Invited position paper to the 1991 European Workshop on Language Generation. Published in Zoch & Horacek (eds) New Concepts in Natural Language Generation: Planning, Realization, Systems, Pinter, London, pgs. 275-278.

# "Does Natural Language Generation Start from a Specification?"

David D. McDonald
Brandeis University

To ask what language generation starts from, one must first answer the question "what is generation". This is not a trivial matter since people with extensive experience in the field will not give the same answer. Does it include text planning? Does it start as far back as the selection of the information to be communicated or even with the initial decision to speak rather than stay silent? Or is it restricted to the operations of a linguistic component that works from near-surface level specifications and does little more than apply a grammar to them to produce well-formed sentences?

To keep this from being simply a matter of semantics in the trivial sense, it is crucial to establish a methodological basis for what language generation should include--what sets its operations off from those generic to perception, thought, planning, conceptualization, or other aspects of human mental activity. I propose that the distinguishing element is the presence of language--either active knowledge of the particulars of the speaker's language that is drawn on in making decisions, or the use of items whose structure or operations are strongly constrained by linguistic principles.

This criterion shifts the question from "how far back does generation go?" to "how far back does knowledge of language go?", which is something that I think we can answer more objectively: If one can devise a way to plan the content and organization of a text without drawing on knowledge about language in the process, then the components involved can reasonably be said to be outside of the generator. If one cannot (which I personally believe is the case), then there will be a slippery slope from processes that primarily engage knowledge of the world and of the interlocutors' situation using only a little knowledge of language, to those that primarily engage knowledge of syntax and morphology using only a little awareness of the situation (e.g. for the construction of intonational contours). Given a slippery slope, it is simpler to consider all of the processes as within "the generator", and to shift our attention away from "initial" specifications that are presumed to be outside of the generator (since there may not be any), and instead to consider the characteristics of the structures that dictate the actions of the generation process in its early stages generally, without attempting to draw any hard boundaries a priori.

Whatever the details of the process may be, we can certainly agree that at the moment the speaker decides to speak--to carry out some action through the use of language--that she is embedded in a situation, and that it is her construal of that situation that has the greatest influence on exactly what she will say. The question then is how the representational elements and dispositions for action that comprise speakers' model of their situation are engaged to bring about the production of an utterance.

Architecturally, there are two general alternatives. The easiest to program but least efficient is to embed the knowledge of how to compose effective utterances in a process that searches the situation to determine which utterance it should build and constructs a representation of its findings. This representation is usually taken as the initial specification for the generator's actions, though in some sense the "initial" specification was actually the model of the situation. The more efficient alternative has the elements of the situation act

independently on their own to group together and produce the utterance directly as the elements are mapped to their progressively more linguistic counterparts. Here there is no initial specification in the sense of a deliberately constructed expression, though there will be intermediate representations of the utterance at later stages of development. Actual generators today are mixes of both approaches, depending on whether control and decision making lie with the representations or with the processes that manipulate them.

As an example of an explicitly constructed specification, consider the example below, from McDonald & Pustejovsky 1985. It was arrived at by working backwards from an actual text (shown below it), with the goal of including enough information in the specification to control the nuances of expression that we saw in the text. We believed that we could construct a generator that could work from expressions like it, though we never wrote the program. The specification is comprised of typed, structured objects taken from the program's model (indicated by enclosing #< >), where the types dictate default realization choices. Keywords (preceded by ":") give the value of realization options.

```
(the-day's-events-in-the-Gulf-tanker-war
   :events-require-certification-as-to-source
   (main-event #<same-event-type_varying-patient
                  #<hit-by-missiles Thorshavet>
                  #<hit-by-missiles Liberian>>
      :unusual #<number-of-ships-hit 2>
      :identify #<ship> )
   (particulars  #<damage-report Thorshavet Oslo-officials>
              #<damage-report Liberian Lloyds> ))
```

*"Two oil tankers, the Norwegian-owned Thorshavet and a Liberian-registered vessel, were reported to have been hit by missiles Friday in the Gulf. The Thorshavet was ablaze and under tow to Bahrain, officials in Oslo said. Lloyds reported that two crewmen were injured on the Liberian ship."*

International Herald Tribune, December 23rd, 1984.

Most of this specification dictates the marked aspects of the text, with the rest of its structure following from defaults. The keyword :unusual forces the fronting of the number of tankers so as to give it a position of textual prominence; :identify is responsible for the inclusion of the ships' names and registries. The merger of the two main events into one clause follows from the fact that the source program is taken to have conceptualized them as a single significant item (given by the object #<same-event-type...>) because having two ships attacked on the same day was unusual at the time the article was written.

The specification as a whole amounts to an identification of the information the utterance is to contain (organized textually as main event and particulars), plus annotation dictating any special ways it is to be realized. It is a self-contained, abstract representation of the eventual utterance, that does not require any later references to the situation since all of the contextual influences have already been factored into it.

There is no question that a specification such as this would be effective--especially within a constrained sublanguage such as this sort of news report. The architectural issue is whether it would ever actually have to be constructed in order to do its work. From the perspective of the second, no-specification direct-action alternative, the same information would be involved, but it would never be marshaled together all at one time as a single expression to only then be passed on to the rest of the generation process. Instead, the terms

would enter incrementally. The information about departures from defaults (the keywords) and other style-dictating directives would be accessible in the generator's ongoing computational state, and the textual organization that gives the utterance its large-scale order would reflect the order and perspective under which the terms were selected as the situation was examined.

An example of this kind of architecture is Jeff Conklin's "Genaro" text planner (Conklin et al. 1983). In the course of describing a picture of a house, Genaro selected successive objects, relations, and properties in order of their relative visual salience. These units where then immediately entered into a linguistic level of representation akin to Meteer's Text Structure (Meteer 1992), where they were given a textual organization according to their relationship to the material already entered and a model of prose style. Genaro also dictated the focus and focus-change status of this level through explicit instructions and state variables.

In an architecture like Genaro's, we have a "message level": a level of representation where one reasons about what information should be included in an utterance and what implicit information to communicate by how it is organized; but we do not have a "message", in the sense of an expression that encodes that information in non-linguistic terms. The only "expressions" in such a generator are the progressively more specific representations of the utterance as they are assembled and incrementally mapped to later levels and eventually spoken, and even these expressions do not have to exist any longer than needed to construct the corresponding portion of the next level.

To my mind, the no-expression, specification-less architecture is an engaging psycholinguistic hypothesis because it narrowly restricts the amount of information that is available to direct the generation process at any one time. By contrast, having a specification of the full utterance available before starting means that potentially arbitrary facts about later portions of the utterance can influence realization choices for early portions, a hypothesis that would have minimal explanatory power because it offers so little constraint. A further difficulty with having a full initial specification is the question of whether it will be expressible (in Meteer's sense), since by definition it is assembled without benefit of any linguistic knowledge of how its parts will be allowed to combine; the space of collective valid realizations is sufficiently sparse that the expression may well fail and have to be replanned, yet people only rarely talk themselves into a corner. And, while it is true that apparent human behavior can only be a weak guide to how to design a generation system, people are nevertheless the only full-scale generators available to us as models, and the more we take from psychological evidence the easier our task will be.

Conklin, E. Jeffery, Kate Erlich, & David McDonald (1983) "An Empirical Investigation of Visual Salience and Its Role in Text Generation", Cognition and Brain Theory, 6(2) Spring 1983, pp. 197-225.

McDonald, David & James Pustejovsky (1985) "TAG's as a Grammatical Formalism for Generation", proc. Association for Computational Linguistics, Univ. of Chicago, July 8-12 1985, pp. 94-103.

Meteer, Marie (1992) Expressibility and the Problem of Efficient Text Planning, Pinter, London.