Generation Systems Should Choose Their Words*

Mitchell Marcus
AT&T Bell Laboratories
Murray Hill, NJ 07974

Almost all current natural language generation (NLG) systems, as contrasted with current NL understanding (NLU) systems, have one somewhat surprising property in common: Most current NLG systems don't use words at all. Such systems operate by incrementally specifying fragments of linguistic structure in a top-down fashion, typically inserting specific lexical items only when the frontier of the structure is encountered. In some important sense, these systems have no real knowledge of lexical semantics and only rarely make lexical choices. Instead they choose between one tree fragment or another, only rarely able to see the leaves for the trees. Somehow the fact that particular words have particular meanings is incidental to the operation of these systems; they use fragments of linguistic structure which eventually have words as their frontiers, but they have little or no explicit knowledge of these words and what they mean. At best, these systems assume that each conceptual primitive corresponds to a particular unique lexical item or phrase, trivializing the problem of lexical choice to one of table lookup, and trivializing the problem of lexical semantics to the claim that the meaning of the word can be represented by the same word in upper case, more or less. While this practice may suffice for generation systems for narrow application domains, it is most certainly wrong from a cognitive point of view.

What makes all this surprising is that current research in generation focusses on such subtle and difficult matters as responding appropriately to the users intentions, correctly structuring the illocutionary force of a generated utterance, correctly utilizing rhetorical structures and the like. While we have undertaken the understanding of such subtle phenomena, and have attempted to build systems which sound highly fluent, we have avoided research on what would make such systems mean the literal content of the words they use.

The alternative, which we must face up to sooner or later, is to attack the closely related problems of lexical semantics and lexical choice directly. In an ideal world, I believe, a NLG system should have available to it a lexicon of words and their meanings, represented in a non-domain specific way, along with a general mechanism which uses this representation to map a message in "Mentalese" into whatever words are appropriate to realizing the meaning embodied in it. One might well expect such a system to somehow compile or invert this lexicon into a representation which allows this process to be executed efficiently, but the meaning of a word (in non-domain specific terms) should be encoded, even if implicitly, in such an inverted representation. Thus, the use of discrimination-net-like mechanisms for lexical choice, for example, is not excluded by the considerations of concern here; such networks, however, ought to be invertable to yield real definitions of lexical content in other than a domain specific way. (On the other hand, discrimination nets impose a rigid order in which choice criteria are examined regardless of the context under consideration, but such difficulties are not the focus of my comments here.) One touchstone for such a system is that it should be capable of absorbing new lexical entries nearly seamlessly, with only a compilation step before such definitions can be efficiently utilized.

* The views expressed here result from ongoing discussion with Robert Rubinoff, to whom the author is greatly indebted. Any shortcomings are my responsibility, of course.

Current systems by and large fail to represent lexical meaning, and back away from most issues of lexical choice. Symptomatic of this lack of concern with words and lexical choice is the structure of the lexicon in such systems. Current NLG systems often lack any lexicon per se, as distinct from a dictionary of the semantic primitives of the "message" domain. A typical case is the TEXT system (McKeown, 1985), whose dictionary is exactly a dictionary of primitives, not words. Its function is to translate primitives in the message into syntactic fragments that usually include at least one word each, but it fundamentally encodes stories about how to encode semantic primitives in very particular configurations for a particular application, not stories about the meanings of words. To see that this is true, consider the translation of the concept "GUIDED" into a fragment of structure consisting of the adjective "guided" and the noun "projectile". Note that the system is entirely unaware of the contribution of the meanings of "guided" and "projectile" to achieve this translation; in this sense, it knows nothing about the meanings of either "guided" or "projectile" per se. (The lexicon proposed in (Mathiesson, 1981) is an exception to this, and might allow a full encoding in KL-ONE of the lexical semantics of a word. However, the structures that Mathiesson presents encode only selectional restrictions and thematic relations of the verb; the examples given do not attempt to encode conceptual meaning.)

A general exception to this is the meaning of "closed class" function words such as determiners like "the" and "a" and relative markers like "who" or "what". Many systems, particularly systems which utilize systemic grammar or one or another unification based approach (e.g. (Appelt 1985)), build up the choice of closed class items by combining binary features such as "definite/indefinite" and "singular/plural", resulting in the end in a fully specified description of one grammatical formative or another. But note that it is only this subpart of the lexicon whose meaning can be represented adequately by sets of independently determined binary features.

Another issue which highlights the lack of lexical knowledge is the limited extent to which knowledge of the lexicon of a particular domain used by a generation component of a full NL interface could be shared with the understanding component. By and large, existing NLU systems use some form of compositional semantics to do analysis, using some restricted form of lexical meaning. Thus, most analysis systems would construct a representation of the meaning of "guided projectile" by modifying the meaning of "projectile" with the meaning of "guided" in classical fashion. If such an analysis system paralleled TEXT, it would match the entire syntactic structure of "guided projectile" to yield one underlying semantic atom. In fact, this is exactly the theoretical position taken by the PHRAN/PHRED project (Wilensky and Arens, 1981; Jacobs, 1985), which views all language understanding and generation as a phrasal process. This view exactly allows Wilensky and his collaborators to use the same lexicons both for generation and analysis, although the position they currently take is that this work is not be viewed as theoretically motivated, as Rubinoff points out (Rubinoff, pc).

This view, however, begs several major questions: 1) Why should the same word be used in different contexts? The pure phrasal view denies that words per se have meanings, and views all uses of words within phrases as essentially idiomatic. Thus, it would seem unclear why "guided" should be used in a wide range of phrases all of which are consistent with the view that "guided" has a particular consistent meaning which it contributes to each of these contexts. If all understanding and generation is phrasal, then the contribution of a particular lexical item to each phrase in which it occurs would be far more idiosyncratic than appears to be the case.

In sharp contrast with the trend in generation, a wide variety of current theories of linguistic competence are increasingly lexicalist. These theories increasingly assume that the locus of much grammatical knowledge is the lexicon itself, and that much grammatical structure follows from constraints on the use of particular lexical items. This trend is all the more striking because these theories, taken together, are widely divergent on most other details of linguistic analysis.

For example, in the Government-Binding framework (Chomsky, 1981), large scale phrase structure is seen as projected upward from the lexical items themselves, with aspects of syntactic structure such as case marking deriving from properties of particular lexical items. In GB, most grammatical properties are viewed as properties of words and grammatical formatives; these properties play themselves out in accordance to the constituency structure of the grammatical tree, but the properties themselves derive from words. It is widely suggested, in fact by Stowell (Stowell 1981) and others, that constituency structure itself is derivative on an interaction of more abstract grammatical properties and the properties of particular lexical items. While these theories by and large fail to explicitly consider issues of meaning deeper than argument structures of particular verbs, recent work suggests that many of these properties of "theta-grids" follows from deeper semantic generalizations and much work is currently going on in this area (e.g. (Levin and Rappaport 1985), (Jackendoff, 1983)).

The framework of Lexicalist-Functional Grammar (Bresnan, 1982) began with the observation that what appeared to be large scale properties of the structural configurations of sentences could be accounted for by local "pre-compiled" statements about the argument structures of particular lexical items, in particular verbs. In LFG, the grammatical structure of a sentence follows, in large measure, from the mutual satisfaction of elaborate sets of constraints inherited from words, with the constituency structure of secondary importance.

Perhaps clearest of all are linguistic theories following from Richard Montague's theories of natural language semantics, including both early GPSG (Gazdar, 1982) and recent work in categorial grammar. In these theories, a very simple grammatical component serves to control the computation of a semantic representation which follows by the composition of lambda-expressions representing word meanings. In such a framework, all meaning per se derives from lexical items and grammatical formatives; syntax merely serves to indicate exactly what lambda-reductions should be performed and in what order.

As a practical matter, current approaches seem extremely well suited for building generators for particular applications in narrow, well defined domains. On the other hand, each new application requires the construction of an application-specific dictionary of translations from the primitives of the messages constructed by the application into fragments of English structure. In the long run, if NLG systems are to be both fluent and quickly portable, they must actually know about both words and their meanings.

References

Appelt, Douglas E., *Planning English sentences*, Cambridge University Press, Cambridge, UK, 1985.

Bresnan, Joan (ed.), *The Mental Representation of Grammatical Relations*, MIT Press, Cambridge, MA, 1982.

Chomsky, Noam, *Lectures on Government and Binding*, Foris, Dordrecht, Holland, 1981.

Gazdar, Gerald, "Phrase structure grammar", in Pauline Jacobsen and Geoffrey K. Pullum, eds., *The Nature of Syntactic Representation*, Reidel, Dordrecht, Holland, 1982.

Jacobs, Paul S., "PHRED: A generator for natural language interfaces", Computer Science Division Technical Report UCB/CSD 85/198, UC Berkeley, Berkeley, CA, 1985.

Jackendoff, Ray, *Semantics and Cognition, MIT Press*, Cambridge, MA, 1983.

Levin, Beth & Malka Rappaport, "On the Formation of Adjectival Passives", Lexicon Project Working Papers Number 2, Center for Cognitive Science, MIT, Cambridge, MA, 1985.

Mathiesson, C.M.I.M, "A grammar and a lexicon for a text-production system" in *Proceedings of the 19th Annual Meeting of the Association for Computational Linguistics*, Stanford, CA, 1981, pp. 49-56.

Kathleen R. McKeown, *Text generation*, Cambridge University Press, Cambridge, UK, 1985.

Stowell, Tim, *Origins of Phrase Structure*, unpublished Ph.D. Thesis, MIT, Cambridge, MA, 1981.

Wilensky, Robert, and Yigal Arens, "PHRAN --A Knowledge-based approach to natural language analysis", ERL Laboratory Memorandum UCB/ERL M80/34, UC Berkeley, Berkeley, CA, 1981.