different paths in taxonomic knowledge could vary in the extent to which instances of the categories share common properties (values for attributes). Therefore, different paths in conceptual knowledge could have different levels as basic. *Basic level is a property of a category in a path of a classification hierarchy, and is defined to be applicable to all instances of the category, whether the instances are typical or not.* For example, *table* is the basic level for all *dining tables*, *seminar room tables*, *pool tables* and so forth.

Among common biological categories like animals, birds, fruit and so on, a telltale empirical indicator of basic level is the *shape* of objects. The experiments of Rosch and her colleagues found that:

- 1. *Shape overlap*: Objects belonging to the same basic level category have a greater degree of shape overlap than do objects belonging to the same superordinate category but different basic level categories. Objects belonging to the same basic level category do not have significantly less overlap than do objects belonging to the same subordinate level category. For example, consider the portion of taxonomic knowledge with *fruit* as the superordinate level. Possible shapes are dispersed more widely among *fruits* than among *apples* (basic level). *apples* have a good degree of shape overlap which is not much more dispersed than the distribution of shapes among *granny smith* apples.

- 2. *Shape averaging*: Category members could be identified from the average of the shapes of two members of a basic level category as well as when averaging shapes from a subordinate level category. But people were much worse at identifying category membership when the shapes were drawn from different basic level categories belonging to a given superordinate category.

6.1.1 The Primacy of Basic Levels

The psychological primacy of basic levels is attributed to a number of special properties of categories. We summarize the significant ones below:

- The basic level is the most general level at which category members have similar overall shapes.
- At the basic level, a mental image can reflect the entire category.
- Basic level is the only level at which category membership can be determined by an overall gestalt perception, without detailed analysis of properties.
- Objects are recognized as members of basic level categories more rapidly than as members of categories at other levels of classification.
- Basic level is the one at which *adults spontaneously name objects.*
- *Labels for basic level categories are unmarked linguistically.*

- *Developmental priority in concept acquisition*: Basic level categories are shaped before categories at other hierarchical levels, in human concept learning.

- *Developmental priority in language acquisition*: There is a corresponding primacy in human language acquisition as well: the first botanical labels children learn are names for basic level categories. Similar results are available in cognitive psychology for other, non-botanical taxonomies. Languages first encode basic level biological categories, and only later, if at all, encode categories superordinate or subordinate to basic level ones.

Basic levels are privileged, or psychologically preferred, because of their *structural* properties, that is, because of the statistical distribution of property values among instances and non-instances of a category. In general, basic level categories are where a number of inference-related abilities are maximized in humans [45]. Mervis and Rosch state that the principles underlying the determination of which hierarchical level is basic are universal. However, for a given domain, the particular level which is found to be basic may not be universal, and may show variations due to cultural and other individual factors.

6.1.2 Basic Level Preference and Canonical Salience

The parallels between basic level preference and canonical salience may be observed in a number of aspects. We list below the important resemblances:

- According to Flavell and Wellman [12], a person is not conscious of the actual working of the process of basic level categorization. This characteristic tallies with our conception of canonical bias as being highly automatic and not being under volitional control.

- The developmental priority of basic levels in concept development as well as language development parallels the development of canonical bias (Osgood's natural salience) in early childhood through perceptual experience.

- Cognitive psychologists refer to the basic level in such terms as *usual*, *natural* and *linguistically unmarked*. Identical terms have been applied in psycholinguistics and linguistics to refer to the preferred status of linguistic constructions.

- The primacy of basic levels is rooted in the structural aspects of taxonomic knowledge, in that it is determined by the statistical nature of attribute values. This aspect is similar to the role of frequency of perceptual experience in determining the natural bias in expressing stative relations, action relations and so forth (chapter 4).

6.2 Entry Level Effects

Jolicoeur *et al.* [29] performed a set of four experiments to investigate the role of basic level. One of their chief interests was to determine the role of *typicality* or *atypicality* of the concrete objects in determining the level of abstraction at which objects are identified. They do not give a definition of typicality, or discuss the factors that contribute to the atypicality of objects. Nevertheless, the objects they identify as *atypical* reveal sufficient clues on the determinants, and can be described as *salient* due to such determinants as *uniqueness* (along with which *atypicality* is grouped in chapter 3), *vividness* and so on.

Jolicoeur *et al.* conducted four experiments. The first two pertained to the levels immediately superordinate and subordinate to the basic level. Their first experiment, which consisted of a picture-naming task, concluded that the activation of the superordinate level occurs after the necessary activation of the corresponding basic level, rather than by a slow feature-matching process. In their second experiment, they concluded that basic level concepts are activated faster than their corresponding subordinate levels. Their results indicated that the identification of the subordinate level required additional perceptual processing that was not required in identifying an object as a member of the basic level category.

6.2.1 Experiments Examining Atypicality Effects

The third and fourth experiments of Jolicoeur *et al.* are of particular interest to us, as they were aimed at examining the notion that the basic level is the level at which instances of that category are first identified (= categorized). Given the definition of basic level as a property of the *category*, and therefore applicable to all instances of that category, it would be worth while to see if there were instances which did not get identified first at the basic level. Since almost all natural categories have many instances which are atypical, it would be worth while to examine the basic level process for atypical instances of the category. If some concrete objects are first identified at a level subordinate to the basic level, then these objects will tend to be named with labels corresponding to the subordinate level rather than to the basic level.

Experiment 3 of Jolicoeur *et al.* involved tasks of naming a number of objects, including both typical and atypical members of the categories. The results indicated that for typical members of the categories, basic level names were preferred. However, for many atypical exemplars, a significant shift in the tendency was observed, towards naming objects at the subordinate level. (cf. *look at that ostrich* example). This tendency was verified in the fourth experiment which sought more direct evidence. The authors maximized the validity of their results by ensuring precise experimental conditions: for example, they ensured that the subjects' delays or difficulties in naming was not due to their not knowing the names of the objects.

Some of the objects (in the pictures) for which the subjects chose names at the level immediately subordinate to the basic level were: *penguin, submarine, ostrich, poodle, peacock*, etc. However, there were some atypical objects for which the basic level names were chosen. There was no tendency to shift the identifier to the subordinate level. Some examples of such objects in the experiments were: *tugboat, Rolls Royce, dachshund*, etc.

Finally, it is important to note that the (a)typicality of a category is evaluated with respect to the immediate superordinate. For example, *ostrich* is an

Category	*Typical Exemplar*	*Atypical Exemplar*
dog	collie	poodle
fish	bass	seahorse
spoon	tablespoon	Japanese spoon
knife	kitchen knife	cleaver
pants	slacks	overalls
shoes	casual shoe	sandal
chair	kitchen chair	rocking chair
table	kitchen table	pool table

Table 6.1: Some Exemplars in Jolicoeur *et al.*'s Experiments

atypical *bird*, *crabapple* is an atypical *apple*, and so on. When the reference level for defining atypicality coincided with the basic level, the shifts towards the subordinate level in naming atypical objects was pronounced: for example, *crabapples* which are atypical *apples*, were named *crabapple*; *apple* is the basic level. When the reference level for defining atypicality was the level superordinate to the basic level, there was little tendency to shift the label away from the basic level for atypical exemplars: for example, *lime*, an atypical *fruit* was named *lime*: *lime* itself is the basic level. Given that basic level and atypicality both contribute to the fast identification and naming of objects, there was a coincidence of these effects in *lime*, which we interpret as a synergistic interaction between the preferences dictated by basic level and by atypicality.

Some of the categories for which Jolicoeur *et al.* studied the effects of atypicality have been shown in 6.1.

One possible way of explaining the shifts in the naming (identifying) tendencies is to assert that for atypical instances, the basic level is defined at the level of towards which the shift occurs. Jolicoeur *et al.* recognize this possibility, and point out that such a re-definition of basic level for atypical exemplars is not appropriate, since basic level is defined for entire categories, and is applicable by definition to all instances of those categories, whether typical or not. Moreover, there is a wealth of evidence that confers primacy on the basic level (summarized earlier in this chapter). Redefining the basic level for specific instances based on observed effects in identifying/naming tasks is therefore not appropriate.

The authors therefore propose the concept of *entry (point) level* as the level at which objects are identified. This level, as Jolicoeur *et al.* note, corresponds in many cases with the basic level; but in several instances, it does not.

6.2.2 Entry Level Preference and Effective Salience

Jolicoeur *et al.* say:

> *"If the object is very distinctive or atypical example of a basic level category, then it may have its own entry point defined at the subordinate level. The notion of entry point, therefore, is an attribute of individual exemplars rather than an attribute of categories."*

This is precisely the distinction between canonical salience and effective salience! Basic level is defined for categories, just as canonical salience is defined in a parametric form over agents, objects and such abstract entities, instances of which are talked about in specific situations.

As the specific level at which particular objects are described, entry point level may or may not coincide with the basic level, in just the same way that all linearization decisions are based on maximum effective salience. An entity with maximum effective salience, chosen in a particular instance, may or may not be the one which has maximum canonical salience.

It is also easy to note that the factor that causes a shift away from the basic level in naming, viz., atypicality, arises in particular instances because of uniqueness, vividness, and such determinants, and coincides with our notion of particular objects in generation possessing *instantial salience* due to a possible combination of several determinants. (Jolicoeur *et al.* do not discuss the determinants of atypicality in detail, or about how atypicality is measured, and thus differ in their exposition from Osgood and his associates who have discussed vividness, speaker motivation and other determinants of salience in very good detail.) We present a probabilistic method of quantifying atypicality, in our illustration later in this chapter.

We now interpret the major findings of experiments 3 and 4 of Jolicoeur *et al.* in terms of salience dynamics, combining them with Rosch *et al.*'s findings on basic level:

1. *The tendency to name objects at the basic level is natural, and is usually quite strong*: Considering the labels corresponding to the levels in a categorization hierarchy as constituting a linguistic choice system in a naming task, basic level preference is a form of canonical salience which biases (usually strongly) the tendency to select the name corresponding to the basic level.

2. *Atypical exemplars of basic level categories tend to shift the preference away from the basic level towards the subordinate level*: Instantial salience facilitates the choice of subordinate names. In competitive interaction with canonical salience, instantial salience sometimes 'wins out'.

3. *Even if the exemplars are atypical, the basic level label may still be chosen*: The strength of instantial salience dictating preference of the subordinate label may not be strong enough to overcome the threshold set up by the (stronger) natural tendency to prefer the basic level label. In competitive interaction with canonical salience, instantial salience loses.

4. *Basic level preference and atypicality sometimes dictate the same effect, and reinforce each other*: canonical salience and instantial salience sometimes interact synergistically in naming tasks.

This interpretation is based on the impressive degree of similarity we see between the main results of Jolicoeur *et al.*'s experiments in cognitive psychology and those of Osgood and Bock and others' work in psycholinguistics. The other parallels between basic level and Osgood's naturalness have been already pointed out, in an earlier subsection. To the best of our knowledge, this resemblance has not been recognized in all literature, in such specific detail as we have found.

Another interesting connection between Osgood's naturalness and basic level is the setting in which psycholinguists and cognitive psychologists found it most convenient to observe them: Psycholinguists found the *Simply Describing* setting most suitable for observing perceptually-rooted naturalness, while cognitive psychologists found basic level effects in *free naming* tasks.

We also take these similarities as adding good support to the belief of salience as a pervasive principle of choice in language production, with similar types of interactions between canonical salience and instantial determinants. A key factor which contributed to establishing this belief was a close examination of experimental results from two disparate disciplines. Moreover, our interest in delineating the categories *canonical salience* and *instantial salience* in the early stage of our work, and (later) *effective salience*, appears to have 'paid off'. Time is now ripe for us (following a brief digression) to explore the applicability of decision theory as a formal vehicle for conveying the name choice criterion!

The Experiential Roots of Communicative Maxims

As noted by Reiter [64], who uses basic level preference in noun generation (using handcoded annotations in a KL-ONE-based knowledge representation), failure to use basic level terms (or more accurately, entry level preference) can have negative effects in communication. For example, consider the pair (from [64])

(a) Don't go swimming; there is a *shark* in the water
(b) Don't go swimming; there's a *tiger shark* in the water

If *shark* is the mutually-shared entry level to refer to an instance of a *tiger shark*, then the speaker, by choosing to say (b) instead of (a), conversationally implicates that (a) could not have been used– that it is somehow relevant that the entity is not just a shark, but a *tiger* shark. Unless such an implicature is intended, and the relevance is inferrable, the speaker should prefer (a) to (b).

If the speaker uses a term which is more general than the entry level, it can be confusing as well:

(a) Don't go swimming; there's an *aquatic creature* in the water.
(b) Don't go swimming; there's a *living entity* in the water.

Both (a) and (b), while entertaining, are hardly enlightening, and are not likely to be used unless the understatement is intended as a figure of speech!

General day-to-day talk involving references to objects, being guided by entry level preference, automatically avoids misleading conversational implicatures and preserves relevance. We are tempted to speculate the following: both basic level preference and occasional violation of it due to factors like atypicality and vividness, etc, knowledge of which is likely to be mutually shared, have connections with experience with objects and their properties. By using these principles for noun selection when referring to objects, 'maxims of communication' are automatically adhered to; misleading conversational implicatures are not frequently generated, and relevance is preserved.

Grice's maxims of communication are mostly discussed in deliberate communicative settings. What would it mean for us to adhere to Grice's maxims in simple day-to-day speech involving references to objects? We conjecture that natural principles of word preference (basic/entry level) automatically ensure 'felicitous' communication. The regularities observed in felicitous communication can be described in terms of highly general maxims; however, these maxims need not be invoked to create mechanisms for ensuring felicitous communication in NLG systems. Adherence to these maxims in rapid conversations and day-to-day speech is at best a side effect. Incidentally, basic level categories also tend to have shorter labels than their subordinates or superordinates (cf. *Be concise!*)

6.3 Modelling Entry Level Preference in Describing Objects

We proposed in the previous section that entry level effects in object naming can be visualized in terms of canonical salience (basic level) and instantial salience (atypicality, etc) interacting to determine effective salience (salience of entry level category).

In this section, we present in detail a model of entry level preference as expected utility maximization. We need three major components for this model:

1. Modelling canonical salience (basic level preference) as a probability distribution, setting up an uncertain environment (= not under speaker's control; inescapable) in which decisions are made. By this model, basic level is the level in the classification hierarchy 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 *category utility*, a measure which was proposed by Gluck and Corter [14] as a predictor of basic levels in taxonomic hierarchies.

2. Utilities as functions of the instantial salience of the objects being named.

3. Effective salience as expected utility, determining the choice of names.

In the rest of this section, we first present a representation for taxonomic concept knowledge which is capable of capturing the dependence of basic level effect on the structural aspects of categorization, that is, on the statistical characteristics of the distribution of values of attributes among instances and non-instances of categories. This representation adapts the probabilistic concept, proposed in cognitive psychology by Smith and Medin [68], to the taxonomic knowledge base.

We then derive in detail the canonical salience distribution on the basis of what we call *differential category utility*, which is computed using probabilistic knowledge about the properties of categories, represented in taxonomic knowledge base. Using the same knowledge of probabilities, along with additional mechanisms, we then quantify instantial salience that arises due to unusual properties (= atypicality). Finally, we present the decision criterion for name selection as expected utility maximization.

Throughout the section, an example involving a small portion of taxonomic knowledge is carried through, to illustrate as well as to develop the details.

6.3.1 Concept Representations

Smith and Medin [68] have developed a probabilistic view of concept representation which has been motivated strongly by the need to explain human categorization and concept recognition, and the use of conceptual knowledge in inference. The view has been particularly propounded in order to tackle some specific, difficult problems, such as the existence of disjunctive concepts with different degrees of disjunctiveness, the difficulty of specifying defining features for most concepts, etc. The probabilistic view is offered as an alternative to the all-or-nothing classical view of concepts and category membership.

In a probabilistic concept representation, a concept is a summary description with properties related to concept membership probabilistically. There are three variations in the probabilistic view: (1) dimensional (quantitative), (2) featural (qualitative) and (3) holistic, each having to do with the way properties of objects are represented.

In the dimensional (quantitative) approach to concept representation, concepts having the same relevant (attribute) dimension can be represented as points in a multidimensional metric space. With this representation, it is possible to pose category membership questions in terms of whether a candidate instance is within some threshold distance of a concept. The concept representation we use is most similar to the dimensional approach. We recommend the book of Smitn and Medin [68] to the reader as a highly readable and lucid exposition of the probabilistic view of concept representation.

The probabilities admit a relative frequency interpretation by which

$$p(A_i = V_{ij}|C_k) = \frac{\#times\ C_k\ was\ observed\ to\ possess\ the\ value\ V_{ij}\ for\ A_i}{\#times\ C_k\ was\ observed} \tag{6.1}$$

Probabilistic Concept

The knowledge base we use for taxonomic knowledge consists of concept representations 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. We use probabilistic concepts in the next chapter as well, in a method for generating comparison statements.

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* (also called an *attribute dimension*) and $\mathcal{V}_i$ is its *probabilized value space*. $\mathcal{V}_i$ is a finite set of n_i pairs $V_{ij} : p_{ij}$, with $p_{ij} \in (0, 1]$, and $\sum_{j=1}^{n_i} p_{ij} = 1$.

Figure 6.1 shows an example of a probabilistic concept.

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.

Conditional probabilities of the form $p(A_i = V_{ij}|C_k)$ are indices of how *predictable* the value V_{ij} is by members of category C_k. For example, if most apples are red, then $p(colour = red|apple)$ will be high, and *apple* is a good predictor for *red*. Equivalently, such conditional probabilities are indices of *intra-class similarity*.

On the other hand, conditional probabilities of the form $p(C_k|A_i = V_{ij})$ are indices of how *predictive* the value V_{ij} is of category C_k. A large value for $p(C_k|A_i = V_{ij})$ would indicate that not many objects in classes other than C_k have value V_{ij} for attribute A_i. Equivalently, such conditional probabilities are indices of *inter-class similarity*.

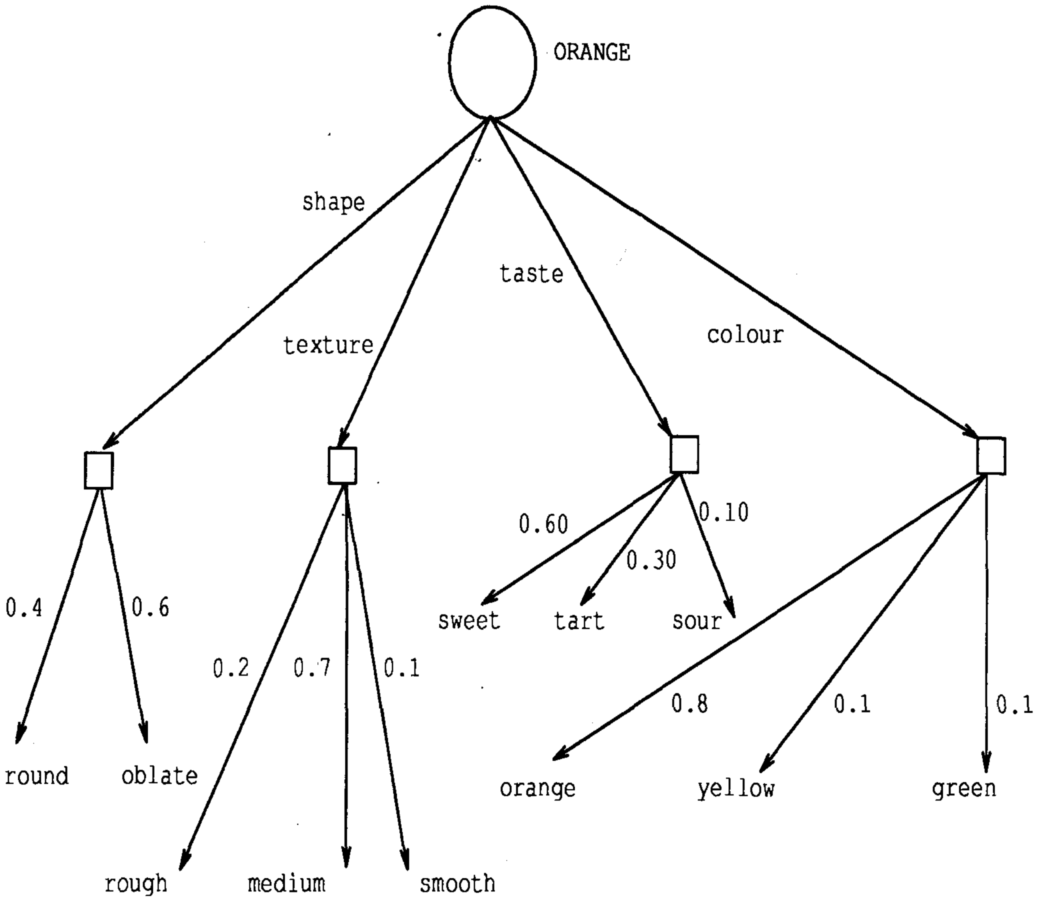

Figure 6.1: The Fruit *ORANGE* as a Probabilistic Concept

Probabilistic Concepts in Taxonomic Knowledge In the taxonomic knowledge base, the probabilistic summary of value distribution for attributes is stored at several levels. Each category in the taxonomic hierarchy is thus a probabilistic concept. The probabilities at the various levels of abstraction are not all equally easy to assess subjectively or estimate empirically. However, we will assume their existence without excessive concern about their precision. Table 6.2 shows an example of probabilistic concept representations in a taxonomic hierarchy. It is to be noted that the probabilities at successive levels of a path in the hierarchy are related to each other. For example, the probability of the red colour of apples summarizes contributions from the probabilities of the red colour of all types of apples. This dependence is demonstrated and used later in our calculations of a measure of atypicality.

6.3.2 Category Utility

Category utility is a measure proposed by Gluck and Corter [14] to predict basic levels in humans empirically. It is a measure of usefulness of categorization which is sensitive to the distribution of values for attributes. Gluck and Corter have proposed category utility in two forms:

- guessing game measure: expected score in a game involving prediction of values for attributes of an object.
- information-theoretic measure: related to the *grouping axiom* of information theory.

The latter has much formal resemblance to the former, and we do not discuss it here.

Category utility is familiar to the machine learning community, in particular to researchers interested in concept learning and categorization. It is used by Fisher [10] as a basis for a system called *COBWEB* that performs (unsupervised) incremental conceptual clustering.

Gluck and Corter state in their paper that their measure performed the best in experiments that attempted to predict basic levels in humans, compared to previously proposed measures which category utility superseded. However, they do not derive the measure, nor do they show how they estimated or elicited the probabilities. They report that the reaction times they obtained in their experiments involving naming tasks matched what category utility calculations predicted; however, we do not know from the paper whether category utility was used as a major component in the experiments. Gluck and Corter propose category utility measures for the instance where a set of objects is partitioned into two groups C and $not{-}C$. They give category utility for a concept with one attribute dimension. In the following derivation, we generalize category utility to a partition of a category into an arbitrary n subcategories, and allow an arbitrary number of attribute dimensions in a concept and value possibilities for attributes. Fisher [10] also uses a more general expression for category utility in his incremental learning task.

The Guessing Game Measure of Category Utility

We assume that the set of objects C is grouped/partitioned into n categories $C_1, C_2, \ldots, C_n$. For any $k \in [1, n]$, recall that the conditional probability $p(A_i = V_{ij}|C_k)$ is a measure of how well the value V_{ij} makes member objects of the class C_k *predictable.*

The scenario we consider here is a game in which a participant guesses the values for various attributes of some object. Assume that the participant knows what the possible values V_{ij} are for the attribute A_i. Assume further that the participant adopts what Gluck and Corter call a *probability-matching strategy*: for an attribute A_i, she guesses a value V_{ij} with a probability equal to the value of $p(A_i = V_{ij})$ in her own knowledge.

Ascribed as a score to the participant in the guessing game, the expression derived below for category utility is related to the expected number of correct guesses. Ascribed as a score to the category C_k, category utility $CU(C_k)$ is interpreted as the potential of the category C_k to enable inferences about itself.

If the category information (C_k) is given to the participant prior to guessing, then the probabilities in question are $p(A_i = V_{ij}|C_k)$. With a probability-matching strategy, 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:

$$\sum_{j=1}^{n_{ik}}[p(A_i = V_{ij}|C_k)]^2 \qquad (6.2)$$

where n_{ik} is the number of possible values of attribute A_i in concept C_k.

If guessing proceeds for every attribute of the object, the score is:

$$\sum_{i=1}^{l}\sum_{j=1}^{n_{ik}}[p(A_i = V_{ij}|C_k)]^2 \tag{6.3}$$

This is the expected number of values of attributes that can be correctly guessed for an object given that it belongs to C_k.

The expected value of the above sum, taken over all categories, gives the *expected score* of the participant who is assumed to have prior information about the distributions of $C_1, C_2, \ldots, C_n$:

$$\sum_{k=1}^{n} p(C_k) \cdot \sum_{j=1}^{n_{ik}}[p(A_i = V_{ij}|C_k)]^2 \tag{6.4}$$

This is the expected number of values of attributes that can be correctly guessed for an arbitrary object given a partition of the collection C of the objects into $\{C_1, C_2, \ldots, C_n\}$.

The contribution of concept C_k to the above score, denoted as $CU_1(C_k)$, is:

$$CU_1(C_k) = p(C_k) \cdot \sum_{j=1}^{n_{ik}}[p(A_i = V_{ij}|C_k)]^2 \tag{6.5}$$

In the above analysis, the probabilities are conditional on C_k since we assumed that the guesser had prior knowledge about the distributions for C_k. If category information were not given, then the analysis proceeds as above, except that the probabilities are unconditional. If guessing takes place for every attribute of the object with no prior knowledge about the categories, the expected number of attribute values that can be correctly guessed is:

$$\sum_{i=1}^{l}\sum_{j=1}^{n_i}[p(A_i = V_{ij})]^2 \tag{6.6}$$

where l is the number of attributes guessed, and attribute A_i has n_i possible values. Compare this expression with expression 6.3 above.

Assuming that the value of attribute A_i is being guessed, we next consider the expected *increase* in the score made possible by knowledge of category information, over the score with no prior knowledge of categories:

$$\sum_{k=1}^{n} p(C_k) \cdot \sum_{j=1}^{n_{ik}} [p(A_i = V_{ij} | C_k)]^2 - \sum_{j=1}^{n_{ik}} [p(A_i = V_{ij})]^2 \tag{6.7}$$

Category utility of the partition $C = \{C_1, C_2, \ldots, C_n\}$ is:

$$CU(C) = \sum_{k=1}^{n} p(C_k) \cdot \sum_{i=1}^{l} \sum_{j=1}^{n_{ik}} [p(A_i = V_{ij} | C_k)]^2 - \sum_{i=1}^{l} \sum_{j=1}^{n_{ik}} [p(A_i = V_{ij})]^2 \tag{6.8}$$

This is a general form of Gluck and Corter's expression for category utility, for an arbitrary partition size n and an arbitrary number l of attribute dimensions.

Given that $\sum_{k=1}^{n} p(C_k) = 1$ for the partition C, the subtrahend in $CU(C)$ above can be written as

$$\sum_{k=1}^{n} p(C_k) \sum_{i=1}^{l} \sum_{j=1}^{n_{ik}} [p(A_i = V_{ij})]^2 \tag{6.9}$$

We can therefore rewrite $CU(C)$ as:

$$CU(C) = \sum_{k=1}^{n} p(C_k) \left(\sum_{i=1}^{l} \sum_{j=1}^{n_{ik}} [p(A_i = V_{ij} | C_k)]^2 - \sum_{i=1}^{l} \sum_{j=1}^{n_{ik}} [p(A_i = V_{ij})]^2 \right) \tag{6.10}$$

The category utility of C_k is finally defined as the individual contribution of C_k to $CU(C)$ above:

$$CU(C_k) = p(C_k) \left(\sum_{i=1}^{l} \sum_{j=1}^{n_{ik}} [p(A_i = V_{ij} | C_k)]^2 - \sum_{i=1}^{l} \sum_{j=1}^{n_{ik}} [p(A_i = V_{ij})]^2 \right) \tag{6.11}$$

A Short Interlude on Notation

- C_{entity}, also called *entity*, is the root node of the concept hierarchy. Every object *IS-A* C_{entity}.
- $p(A_i = V_{ij})$ is equivalent to $p(A_i = V_{ij} | C_{entity})$.
- $sup(C_k)$ is the immediate superordinate of C_k. We will also write it sometimes as $C_{sup(k)}$.

- A partition of a category into subcategories is written as $C_{sup(k)} = \{C_1, C_2, \ldots, C_k\}$.
- $p(C_{entity}) = 1$.

An observation on the unconditional attribute probabilities: Given that $p(A_i = V_{ij}) = p(A_i = V_{ij}|C_{entity})$, in the guessing game, the knowledge used is that of possible values that A_i can have in *any* object. We intuit that in total absence of category information, probability distributions for attribute values tend to be more dispersed than when category information is given. (By dispersion we mean the number of possible values, as well as the flatness of the distribution curve). That is, when I am told, "*I have an object on mind; can you guess its colour?*", all possible values on the *colour* dimension are candidate guesses. When some sort of category information is presumed or given, as in, say, "*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 some 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. We can state the intuition expressed in the previous paragraph more formally as follows: for any C_k more specific than C_{entity}, $\sum_j [p(A_i = V_{ij})]^2$ will tend to be less than $\sum_j [p(A_i = V_{ij}|C_k)]^2$. The further down we go in the concept hierarchy towards more specific C_k's, the greater this difference tends to be.

This difference is accentuated when we consider the entire makeup of an object with several attributes, and add up the quantities corresponding to all attributes. That is,

$$\sum_i \sum_j [p(A_i = V_{ij})]^2 \ll \sum_i \sum_j [p(A_i = V_{ij}|C_k)]^2 \tag{6.12}$$

Differential Category Utility

A remark on the concept probabilities: As may be noted in our expression for category utility, the calculations also involve the probabilities of the concepts, $p(C_k)$. Are these probabilities unconditional? How are they obtained? Neither Gluck and Corter [14] nor Fisher [10] explain how to get (or how they estimate) the unconditional probabilities. We think that in most judgments, $p(C_k)$ is really $p(C_k|C_{sup(k)})$. In other words, conditional probabilities of concepts are naturally and conveniently assessed in the context of their immediate superordinates. (In several analyses we have seen so far, including Fisher's, this seems to be the case, though no explicit details are given in the publications.) We used this intuition to guide our estimation of probabilities for the illustration to follow.

Given that most judgments of concept probabilities are conditional on their superordinates (by the observation in the paragraph above), we examine the expression (6.11) for category utility: we notice a mismatch between the concept probability and the subtrahend in the term of probability squares! While the former is (according to the previous paragraph) estimated with respect to the superordinate concept, the latter (the squared probabilities in the subtrahend) is unconditional- hence there is a mismatch in the conditions in which the probabilities are evaluated. Such a measure cannot purport to be correct. We do not have any indication in the literature on how researchers tackled this problem, or whether they perceived this as a problem at all. We take the attribute probabilities in the subtrahend square term of category utility (6.11) as conditional upon the immediate superordinate as well.

Moreover, the locations of the basic level can vary between hierarchies, and is not necessarily (and in fact, is hardly ever) the most specific level for natural concepts. If we use $CU(C_k)$ (generalization of Gluck and Corter's) directly as a measure of likelihood of choosing the label of C_k, we suspect that it will always choose the most specific label available. This is because as we go down the concept hierarchy, the first sum of squares term in the expression for category utility will keep increasing. The sum of probability squares is a value between 0 and 1, and it increases as the distribution becomes less dispersed. $CU(C_k)$'s, as indices of amount of information, will increase as we tend towards more specific concepts in a hierarchy. It appears that what is maximized at the basic level is not the amount of information (*alias* the ability to support inference) in absolute terms, but rather the differential amount relative to the superordinate.

We have some evidence from cognitive psychology to support our line of thinking:

- Rosch and Mervis [65], pp. 586:

  > "Categories below the basic level are subordinates (e.g. lawn chair, sports car). Subordinates are also bundles of predictable attributes and functions, but contain *little more information than* the basic level object to which they are subordinate". (our emphasis).

- Mervis and Rosch [45]: Basic level is *the most general level* at which category members have similar overall shapes, and a mental image can reflect the entire category.

Basic level is thus the most abstract (general) category at which instances have many common properties. Consider a traversal of the hierarchy starting at the root (C_{entity}): as we go down, not only does category utility increase, but the rate at which it increases, increases. Soon, there comes a point where *the rate of increase decreases* (though category utility itself continues to increase). Basic level is therefore level C_k where the difference between $CU(C_k)$ and $CU(C_{sup(k)})$ is maximum.

Differential Category Utility: The differential category utility of a concept C_k is:

$$DCU(C_k) = p(C_k|C_{sup(k)}) \left(\sum_{i=1}^{l} \sum_{j=1}^{n_{ik}} [p(A_i = V_{ij}|C_k)]^2 - \sum_{i=1}^{l} \sum_{j=1}^{n_{ik}} [p(A_i = V_{ij}|C_{sup(k)})]^2 \right) \tag{6.13}$$

That is, while category utility is defined as the expected *increase* in score made possible by knowledge of C_k *over the score with no prior category knowledge whatsoever*, **differential category utility** (DCU) is defined as the expected *increase* in score made possible by knowledge of C_k *over the score with knowledge only of* $C_{sup(k)}$*, the superordinate of* C_k.

C_{entity} **as a 'boundary' case:** According to the definition of CU as the expected increase in score with respect to unconditional probabilities, $CU(C_{entity}) = 0$. In the definition of differential category utility, we treat this as a special 'boundary' condition. C_{entity} has no superordinate. For use in the expression for $DCU(C_{entity})$, we treat $sup(C_{entity}) = C_{entity}$. The hierarchy is clamped above at that level. So, the square terms in the difference cancel out, and we have $DCU(C_{entity}) = 0$.

6.3.3 Canonical Salience from Differential Category Utility

Now we proceed to use DCU in computing the canonical probabilities of basic level preference. We also launch our detailed illustration here, the computations with which will be carried through the section.

To start with, figure 6.2 depicts the example of a concept hierarchy we will discuss.

The root node in the hierarchy is C_{entity}, and the leaf nodes shown (*macintosh*, etc) are the most specific object classes. Each category in the hierarchy is a probabilistic concept.

The probabilities themselves are not shown in figure 6.2, but are tabulated separately for convenient viewing in tables 6.2 and 6.3. The tables represent the two kinds of probabilities we need:

- property probabilities in the hierarchy.
- concept/category probabilities

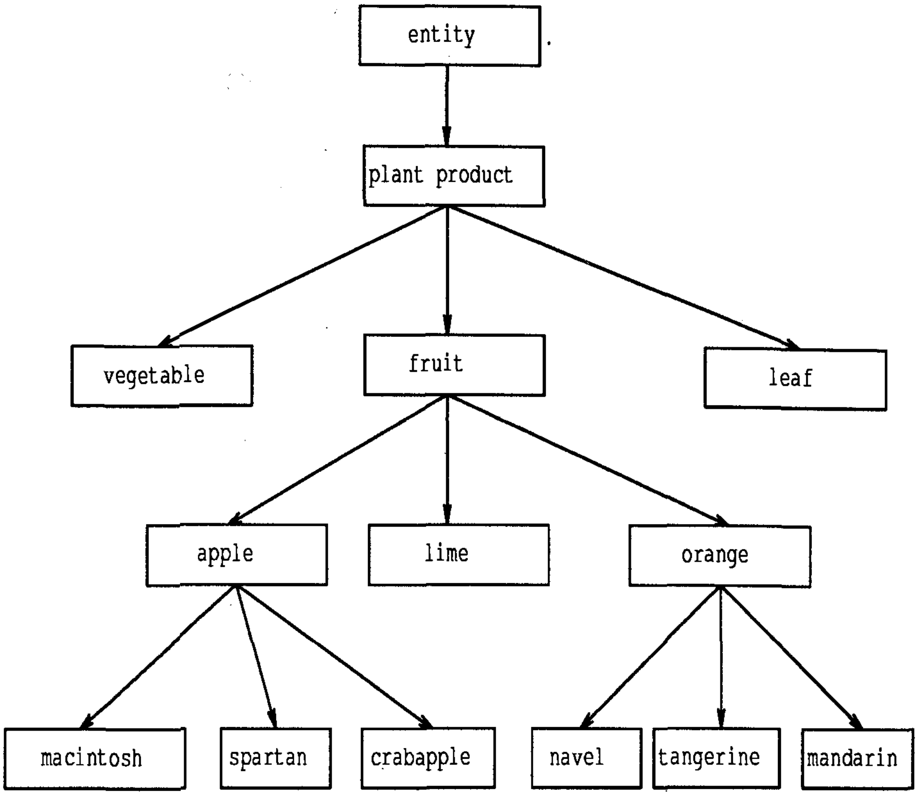

Figure 6.2: A Portion of a Hierarchy of Plant Products

Property probabilities: $p(A_i = V_{ij}|C_k)$ are shown in table 6.2. These values were estimated completely subjectively. For simplicity, when the distributions are too dispersed, and the probabilities were assumed to be uniformly distributed, we just list the height of the distribution and specify, for example, "uniform, 0.04" to indicate that there are 25 different possible values which are equiprobable. The rows are ordered from the most general level at the top to the most specific level at the bottom of the table. The two rows at the bottom, *macintosh* and *crabapple*, are siblings.

Concept probabilities: $p(C_k)$'s we consider are $p(C_{entity}) = 1$ for the

concept	*colour*	*shape*	*size*	*taste*	*texture*
entity	uniform 0.04	uniform 0.025	uniform 0.01	uniform 0.05	uniform 0.05
plant product	uniform 0.05	uniform 0.04	uniform 0.04	uniform 0.1	uniform 0.1
fruit	uniform 0.075	round:0.4 6 others of 0.1 each	uniform 0.1	sweet:0.5 5 others of 0.1 each	smooth:0.35 medium:0.45 rough:0.20
apple	red:0.75 green:0.15 yellow:0.10	round:0.95 oblate:0.05	small:0.2 medium:0.7 large:0.1	sweet:0.9 sour: 0.1	smooth:0.9 medium:0.1
macintosh	red:0.9 green:0.1	round:0.95 oblate:0.05	medium:0.9 large: 0.1	sweet:0.9 sour:0.1	smooth:0.9 medium:0.1
crabapple	yellow:0.75 red:0.25	round:0.95 oblate:0.05	small:0.9 medium:0.1	sweet:0.01 sour:0.99	smooth:0.1 medium:0.9

Table 6.2: Property Probabilities $p(A_i = V_{ij}|C_k)$ in Illustration

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. Alternatively, we can visualize these conditional probabilities as numerical weights on the respective branches (*IS-A* link) emanating from $C_{sup(k)}$ to C_k. The concept probabilities which will be used in the ensuing computations have been tabulated in table 6.3.

Paths in the Hierarchy

We illustrate the computation of category utility for concepts in two paths in the plant product hierarchy under discussion. The concepts in the paths specified below are sequenced from the most general to the most specific.

$$path_1 = entity \rightarrow plant\ product \rightarrow fruit \rightarrow apple \rightarrow macintosh \quad (6.14)$$

concept	*conditional probability*	*probability value*
entity	$p(entity)$	1.00
plant product	$p(plant\ product \mid entity)$	0.33
fruit	$p(fruit \mid plant\ product)$	0.10
apple	$p(apple \mid fruit)$	0.05
macintosh	$p(macintosh \mid apple)$	0.15
crabapple	$p(crabapple \mid apple)$	0.10

Table 6.3: Concept Probabilities $p(C_k|C_{sup(k)})$ in Illustration

and

$$path_2 = entity \rightarrow plant\ product \rightarrow fruit \rightarrow apple \rightarrow crabapple \quad (6.15)$$

Computing Differential Category Utility

From the probabilities listed in the tables 6.2 and 6.3, (which in turn depict representations in taxonomic concept knowledge), we compute differential category utility for a concept according to our definition above (6.13) Given knowledge about C_k, we compute $DCU(C_k)$ (category utility of C_k) as the increase in score relative to being given knowledge about the superordinate of C_k.

We should compute the following quantities.

1. $t_1(k) = \sum_i \sum_j [p(A_i = V_{ij}|C_k)]^2$ (from table 6.2)
2. $t_2(k) = \sum_i \sum_j [p(A_i = V_{ij}|C_{sup(k)})]^2$ (from the same table)
3. $d(k) = t_1(k) - t_2(k)$
4. $DCU(C_k) = p(C_k|C_{sup(k)}) \cdot d(k)$ (from table 6.3 and expression (3) above)

We list the results of these computations in table 6.4. Note the special case of C_{entity}. For other concepts, t_2 is the difference between its own t_1-term and the t_1-term of its immediate superordinate. The conditional concept probabilities are reproduced here from the previous table (6.3).

C_k	$t_1(k)$	$t_2(k)$	$d(k)$	$p(C_k \mid C_{sup(k)})$	$DCU(C_k)$
entity	0.175	0.175	0.000	1.00	0.000
plant product	0.330	0.175	0.155	0.33	0.052
fruit	1.060	0.330	0.730	0.10	0.073
apple	3.680	1.060	2.620	0.05	**0.131**
macintosh	4.185	3.680	0.505	0.15	0.076
crabapple	4.150	3.680	0.470	0.10	0.047

Table 6.4: Differential Category Utilities $DCU(C_k)$

Canonical Salience in Object Description Generation

We visualize basic level preference as arising from prominence in a probability distribution, with the basic level corresponding to the maximum (most probable value). The labels associated with the concepts in a path of a hierarchy constitute a linguistic choice system. Canonical salience of a category in a path is the likelihood of its associated label being chosen to describe an object classified under that path of the concept hierarchy in a free-naming situation.

The typical patterns of distribution, as noted by cognitive psychologists, are as shown in figure 6.3. In this figure, the nodes are ordered from the most general level at the top to the most specific one at the bottom. The letter 'B' identifies the location of the basic level, where the canonical probability distribution is maximum. The curves shown are envelopes of discrete distributions.

Probabilistic definition of canonical salience in the free-naming of objects: Given a set of concepts $\{\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_n\}$ constituting a path in taxonomic concept knowledge such that $\mathcal{C}_k = \mathcal{C}_{sup(k+1)}$ for $k = 1 \ldots (n-1)$, define:

$$\mathcal{P}_{\mathcal{B}}(\mathcal{C}_k) = \frac{DCU(\mathcal{C}_k)}{\sum_{i=1}^{k} DCU(\mathcal{C}_k)} \tag{6.16}$$

so that $\mathcal{P}_{\mathcal{B}}(\mathcal{C}_k) \in [0,1]$ and $\sum_{i=1}^{n} \mathcal{P}_{\mathcal{B}}(\mathcal{C}_k) = 1$. The canonical salience of concept $\mathcal{C}_k$ is $\mathcal{P}_{\mathcal{B}}(\mathcal{C}_k)$.

(Note: We reserve the expression (with 'calligraphic $\mathcal{C}$') $\{\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_n\}$

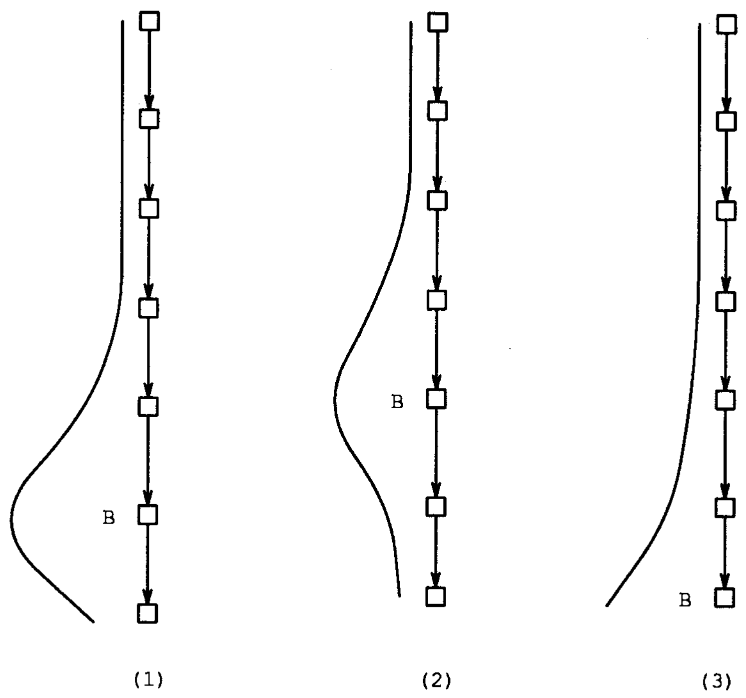

Figure 6.3: Patterns of Basic Level Location in Taxonomic Hierarchies

to denote a collection of concepts constituting a path in a taxonomy, and $\{C_1, C_2, \ldots, C_n\}$ to denote a collection of sibling-concepts belonging to a common superordinate concept C).

We tabulate the probability distributions of $\mathcal{P}_{\mathcal{B}}(\mathcal{C}_k)$ for the *macintosh* and *crabapple* paths in tables 6.5 and 6.6 respectively. We also tabulate alongside, in both cases, the category utilities $CU(\mathcal{C}_k)$, calculated according to the expression (6.11), and $DCU(\mathcal{C}_k)$, calculated according to (6.13). Notice the pattern of increase of $CU(\mathcal{C}_k)$ in these tables: as we go down the hierarchy from C_{entity}, CU takes the largest leap from *fruit* to *apple*; thereafter, the

$\mathcal{C}_k$	$CU(\mathcal{C}_k)$	$DCU(\mathcal{C}_k)$	$\mathcal{P}_{\mathcal{B}}(\mathcal{C}_k)$
entity	0.000	0.000	0.000
plant product	0.155	0.052	0.156
fruit	0.855	0.073	0.220
apple	3.505	0.131	**0.395**
macintosh	4.010	0.076	0.229

Table 6.5: Canonical Probabilities for the *macintosh* Path

$\mathcal{C}_k$	$CU(\mathcal{C}_k)$	$DCU(\mathcal{C}_k)$	$\mathcal{P}_{\mathcal{B}}(\mathcal{C}_k)$
entity	0.000	0.000	0.000
plant product	0.155	0.052	0.172
fruit	0.855	0.073	0.241
apple	3.505	0.131	**0.432**
crabapple	3.975	0.047	0.155

Table 6.6: Canonical Probabilities for the *crabapple* Path

increase is less, whether we go to *macintosh* or to *crabapple*. On the other hand, the differential category utility peaks at *apple* for both hierarchies, and predicts *apple* as the basic level.

6.3.4 Instantial Saliences and Utilities

According to the experimental results of Jolicoeur *et al.* [29], 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.

However, we can easily think of instances where preference shifts *upwards*, to a more general level: e.g., when the name is not known or sufficiently familiar to the speaker, or when she has trouble recalling the name, she may say, for

object	*context of evaluation of salience*
crabapple	apple
ostrich	bird
lime	fruit
peacock	bird
pool table	table

Table 6.7: Conceptual Context of Salience Evaluation

example, *pass me that thing!* or *who is that person?*, *thing* and *person* being general and all-inclusive names for the respective categories. We, however, limit our discussion to salience-induced shifts of preference in object-naming tasks towards subordinate levels of taxonomic hierarchies.

Atypicality of objects is assessed in the context of their immediate superordinate. In table 6.7 we list some (atypical) objects and the context in which they are salient.

More than one subconcept of a concept can be atypical: Consider a set of concepts $\{C_1, C_2, \ldots, C_n\}$ grouped under C. One or more of the C_k's could be salient in C. For example, in

$$bird = \{robin, raven, sparrow, peacock, ostrich, \ldots\}$$

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 saliences of the concepts C_k are 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 $S(peacock)$ and $S(ostrich)$ to be high (> 0.5).

The choice between C_k and C in a free-naming task is governed in part by a utility function u which apportions the number 1 between C_k and C as follows:

$$\begin{aligned}
&if\ S(C_k) > 0.5\ then \\
&\quad u(C_k) = S(C_k) \\
&else \\
&\quad u(C_k) = 0.5
\end{aligned} \tag{6.17}$$

and

$$u(C) = 1 - u(C_k) \tag{6.18}$$

Moreover, since the context in which the salience 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 $u(bird) = 0.2$.

Unlike in the problem of syntactic linearization (chapter 5), it seems that in the present problem, the utility functions have to 'know something' about the canonical distribution, e.g., that 'bird' is a basic level term. At first glance, this makes no sense. However, this is not a problem, because saliences of objects (= saliences of the most specific categories) are evaluated with respect to their immediate superordinates, without regard to whether those superordinates are basic levels or not. In the case of *crabapple*, it just so happens that its immediate superordinate, viz., *apple*, in the context of which *crabapple* is salient, is a basic level. On the other hand, consider *lime* which is the most specific category in its path: though *lime* is atypical (*alias* salient) for a *fruit* (its immediate superordinate), the basic level is *lime* itself, not *fruit*.

Quantifying Instantial Salience

We proposed that the function S evaluate the instantial saliences of the concepts C_i 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 will illustrate a calculation here which evaluates salience caused by atypicality using probabilistic knowledge in the concept taxonomy. Let us consider for our illustration the *crabapple* and the *macintosh* paths in the *apple* hierarchy. The relevant probabilistic attributes of the *apple* hierarchy have been tabulated earlier in this chapter in table 6.2.

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 representations of crabapple and apple.

As a first attempt, we considered computing this index as the difference in value probabilities for the respective attributes (taste or size), read off the concept representations of crabapple and apple, as follows:

$$p(taste = sour|crabapple) - p(taste = sour|apple) = 0.9 - 0.1 = 0.8 \quad (6.19)$$

and

$$p(size = small|crabapple) - p(size = small|apple) = 0.9 - 0.2 = 0.7 \quad (6.20)$$

However, these expressions are flawed because of a 'discounting' defect, explained below.

In hierarchically-structured probabilistic concept representations, the property probabilities at level $C_{sup(k)}$ reflect a summary of distributions of values in the subconcepts C_k into which $C_{sup(k)}$ is partitioned. That is, $p(A_i = V_{ij}|C_{sup(k)})$ is composed of contributions $p(A_i = V_{ij}|C_k)$ from all of its subconcepts, each one of which is potentially non-zero. In fact, it might even be that *all* of the contribution to the probability in $C_{sup(k)}$ is from a single subconcept C_k! Therefore, for instance, when we calculate the degree of unusualness/atypicality of the size of crabapple among apples as the difference in probability, given above (in 6.20), we end up unnecessarily subtracting the contribution of crabapples to the probability of small size of apples. The expression (6.19) is similarly defective. We will avoid the problem if we isolate the contribution of crabapples to the probability of small size of apples, and add it as a compensating term in our calculation of atypicality.

The relation between attribute (here, *size*) probabilities in apple and those in its subconcepts can be expressed as follows:

$$\begin{aligned} &p(size = small|apple) = \\ &\quad p(size = small|macintosh) \cdot p(macintosh|apple) + \\ &\quad p(size = small|crabapple) \cdot p(crabapple|apple) + \\ &\quad p(size = small|spartan) \cdot p(spartan|apple) + \ldots \end{aligned} \tag{6.21}$$

The contribution of crabapple to the probability of small size of apples is:

$$p(size = small|crabapple) \cdot p(crabapple|apple) \tag{6.22}$$

Both probabilities in this product are available in the *crabapple* representation.

We can now give an exact probabilistic measure of the degree of salience of the smallness of crabapples among apples:

$$\begin{aligned} &p(size = small|crabapple) - \\ &\quad [p(size = small|apple) - \\ &\quad p(size = small|crabapple) \cdot p(crabapple|apple)] \end{aligned} \tag{6.23}$$

In general, 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:

$$S_p(C_k, A_i) = p(A_i = V_{ij}|C_k) - [p(A_i = V_{ij}|C) - p(A_i = V_{ij}|C_k) \cdot p(C_k|C)] \tag{6.24}$$

A_i	$S_p(crabapple, A_i)$
taste	0.989
shape	0.095
texture	0.890
colour	0.725
size	0.790

Table 6.8: Property Saliences Arising from Atypicality in *crabapple*

Using the probabilities in the representations shown in tables 6.2 and 6.3, we get

$$S_p(crabapple, taste) = 0.989 \tag{6.25}$$

$$S_p(crabapple, size) = 0.790 \tag{6.26}$$

and

$$S_p(crabapple, colour) = 0.725 \tag{6.27}$$

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 example, $S_p(crabapple, shape) = 0.095$. That is, 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 are 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 calculations to follow.

Table 6.8 shows the probabilistic property salience (S_p) values for *crabapple*.

It may be noted that the saliences of all the properties in *macintosh* are below the threshold θ_s (= 0.5), as is the salience of the shape property of *crabapple*.

A_i	$w_0(A_i)$
taste	0.50
shape	0.20
texture	0.20
colour	0.05
size	0.05

Table 6.9: Intrinsicness Weights $w_0(A_i)$

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:

$$S(C_k) = \sum_i w(A_i) \cdot S_p(C_k, A_i) \tag{6.28}$$

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, even if probabilistic computations may indicate otherwise. 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 [32] (and presented in chapter 3). In the next chapter, we use property intrinsicness for ruling out anomalous similes. It is to be noted that w is a score assigned to the *attribute* (e.g., colour), not to the *value* (e.g., red) of the property.

Table 6.9 shows the weights $w_0(i)$ attached to the five properties under consideration. Note that $\sum_i w_0(A_i) = 1$. When some of the properties get left out of the salience calculations because of insufficient score (= falling below θ_s), the relevant weights are re-normalized as w_i, as the following calculations will illustrate.

A_i	$w(A_i)$	$S_p(crabapple, A_i)$
taste	0.625	0.989
texture	0.250	0.890
size	0.0625	0.790
colour	0.0625	0.725

Table 6.10: Renormalized Intrinsicness Weights and Property Saliences in *crabapple*

For the *crabapple* path, the *shape* property gets excluded from salience calculations by the threshold θ_s. The weights w_0 for the four other attributes that are counted in the salience computation are expressed in the normalized form as w, calculated as

$$w(A_i) = w_0(A_i) / \sum_{i:S_p(crabapple,A_i)\geq\theta_s} w_0(A_i) \tag{6.29}$$

so that $\sum_{i:S_p(crabapple,A_i)\geq\theta_s} w(A_i) = 1$.

For *crabapple*, we get the table (6.10) of w_i's and S_p's, the former quantity being derived from $w_0(A_i)$ and normalized as shown above (6.29) and the latter coming from the probabilistic computations shown in table 6.8.

Computing the salience S: Using these values, the probabilistic property salience of crabapple, indicating its degree of atypicality among apples, is computed as

$$S(crabapple) = \sum_i w(A_i) \cdot S_p(crabapple, A_i) = 0.935 \tag{6.30}$$

which is a high score, indicative of how well crabapple stands out among apples, given probabilistic knowledge of simple perceptual properties.

Evaluating the utility function: The utilities $u(crabapple)$ and $u(apple)$, which represent the instantial saliences of *crabapple* and *apple* respectively, are now calculated from $S(crabapple)$ according to expressions 6.17 and 6.18 given earlier. These utilities will interact with the probabilities $\mathcal{P}_\mathcal{B}$ derived from differential category utility (6.16) in choosing the label as a referring expression

C_k	$u(C_k)$
entity	0.000
plant product	0.000
fruit	0.000
apple	0.065
crabapple	0.935

Table 6.11: Utilities for the *crabapple* Path

in a free-naming task. Since $S(crabapple) > 0.5$,

$$u(crabapple) = S(crabapple) = 0.935 \tag{6.31}$$

and

$$u(apple) = 1 - u(crabapple) = 0.065 \tag{6.32}$$

The utilities for all concepts in the *crabapple* path are tabulated in table 6.11.

The utilities for the higher concepts (*fruit, plant product* and *entity*) are zero, since the salience judgment of *crabapple* is carried out relative to *apple*. When the model is sophisticated in the future to consider other factors in the generation context that could motivate the choice of the higher, more abstract terms, the utility function will be refined so as to be able to assign appropriate non-zero utilities to the higher concepts.

Computations for the macintosh path: For *macintosh*, we do not get any appreciable salience score, as table 6.12 shows.

Since the S_p's for all the attributes fall below the threshold, $S(macintosh) = 0$. Therefore, the utility function u assigns equal scores of 0.5 to *macintosh* and *apple*, indicating that there is nothing particularly striking or prominent in *macintosh* in the context of *apples*. The utilities for all concepts in the *macintosh* path are tabulated in table 6.13.

A_i	$S_p(macintosh, A_i)$
taste	0.135
shape	0.1425
texture	0.135
colour	0.285
size	0.335

Table 6.12: Property Saliences Arising from Atypicality in *macintosh*

$\mathcal{C}_k$	$u(\mathcal{C}_k)$
entity	0.000
plant product	0.000
fruit	0.000
apple	0.500
macintosh	0.500

Table 6.13: Utilities for the *macintosh* Path

6.3.5 Entry Level Effect as Expected Utility Maximization

In this subsection we derive and compute the effective salience of a potential choice (label in the naming task) as expected utility. 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 chapter 5. We are given:

- an object classified under a set of concepts $\{\mathcal{C}_1, \mathcal{C}_2, \ldots \mathcal{C}_n\}$ representing a path in taxonomic concept knowledge such that $\mathcal{C}_k = \mathcal{C}_{sup(k-1)}$ for $k = 1 \ldots (n-1)$.
- the probabilities $\mathcal{P}_{\mathcal{B}}(\mathcal{C}_k) \in [0, 1]$ for $k = 1 \ldots n$, representing the canonical salience of the concepts, and

	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	$u(C_n)$

(6.33)

Table 6.14: Matrix of $\mathcal{U}(A_{C_i}, S_j)$

- the utilities $u(C_k) \in [0,1]$ for $k = 1 \ldots n$, standing for the instantial salience of the concepts.

In decision theoretic terms, we have:

- a set of actions $A = \{A_{C_1}, A_{C_2}, \ldots, A_{C_n}\}$, 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: $S = \{S_1, S_2, \ldots, S_n\}$ where S_i is the state in which $u(C_i) = max(\{u(C_1), u(C_2), \ldots u(C_n)\})$

- 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 $i = 1 \ldots n, j = 1 \ldots n$.

- 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.

- An uncertain environment of decision making, characterized by the probability distribution $\mathcal{P}_{\mathcal{B}}$ with n values, one corresponding to each state. We can transfer the distribution on to the set of states and write: $p(S_j) = \mathcal{P}_{\mathcal{B}}(C_j)$.

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 6.14.

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.

$\mathcal{C}_k$	$\mathcal{P}_\mathcal{B}(\mathcal{C}_k)$	$u(\mathcal{C}_k)$	$EU(\mathcal{C}_k)$
entity	0.000	0.000	0.000
plant product	0.156	0.000	0.000
fruit	0.220	0.000	0.000
apple	0.395	0.500	**0.198**
macintosh	0.229	0.500	0.115

Table 6.15: Expected Utilities for the *macintosh* Path

Given the utilities tabulated in the matrix 6.14, we get:

$$EU(A_{\mathcal{C}_i}) = \sum_{j=1}^{n} p(S_j) \cdot \mathcal{U}(\mathcal{C}_i, S_j) = \mathcal{P}_\mathcal{B}(\mathcal{C}_i) \cdot u(\mathcal{C}_i) \tag{6.34}$$

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

Illustration

Continuing with our illustration of the *macintosh* and *crabapple* paths in the apple hierarchy, we calculate the expected utilities and display them in the tables 6.15 and 6.16. The probabilities are as in tables 6.5 and 6.6, and the utilities are according to the calculations 6.13 and 6.11. In the *macintosh* path, *apple* scores maximum expected utility, and is chosen as the label to refer to a *macintosh* apple in a free naming task.

For the *crabapple* path, as shown in table 6.16, *crabapple* attains maximum expected utility among all the concepts in the path, and is hence chosen as a referring expression. Although the basic level preference for *apple* is fairly

$\mathcal{C}_k$	$\mathcal{P}_\mathcal{B}(\mathcal{C}_k)$	$u(\mathcal{C}_k)$	$EU(\mathcal{C}_k)$
entity	0.000	0.000	0.000
plant product	0.172	0.000	0.000
fruit	0.241	0.000	0.000
apple	0.432	0.115	0.028
crabapple	0.065	0.935	**0.145**

Table 6.16: Expected Utilities for the *crabapple* Path

strong, it is overridden by a stronger preference for *crabapple* arising from its atypicality, thus bringing about a very tangible *entry level effect*, quite reminiscent of the passive voice selection shown in chapter 5!

Chapter 7

Property Salience in Comparisons

The contents of this chapter represent three dimensions of our interest in salience:

- formalizing some of the conceptions of salience, enumerated through the determinants in chapter 3
- formalizing the interactions in another language generation phenomenon in which salience has been implicated to play a role
- examining the suitability of conveying the salience interactions in the language generation phenomenon in terms of decision-theoretic criteria

The first two of these were part of our goals at the outset. The last dimension was activated at a much later stage as a worth while goal to pursue! We struck it rich on all three counts in the research reported in this chapter, also recently published in a condensed form in our recent paper [62].

The language generation phenomenon described in this chapter is the generation of comparison sentences of the form *A is like B*. Syntactically , these sentences are hardly daunting, and may not appear as a ***language generation*** problem. In fact, the knowledge that is brought to bear on the selection of the object of comparison is primarily conceptual. However, this hardly invalidates comparison generation as problem pertinent to language generation. Many language generation problems involve intensive dependence on conceptual knowledge. Communicating through comparisons, both literal and metaphorical, is common in day-to-day speech.

The task of communicating through comparisons crucially involves knowledge of saliences of object properties. The main determinants of salience formalized in this chapter are a combination of *pervasiveness* and *uniqueness* discussed in chapter 3: a property value of an object has high salience if it is pervasive among the instances of that object, and unique among other objects classified along with that object.

The chief aspects of our method of comparison generation are:

- *probabilistic concept representations* in the knowledge base, embedded in an associative network
- property salience modelled in terms of information-theoretic *redundancy* and *normalized redundancy*
- decision-theoretic re-formalization of redundancy as expected utility
- two different *cost* measures to avert miscommunication and anomaly, based on probabilistic knowledge and property intrinsicness
- *net expected utility* as the decision criterion

An interesting aspect of this research was that the formalization of property salience in information theoretic terms led automatically to a decision theoretic conception of the process through mathematical equivalence, rather than through subjective or heuristic assignment of utilities. The underlying probabilities, however, are subjectively assessed.

7.1 Introduction to Comparison Generation

In this chapter we present a method for generating simple comparison sentences with the aid of the notion of salience. 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*.

Iwayama *et al.* [27] used salience (based on information theory) computationally in their metaphor understanding system. We intuit that the role of salience in NLG should be correspondingly significant. In this chapter, we make that intuition precise and usable in language generation.

The main source of knowledge that controls decision making is a probabilistic conception of salience of essential, empirically observable properties among concrete objects. By *essential* properties we refer to those that are contrasted from *contingent properties* in the usual sense. Contingent properties are not represented; however, we do not claim that they are not pertinent in the comparisons that humans generate in their day-to-day speech. By *empirically observable* we mean simple perceptual properties like colour, taste, shape, texture and so on. 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*. Net expected utility combines the interpretation of utility as a positive score (reward) with that of cost as a negative score (penalty). The method proposed has been implemented in a generation system written in CProlog, which generates *Mary's* *cheeks are like apples* and a few of its variants.

7.2 Knowledge Representation

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.

These probabilistic representations were used in the previous chapter in modelling entry level preference in naming objects. For convenience, we reproduce below the definition of a probabilistic concept from chapter 6:

> 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: $p(colour = green|mango) = 0.15$. In our implementation, these representations are embedded in an associative network.

Generic representations of potential A-terms: The knowledge base also includes probabilistic, generic representations of concepts whose instances may be described in terms of a comparison. For example, probabilistic representation of *cheeks* is in the knowledge base. The generator may use it in communicating about *Mary's cheeks*. A diagram of a probabilistic concept is given in the previous chapter (figure 6.1).

Properties as Abstract Concepts

Attributes like *colour*, *texture* and so on are represented as abstract concepts, with links directed towards their instances (possible values) in any of their extensions. For example, *red* is a *colour*. Figure 7.1 (A) shows a portion of the knowledge base with attributes as abstract concepts.

Property values (e.g., *round*) are represented in the network as abstract concepts, with links directed towards their extensions (like *apples* and so on). Figure 7.1 (B) shows a portion of the knowledge base with property values as abstract concepts with links to their extensions.

Given the generator input, the properties specified in the generator input are first matched with the abstract concepts in the knowledge base, and through the *extensions* links, concepts possessing the properties (candidate B-terms) are accessed.

7.3 Modelling Property Salience

We introduce this section by summarizing the considerations about property salience in comparison generation. We discuss them in detail in the rest of this section.

- The concept (*apple*) is a candidate example of the property *colour:red* if the **most probable value** of the colour of apple is red.
- The number of other possible colours of apple (*green, yellow*) have the effect of either enhancing or suppressing the salience of *red* in apples.
- If *red* has high probability, and there are very few other possible values, salience of *red* is high.
- Even if *red* was the most probable colour, if several other colours are equally or nearly equally probable, *red* will not stand out much, and its salience will be low.

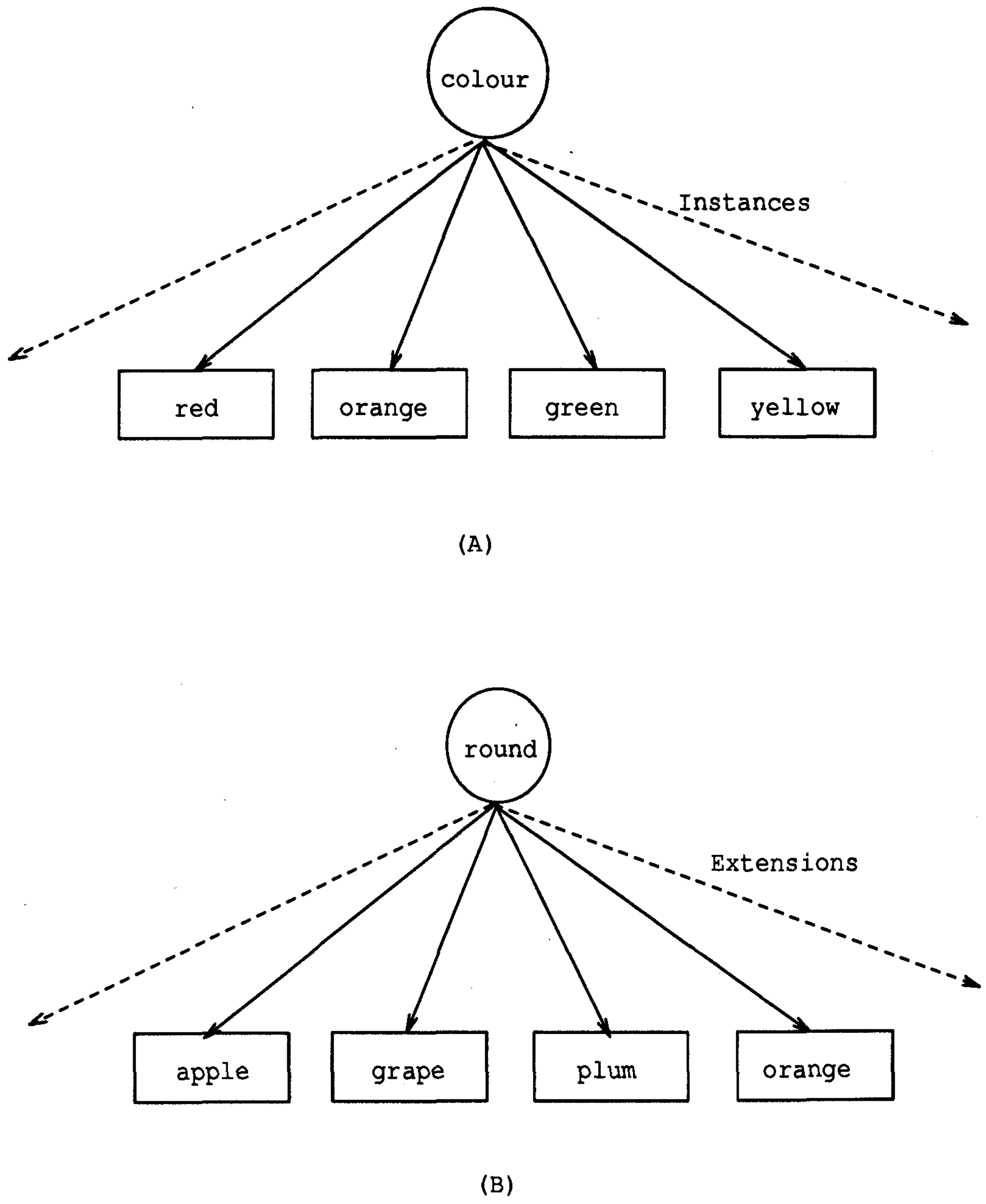

Figure 7.1: Attributes and Values as Abstract Concepts

- For *apple* to be selected as the example of *red*, the redness of *apple* should be more salient than the redness of say, *strawberry* and other candidate *fruit* which are mostly red.

7.3.1 Salience of the Most Probable Value

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 (maximum) 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. 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 *et al.* [27] 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 chapter. 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

$$H = \sum_{i=1}^{n} p_i log_c \frac{1}{p_i} \tag{7.1}$$

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

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 1.3527. If the colours were equiprobable, H would have been maximum, at 1.5850 (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

$$H_{rel} = \left(if\ n = 1\ then\ 0\ else\ \frac{H}{log_2 n} \right) \tag{7.2}$$

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

$$R = 1 - H_{rel} \tag{7.3}$$

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 0.3350. By comparison, the colour of *orange*, with a distribution of *[orange:0.80, yellow:0.10, green:0.10]*, has a (greater) R of 0.4183. As a final example, the colour of *grapefruit* with *[yellow:0.75, pink:0.25]*, has $R = 0.1887$. 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

$$NR(C_k, P) = \frac{R(C_k, P)}{\sum_{i=1}^{n} R(C_i, P)} \tag{7.4}$$

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 set 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.

Interpretations of information-theoretic quantities in terms of salience: redundancy and normalized redundancy are two major measures corresponding to two different contexts in which salience is evaluated in comparison generation:

- redundancy: among apples, how prominently does smoothness stand out, among the possible textures that apples can have?
- normalized redundancy: how prominently does the smooth texture of apples stand out among all objects (in our case, all *fruit*) which mostly have smooth texture?

Remarks: probabilistic salience depends on the dispersion of values in the context of possible values– the smaller the context, the smaller the magnitude of other candidates, and the larger the magnitude of the given candidate, the larger is the salience of the given candidate. In the case of redundancy, the magnitudes are probabilities of values. In the case of normalized redundancy, the magnitudes are redundancies of the distributions considered.

7.4 Decision Making in Comparison Generation

Although we could state the decision criterion in terms of information-theoretic criteria and proceed with the implementation, the mathematical equivalence that we found between redundancy and expected utility opened up the possibility of decision theory entering the problem (as a solution!). As a result, we were able to develop the algorithm much further, and bring in notions of penalty/cost to suppress anomalous comparisons.

7.4.1 Formal Equivalence of Redundancy and Expected Utility

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.

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

$$
\begin{aligned}
& 1 - \frac{1}{log_2 n} \cdot \sum_{i=1}^{n} p_i log_2 \frac{1}{p_i} \\
&= \frac{log_2 n + \sum_{i=1}^{n} p_i log_2 p_i}{log_2 n} \\
&= \frac{\sum_{i=1}^{n} p_i log_2 n + \sum_{i=1}^{n} p_i log_2 p_i}{log_2 n} \\
&= \sum_{i=1}^{n} p_i \cdot \left(\frac{log_2 (n p_i)}{log_2 n} \right) \qquad (7.5)
\end{aligned}
$$

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-theoretic problem of maximizing expected utility.

Figure 7.2 shows the variation of u_i with p_i for different values of n. When $n = 1$, p_i is also 1, and the reward u_i is maximum at 1. Within a given sample space of n items, the reward u_i increases monotonically with probability p_i. That is, the more the probability p_i of item i is in the sample space, the less of $1 - p_i$ will be divided among others, making item i more prominent. As n increases, the u_i curve cuts the p_i axis earlier, viz., at $\frac{1}{n}$. It can be seen from the figure that for a given $p_i < 1$ (draw an imaginary vertical line at the desired p_i), u_i increases with n.

Figure 7.3 depicts the variation of u_i with n for $p_i = 0.5$. The reward associated with selecting the item with a probability of p_i is more in larger sample spaces, as $1-p_i$ will have to be divided among more items, thus making item i stand out better in the sample space. The curves corresponding to $p_i > 0.5$ will lie above and to the left of the curve in this figure, and those corresponding to $p_i < 0.5$ will lie below and to the right, indicating that for a given n (draw an imaginary vertical line at the desired n), the reward u_i increases with p_i.

7.4.2 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: based on the value of P in the generator input, a search is initiated in the knowledge base and is directed towards concepts in which P occurs redundantly. Such concepts form a set C. 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. Table 7.1 shows the EUs computed in our implementation for the input entity *Mary's cheeks* and the descriptor *[shape:round]*. Note that the EUs add up to 1. The generator outputs *Mary's cheeks are like apples.*

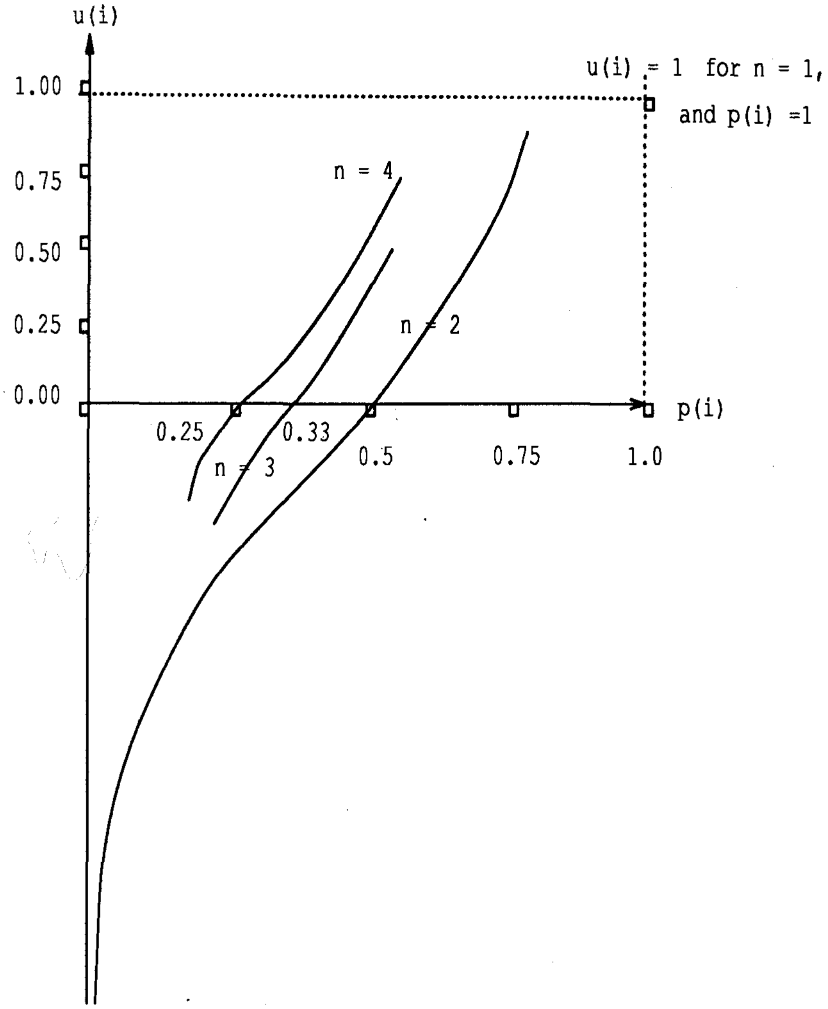

Figure 7.2: $u_i = (log_2\ np_i/log_2 n)\ vs\ p_i$ for $n = 1, 2, \ldots$

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

Table 7.1: Expected Utilities of Candidate B-terms for *[shape:round]*

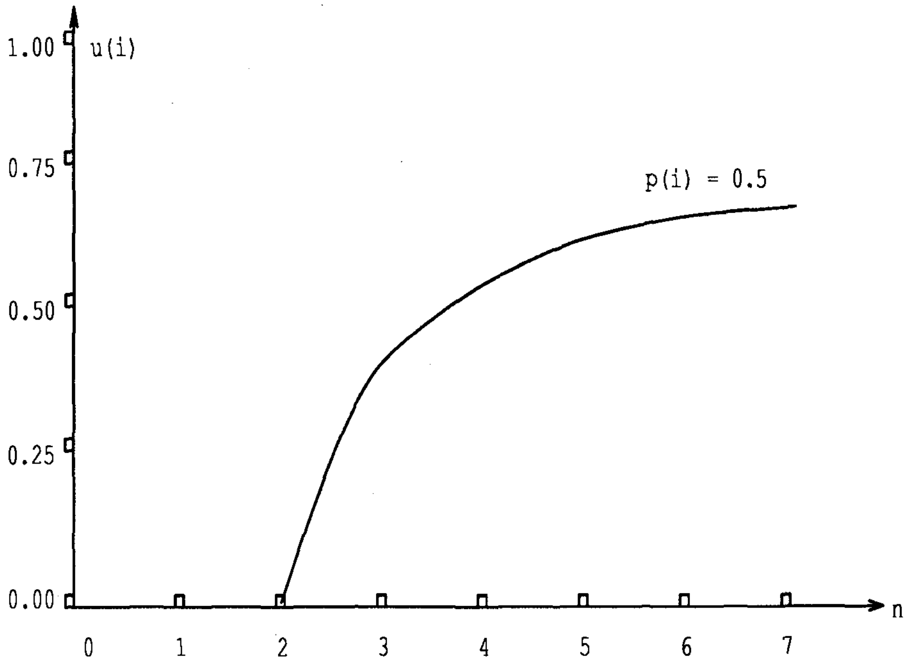

Figure 7.3: Envelope of $u_i = (log_2\ np_i/log_2 n)\ vs\ n$ for $p_i = 0.5$

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 7.2: A Portion of a Matrix of *EU*s

7.4.3 Comparisons Describing Two or More Properties

The problem of communicating more than one property about an entity is more complex than the one-property case discussed above. For example, consider communicating the fact that Mary's cheeks are red *and round*. It is not very likely (or reasonable to insist) that an object which has maximum redundancy for the colour of interest (red) will also be the one with the most redundant shape of interest (say round). What, then, is a good example of an object which is red, smooth and round?

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 7.2 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.

7.4.4 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* 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* is not quite 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).

7.4.5 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 described in it (here, *colour*) the *exception property*, 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. Sentences get lengthier as well, thus defeating one of the prime incentives of communicating through a comparison, viz., economy of expression. One exception is fairly common, as not always do we find *one* best example of all properties we want to communicate about an entity.

7.4.6 Quantifying the Cost of Zero Credit

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 to reflect the penalty for the difficulty of distorting one attribute value into another, and the less the cost the better. The cost will be less if it is in some way *easier* to 'distort' a red apple into a yellow apple.

For the input properties of *[colour:<u>yellow</u>, texture:smooth, shape:round]*, we examined the candidate B-term concepts to see what the exception properties

candidate B-term	*distorted B-term*	does it exist in KB?	with what probability?
apple	*yellow apple*	yes	0.10
lemon	*smooth lemon*	yes	0.20
plum	*yellow plum*	yes	0.10
grapefruit	*no distortion needed*		
grape	*yellow grape*	no	0
banana	*round banana*	no	0
peach	*smooth peach*	yes	0.05

Table 7.3: Exception Properties for *yellow, smooth, round*

are, and whether the concepts 'distorted into' possessing the exception property existed in the knowledge base, in the probabilistic representation. Table 7.3 shows what we found.

Probabilistic Cost of Zero Credit

We propose the following as a probabilistic measure of cost for the running example:

$$\begin{aligned} &Cost(apple, colour) = \\ &p(colour = red|apple) \cdot (1 - p(colour = yellow|apple)) \end{aligned} \tag{7.6}$$

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

$$\begin{aligned} &Cost(C, P) = \\ &p(P = MPV|C) \cdot (1 - p(P = EV|C)) \end{aligned} \tag{7.7}$$

The cost is also a measure of the degree of anomaly or miscommunication.

Observe another important aspect of our method: an object in the knowledge base is a candidate object of comparison if it has some positive credit for one or more properties to be communicated. We choose candidate objects first based on their positive credit (TEU); later, we adjust the credits with penalties (costs). The entire knowledge base provides the potential space in which objects with positive credit (TEU) are found. On the other hand, mismatch is evaluated only within the (far smaller) context of objects which possess some degree of resemblance with the A-term.

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 by Langacker [32] 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!

Total Expected Cost of Zero Credit

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.

Expected Cost of Zero Credit The *expected cost* (EC) of an entry in the TEU matrix for the property P of concept C is therefore

$$p_{zc} \cdot Cost(C, P) + (1 - p_{zc}) \cdot 0 \tag{7.8}$$

where p_{zc} is the probability that the credit to TEU from is 0 from an arbitrary input property P.

Evaluating p_{zc} Recall that an object in the knowledge base is a candidate object of comparison if it has some positive credit (EU) for one or more properties to be communicated. Given that an object is a potential B-term, we know that at least one of the input properties must have contributed to TEU. p_{zc} is the probability that a certain input property did not contribute to TEU, given that the concept is a candidate B-term. This condition can be modelled as an n-input OR-gate. p_{zc} is the probability that an input of the OR-gate is zero, given that the gate output is 1.

Assume that the OR-gate inputs are $A_1, A_2, \ldots, A_n$, and the output is called C. We have, by definition:

$$p_{zc} = p(A_i = 0 | C = 1) \tag{7.9}$$

for some $i \in [1, n]$.

Assume that the gate inputs are independent, and that each one of them, in isolation, is equally likely to be 1 or 0. That is, for any $i \in [1, n]$,

$$p(A_i = 0) = p(A_i = 1) = 0.5 \tag{7.10}$$

By Bayes' rule, we can write:

$$p(A_i = 0 | C = 1) \cdot p(C = 1) = p(C = 1 | A_i = 0) \cdot p(A_i = 0) \tag{7.11}$$

The gate output is 1 for all but one (the one where all inputs are 0) of the input combinations. That is,

$$p(C = 1) = \frac{2^n - 1}{2^n} \tag{7.12}$$

When one of the inputs (say, A_i) is given as 0, the gate output is 1 in all but one of the possible configurations of the other $n - 1$ inputs. That is,

$$p(C = 1|A_i = 0) = \frac{2^{n-1} - 1}{2^{n-1}} \tag{7.13}$$

Substituting the probabilities in Bayes' equality above, we get:

$$p(A_i = 0|C = 1) \cdot \frac{2^n - 1}{2^n} = \frac{2^{n-1} - 1}{2^{n-1}} \cdot \frac{1}{2} \tag{7.14}$$

Hence we get:

$$p(A_i = 0|C = 1) = p_{zc} = \frac{2^{n-1} - 1}{2^{n-1}} \tag{7.15}$$

When n properties are intended to be communicated, the expected cost of zero credit EC is therefore:

$$\begin{aligned} p_{zc} \cdot Cost(C, P) + (1 - p_{zc}) \cdot 0 = \\ \left(\frac{2^{n-1} - 1}{2^n - 1}\right) \cdot Cost(C, P) \end{aligned} \tag{7.16}$$

Total Expected Cost We can now compute the *total expected cost* (TEC) of a candidate example as the sum of individual ECs for each property.

Decision Criterion with Cost Incorporated

At this stage: The decision criterion is now one of maximizing

$$TEU - \alpha \cdot TEC \tag{7.17}$$

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.

7.4.7 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 re-labelled as TEC_{A-B}.

7.4.8 Net Expected Utility as Decision Criterion

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

$$NEU = TEU - \alpha_1 \cdot TEC_{A-B} - \alpha_2 \cdot TEC_{B-A} \qquad (7.18)$$

where α_1 and α_2 are non-negative.

With NEU, for the input properties *yellow* and *smooth*, we generate *Mary's cheeks are like lemons* for $\alpha_1 = \alpha_2 = 1$.

Relation to Similarity Metrics

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:

- Tversky's metric in his *contrast model* of similarity, given in [73]:

$$s(a,b) = \theta f(A \cap B) - \alpha f(A - B) - \beta f(B - A) \tag{7.19}$$

  This expression evaluates the degree of similarity between two objects a and b as a linear combination of the measures of their common and distinct features. A and B represent the property sets of a and b respectively, and f is a measure of salience.

- Ortony's metric, proposed as an *imbalance model* of similarity, given in [52], and [54], elegantly superseding Tversky's:

$$s(a,b) = \theta f^b(A \cap B) - \alpha f(A - B) - \beta f(B - A) \tag{7.20}$$

  In the imbalance model, the measure of the shared attributes $f^b(A \cap B)$ is evaluated with respect to the B-term. Refer to [54] for a discussion comparing the two models.

Our expression for NEU, while on the surface resembles Tversky's metric, is much more like Ortony's, since the TEU is evaluated with respect to the B-term. However, our NEU is a bit different from Ortony's as well: while Ortony evaluates the measure with respect to all properties in the B-term concept, we evaluate only with respect to properties in the generator input. Other properties in the knowledge base could reinforce or interfere with the similarity judgment.

7.5 Discussion

The measure of salience used in this chapter 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. To the best of our knowledge, this problem has not been considered in NLG so far. We conclude this chapter with short discussions on related aspects.

7.5.1 Comparison with Comparison Comprehension

The problem in natural language understanding may be stated as follows: given A *is like* B, and probabilistic conceptual knowledge on (generic) A and B, infer the properties of B that the generator wished to communicate. The problem in NLG (as developed in the chapter) is: given A and a set of properties of A to be communicated, select B. The set of properties is a list of attribute:value pairs. In NLU one searches for the grounds of a given comparison. In NLG, given the grounds to be communicated, one searches for an object of comparison.

While a lot of knowledge that is brought to bear on the two problems is common (for example, probabilistic conceptual knowledge, property salience measure), the problems faced in modelling the processes are different. For example, in NLU, given that Mary's cheeks are like apples, several properties of both cheeks and apples will be activated, and the problem is to use knowledge of salience in a matching process to determine which properties of cheeks were meant by the speaker. Threshold definitions on salience are required to cut off irrelevant properties that may be activated.

In generation, the problem is to decide what constitutes a good object of comparison. Problems arise due to the possibility of miscommunication and anomalous communication. Appropriate measures are needed to suppress inappropriate comparisons. Moreover, while the portion of the knowledge base in which a comparison is interpreted is quite localized (to the source and target concepts), generation of a comparison involves a search through a potentially-large portion of the knowledge base; appropriate preference criteria are required to control the direction in which the generator goes looking for an object of comparison in the knowledge base.

7.5.2 Comparison with the Work of Iwayama et al.

As we mentioned earlier in this chapter, our work takes its main comparison sentence *Mary's cheeks are like apples* from the work of Iwayama *et al.* [27] on metaphor comprehension. Their *amount of information* (AIP) corresponds to our *redundancy*, while their *difference property* (DP) corresponds to our *normalized redundancy*. However, they use the product $AIP \cdot DP$ as a measure of salience. It is not clear why the product should be used, rather than just DP. In NLG, using DP yields the same preference ordering as using $AIP \cdot DP$. Hence, the multiplication seems unnecessary.

While Iwayama *et al.* depend only on probabilistic salience, we use the non-probabilistic heuristic of property intrinsicness as well. Furthermore, we have enriched our formalization by relating redundancy to expected utility.

7.5.3 Exception Clauses in Comparisons

In human speech, exception clauses in comparisons are not used solely as a compensatory mechanism to avert miscommunication. Deliberate comparisons with exceptions are useful in giving new information, as in *A zebra is like a horse, except that it has stripes*, when confronted with a child's question of what a zebra is (Miller [47]). Comparisons with exceptions as a device for enabling the hearer to assimilate new knowledge. Similar strategies are used by McCoy [37] in correcting user misconceptions based on the user model.

Other kinds of exception clauses indicating not a mismatch of property labels, but one of *degrees* or *intensities*. For example,

Mary's cheeks are like apples, except...they are a lot redder.
Mary's cheeks are like apples, except..they are not nearly as smooth.

7.5.4 Additional Knowledge for Generating Comparisons

Influence of Other Determinants of Salience

First, other determinants of salience like *vividness* and *imageability* 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.

Preference for Concrete Objects

Preference for concrete objects can help generate *indirectly grounded* comparisons like *proud as a peacock*: though peacocks are not quite proud (see general discussion of Ortony *et al.* in [54]), 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.

Accommodating Extensions in Our Model We are using normalized redundancy as a measure of salience, assuming that the context of possible objects is given as $C_{sup(k)}$, the superordinate of C_k in the *IS-A* hierarchy:

$$NR(C_k, P) = \frac{R(C_k, P)}{\sum_{i=1}^{n} R(C_i, P)} \tag{7.21}$$

When applied to apples, for instance, the tacit assumption is that $p(fruit) = 1$.

In a more sophisticated model of selection, we will refine the above expression for NR as:

$$NR(C_k|C_{sup(k)}, P) \cdot p(C_{sup(k)}) \qquad (7.22)$$

$p(C_{sup(k)})$ is a meaningful term through which we can bring in some other independent variables governing preference. This probability will indicate the degree of preferability of searching for an object of comparison among (say) fruits in our generator.

7.5.5 Comparisons, Similes and Metaphor

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.

The 'distance' between the A-term and B-term concepts in the knowledge base is an index of the degree of metaphoricity. If the A-term and B-term concepts are close to each other, and in the limit, under the same parent node in the *IS-A* hierarchy, the comparison is literal. The more remote they are, the more tangible the metaphoricity is.

Another good way of assessing the metaphoricity of a comparison is to examine the elided form of the comparison. Observe the acceptability of the elided comparisons in the following pairs: Compare

(a) Her cheeks are like her sister's cheeks
(b) * Her cheeks are her sister's cheeks

(a) Her cheeks are like apples
(b) ?? Her cheeks are apples

(a) He's like a headache
(b) He's a headache.

There appears to be a continuous cline from literalness to metaphoricity.

Property Independence Assumption

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 [15]. 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.

Chapter 8

Conclusion

We embarked on this research, guided solely by the intuition that the salience heuristic of Conklin and McDonald should be applicable in general to a variety of NLG decisions. We thought that salience was sufficiently significant as an object of theoretical exploration unto itself. However, we had no details on what shape a theory of salience would take. All that was known about salience in NLG apart from the work of Conklin and McDonald lay in brief, scattered references.

Our exploration has paid off, in the sense of offering us with a very valuable pair of categories, viz., canonical salience and instantial salience, and revealing the nature of their synergistic and competitive interactions. These categories have admitted a seductively simple formalization by means of decision theoretic elements: everything seemed to fall in place instantly, all at once! Canonical saliences were probabilities, instantial saliences were utilities, determinants of salience were utility functions, and effective salience was expected utility. Several of the findings are more in the nature of discoveries rather than inventions–which, personally, was quite rewarding.

Because of lack of standard methodological guidelines to approach salience, we invented our own methodology which suited the investigation as well as our temperament. The aim of the effort was to mould a theory with long-term significance and endurance; therefore, it was quite important to ask the 'right' questions and seek the 'right' distinctions initially. However, until some results are obtained, the 'rightness' of questions or methodology cannot be affirmed. We believe that we have been rewarded with several interesting results.

Delineating canonical salience and instantial salience first helped us focus on their interactions. Seeking explicit psycholinguistic support helped us arrive at a formal model of the interactions, which is subtle, yet mathematically very simple. With the canonical-salience interactions and decision theory, we were able to 'travel to' another phenomenon, (viz., basic level- entry level effects in object naming) new to us, and new to salience, where we were able to use our distinction and formal framework to flesh out the phenomenon computationally.

We then formalized salience of properties of objects in comparison generation using information theory, and found that decision theoretic nature of interactions revealed itself through mathematical equivalence. Throughout the work we have woven complex threads connecting works in psycholinguistics, cognitive psychology, computational linguistics and so on, proving disparate experiments as arriving at similar implications for salience, and using potentially-useful clues, wherever they came from. Apart from focussing on the interactions, we have also modelled some determinants of salience formally: aspects like property salience and atypicality, using information theory and probability theory.

Much remains to be done, largely because much has been opened up:

- *Empirical and Experimental*: Interesting corpus work can be conducted with large bodies of real text to obtain the canonical probabilities and to explore the question of basicness of constructs. Work in the systemic linguistic tradition such as those of Halliday [21] and Nesbitt and Plum [49] already relates corpus work as a source of knowledge about relative frequencies, which can shed light on the probabilities. The systemicists have related probability to variation in language. However, they have not explored the computational implications, nor have they used probabilities in language generation, and much less, view them as psycholinguistically significant. In addition to empirical work with linguistic data, psycholinguistic experiments could also be performed to test some of the ideas presented here. However, we are not completely familiar with the state of art in experimental psycholinguistics and what its powers and limits are. Some of the well-developed techniques employed in decision analysis to elicit preferences (of human decision makers) for constructing utility functions could be applied to assess subjective judgments of salience levels for use in algorithms. Decision theory does not deal with subjectivity by pretending that it does not exist, but rather accepts it as a common fact in decision making, and attempts to provide techniques that can help improve the correct use of subjective numerical estimations.

- *Technical*: Work can begin on *structuring* the decisions on the basis of the probabilities and utilities that enter into them. Such work will yield new architectures for generation. Not making any architectural commitments at the start has helped us study salience for its own sake. At this stage, after proposing decision theory as the framework, the architectural implications of our findings for computer generators can be explored. The multitude of numbers that one has to deal with in a quantitative approach can be a deterrent to ventures of large-scale implementations; however, the profusion of interacting numbers is (mercifully) manageable in more 'linguistic' tasks like linearization than, say, comparison generation, where the 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.

- *Theoretical*: The question that comes to mind at this stage is: *so, where* is *the theory?* At this stage, the theory of salience has an easily recognizable content, but not so easily recognized a shape. The theory is contained in the terms and distinctions we have developed through the thesis: determinants, canonical-, instantial- and effective saliences, synergistic and competitive interactions, and other elements like salience propagation which we mentioned only fleetingly. A candidate vehicle to convey the theory has been obtained through extensive arguments and clue-seeking: decision theory. Various manifestations of salience have been shown to have mathematical interpretation in terms of probability, utility, expected utlity, total expected utility and net expected utility-the last two were proposed in chapter 7. Total expected utility represents an aggregation of effective salience scores from independent sources. Net expected utility incorporates penalty for interference from 'unwanted' salience. Decision theory is a vast area with varied dimensions, and we have only scratched its surface. Decision theory has elaborate facilities for studying or conveying the varied facets of preference and uncertainty in decision making situations. A continuation of this work will proceed to put together the salience terms and distinctions as an organized theory on the one hand, and explore the suitability of other aspects of decision theory to model salience dynamics in NLG, on the other.

References

[1] H. Alshawi. *Memory and Context for Language Interpretation.* Cambridge University Press, 1987.

[2] W. Claassen. Generating Referring Expressions in a Multimodal Environment. In *Proceedings of the Sixth International Workshop on Natural Language Generation*, pages 247–262, 1992.

[3] E.J. Conklin. Data-driven Indelible Planning of Discourse Generation Using Salience. COINS Technical Report 83-13, University of Massachusetts, 1983.

[4] 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.

[5] W.E. Cooper and J.R. Ross. *World Order*, In *Functionalism*, Grossman, R.E., James San, L. and Vance, T.J., Editors, pages 63–111. Chicago Linguistic Society, 1975.

[6] R. Dale. *Generating Referring Expressions in a Domain of Objects and Processes.* PhD thesis, University of Edinburgh, 1988.

[7] Y. Falk. Constituency, Word Order, and Phrase Structure Rules. *Linguistic Analysis*, 11:331–360, 1983.

[8] G. Fauconnier. *Mental Spaces: Aspects of Meaning Construction in Natural Language.* MIT Press, 1985.

[9] C.J. Fillmore. *The Case for Case Reopened*, In *Syntex and Semantics 8: Grammatical Relations*, Cole, P. and Sadock, J.M., Editors, pages 59–81. Academic Press, 1977.

[10] D.H. Fisher. Knowledge Acquisition via Incremental Conceptual Clustering. *Machine Learning*, 2:139–172, 1987.

[11] S.T. Fiske and S.E. Taylor. *Social Cognition.* Addison-Wesley, 1984.

[12] J.H. Flavell and H.M. Wellman. *Meta-memory*, In *Perspectives on the Development of Memory and Cognition*, Kail, Jr. R.V. and Hagen, J.W., Editors, pages 3–33. Erlbaum, 1976.

[13] G.B. Flores d'Arcais. *Some Perceptual Determinants of Sentence Construction*, In *Studies in Perception*, Flores d'Arcais, G.B., Editor. Martello-Guinti, Milan, 1975.

[14] 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.

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

[16] C.H. Haagen. Synonymity, Vividness, Familiarity, and Association Value Ratings of 400 Pairs of Common Adjectives. *Journal of Psychology*, 27:453–463, 1949.

[17] E. Hajičova, P. Kubon, and V. Kubon. Hierarchy of Salience and Discourse Analysis and Production. In *Proceedings of the 13th COLING, Helsinki, Volume 3*, pages 144–148, 1990.

[18] E. Hajičova and J. Vrbova. On the Salience of the Elements of the Stock of Shared Knowledge. *Folia Linguistica*, 15:291–303, 1981.

[19] E. Hajičova and J. Vrbova. *On the Role of the Hierarchy of Activation in The Process of Natural Language Understanding*, In *COLING'82*, Horecky, J., Editor, pages 107–113. North-Holland, 1982.

[20] M.A.K. Halliday. Notes on Transitivity and Theme in English. *Journal of Linguistics*, 3:199–214, 1967.

[21] M.A.K. Halliday. *Towards Probabilistic Interpretations.* Paper given at the Sixteenth International Systemic Congress, Hanasaari, Finland, 1989. To appear in Eija Ventola (Ed), Selected Papers from the Sixteenth International Systemic Congress.

[22] R. Hasan. *The Ontogenesis of Ideology: An Interpretation of Mother-Child Talk*, In *Semiotics, Ideology, Language*, Threadgold, T. *et al.*, Editors, pages 125–146. Sydney Association for Studies in Society and Culture, 1986.

[23] B.J. Haslett. *Communication: Strategic Action in Context.* Lawrence Erlbaum Associates, 1987.

[24] M. Henrion, J.S. Breese, and E.J. Horvitz. Decision Analysis and Expert Systems. *AI Magazine*, 12(4):64–91, 1991.

[25] A Herskovits. *Language and Spatial Cognition.* Cambridge University Press, 1986.

[26] E. Hovy. *Generating Natural Language under Pragmatic Constraints.* Lawrence Erlbaum Associates, 1988.

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

[28] P.N. Johnson-Laird. The Interpretation of the Passive Voice. *Quarterly Journal of Experimental Psychology*, 20:69–73, 1968.

[29] P. Jolicoeur, M. Gluck, and S. Kosslyn. Pictures and Names: Making the Connection. *Cognitive Psychology*, 16:243–275, 1984.

[30] R.L. Keeney and H. Raiffa. *Decisions with Multiple Objectives: Preferences and Value Tradeoffs.* John Wiley and Sons, 1976.

[31] G. Lakoff and M. Johnson. *Metaphors We Live By.* University of Chicago Press, 1980.

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

[33] G. Leech. *Principles of Pragmatics.* Longman, London, 1983.

[34] R.D. Luce and P. Suppes. *Preference, Utility, and Subjective Probability,* In *Handbook of Mathematical Psychology, Volume III*, Luce, R.D., Bush, R.R. and Galanter, E., Editors, pages 249–410. John Wiley and Sons, 1983.

[35] W. Mann and C. Matthiessen. Nigel: A Systemic Grammar for Text Generation. Technical Report ISI/RR-83-105, USC/Information Sciences Institute, 1983.

[36] W.C. Mann. An Overview of the Penman Text Generation System. In *Proceedings, AAAI'83*, pages 261–265, 1983. Also appears as USC/Information Sciences Institute RR-83-114.

[37] K. McCoy. Reasoning on a Highlighted User Model to Respond to Misconceptions. *Computational Linguistics*, 14:52–63, 1988.

[38] D.D. McDonald. Description Directed Control: Its Implications for Natural Language Generation. *Computers and Mathematics*, 9(1):111–129, 1983.

[39] D.D. McDonald. Type-driven Suppression of Redundancy in the Generation of Inference-rich Inputs. In *Proceedings of the Sixth International Workshop on Natural Language Generation*, page 73-88, 1992.

[40] D.D. McDonald and E.J. Conklin. Salience as a Simplifying Metaphor in Natural Language Generation. In *Proceedings of AAAI'82*, pages 75–78, 1982.

[41] D.D. McDonald and J.D. Pustejovsky. TAG's as a Grammatical Formalism for Generation. In *Proceedings of the 23rd Annual Meeting of the ACL, Chicago*, pages 94–103, 1985.

[42] D.D. McDonald, M.M. Vaughan, and J.D. Pustejovsky. *Factors Contributing to Efficiency in Natural Language Generation*, In *Natural Language Generation: New Results in Artificial Intelligence, Psychology and Linguistics*, Kempen, G., Editor, pages 159–181. Martinus Nijhoff, 1987. Proceedings of the Third International Workshop on Natural Language Generation, Nijmegen, 1986.

[43] K.R. McKeown. *Text Generation: Using Discourse Strategies and Focus Constraints to Generate Natural Language Text.* Cambridge University Press, 1985.

[44] I.A. Mel'čuk. *Meaning-Text Models: A Recent Trend in Soviet Linguistics.* The Annual Review of Anthropology, 1981.

[45] 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.

[46] M.W. Meteer, D.D. McDonald, S.D. Anderson, D. Forster, L.S. Gay, A.K. Huettner, and P. Sibun. Mumble-86: Design and implementation. COINS Technical Report 87-87, Univerity of Massachusetts, 1987.

[47] G.A. Miller. *Images and Models, Similes and Metaphors*, In *Metaphor and Thought*, Ortony, A., Editor, pages 202–250. Cambridge University Press, 1979.

[48] G.A. Miller and P.N. Johnson-Laird. *Language and Perception.* The Belknap Press of Harvard University Press, 1976.

[49] C. Nesbitt and G. Plum. *Probabilities in a Systemic-Functional Grammar: The Clause Complex in English*, In *New Developments in Systemic Linguistics, Volume 2*, Fawcett, R.P. and Young, D.J., Editors, pages 6–38. Pinter Publishers Limited, 1988.

[50] S. Nirenburg and I. Nirenburg. A Framework for Lexical Selection in Natural Language Generation. In *Proceedings of the 12th COLING, Budapest*, pages 471–475, 1988.

[51] R. Nisbett and L. Ross. Assigning Weights to Sata: The 'Vividness Criterion'. In *Human Inference: Strategies and Shortcomings of Social Judgment*, pages 43–62. Prentice-Hall, 1980.

[52] A. Ortony. Beyond Literal Similarity. *Psychological Review*, 86(3):161–179, 1979.

[53] A. Ortony. *Similarity in Similes and Metaphors*, In *Metaphor and Thought*, Ortony, A., Editor, pages 186–201. Cambridge University Press, 1979.

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

[55] C.E. Osgood. *Salience Dynamics and Unnaturalness in Sentence Production*, In *Lectures on Language Performance*, Osgood, C.E., pages 185–210. Springer-Verlag, 1980.

[56] C.E. Osgood. *Where do sentences come from?*, In *Semantics: An Interdisciplinary Reader in Philosophy, Linguistics and Psychology*, Steinberg, D.D. and Jakobovits, L.A., Editors, pages 497–529. Cambridge University Press, 1971.

[57] C.E. Osgood. *Lectures on Language Performance.* Springer-Verlag, 1980.

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

[59] C.E. Osgood, W. May, and M.S. Miron. *The Measurement of Meaning.* University of Illinois Press, 1957.

[60] T. Pattabhiraman and N. Cercone. Selection: Salience, Relevance and the Coupling between Domain-level Tasks and Text Planning. In *Proceedings of the Fifth International Workshop on Natural Language Generation, Dawson, PA, 1990*, pages 79–86, 1990.

[61] T. Pattabhiraman and N. Cercone. Salience in Natural Language Generation. In *Proceedings of the IJCAI-91 Workshop on Decision Making throughout the Generation Process, Sydney*, pages 34–41, 1991.

[62] 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.

[63] C. Pollard and I.A. Sag. *Information-Based Syntax and Semantics, Volume 1: Fundamentals.* CSLI, 1987. CSLI Lecture Notes No. 13.

[64] E. Reiter. A New Model of Lexical Choice for Nouns. *Computational Intelligence*, 7(4):240–251, 1991.

[65] E. Rosch and C.B. Mervis. Family Resemblances: Studies in the Internal Structure of Categories. *Cognitive Psychology*, 7(4):573–604, 1975.

[66] I.G. Sarason and B.R. Sarason. *The Importance of Cognition and Moderation Variables in Stress*, In *Toward a Psychology of Situations*, Magnusson, D., Editor, pages 195–210. Lawrence Erlbaum, 1981.

[67] L.J. Savage. *Foundations of Statistics.* Wiley, 1954.

[68] E.E. Smith and D.L. Medin. *Categories and Concepts.* Harvard University Press, 1981.

[69] S.N. Sridhar. *Cognition and Sentence Production.* Springer-Verlag, 1988. Springer Series in Language and Communication, 22.

[70] J. Svartvik. *On Voice in the English Verb.* Mouton, The Hague, 1966.

[71] L. Talmy. *Figure and Ground in Complex Sentences*, In *Universals of Human Language, Volume 4*, Greenberg, J.H. *et al.*, Editors, pages 625–649. Stanford University Press, 1978.

[72] S.E. Taylor and J.V. Wood. *The Vividness Effect: Making a Mountain out of a Molehill?*, In *Advances in Consumer Research, Volume 10*, Bagozzi, R.P. and Tybout, A.T., Editors, pages 540–542. Association for Consumer Research, Ann Arbor, MI, 1983.

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

[74] H. Uszkoreit. Constraints on Order. Technical Report CSLI-86-46, CSLI, Stanford, CA, 1986.

[75] H. Uszkoreit. Linear Precedence in Discontinuous Constituents: Complex Fronting in German. Technical Report CSLI-86-47, CSLI, Stanford, CA, 1986.

[76] J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior.* Princeton University Press, 1947.

[77] D. von Winterfeldt and W. Edwards. *Decision Analysis and Behavioral Research.* Cambridge University Press, 1986.

[78] D. Waltz. *Generating and Understanding Scene Descriptions*, In *Elements of Discourse Understanding*, Joshi, A. *et al.*, Editors, pages 266–282. Cambridge University Press, 1982.

[79] L. Wanner and J.A. Bateman. A Collocational Based Approach to Salience-Sensitive Lexical Selection. In *Proceedings of the Fifth International Workshop on Natural Language Generation, Dawson, PA, 1990*, pages 31–38, 1990.

[80] P.L. Yu. *Multiple-Criteria Decision Making: Concepts, Techniques and Extensions.* Plenum Press, 1985.