A New Model for Lexical Choice for Open-Class Words

Ehud Reiter†

Aiken Computation Laboratory
Harvard University
Cambridge, Mass 02138

Abstract

The lexical choice process should be regarded as a constraint satisfaction problem: the generation system must choose a lexical unit that is *accurate* (truthful), *valid* (conveys the necessary information), and *preferred* (maximal under a preference function). This constraint-based architecture allows a clean separation to be made between what the system *knows* of the object or event, and what the system wishes to *communicate* about the object or event. It also allows lexical choices to be biased towards *basic-level* (Rosch 1978) and other preferred lexical units.

1. Introduction

Lexical choice for open-class words has typically been regarded as a matching or classification problem. The generation system is given a semantic structure that represents an object or event, and a dictionary that represents the semantic meanings of the *lexical units* (Zgusta 1971) of the target language; it then chooses the lexical unit (or set of lexical units) that best matches the object or event. This paper proposes an alternative lexical choice architecture, in which the lexical choice process is regarded as a constraint satisfaction problem: the generation system must choose a lexical unit that is *accurate* (truthful), *valid* (conveys the necessary information), and *preferred* (maximal under a preference function).[1] This constraint-based architecture is more robust than classification systems. In particular, it allows a clean separation to be made between what the system *knows* of the object or event, and what the system wishes to *communicate* about the object or event; and it allows lexical choices to be biased towards *basic-level* (Rosch 1978) and other preferred lexical units.

Throughout this paper, it will be assumed that both lexical units and objects/events are represented as classes in a KL-ONE type taxonomy (Brachman and Schmolze 1985). For example, the lexical unit *Bachelor* might be represented as the generic class (*Human* with role value restrictions *Sex:Male, Age-status:Adult, Married:False*); and the object *Terry* might be represented as the individual class (*Human* with role fillers *Sex:Male, Eye-color:Brown, Birthplace:Chicago, Employer:IBM*, ...). Default attributes as well as definitional information can be associated with lexical units; this is essential for making appropriate lexical choices (Section 5). Figure 1 shows a sample taxonomy that will be used for most of the examples in this paper. Lexical units (e.g., *Bachelor*) are shown in bold font, while objects (e.g., *Terry*) are shown in italic font. Role value restrictions (VR's), such as *Sex:Male* for *Man*, are listed textually instead of displayed graphically, to simplify the complexity of the diagram; default attributes (e.g., *Can-fly:True* for *Bird*) are listed in italic font. Basic-level classes (e.g., *Man*) are underlined.

Section 2 of the paper discusses classification-based systems and some of the problems associated with them. Section 3 introduces the proposed constraint-based system; Section 4 looks in more detail at the *lexical preferences* used by the system; and Section 5 briefly discusses the need for default attributes in the semantic representations of lexical units. The constraint-based lexical choice system has been incorporated into the FN system (Reiter 1990), which generates certain kinds of natural language object descriptions. FN uses some additional *preference rules* that primarily affect NP formation; these rules are not discussed in this paper.

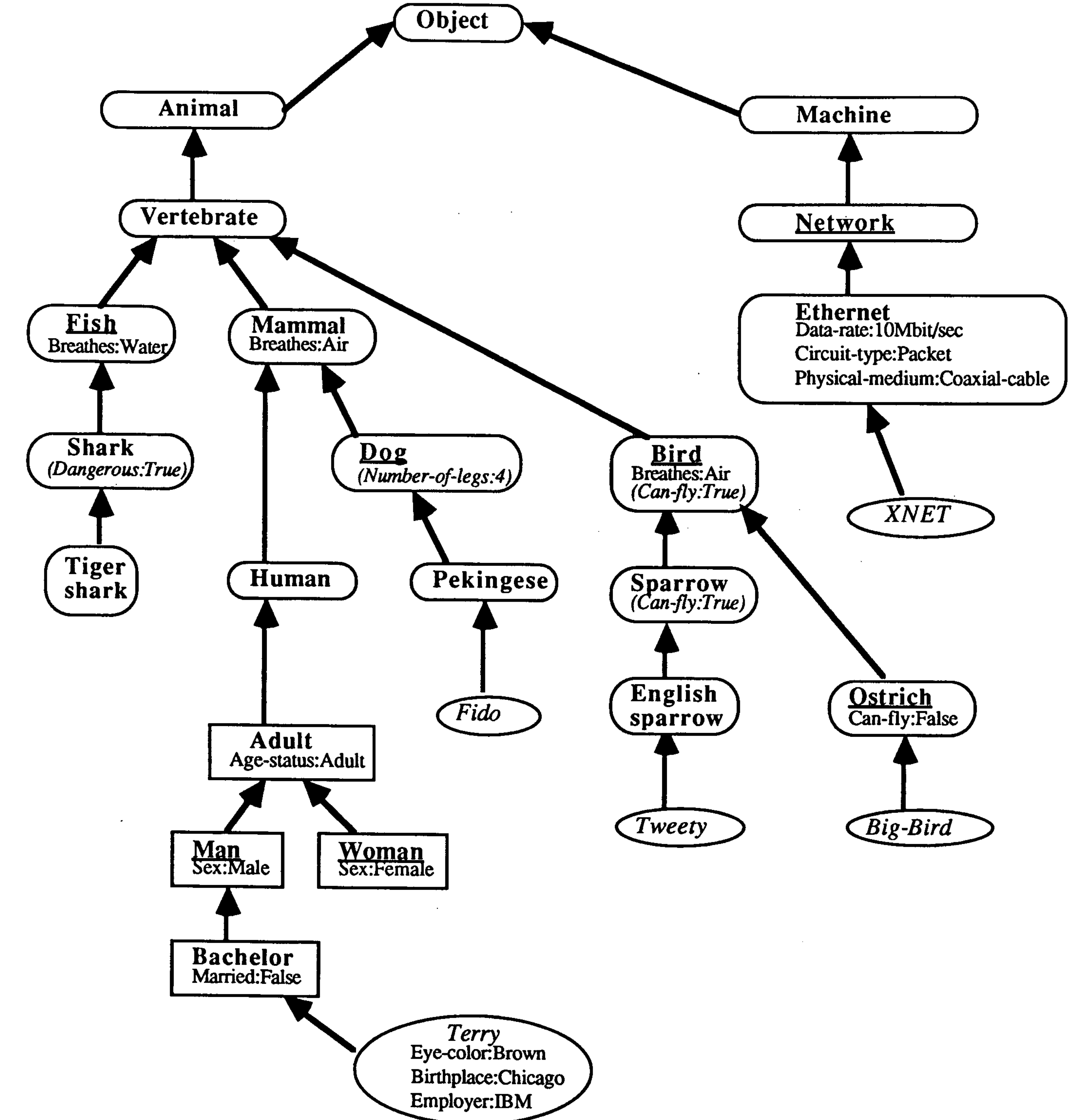
2. Lexical Choice as Classification

The two major approaches (to date) for lexical choice have been *discrimination nets* and *structure mapping systems*. Both of these approaches can be regarded as classification/matching architectures, where a classifier is given an object or event, and is asked to find an appropriate lexical unit that fits that object or event.

Discrimination nets (e.g., Goldman 1975; Pustejovsky and Nirenburg 1987) are basically decision trees. They are typically used as high-speed 'compiled' classifiers that select the most specific lexical unit that subsumes

† Currently at the Department of Artificial Intelligence, University of Edinburgh, 80 South Bridge, Edinburgh EH1 1HN, Scotland. E-mail: reiter@aipna.edinburgh.ac.uk

[1] This paper does not examine the kind of collocational and selectional constraints discussed by Cumming (1986) and Nirenburg and Nirenburg (1988).



(Key)

Primitive Class

Defined Class

Individual

Basic Level Class — defining role VR

Lexical Unit Class — *(default role filler)*

Object Class

is subsumed by →

Figure 1: Objects and Lexical Units in a Taxonomy

the target object or event. For instance, looking at some of Goldman's examples, the event *Ingest(John,Milk027)*, which can be represented in KL-ONE as (*Ingest* with VR's *actor:John* and *theme:Milk027*), has as its most specific subsuming lexical unit (*Ingest* with VR *theme:Liquid*), and thus is lexically realized as "drink". Similarly, the action *Ingest(Bear036,Fish802)*, which can be represented in KL-ONE as (*Ingest* with VR's *actor:Bear036* and *theme:Fish802*), has (*Ingest* with VR's *agent:Non-human-animal* and *theme:Solid*) as its most specific subsumer in a taxonomy of German lexical units, and thus is realized, in German, as "fressen".

Structure-mapping systems (e.g., Jacobs 1987; Iordanskaja *et al.* 1988; note that different terminology is used in different papers) take as input a semantic structure that needs to be communicated to the user, search for pieces of the input structure that are equivalent to lexical units, and then replace the matched structure by the corresponding lexical unit. The matching and substitution process continues until the semantic structure has been completely reformulated in terms of lexical units. For example, the structure (Human (:sex male) (:age-status adult) (:wealth high)) might be mapped into the structure ("man" (:attribute "rich")), and hence lexically realized as "rich man". In KL-ONE terms, the matching process can be considered to be a search for a class definition that uses only classes and role VR's that can be realized as lexical units; e.g., the above example essentially redefines the class (*Human* with role VR's *Sex:Male, Age-status:Adult, Wealth:High*) as the equivalent class ("*man*" with VR "*rich*"), where the lexical unit "*man*" represents the class (*Human* with role VR's *Sex:Male, Age-status:Adult*), and the lexical unit "*rich*" is equivalent to the role VR *Wealth:High*.

Recently, the machine translation group at CMU has proposed an alternative lexical choice system that is based on a variant of *nearest neighbor classification* (Center for Machine Translation 1989; Nirenburg *et al.* 1987). In the CMU system, both objects and lexical units are treated as points or regions in a feature space, and the classifier works by choosing the lexical unit that is closest to the target object, using a fairly complex distance (matching) metric (collocation constraints are also taken into consideration). For example, the object (*Human* with VR's *Sex:Male* and *Age:13*) would be judged closest to the lexical unit (*Human* with VR's *Sex:Male* and *Age:range(2,15)*), and thus would be realized as "boy".

All of the above classification-based lexical-choice architectures[2] suffer from two basic flaws:

- they do not allow a clean separation to be made between what the system *knows*, and what it wishes to *communicate*;
- they do not provide a clean mechanism for allowing the lexical choice process to be biased towards *preferred* lexical units.

These failures may lead classification-based systems to choose inappropriate lexical units that carry unwanted *conversational implicatures* (Grice 1975), and therefore mislead the user.

2.1. One Input vs Two Inputs

Classification-based systems take as their input a single set of attributes about the object/event being lexicalized, and use this set of attributes to select a matching classification. However, lexical choice systems should look at *two* input sets of attributes: the set of object/event attributes that are relevant and need to be conveyed to the user, and the set of attributes that constitute the system's total knowledge of the object/event being lexicalized.

A lexical choice system that looks only at the system's domain knowledge about the object/event, and ignores the set of relevant attributes, may choose inappropriate lexical items that carry unwanted *relevance* conversational implicatures. In particular, a system that simply selects the most specific lexical unit that subsumes the object/event (as many discrimination net systems do) may mislead the user by choosing lexical units that are too specific. For example, consider the following exchange:

1) A: "Is Terry a woman?"
2a) B: "No, Terry is a *man*"
2b) B: "No, Terry is a *bachelor*"

B's communicative goal is simply to inform A that *Terry* has the attributes {*Human, Age-status:Adult, Sex:Male*}, so utterance (2a) is an appropriate response. A lexical choice system that simply selected the most specific lexical unit that subsumed *Terry* would generate utterance (2b), however. Utterance (2b) is inappropriate, and would probably lead A to infer the (incorrect) conversational implicature that B thought that Terry's marital status was relevant to the conversation.

A lexical choice system that looks only at the attributes being communicated, and ignores the system's

[2] Individual lexical-choice systems can, of course, be augmented with special code that addresses some of these issues; the claim is that the classification-based lexical-choice architectures do not easily or naturally deal with these problems.

general domain knowledge about the object/event, may also make inappropriate lexical choices that lead to unwanted conversational implicatures. For example, suppose A wished to communicate to B that XNET was a *Network* with the attributes {*Data-rate:10Mbit/sec, Circuit-type:Packet-switched*}. Consider three possible lexicalizations:

3a) "XNET is a *network*"
3b) "XNET is a *10 Mbit/sec packet-based network*"
3c) "XNET is an *Ethernet*"

Utterance (3c) is the most appropriate utterance (assuming the user has some domain knowledge about Ethernets). Utterance (3a), however, would be generated by a system that simply chose the most specific lexical unit that subsumed {*Network, Data-rate:10Mbit/sec, Circuit-type:Packet-switched*}.[3] This utterance fails to fulfill the communicative goal of informing the reader that the network has the attributes {*Data-rate:10Mbit/sec, Circuit-type:Packet-switched*}, and is therefore unacceptable. Utterance (3b) would be generated by a structure-mapping system that chose a lexical unit according to the above strategy, and then added explicit modifiers to communicate attributes that were not implied by the lexical class.[4] This utterance successfully communicates the relevant information, but it also implicates, to the knowledgeable hearer, that *XNET* is *not* an Ethernet — because if it was, the knowledgeable hearer would reason, then the speaker would have used utterance (3c).

2.2. Preferred Lexical Units

Certain lexical units, in particular those that represent *basic-level classes* (Rosch 1978), are preferred and should be chosen whenever possible. Cruse (1977) and others have suggested that the failure to use a basic-level class in an utterance will conversationally implicate that the basic-level class *could not* have been used. For example, consider the following utterances:

4) A: "I want to flood room 16 with carbon dioxide"
5a) B: "Wait, there is an *animal* in the room"
5b) B: "Wait, there is a *dog* in the room"
5c) B: "Wait, there is a *Pekingese* in the room"

Assume the object in question is *Fido*, and A's communicative goal is simply to inform B that *Fido* has the attributes {*Animal, Breathes:Air*}, and hence would be adversely affected if the room was flooded with carbon dioxide. Utterances (5a), (5b), and (5c) all fulfill this communicative goal (assuming that *Breathes:Air* is a default attribute of *Animal*), but utterance (5b) is preferred because *Dog* is a basic-level class. Utterance (5a) is odd because the use of the *superordinate* class *Animal* implicates, according to Cruse's hypothesis, that the animal in question is not a *Dog, Cat*, or other commonly known type of animal (or at least the speaker does not know that the animal is a member of one of these species); utterance (5c) is odd because the use of the *subordinate* class *Pekingese* implicates that it is somehow relevant that the animal is a Pekingese and not some other kind of dog. If both of these implicatures are incorrect, the speaker should choose the lexical unit *Dog* if he wishes to avoid misleading the hearer.

It should be pointed out that the strategy of simply always picking a basic-level class that subsumes the object/event will not work, because it ignores the system's communicative goals. For instance, a system that followed the basic-level strategy would, in the situation of example 3, generate utterance (3a) or (3b). Both of these are inappropriate and implicate, to the knowledgeable user, that utterance (3c) could not have been used, i.e., that XNET is not an *Ethernet*.

3. Lexical Choice as Constraint Satisfaction

The above problems can be avoided by regarding lexical choice as a constraint-satisfaction task instead of a classification task. More precisely, the task of choosing an appropriate open-class lexical unit should be formalized as follows:

Input:

- *Entity*: a taxonomy class that represents the system's knowledge of the object or event being lexicalized.
- *To-Communicate*: a set of predicates (attributes) that represent the relevant information about the object that needs to be communicated to the user.

[3] Another possibility is choosing the most general lexical unit that is subsumed by the attributes being communicated. However, this cannot be done by a system that ignores the object and only looks at the attributes being communicated, because such a system would not know which lexical units accurately described the object. For example, if there were two classes *Ethernet* and *Applenet* that had the attributes {*Network, Data-rate:10Mbit/sec, Circuit-type:Packet-switched*}, the system could only decide whether to generate "Ethernet" or "Applenet" by determining which of these classes subsumed the object being described (e.g., "Ethernet" should be used to describe XNET). See also example 5, where the most appropriate lexical unit that informs the hearer that *Fido* has the attributes {*Animal, Breathes:Air*} is "dog", not "mammal" or "animal".

[4] In this example, the 'lexical choice' system is assumed to capable of forming a complete NP. In general, it is often difficult to separate the task of selecting a single word from the task of forming a complete phrase.

Output: A lexical unit *Lex* that is a member of the knowledge-base taxonomy, and that satisfies the following constraints:

- *Accurate*: *Lex* must be a truthful description of *Entity*. Formally, *Lex* must subsume *Entity*.
- *Valid*: The use of *Lex* in an utterance must inform the user that the predicates in *To-Communicate* hold for *Entity*. Formally, every predicate in *To-Communicate* must either be inferrable from the definition of *Lex* (e.g., subsume *Lex*), or be a default attribute that is associated with *Lex*.
- *Preferred*: *Lex* must be a maximal element of the set of accurate and valid lexical units under certain *lexical preference rules* (Section 4).

In other words, the lexical choice system is given two inputs, which represent the system's knowledge of the object or event, and the relevant information about that object or event that needs to be communicated to the user; and is expected to produce as its output a maximal lexical unit (under the lexical preference rules) that is truthful and conveys the relevant information.

The constraint-based system makes appropriate lexical choices in each of the previous examples:

- *Entity* = *Terry*, *To-Communicate* = {*Human*, *Sex:Male*} (example 2). Both *Man* and *Bachelor* are accurate and valid lexical units. *Man* is chosen, because it is basic-level and therefore preferred.
- *Entity* = *XNET*, *To-Communicate* = {*Network*, *Data-rate:10Mbit/sec*, *Circuit-type:Packet-switched*} (example 3). *Ethernet* is chosen, because it is the only accurate and valid lexical unit.
- *Entity* = *Fido*, *To-Communicate* = {*Animal*, *Breathes:Air*} (example 5). Accurate and valid lexical units include *Animal*, *Mammal*, *Dog*, and *Pekingese*. *Dog* is chosen, because it is basic-level.

4. Preferences Among Lexical Classes

If several lexical units are accurate and valid, a set of *lexical preferences rules* is used to select the lexical unit the system will utter. The preference for basic-level classes was previously mentioned (Section 2.2), but it is complicated by *entry-level* effects (Section 4.1). Additional lexical preferences include the *length/subset* preference (Section 4.2). Combined, the lexical preference rules impose a *lexical preference hierarchy* on the lexical units in the knowledge base. Figure 2 shows part of the lexical preference hierarchy that is associated with the knowledge base of Figure 1.

4.1. Basic-Level vs Entry-Level Preferences

Hirschberg (1985) has suggested that it may be better to use Jolicoeur *et al.*'s (1984) notion of *entry level* classes instead of Rosch's *basic level* classes. The difference is that under the entry-level hypothesis, which category is unmarked (i.e., which category may be used without generating a conversational implicature) may depend on how atypical the object is. For example, consider:

(the speaker points to a robin)
7a) "Look at the *bird*"
7b) "Look at the *robin*"

(the speaker points to an ostrich)
8a) "Look at the *bird*"
8b) "Look at the *ostrich*"

Under the basic-level hypothesis, a category is either basic-level or it is not, and if it is basic-level, then it is always the unmarked way of referring to any object that belongs to it. Therefore, under this hypothesis utterances (7a) and (8a) are both unmarked and carry no conversational implicatures, since *Bird* is a basic-level category for most urban Americans. Under the entry-level hypothesis, in contrast, while a basic-level category is the unmarked way of referring to 'normal' members of the category, it may not be the unmarked way of referring to atypical members. Instead, a more specialized category may be the unmarked way of referring to atypical members. Thus, under the entry-level hypothesis, even if utterance (7a) was the unmarked way of referring to robins (which are typical birds), utterance (8b) could still be the unmarked way of referring to ostriches (which are atypical birds).

The lexical-choice system can allow for entry-level effects if it allows any lexical unit to be marked as basic-level in the taxonomy, but then only considers the lowest such marked class to be a true basic-level (and hence lexically-preferred) class for an object. More precisely, if an object has two subsumers A and B that are both marked as basic-level classes, and A subsumes B, then the system should only treat B as a lexically-preferred class for the object. For example, in Figure 1 *Bird* and *Ostrich* are both marked as basic-level. Therefore, the lexical-choice system should treat *Bird* (but not *Sparrow*) as a lexically-preferred class for *Tweety* (a *Sparrow*), and *Ostrich* (but not *Bird*) as a lexically-preferred class for *Big-Bird* (an *Ostrich*).

4.2. Length/Subset Preferences

A lexical unit A is almost always preferred over a lexical unit B if A's surface form uses a subset of the words used by B's surface form (this can be considered

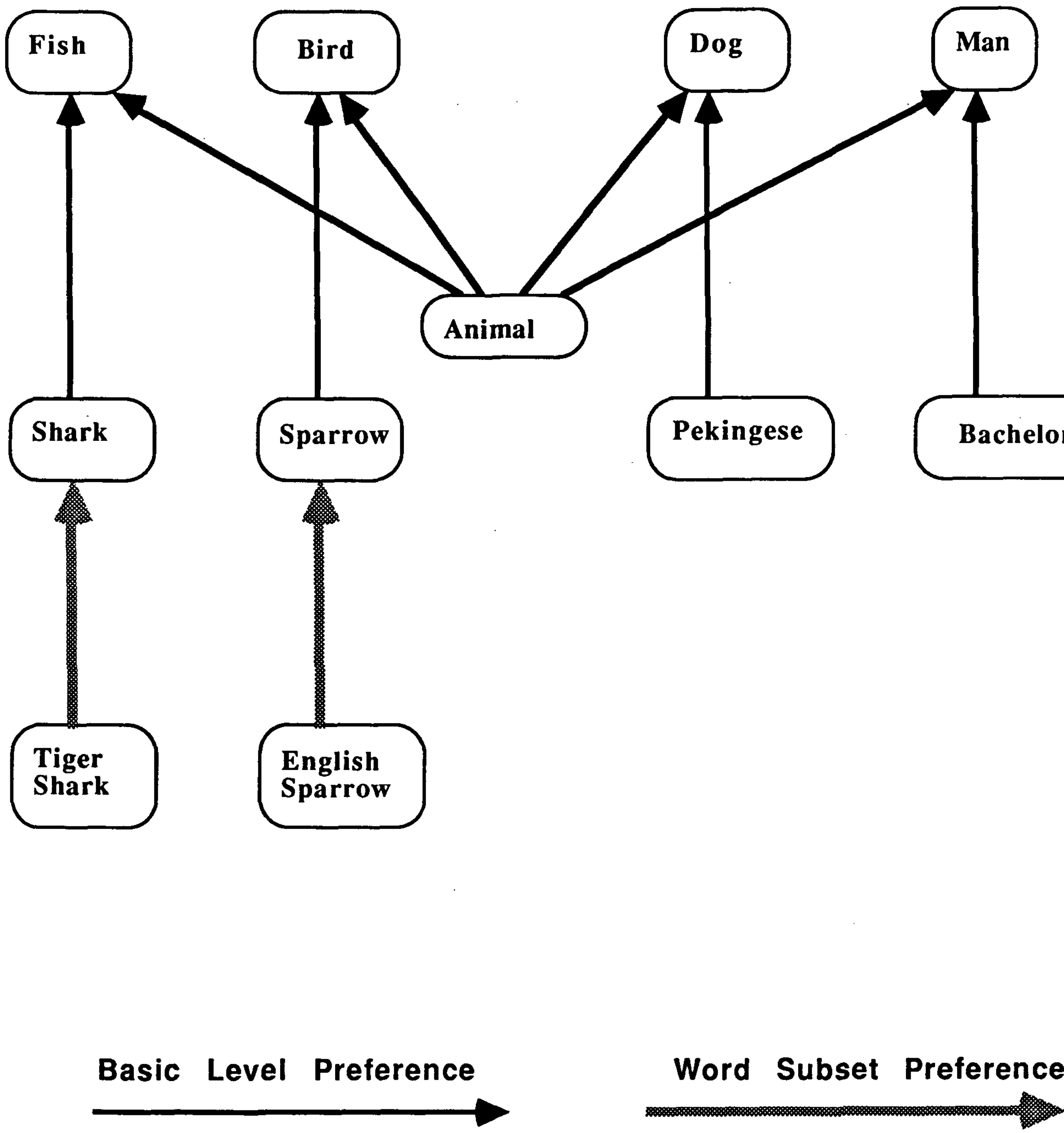


Figure 2: Some of the Lexical Preferences from Figure 1

to be a consequence of Grice's *maxim of quantity* (Grice 1975)). Consider, for example,

9a) "Don't go swimming; there is a *shark* in the water"
9b) "Don't go swimming; there is a *tiger shark* in the water"

According to the subset lexical preference rule, lexical unit *Shark* is preferred over lexical unit *Tiger-shark*. Therefore, the use of utterance (9b) carries the conversational implicature that utterance (9a) could not be used, i.e., that it was relevant that the animal was a *Tiger-shark* and not some other kind of *Shark*. A hearer who heard utterance (9b) might infer, for example, that the speaker thought that tiger sharks were unusually dangerous kinds of sharks. If no such implicature was intended by the speaker, then he should use utterance (9a), not utterance (9b).

A stronger version of this preference rule would be to prefer lexical unit A to lexical unit B if A's surface form used fewer open-class words than B's surface form. This would, for example, correctly predict that *Dog* is preferred over *Great-Dane*, and that *Flower* is preferred over *Rocky-Mountain-iris*. This preference is usually accurate, but it does fail in some cases. For example, it is questionable whether *Porsche* is preferred over *Sports-car*, and doubtful whether *Mammal* is preferred over *Great-Dane*.

There are cases where the basic-level preference conflicts with (and takes precedence over) both the subset and the length preferences. Such conflicts are probably rare, because psychological and linguistic findings suggest that basic-level classes are almost always lexically realized with single words (Rosch 1978; Berlin *et al.* 1973). However, there are a small number of basic-level classes that have multi-word realizations, and this can lead to conflicts of the above type. Consider, for example,

10a) "Joe has a *machine*"
10b) "Joe has an *appliance*"
10c) "Joe has a *washing machine*"

Washing-machine is probably basic-level for most Americans. Therefore, utterance (10c) is preferred over utterances (10a) and (10b), despite the fact that the length preference suggests that utterances (10a) and (10b) should be preferred over utterance (10c), and the subset preference suggests that utterance (10a) should be preferred over utterance (10c).

4.3. Other Lexical Preference

There are lexical preferences that are not captured by either the basic-level preference or the subset/length preference. For example, suppose the speaker wished to refer to two animals, a horse and a cow. Consider the difference between

11a) "Look at the *animals*"
11b) "Look at the *mammals*"
11c) "Look at the *vertebrates*"

None of the above are basic-level classes (*Horse* and *Cow* are basic-level for most urban Americans). Therefore, neither the basic-level nor the length/subset rules indicate any preferences among the above. However, it seems clear that utterance (11a) is much preferable to utterance (11b), and that utterance (11b) is probably preferable to utterance (11c). In addition, the use of utterances (11b) or (11c) seems to implicate that utterance (11a) could not have been used.

5. Default Attributes

One final point is that the representation of the semantics of lexical units must include default attributes as well as definitional information. These defaults may represent domain knowledge (e.g., birds typically fly) or useful conventions that have evolved in a particular environment (e.g., most computers at Harvard's Aiken Computation Lab run the UNIX operating system). Systems that ignore default attributes may make inappropriate lexical choices, and therefore generate utterances that carry unwanted conversational implicatures.

For example, if *To-Communicate* was {*Bird*, *Can-fly:True*}, and *Entity* was *Tweety*, consider the difference between

12a) "Tweety is a *bird*"
12b) "Tweety is a *bird that can fly*"

If the generation system ignored default attributes, it would have to generate something like utterance (12b). Utterance (12b) sounds odd, however, and a person who heard it might infer unwanted and unintended conversational implicatures, e.g., that some other bird under discussion was not able to fly. Utterance (12a) is much better, but it can only be generated by a generation system that takes into consideration the fact that *Can-fly:True* is a default attribute of *Bird*.

For another example, suppose an NLG system wished to inform a user that a particular computer was a VAX that ran the UNIX operating system and the Latex text processor (i.e., *To-Communicate* = {*VAX*, *Operating-*

system:UNIX, Available-software:Latex}). Consider two possible utterances:

13a) "Hucl is a *VAX that runs Latex*"
13b) "Hucl is a *UNIX VAX that runs Latex*".

Utterance (13a) is acceptable, and indeed expected, if the user thinks that *Operating-system:UNIX* is a default attribute of *VAX*'s in the current environment (e.g., at the Aiken Computation Lab). In a different environment, where users by default associate *Operating-system:VMS* with VAX's, utterance (13a) would be misleading and unacceptable, and utterance (13b) should be generated.

6. Conclusion

This paper has proposed a lexical choice system that searches for lexical units that are *accurate, valid*, and *preferred* with respect to the information the generation system wishes to communicate (*To-Communicate*), and the object or event being lexicalized (*Entity*). This system is more robust than discrimination nets and other existing classification-based lexical choice systems, and in particular is less likely to make inappropriate lexical choices that lead human readers to infer unwanted conversational implicatures. The improved performance is largely a consequence of the fact that the system allows a clean separation to be made between what the system *knows*, and what it wishes to *communicate*; and the fact that the system allows lexical choice to be biased towards *preferred* lexical units.

Acknowledgements

Many thanks to Joyce Friedman, Barbara Grosz, Chris Mellish, Stuart Shieber, and Bill Woods for their help. This work was partially supported by a National Science Foundation Graduate Fellowship, an IBM Graduate Fellowship, and a contract from U S West Advanced Technologies. Any opinions, findings, conclusions, or recommendations are those of the authors and do not necessarily reflect the views of the National Science Foundation, IBM, or U S West Advanced Technologies.

References

Berlin, B.;Breedlove, D.; and Raven, P. 1973 General Principles of Classification and Nomenclature in Folk Biology. *American Anthropologist* 75:214-242.

Brachman, R. and Schmolze, J. 1985 An overview of the KL-ONE knowledge representation system. *Cognitive Science* 9:171-216.

Center for Machine Translation 1989 *KBMT-89 Project Report*. Carnegie-Mellon University.

Cruse, D. 1977 The pragmatics of lexical specificity. *Journal of Linguistics* 13:153-164.

Cumming, S. 1986 The Lexicon in Text Generation. ISI Research Report ISI/RR-86-168. Information Sciences Institute, University of Southern California.

Goldman, N. 1975 Conceptual generation. In R. Schank and C. Riesbeck (Eds.), *Conceptual Information Processing*. American Elselvier: New York.

Grice, H. 1975 Logic and conversation. In P. Cole and J. Morgan (Eds.), *Syntax and Semantics: Vol 3, Speech Acts*, pg 43-58. Academic Press: New York.

Hirschberg, J. 1985 *A Theory of Scalar Implicature*. Ph.D thesis. Report MS-CIS-85-56, LINC LAB 21, Department of Computer and Information Science, University of Pennsylvania.

Iordanskaja, L.; Kittredge, R.; Polguere, A. 1988 Implementing a Meaning-Text Model for Language Generation. Presented at COLING 1988 (not in proc.).

Jacobs, P. 1987 Knowledge-Intensive Natural Language Generation. *Artificial Intelligence* 33:325-378.

Jolicoeur, P.; Gluck, M.; and Kosslyn, S. 1984 Pictures and Names: Making the Connection. *Cognitive Psychology* 16:243-275.

Nirenburg, S. and Nirenburg, I. 1988 A Framework for Lexical Selection in Natural Language Generation. *Proceedings of the 12th International Conference on Computational Linguistics* (2):471-475.

Nirenburg, S.; Nyberg, E.; and Kenschaft, E. 1987 Inexact Frame Matching for Lexical Selection in Natural Language Generation. Unpublished memo, Center for Machine Translation, Carnegie-Mellon University.

Pustejovsky, J. and Nirenburg, S. 1987 Lexical selection in the process of natural language generation. In *Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics*: pages 201-206.

Reiter, E. 1990 Generating Descriptions that Exploit a User's Domain Knowledge. In R. Dale, C. Mellish, and M. Zock (Eds.), *Current Research in Natural Language Generation*. Academic Press: London. Forthcoming.

Rosch, E. 1978 Principles of Categorization. In E. Rosch and B. Lloyd (Eds.), *Cognition and Categorization*. Lawrence Erlbaum: Hillsdale, NJ.

Zgusta, L. 1971 *Manual of Lexicography*. Academia Press: Prague.