

What is Decided and What Just Happens?

David D. McDonald

Content Technologies, Inc.
14 Brantwood Road, Arlington MA 02174-8004 U.S.A
(617) 646-4124 MCDONALD@BRANDEIS.EDU

1. What counts as a decision ?

As a field we tend almost universally to view generation as "a goal-directed decision-making process" (e.g. McDonald 1980). We catalog the alternative choices that a speaker can make, study the contexts that utterances occur in and how they constrain these choices, and we look for systematic relationships between these choices that can explain how utterances take the form that they do or how a particular processing architecture can weave its way through them efficiently. Our generators make choices, but are they actually deciding anything?

As we use the word 'decision' in everyday life, it refers to choices that we can consciously deliberate over, choices where some of the outcomes are truly optional. Should I pick up the milk at the corner store, which is close but has higher prices, or should I instead go further away to the supermarket where I can also pick up a newspaper? Should I take a job in industry or hold out for an academic position? Decisions like these call for planning. We lay out the alternatives without yet acting on any of them. We compare their consequences, mentally simulating what would happen. These processes are largely conscious and can take a significant amount of time roughly proportional to the number of alternatives and the complexity of their interaction.

By contrast, the processes that are engaged when people produce an utterance have a very different character. Generation in people is highly constrained, automatic, almost entirely unconscious, and shows no effects of utterance length in its time course --- more like walking than deciding what college to go to.

Given this difference in the character of the behavior, the question of whether generation involves making decisions is not just a matter of terminology. Can you be said to be making a decision when there are no options? Are you making a decision if you are just carrying out orders? A generator whose processing architecture is very specifically designed to fit its one task may well have no other options and just be carrying out orders. On the other hand, a generator that employs a general purpose planning algorithm with no constraints on where it can look for information and no limits on its algorithms other than what its designer happens to have programmed for it so far may have decision-making, in the ordinary sense, as its primary activity.

1.1 Grammaticality

To make this difference more apparent we can ask whether a generator 'decides' to be grammatical. The answer is 'yes' if it is possible for it to formulate a set of actions that will lead to ungrammatical utterances (e.g. agreement failures or incorrect order or case assignments). The answer is 'no' if the generator's architecture makes that literally impossible.

Most mature realization components in fact have this property of “inescapable grammaticality”. Barring a mistake by the grammarian, there is no path through a systemic grammar that will yield an inconsistent set of features and from them an ungrammatical realization. Similarly in Mumble, all of the surface elements of any utterance are predefined as locally grammatical trees in TAG tree families, and their composition is regulated top down so that no tree is substituted or adjoined into a context that would make the utterance ungrammatical.

While Mumble intrinsically cannot formulate an ungrammatical utterance, this is not the whole story, since Mumble is only a realization component and not the whole generator. Mumble is driven from a “specification” of the set of lexicalized trees to be used, the points where they are to be combined, and the particulars of their non-phrasal properties such as number or tense.¹ As it happens, there is nothing in Mumble itself that prevents an earlier processing stage from assembling a bad specification. That is to say that while the objects and composition rules that make up surface structures are constrained by prior design so that no ungrammatical combinations are possible, there is no such constraint on the specifications that spell out what objects (trees) and compositions to use in a given case. (If an incompatible element occurs in a specification it will be ignored and nothing realized for it.)

Ensuring that only good specifications are passed to Mumble was one (among a great many) of the motivations behind Meter’s “Text Structure” (1990), the level of representation just prior to Mumble in her Spokesman generation system. Text Structure is a constructed representation built using a fixed vocabulary of constituent types and rules of composition. The representational units at Text Structure mirror those of the surface syntax (as construed by a TAG), e.g. head and arguments, matrix and adjuncts; but they abstract away from particular syntactic categories so that the same kind of relations can be seen to hold over very different scales of text from noun phrases to paragraphs. The units are typed, and these types are semantic rather than syntactic. They express the major distinctions among the kinds of things that the language supports systematically in its morpho-syntax, e.g. singular/plural, completed/ongoing, object/process (see, e.g., Talmy 1987).

Operationally these types define projections onto the linguistic specifications that are input to Mumble. They are reifications of the legal combinations of specification elements, with the consequence that a well-formed Text Structure must by necessity lead to a good specification that can be grammatically realized in its entirety, since there are no mechanisms in Text Structure that could be used to form a bad specification.

Text Structure is presently built directly by incremental additions rather than in large chunks through the execution of some yet more abstract specification. It ensures, through its predefined vocabulary of semantic types and its type-sensitive constraints on composition, that any information entered into Text Structure will be expressible as a grammatical utterance. Nevertheless there is the obvious question of how is a speaker guaranteed that anything he, she, or it might want to say can be successfully expressed in Text Structure. This has no answer as yet, and the answer might well turn out to be that there is no such guarantee, since people cannot in fact express every thought they try to communicate.

With a mechanical speaker, where we are permitted to fashion the category structure and thought processes to our own design, we can probably guarantee the expressibility of individual units (though not arbitrary combinations of them) by designing the conceptual

¹ Specifications of this sort are specific to Mumble-86, the most recent version of the system (Meter et al. 1987). The input to earlier versions of the system was not nearly as constrained. Only in the current version is the grammaticality of the output text a consequence of the notation of the system and not just the discipline of the human developer.

categories to be a more specific superset of the semantic categories evidenced in language and represented at Text Structure (i.e. the mental categories/types are to be “subsumed” by those at Text Structure (Woods 1991; cf. Matthiessen 1987). Since the mental categories will be more numerous and more specific, projecting units with these categories onto constituents at Text Structure will require some information to be lost, and so by definition will require some choices to be made. These choices could, for example, come down to adopting a perspective: If we have an entity which both involves a process and about which we have predicated some properties, then we have the option to construe it as an event, emphasizing the process, or as one of the semantically available kinds of object, emphasizing its capacity to enter into predications.

2. Problems and Complications

The picture that I have tried to paint is one where the generator just ‘does the right thing’ (McDonald 1988). Decisions, in the sense of deliberations over accumulated information and weighing the consequences of alternatives, do not take place in the generator but rather in the mental attitudes of the speaker, where the principal influence is in how an item has been categorized.

When we look at naturally occurring texts we quickly get evidence that differences in semantic categories do not have to stem strictly from mental attitudes but may be accidental (i.e. inconsequential to intended meaning or nuance) and can apparently follow from textual constraints.

2.1 Textually-forced categorization decisions

Consider the two texts below. Both are paragraphs excerpted from long, signed articles in the Wall Street Journal on the same day, September 27, 1990. The prior context is comparable in both cases: the authors are explaining the factors that recently influenced traders on the stock market and the oil futures market respectively. What is interesting here is the information that they report in common (shown in bold); what appears in the first excerpt as noun phrases (NPs), thereby reflecting an object perspective on the events, is given in the second as clauses, reflecting event perspectives.

1. *Adding to the real worries yesterday were a series of rumors that all proved untrue. Traders reported the market was hurt by talk of: a coup in the Soviet Union; **an ultimatum from President Bush for Iraq to evacuate Kuwait; air battles in the Middle East; and the storming of the U.S. embassy in Kuwait.** None of the rumors was substantiated.*
2. *There were rumors -- all denied -- that **an Iraqi plane had been shot down in the Persian Gulf and that President Bush had issued a new ultimatum to abandon Kuwait.***

I submit that there is virtually no difference in selected information in the case of the ultimatum, and no interesting difference in the case of the air battle / downing of a plane since all that is involved is a difference in level of detail. The particulars of the contents could be swapped between the excerpts and no one would have noticed. Swapping the form that that content received, however, would have yielded quite jarring texts, and I strongly doubt that any journalist would permit it.

If mental perspective were all that was involved then we could expect NPs and clauses to be freely mixed in situations like these according to which perspective was most informative or familiar in each case. What then is the source of the uniformity? A likely source is the preferred complement patterns for the “preambles” of the two lists. In the first excerpt there is a constraint from the head phrase ‘...talk of’, which strongly subcategorizes for a list of NPs. The existential in the second excerpt ends with ‘rumors’,

which somewhat strongly (to my ear) requires the complements that spell out what the rumors are to be given as 'that' clauses.

The question to ponder is which came first: the preambles with their preferences for the form of their complements, or the information in those complements (i.e. the rumors), with its individual preferences for object versus process perspective. Which realization decision imposed its influence on the other?

We will take it that the semantic type by which some information is to be realized must be committed to as or before the information is entered into the Text Structure. Consequently, if the preambles came first it means that they were entered into the text structure before the lists of rumors, that they thereby defined the Text Structure sites at which the lists would be entered, and that with those sites came sets of constraints that impose the preambles' preferences. For this ordering to work, the lists of rumors must have been uncommitted to their semantic perspective before they were introduced, otherwise the constraints from the preambles could rule them out (if they had unknowingly been given incompatible perspectives) and the realization of those rumors would block. To use the preambles the incompatible rumors would have to be given an alternative perspective, if one is available, or the information would have to be reconceptualized, i.e. a different set of lexical choices found that would support a compatible perspective, which would likely entail a different construal of what the germane information actually was.

We should also consider whether there is any influence from the fact that the clause case only involved two rumors, while in the NP case the list is twice as long. Is there a textual preference for listing NPs rather than clauses,² and can this preference coerce otherwise malleable information perspectives? The journalists were presumably well aware of the rumors before they started writing, and in particular would probably have decided how many to report. This commitment to the identity of the rumors seems on its face to be an act of selection, which normally for a generator like Spokesman has simultaneously involved committing to the semantic types by which the rumored events would be expressed. However given the discussion above, it seems more profitable to separate out those two commitments: deciding at an early point that the information should be included, and only later, once its context in the utterance has become clear, establishing how it is to be construed/lexicalized. This raises the issue of how this first decision is to be recorded --- what sort of representation it should be given and what sort of work that representation should do --- which we leave for later work.

2.2 Missing logical arguments: deliberate or unconscious?

Journalists, possibly because they are usually summarizing and reporting rather than developing their articles from scratch, can produce very dense, inferentially loaded texts.³ Consider the following excerpt from an article in the Who's News column of the Journal from February 14, 1991.

After almost nine tumultuous years, George L. Ball resigned yesterday as chairman and chief executive officer of Prudential-Bach Securities Inc., the nation's fourth-largest securities firm.

[three intervening paragraphs on the implications and causes of his departure]

² I have employed such a preference in Mumble for organizing paragraphs describing the concept nodes in a KL-One network (McDonald 1980). If there were more than three roles associated with a concept then the paragraph was structured as a summary sentence listing the roles followed by a series of conjoined, heavily reduced clauses giving each role's value restriction. If there were three or fewer roles the information was mixed together.

³ Journalists also may have revised their text several times before the articles are published, which could involve processes quite different from the rapid, off-the-cuff processes we have focused on. By the same token, they are skilled professional writers, and may well usually 'get it right the first time'.

Robert A. Beck, a 65-year-old former Prudential chairman who originally bought the brokerage firm, was named chief executive of Prudential-Bache until a replacement is named. Mr. Ball will remain as interim chairman.

We will focus here on the phrase ‘*a replacement*’ in the second to last sentence. Assuming we understand the passage at all, then from the context we know that the phrase refers to some future event whereby Mr. Beck will be replaced as CEO of Prudential Bache by some as yet unknown person. This means that we have fleshed out our representation of the information conveyed by the NP to include two people, a title, and a company --- four logical arguments to ‘*a replacement*’ all added just on the basis of inference. Is this something the journalist planned for us to do; or did it just happen, much the same way that Mumble just happens to produce grammatical text ?

A key issue here is what is the speaker’s own internal conception underlying the phrase. Is it the event, in which case the generator will be doing some sort of ‘argument suppression’, or is it only the concept of ‘being a replacement’ without any of the logically accompanying arguments. If the later, then extensive checks will be required to establish whether the concept/phrase can be successfully used in that textual context, and it is not obvious how such a search would be bounded. The time-course alone suggests that we would want to call such a process ‘planning’.

If we assume that the speaker’s source was the whole event, e.g. that they employed the strategy of describing a future individual by referring to the event that will establish who they are (cf. ‘*the winner of the race*’), then can the speaker be said to have decided to do anything beyond using that strategy and selecting the event to be used? In a slightly different context, with different individuals and relations present in the nearby text, a different combination of arguments would have to be suppressed, yielding, e.g. ‘*Mr. Beck’s replacement*’ or even ‘*Mr. Beck’s replacement at Pru-Bache*’.

Under this assumption, the suppression and the resulting NPs are a consequence of the textual context, not the speaker’s selection. Which arguments are realized cannot be determined before all the text up to the point where the NP is to appear has been realized, including at least the lexical choices and the final order and syntactic shape. (This is equivalent to the point where the linguistic specification for Mumble has been formed by reading out the representation at Text Structure.)

I find this argument suppression assumption the most satisfying, because it should lend itself to a description-directed account. Working out how to implement it is an ongoing project as this is written. The goal is a design that makes the suppression as automatic and inescapable as Mumble’s grammaticality.

3. Conclusions

I have tried to show that it is quite possible to have a generator that carries out its processing without ever making a decision in the normal sense of the word. One employs a highly constrained design, where the character of the outcome is implicitly regulated by the notation, along with largely precomputed set of alternatives so organized that once the situation has been identified a single choice can be taken and directly acted upon without having to wait to consider the possible implications of other alternatives.

I then pointed out two potential problems with this neat picture that complicate it by adding what amount to timing constraints: information that strongly contributes to the choices but which only becomes evident at a late stage once most of the textual context has been established. How to accommodate these constraints while retaining the considerable efficiency of the straight-line, no-decision approach is not yet clear, but then no one has expected our field to run out of interesting research problems any time soon.

4. References

- Matthiessen, Christian (1987) "Notes on the organization of the environment of a text generation grammar", in Gerard Kempen (ed.) **Natural Language Generation**, Martinus Nijhoff, pgs. 253-278.
- McDonald, David (1980) **Natural Language Production as a Process of Decision-making under Constraints**, unpublished Ph.D. Dissertation, MIT Artificial Intelligence Laboratory, August 1980.
- Meteer, Marie (1990) **The "Generation Gap": the Problem of Expressibility in Text Planning**, Ph.D. Dissertation, Dept. of Computer & Information Science, University of Massachusetts at Amherst.
- _____, David McDonald, Scott Anderson, David Forster, Linda Gay, Alison Huettner & Penelope Sibun (1987) **Mumble-86: Design and Implementation**, TR #87-87 Dept. Computer & Information Science, UMass., September 1987.
- Talmy, Leonard (1987) "The Relation of Grammar to Cognition", in B. Rudzka-Ostyn (ed.) **Topics in Cognitive Linguistics**, John Benjamins.
- Woods, William (1991) "Understanding Subsumption and Taxonomy: A Framework for Progress", in John Sowa (ed.), **Principles of Semantic Networks: Explorations in the Representation of Knowledge**, Morgan Kaufmann.