Pragmatics and Natural Language Generation*

Eduard H. Hovy

Information Sciences Institute of the University of Southern California, 4676 Admiralty Way, Marina del Rey, CA 90292-6695, USA

ABSTRACT

This paper addresses the question "why and how is it that we say the same thing differently to different people, or even to the same person in different circumstances?" We vary the content and form of our text in order to convey more information than is contained in the literal meanings of our words. This information expresses the speaker's interpersonal goals toward the hearer and, in general, his or her perception of the pragmatic aspects of the conversation. This paper discusses two insights that arise when one studies this question: the existence of a level of organization that mediates between communicative goals and generator decisions, and the interleaved planning-realization regime and associated monitoring required for generation. To illustrate these ideas, a computer program is described which contains plans and strategies to produce stylistically appropriate texts from a single representation under various settings that model pragmatic circumstances.

1. Introduction

1.1. The problem

In the fifties, generation was considered the easy part of natural language processing. After all, it is straightforward to write a generator that produces impressive text by associating a sentence template (or some equivalent general grammatical form) with each representational type and then using a grammar to realize the template into surface form.

Unfortunately, with this simplistic view of generation, generators are not sensitive to anything but the input items, and therefore produce the same output to all hearers in all circumstances. However, when we produce lan-

guage, we tailor our text to the hearer and to the situation. This enables us to include more information than is contained in the literal meanings of our words; indeed, the additional information often has a stronger effect on the hearer than the literal content has. This information is carried by both the content and the form of the text. As speakers and hearers, from various ways of expressing a single underlying conceptualization, we make various interpretations of the speaker, his or her goals, the hearer, and the conversational circumstances. These interpretations are governed by rules. Speakers use the rules to determine how to say what they want to say. Generator programs with any real flexibility of expression require such rules too.

What types of additional information can speakers convey? Consider the differences evident in the speaker's point of view in the following four descriptions of an event that occurred at Yale University in April 1986:

> (1) "On April 4, concerned Yale students constructed a shantytown on Boesak Plaza as a reminder to those in Woodbridge Hall (and all over campus and the community) that Yale is complicit [sic] with the system of apartheid that creates shantytowns where thousands of blacks are forced to live in squalor and fear. The shantytown, Winnie Mandela City, served as a focal point of education concerning South Africa and Yale's investments there. At 5:30 am on April 14 the Yale Administration had the shantytown torn down and had 76 students and community members who were defending the shanties arrested. After a huge outcry, the Administration allowed the shanties to be rebuilt. We will not be silenced; we will continue to challenge the University on their moral failure." (From: protester literature; the protesters renamed the plaza after the South African churchman Allan Boesak)

> (2) "On April 4, a small group of students took over Beinecke Plaza and built some shanties; they wanted to force Yale to sell its stocks in companies with branches located in South Africa. The University asked the students to move the shanties to another location, but the students refused. The University then granted them permission to occupy the plaza until the end of the week, so that they could be there to be seen by the University's trustees, the Yale Corporation, at their meeting. But even after the meeting, the students refused to leave the plaza, and police had to clear the shanties. Later, the University relented, and gave them permission to rebuild the shanties. It also announced that it would send a fact-finding mission to South Africa." (Speaker: anti-divestment student)

> (3) "On April 4, students at Yale built a symbolic shantytown to protest their school's investments in companies doing business in

South Africa. The college ordered the shanties destroyed. The police arrested 76 protesters when the shantytown was torn down. Local politicians and more than 100 faculty members criticized the action. A week after it had ordered the removal of the shantytown—named Winnie Mandela City, after the South African foe of apartheid—the shantytown was reconstructed and the Administration agreed to allow it to remain standing. Concurrently, Yale announced that its trustees, the Yale Corporation, would soon send a fact-finding mission to South Africa to investigate the actions of corporations in which it owns between $350 million and $400 million of stock." (From: *New York Times*, Sunday, 27 April, 1986, Connecticut section)

(4) "Some students erected a shantytown to protest Yale's investments in companies that have operations in South Africa. The University tore it down and arrested several of them. The students continued to demonstrate and finally the University said they could put up the shantytown again. The University said it would investigate its investments in South Africa." (Speaker: neutral student)

Clearly, the first two speakers incorporate strongly their opinions about the shantytown issue; the second two speakers seem more neutral but differ in level of formality. In order for generator programs to produce similarly varied, information-bearing text, such programs must have some means of representing relevant characteristics of the hearer, the conversation setting, and their interpersonal goals. These are the *pragmatic* concerns. In addition, they must contain choice points in the grammar that enable topics to be said in various ways. These are the *syntactic* concerns. Finally, they require criteria by which to make the decisions so that the choices accurately reflect the pragmatic aspects and convey appropriate additional information. These are called here the *rhetorical* concerns.

1.2. What was learned from this work

This paper describes how the program PAULINE (Planning And Uttering Language In Natural Environments) produces stylistically appropriate texts from a single story representation under various settings that model pragmatic circumstances.

PAULINE addresses simultaneously a wider range of problems than has been tried in any single language generation program before (with the possible exception of [21]). It contains about 12,000 lines of LISP code, and produces varied and sophisticated text. As is to be expected, no part of PAULINE provides a satisfactorily detailed solution to any problem; to a larger or smaller degree, each of the questions it addresses is solved by a set of simplified, somewhat ad hoc methods. In fact, some methods have been studied in much

greater detail by other NLP researchers. Others remain as projects for the future.

However, this does not invalidate the content of the work. This research uncovered two principal insights about the nature of language generation that do not depend directly on the details; they will hold for any language generator sophisticated enough to try to achieve a number of communicative goals in a single text. While the details of any particular module will not be defended too hard, nor even the exact extent of each module, the following will be defended to the end: the existence of a level of organization mediating between communicative goals and generator decisions, containing entities called here rhetorical goals; and the monitoring of the modules' operation in an interleaved planning-realization regime. The lessons learned here are going to apply to any large and complex enough generator—human or computer.

PAULINE generated text in three different domains. One of the domains was the above mentioned shantytown affair. To model it, about 120 representation structures, embedded in a property-inheritance network, were used. From this set, given the same input but different communicative goals every time, PAULINE produced over 100 different paragraphs, ranging from a sentence to a longish paragraph. Showing only some goals and settings, a few texts are (all the texts in sanserif capitals in this paper were generated by PAULINE):

Example 1.1. Passerby describing shantytown issue

- Time: *some.*
- Tone of interaction: *informal.*
- Speaker's opinions: *neutral.*
- Depth of acquaintance: *strangers.*
- Goal to affect hearer's opinions: *none.*

> YALE UNIVERSITY PUNISHED A NUMBER OF STUDENTS FOR BUILD-
> ING A SHANTYTOWN ON BEINECKE PLAZA BY ARRESTING 76 STU-
> DENTS AND TEARING THE SHANTYTOWN DOWN ONE MORNING IN
> APRIL. THE STUDENTS WANTED YALE TO DIVEST FROM COM-
> PANIES DOING BUSINESS IN SOUTH AFRICA. FINALLY, THE UNI-
> VERSITY GAVE IN AND ALLOWED THE STUDENTS TO REBUILD THE
> SHANTYTOWN.

Example 1.2. Newspaper article

- Time: *much.*
- Tone of interaction: *formal.*
- Speakers' opinions: *neutral.*
- Depth of acquaintance: *strangers.*
- Goal to affect hearer's opinions: *none.*

IN EARLY APRIL, A SHANTYTOWN—NAMED WINNIE MANDELA
CITY—WAS CONSTRUCTED ON BEINECKE PLAZA BY SEVERAL STU-
DENTS, SO THAT YALE UNIVERSITY WOULD DIVEST FROM COM-
PANIES DOING BUSINESS IN SOUTH AFRICA. AT 5:30 AM ON APRIL
14, IT WAS DESTROYED BY OFFICIALS; ALSO, AT THAT TIME, THE
POLICE ARRESTED 76 STUDENTS. THE STUDENTS REQUESTED
THAT YALE GIVE THEM PERMISSION TO REASSEMBLE THE
SHANTYTOWN WHILE SEVERAL LOCAL POLITICIANS AND FACULTY
MEMBERS EXPRESSED CRITICISM OF YALE'S ACTION. FINALLY,
THE UNIVERSITY PERMITTED THE STUDENTS TO RECONSTRUCT IT
AND, CONCURRENTLY, YALE UNIVERSITY ANNOUNCED THAT A
COMMISSION WOULD GO TO SOUTH AFRICA IN JULY TO EXAMINE
THE SYSTEM OF APARTHEID.

Example 1.3. Protester literature

– Time: *much.*
– Tone of interaction: *formal.*
– Speaker's opinions: *for protesters.*
– Depth of acquaintance: *strangers.*
– Goal to affect hearer's opinions: *switch.*

AS A REMINDER TO YALE UNIVERSITY TO DIVEST FROM COM-
PANIES DOING BUSINESS IN SOUTH AFRICA, A LARGE NUMBER OF
CONCERNED STUDENTS ERECTED A SHANTYTOWN—NAMED WIN-
NIE MANDELA CITY—ON BEINECKE PLAZA IN EARLY APRIL. THE
LOCAL COMMUNITY EXPRESSED SUPPORT FOR THE STUDENTS'
ACTION. THE UNIVERSITY TOLD THE STUDENTS TO ERECT THE
SHANTYTOWN ELSEWHERE. LATER, AT 5:30 AM ON APRIL 14, THE
SHANTYTOWN WAS DESTROYED BY OFFICIALS; ALSO, AT THAT
TIME, THE POLICE ARRESTED 76 STUDENTS. THE STUDENTS RE-
QUESTED THAT YALE UNIVERSITY GIVE THEM PERMISSION TO
REASSEMBLE IT ON BEINECKE PLAZA; ALSO, AT THAT TIME, SEV-
ERAL LOCAL POLITICIANS AND FACULTY MEMBERS EXPRESSED
CRITICISM OF YALE'S ACTIONS. FINALLY, YALE PERMITTED THEM
TO RECONSTRUCT THE SHANTYTOWN AND, CONCURRENTLY, THE
UNIVERSITY ANNOUNCED THAT A COMMISSION WOULD GO TO
SOUTH AFRICA TO EXAMINE THE SYSTEM OF APARTHEID IN JULY.

Example 1.4. Official Yale literature

– Time: *much.*
– Tone of interaction: *formal.*
– Speaker's opinions: *for university.*

- Depth of acquaintance: *strangers.*
- Goal to affect hearer's opinions: *switch.*

> IN EARLY APRIL, A SMALL NUMBER OF STUDENTS WERE INVOLVED
> IN A CONFRONTATION WITH YALE UNIVERSITY OVER YALE'S IN-
> VESTMENT IN COMPANIES DOING BUSINESS IN SOUTH AFRICA.
> THE STUDENTS CONSTRUCTED A SHANTYTOWN—NAMED WINNIE
> MANDELA CITY—ON BEINECKE PLAZA IN ORDER TO FORCE THE
> UNIVERSITY TO DIVEST FROM THOSE COMPANIES. YALE RE-
> QUESTED THAT THE STUDENTS ERECT IT ELSEWHERE, BUT THEY
> REFUSED TO LEAVE. THE UNIVERSITY INTENDED TO BE REASON-
> ABLE. THE UNIVERSITY GAVE IT PERMISSION TO EXIST UNTIL THE
> MEETING OF THE YALE CORPORATION, BUT EVEN AFTER THAT
> THE STUDENTS STILL REFUSED TO MOVE. AT 5:30 AM ON APRIL 14,
> OFFICIALS HAD TO DISASSEMBLE THE SHANTYTOWN. FINALLY,
> YALE, BEING CONCILIATORY TOWARD THE STUDENTS, NOT ONLY
> PERMITTED THEM TO RECONSTRUCT IT, BUT ALSO ANNOUNCED
> THAT A COMMISSION WOULD GO TO SOUTH AFRICA IN JULY TO
> EXAMINE THE SYSTEM OF APARTHEID.

This paper has the following structure: Starting below, the first principal finding of this work is outlined: any generation system sophisticated enough to operate in service of communicative goals will have to maintain an inter-mediate level of goals and strategies, called here rhetorical goals. To illustrate this, the program's pragmatic settings and syntactic choices are described, followed by a description of the way these must be linked using the rhetorical goals. Next, the second principal lesson is described: any generation program flexible enough to operate under a number of communicative goals (including possibly inconsistent ones) simultaneously will have to monitor the effects of its individual utterance components under an interleaved planning-realization regime. This description includes a brief overview of PAULINE. Finally follow descriptions of three rhetorical goals, their associated strategies, and the ways they affect the text.

2. Incorporating Pragmatics in a Generator

2.1. What is pragmatics?

Before representing the shantytown affair, I collected newspaper articles, protester and university literature, descriptions from fellow-students, etc., to see what was typically said. The question then arose: how could one get a generator to produce such a variety of paragraphs? It seemed fairly clear that with language one communicates more than simply the factual information—one also conveys interpersonal information, and the text reflects one's goals

and internal state, the effects one wishes to have on the audience, and the situation and medium of communication.

On the subject of "pragmatics," which deals with these questions, there is a great deal of literature, little of it specific enough to be of much use in a computer program. Even defining what pragmatics is, or whether it is distinct from semantics at all, is still a hotly debated topic (see [16, 32, 33, 55, 56, 61, 67]). Roughly, though,

- syntax = relations between words in a sentence,
- semantics = relations between expressions and their designata,
- pragmatics = reference to interlocutors and context in communication.

The AI natural language processing research that has taken pragmatics into account concentrated mainly on generation (pragmatics in parsing is a wide open problem); of this, most work has been done not on the general ways in which pragmatic information can be conveyed, but on specific individual types of pragmatic information used by speakers. Cohen [22] studied the effect of the hearer's knowledge on the selection of appropriate speech act (say, REQUEST versus INFORM OF WANT). Appelt [2] described reasoning about the hearer's knowledge in order to plan the inclusion and organization of topics. The effect of hearer knowledge on user instruction and on object description is described by Woolf and McDonald [98] and Paris [79] respectively; the explanation generator of Swartout [93] had a switch distinguishing between two types of hearer knowledge (either programmer or medical expert). In [59], Jameson describes a program that selects appropriate utterances in job interview situations. Bienkowski describes automatic elaboration of basic text in [7]. McCoy [71] discusses perspective selection. Much related work on the structure of discourse uses some pragmatic information, such as the work of Grosz and Sidner [41, 42]. Mann and Matthiessen's systemic generator Penman [69, 70] contains a number of oracles that would be based on pragmatic criteria. Clippinger's program [21] contained a module that reasoned about the pragmatic effects of its generation goals.

A number of general classifications of speaker intent have been made by AI researchers. A very general discussion of speaker goals is given in [13]. In [60], Johnson and Robertson use goals to model a speaker having a conversation. In [88], Schank and some students analyze the different speaker intents underlying a statement. Other goal classifications can be found in [15, 26, 96]. Furthermore, much sociological and psycholinguistic work has been done in this regard. Bloomfield mentions the effects of emotional relations on sentences [9]; Gazdar lists general pragmatic constraints on sentences in [33]; in [14] Bühler names some pragmatic aspects of conversations; Jakobsen extends this list in [58]. Both Grimes [40] and van Dijk [94] identify a number of pragmatic features and discuss what roles they play in the topic selection, focus, and realization. In the tradition of systemic grammar (see, say, Halliday's work in

[43, 44]), interesting recent work can be found in [30, 38]. The effects of context on utterances is studied in [18–20]; Gibbs discusses the effects of context on the processing of indirect requests [34, 35]. In [76–78], Osgood and others discuss effects of notions such as "naturalness" and "vividness."

In an attempt to subsume the work mentioned above and to include such aspects as opinions and social relationships (which pragmatics traditionally does not cover), the following categorization was derived:

- conversational atmosphere (physical setting),
- interlocutors' personal characteristics (factual knowledge, opinions, emotional states, interpersonal relationships, etc.),
- speaker's goals with respect to the hearer (effect on future behavior, opinions, relative status, etc.).

2.2. Pragmatics in PAULINE

In order to study the relationship between pragmatic considerations and computer language generation, one requires something concrete enough to program. To characterize the pragmatics of its conversation, PAULINE used a list of features based upon the abovementioned categorization. The precise names and values of these features are not a serious claim; the (pragmatic!) justification is that they are the kinds of features necessary for language generation. Any language processing program addressing these questions will have features that, on some level, resemble them. In this representation of pragmatics, each feature was given a fixed number of possible values, usually lying on a scale. In a few cases, features were conflated and the result merely given a set of distinct values; this could eventually be refined. PAULINE's characterization of the *conversation setting* and *interlocutor characteristics* is:

- conversational atmosphere (setting):
  - time: *much, some, little,*
  - tone: *formal, informal, festive,*
  - conditions: *good, noisy;*
- speaker:
  - knowledge of the topic: *expert, student, novice,*
  - interest in the topic: *high, low,*
  - opinions of the topic: *good, neutral, bad,*
  - emotional state: *happy, angry, calm;*
- hearer:
  - knowledge of the topic: *expert, student, novice,*
  - interest in the topic: *high, low,*
  - opinions of the topic: *good, neutral, bad,*
  - language ability: *high, low,*
  - emotional state: *happy, angry, calm;*

- speaker–hearer relationship:
  - depth of acquaintance: *friends, acquaintances, strangers,*
  - relative social status: *dominant, equal, subordinate,*
  - emotion: *like, neutral, dislike.*

In addition, PAULINE can have the following *interpersonal goals*:

- hearer:
  - affect hearer's knowledge: *teach, neutral, confuse,*
  - affect hearer's opinions of topic: *switch, none, reinforce,*
  - involve hearer in the conversation: *involve, neutral, repel,*
  - affect hearer's emotional state: *anger, neutral, calm,*
  - affect hearer's goals: *activate, neutral, deactivate;*
- speaker–hearer relationship:
  - affect hearer's emotion toward speaker: *respect, like, dislike,*
  - affect relative status: *dominant, equal, subordinate,*
  - affect interpersonal distance: *intimate, close, distant.*

Assuming that generation is influenced by pragmatic features of this kind, the question is: how? To answer this, we must consider what the generation process itself involves.

2.3. Decisions generators have to make

Any system that produces a sentence must perform a large number of tasks, each with characteristic decisions. These decisions range from having relatively wide-range syntactic effect, such as the selection of sentence themes (hereafter called *topics*) and their organization into phrases and sentences, to having relatively localized effect such as word choice:

- *topic collection*: for a given topic, collect its aspects and related items as candidate topics for other sentences;
- *topic organization*: find appropriate groupings and interpretations of the candidate topics; find appropriate ways to juxtapose them in multi-predicate phrases; find ways of expressing relationships among them;
- *sentence organization*: for each topic, select an appropriate subject, pre-sentence adjunct groups (adverbials, circumstantials, etc.), verb, predicate constituents, etc., and order them;
- *content and organization of sentence constituents*: determine and order the contents of adjunct groups and noun groups within sentences;
- *word choice*: select appropriate words and phrases.

Whatever the nature of the generator and the implementation of the grammar, all generators have to perform these tasks. The simplest programs perform them by having only one available option. However, as soon as a generator is given the ability to realize a topic in more than one way, it has to

be able to make its choice in a principled manner. Since different realizations carry different pragmatic effects, the pragmatic aspects of conversations *must* help determine the speaker's text by influencing the generator's decisions.

This argument seems plausible but runs into problems. Inevitably, attempts to write down rules that relate pragmatic aspects to generator production decisions become bogged down in minutiae and produce rules with very little credibility. For example, how can one determine sentence length from such features as listed above? What syntactic effects are manifested when the speaker is socially dominant over the hearer? Does the speaker's goal to be friendly have any bearing on active versus passive sentences? Most pragmatic aspects do not influence the decisions directly, since they are simply too general to be attuned to the requirements of language production.

2.4. Rhetorical goals

The solution proposed here is that speakers use a number of goals and associated strategies that act as intermediaries between, on the one hand, the speaker's interpersonal goals and perception of other pragmatic aspects of the conversation, and, on the other, the syntactic decisions the realization component has to make. The claim is that any generation system sophisticated enough to be controlled by communicative goals similar to those listed above and that has to make syntactic decisions sensitive to those goals will have to incorporate some such intermediate level of description. These goals will be called *rhetorical goals*. Two groups of strategies they control are presented here: the strategies of opinion and the strategies of style.

(1) *Rhetorical goals of opinion* achieve their results by a number of techniques, such as judicious topic collection, the appropriate juxtaposition of topics into phrases, the inclusion of stress words, adjectives, and adverbs, and the use of slanted words. These goals are described in Section 6.

(2) *Rhetorical goals of style* achieve their results by causing characteristic stylistic effects in the text. Through style, the speaker can communicate additional information that the hearer can interpret and respond to. Classifying all the possible styles of text is an impossible task: one can imagine text characteristics that fit almost any adjective! But certain features of text, such as formality and force, are generally accepted as stylistic. A study of some of the major handbooks of good writing [6, 8, 23, 46, 68, 81, 92, 95, 97] indicates that the authorities agree on a few common broad-based features in their discussions of style. These aspects they usually describe in terms of the characteristics of finished paragraphs of text.

Unfortunately, such a descriptive approach is of very little use in a theory of language production, since it never makes clear why and how each style is formed out of words; nor does it indicate any systematicity behind the classification of styles. In contrast to such descriptions, a functional approach is

to describe styles in terms of the decisions a generator has to make (as listed above). Certain types of decisions group together and form stylistically coherent text, and other types, when grouped, produce text that is stylistically incoherent or odd. The coherent groupings provide criteria for making generator decisions; they are the techniques for achieving rhetorical goals—and through them, high-level communicative goals (examples are given later). These criteria, then, form the link between the syntactic concerns of the generator and the rhetorical goals.

PAULINE uses the following stylistic rhetorical goals, with values along the indicated ranges:

- formality (*highfalutin, normal, colloquial*): highfalutin language is used for speeches;
- simplicity (*simple, normal, complex*): simple text has short sentences and easy words;
- timidity (*timid, reckless*): willingness to spend time to consider including opinions;
- partiality (*impartial, implicit, explicit*): how explicitly you state your opinions;
- detail (*details only, interpretations, both*): too many details can be boring to nonexperts;
- haste (*pressured, unplanned, somewhat planned, planned*): when there's little time. . . ;
- force (*forceful, normal, quiet*): forceful text is energetic and driving;
- floridity (*dry, neutral, flowery*): flowery text contains unusual words;
- color (*facts only, with color*): colorful text includes examples and idioms;
- personal reference (two ranges, for speaker and hearer): amount of direct reference to the interlocutors;
- openmindedness (*narrow-minded, openminded*): willingness to consider new topics;
- respect (four values): being arrogant, respectful, neutral, or cajoling.

Of course, it is impossible to list all possible styles. Every speaker has an idiosyncratic set of techniques, often tailored to particular hearers, for using language to achieve his or her interpersonal goals. Thus, this work should not be interpreted as claiming to describe exhaustively any language user's stylistic knowledge. Rather, it is intended as a description of the general *function of style* in a generator—the expression of rhetorical goals, which in turn serve the speaker's general communicative goals in the text; and of a useful *method of definition of style*—as constraints on the decisions the generator has to make.

2.5. The link with communicative goals

Having introduced the link between rhetorical goals and generator decisions, the question remains: what is the relation between rhetorical goals and the pragmatic aspects of communication? Sophisticated multifunctional generation

would be simple if a one-to-one correspondence existed between rhetorical and pragmatic goals. In this case, each rhetorical goal would simply be the repository for the generator-specific knowledge required to express its pragmatic partner: for example, the goal to cheer up the hearer would result in happy text, regardless of what other pragmatic conditions held for the conversation. Then there would be no reason for the separate existence of rhetorical goals. However, the pragmatic aspects of conversations are not independent, and this fact necessitates the existence of rhetorical goals as entities distinct from pragmatic features. To illustrate, note that a single rhetorical goal can express opposite pragmatic aspects under different conditions. For example, if the speaker has the goal to make the hearer feel close, he or she may activate a rhetorical goal to be humorous. Usually this will work well, but it will backfire if the hearer has just heard of his mother's death. In this case, the speaker's best strategy is to activate the rhetorical goals to be serious and slightly formal—which, under normal circumstances, would tend to alienate the hearer. Hence, combinations of rhetorical goals act *in concert* to produce pragmatic effects. For another example, the following three goals together

- formality: *low*,
- force: *high*,
- partiality: *high*,

have an effect on the text that can be characterized as no-nonsense; the style is distinctly pragmatic and clearly distinguishable from text produced under the combined opposite goals, which can be called *blather*. Thus, though rhetorical goals are activated by configurations of pragmatic aspects, they must exist separate from them in a generator.

Rhetorical goals, then, are the ways the speaker's pragmatic goals can index to (and can determine the application of) his or her stylistic and opinion-manipulative techniques, which control the decisions of the realization process. The advantages of defining and using rhetorical goals are that they enable one to make explicit, collect, and organize many generator strategies and design characteristics that most generators have left implicit or have avoided altogether.

The last three sections of the paper discuss the rhetorical goals detail, formality, and partiality. Others are described in [52].

3. Monitoring Interleaved Planning and Execution

This section briefly describes the second principal insight uncovered by the development of PAULINE. Any generation system that is controlled by a number of communicative goals simultaneously will have to operate under an execution-monitoring regime to orchestrate the harmonious execution of the

various plans and strategies needed to produce text. This requirement springs from the diverse nature of the communicative goals underlying generation: some goals can be explicitly planned for, achieved, and forgotten, while other goals, those for which no plans can be formulated, require intermittent satisfaction throughout the text. Plans for the former type of goal usually mandate long ranges of text, while strategies for the latter act over short (typically, single word) ranges of text, usually as a selection from some number of alternatives. The former type of planning, called here top-down or prescriptive planning, has been studied extensively in the AI planning community; the latter type of planning, called here restrictive planning, has only lately started to emerge.

3.1. The trouble with top-down planning

Our current understanding of language generation includes two phases: text planning and text realization. In some generation work, no text planning ever occurred: [25, 36, 57, 80, 90]. In other work, no serious attention was given to realization (such as Cohen's work on selecting appropriate speech acts [22]).

The issue of interaction between planning and production phases has been addressed in two principal ways. With the *integrated* approach, planning and generation is one continuous process: the planner-realizer handles syntactic constraints the same way it treats all other constraints (such as focus or lack of requisite hearer knowledge), the only difference being that syntactic constraints tend to appear late in the planning-realization process. Typically, the generator is written as a hierarchical expansion planner (see [83])—this approach is exemplified by KAMP, Appelt's planner-generator [1, 2]. With the *separated* approach, planning takes place in its entirety before realization starts; once planning is over, the planner is of no further use to the realizer. This is the case in the generation systems of [7, 71, 72, 74, 79].

Neither approach is satisfactory. Though conceptually more attractive, the integrated approach makes the grammar unwieldly (it is spread throughout the plan library) and is slow and impractical—after all, the realization process proper is not a planning task—and furthermore, it is not clear whether one could formulate all text planning and realization tasks in a sufficiently homogeneous set of terms to be handled by a single planner. On the other hand, the separated approach typically suffers from the stricture of a one-way narrow-bandwidth interface; such a planner could never take into account fortuitous syntactic opportunities—or even be aware of any syntactic notion! Though the separation permits the use of different representations for the planning and realization tasks, this solution is hardly better: once the planning stage is over, the realizer has no more recourse to it; if the realizer is able to fulfill more than one planner instruction at once, or if it is unable to fulfill an instruction, it has no way to bring about any replanning. Therefore, in practice, separated

generators perform only planning that has little or no syntactic import—
usually, the tasks of topic choice and sentence order. (This argument is made
more fully in [47, 73].)

Furthermore, both these models run counter to human behavior: When we
speak, we do not try to satisfy only one or two goals, and we operate (often,
and with success) with conflicting goals for which no resolution exists. We
usually begin to speak before we have planned out the full utterance, and then
proceed while performing certain planning tasks in bottom-up fashion.

3.2. A solution: Interleaved (limited-commitment) planning

A better solution is to perform *limited-commitment* planning—to defer plan-
ning until necessitated by the realization process. The planner need assemble
only a partial set of generator instructions—enough for the realization compo-
nent to start working on—and can then continue planning when the realization
component requires further guidance. This approach interleaves planning and
realization and is characterized by a two-way communication at the realizer's
decision points. The advantages are: First, it allows the separation of planning
and realization tasks, enabling them to be handled in appropriate terms. (In
fact, it even allows the separation of special-purpose planning tasks with
idiosyncratic representational requirements to be accommodated in special-
purpose planners.) Second, it allows planning to take into account unexpected
syntactic opportunities and inadequacies. Third, this approach accords well
with the psycholinguistic research of [24, 27, 62, 66, 77, 82]. This is the ap-
proach taken in PAULINE.

But there is a cost to this interleaving: the type of planning typically
activated by the realizer differs from traditional top-down planning. There are
three reasons for this:

(1) Top-down planning is prescriptive: it determines a series of actions over
an extended range of text. However, when the planner cannot expand its plan
to the final level of detail—remember, it doesn't have access to syntactic
information—then it has to complete its task by planning in-line, during
realization. After in-line planning culminates in a decision, realization process-
ing continues until the next unprovided-for decision point. Unfortunately,
unlike standard plan steps, the options at this point need not work toward the
same goal as before! In an in-line planning regime, subsequent planning
decisions need not have any relation with each other, nor any common goal.
The planner has no way to guess even remotely what the next set of options
and satisfiable goals might be.

(2) In-line planning is different for a second reason: it is impossible to
formulate workable plans for common speaker goals such as pragmatic goals.
A speaker may, for example, have the goals to impress the hearer, to make the
hearer feel socially subordinate, and yet to be relatively informal. These goals
play as large a role in generation as the speaker's goal to inform the hearer

about the topic. However, they cannot be achieved by constructing and following a top-down plan—what would the plan's steps prescribe? Certainly not the sentence "I want to impress you, but still make you feel subordinate"! Pragmatic effects are best achieved by making appropriate subtle decisions during the generation process: an extra adjective here, a slanted verb there. Typically, this is a matter of in-line planning.

(3) A third difference from traditional planning is the following: Some goals can be achieved, flushed from the goal list, and forgotten. Such goals (for example, the goal to communicate a certain set of topics) usually activate prescriptive plans. In contrast, other goals cannot ever be fully achieved. If your are formal, you are formal throughout the text; if you want to be friendly, arrogant, or opinionated, you remain so—you cannot suddenly become "friendly enough" and then flush that goal. These goals, which are pragmatic and stylistic in nature, are well suited to in-line planning.

Generation, then, requires two types of planning. Certain tasks are most easily performed in top-down fashion (that is, under guidance of a hierarchical planner, or of a schema or script applier), and other tasks are most naturally performed in a bottom-up, selective, fashion. That is, some tasks are *prescriptive*—they act over and give shape to long ranges of text—and some are *restrictive*—they act over short ranges of text, usually as a selection from some number of alternatives. Prescriptive strategies are formative: they control the construction and placement of parts in the paragraph and the sentence; that is, they make some commitment to the final form of the text (such as, for example, the inclusion and order of specific sentence topics). Restrictive strategies are selective: they decide among alternatives that were left open (such as, for example, the possibility of including additional topics under certain conditions, or the specific form of each sentence). A restrictive planner cannot simply plan *for*, it is constrained to plan *with*: the options it has to select from are presented to it by the realizer.

3.3. Planning restrictively: Monitoring

Restrictive planning is most appropriate for handling conflicting goals. In generation, situations such as the following are typical:

> rhetorical goal haste with value *high* $\Rightarrow$ start the next sentence immediately,
> rhetorical goal simplicity with value *low* $\Rightarrow$ wait for additional topics so as to conjoin two or more sentences;

or

> rhetorical goal simplicity with value *high* $\Rightarrow$ don't passivize; make simple sentences,
> rhetorical goal partiality with value *high* $\Rightarrow$ be partial, so suppress contentious parts, passivize to avoid a contentious actor;

or

> rhetorical goal partiality with value *high* ⇒ be partial; exclude a
> sentence topic that expresses a contrary opinion,
> rhetorical goal detail with value *high* ⇒ include all the details you
> know.

Fortunately, language is a medium that accommodates the temporary
nonachievement of some goals in favor of others. But some overall orchestra-
tion is obviously required to ensure that no crucial goal is violated or is ignored
for too long. The process of monitoring (called execution monitoring in the
planning literature; see, say, [11, 28, 31, 75, 83]) consists of the following:

- checking, updating, and recording the current satisfaction status of each
  goal;
- determining which goal(s) each option will help satisfy, to what extent, and
  in what ways;
- determining which goal(s) each option will thwart, to what extent, and in
  what ways;
- computing the relative priority of each goal in order to resolve conflicts (to
  decide, say, whether during instruction to change the topic or to wait for a
  socially dominant hearer to change it).

When the planner is uncertain about which long-term goals to pursue and
which sequence of actions to select, the following strategies (the first three of
which are from [29]) are useful:

- prefer *common intermediate* goals (subgoals shared by various goals);
- prefer *cheaper* goals (more easily achieved goals);
- prefer *discriminative intermediate* goals (goals that most effectively indicate
  the long-term promise of the avenue being explored);
- prefer *least-satisfied* goals (goals furthest from achievement);
- prefer *least-recently satisfied* goals (goals least recently advanced);
- use a *combination of the latter two* strategies (a goal receives higher priority
  the longer it waits and the fewer times it has been advanced).

This aspect of generation—goal conflict resolution, monitoring, the recogni-
tion and repair of mistakes—is a result of the dynamic interplay of goals and
opportunities of expression. It arises only upon consideration of a number of
communicative goals simultaneously seeking to direct the generation process.
This issue *must* be addressed by any sophisticated generator.

3.4. The architecture of PAULINE

Prescriptive planning is mostly performed during topic collection and topic
organization and restrictive planning is mostly performed during realization.
Restrictive planning is implemented in PAULINE in the following way: None of

the program's rhetorical goals are ever fully achieved and flushed; they require decisions to be made in their favor throughout the text. PAULINE simply tallies the number of times each such goal is satisfied by the selection of some option (of course, a single item may help satisfy a number of goals simultaneously). For conflict resolution, PAULINE uses the least-satisfied strategy, based on the lowest total satisfaction status. In order to do this, it must know which goals each option will help satisfy. Responsibility for providing this information lies with whatever produces the option: either the lexicon or the language specialist functions in the grammar.

PAULINE has the architecture shown in Fig. 1. Its input is represented in a standard case-frame-type language based on conceptual dependency [84, 85, 87] and is embedded in a property-inheritance network (see [10, 17]). The shantytown example consists of about 120 elements. No intermediate

Input Topics

Topic Collection	– topic collection: CONVINCE RELATE DESCRIBE	
Topic Organization	– interpretation – new topics – juxtaposition – ordering	RHETORICAL
Realization	– sentence type – organization – constituents – words	GOALS & STRATEGIES

Input:
Pragmatic
Aspects of
Conversation

Text

Information flow: growth of text specification

Inquiry and response: control of generator decisions
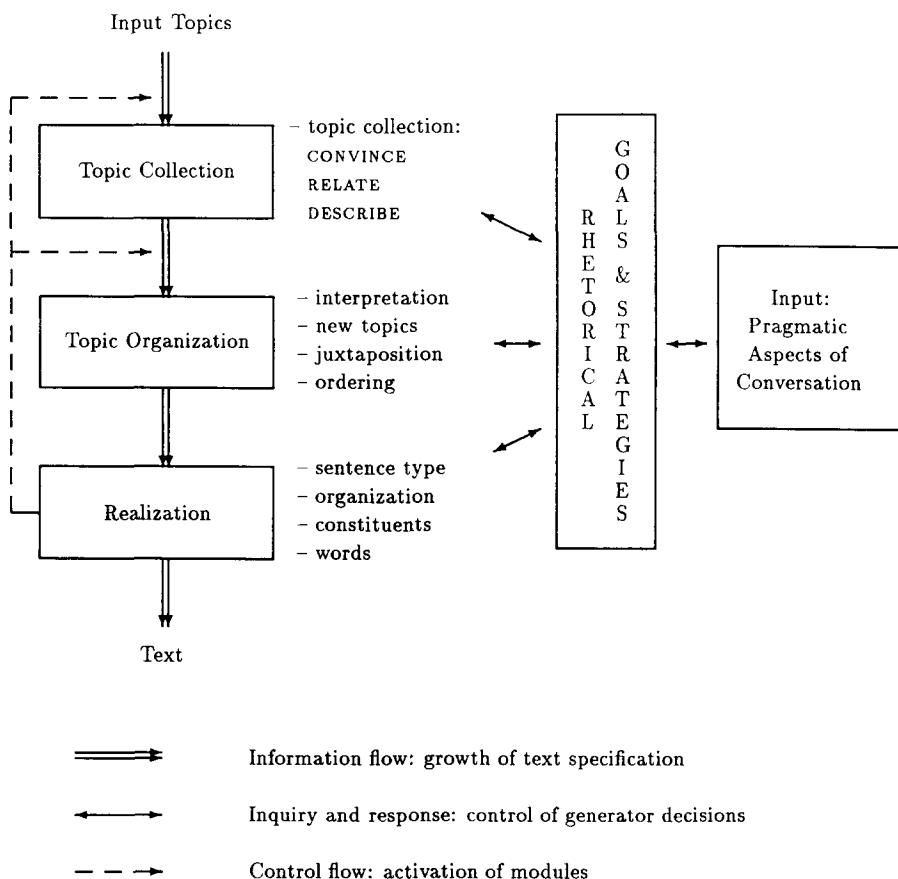
Control flow: activation of modules

Fig. 1. Program architecture.

representation (say, one that varies depending on the desired slant and style) is created. Its grammar is described in [53]. The program consists of about 12,000 lines of T, a scheme-like dialect of LISP developed at Yale. It takes about 2 minutes to generate an average-length shantytown paragraph (examples appear in the next three sections). A full description of the program appears in [52].

4. Formality

4.1. Introduction

In language, the level of formality is one of the strongest carriers of nonliteral information we use. This level reflects the level of formality of the conversational setting (for instance, a burial or a party) and of the interpersonal distance between the interlocutors. But what does it mean for language to "seem relaxed" or to "be formal"? No single item in the language defines the level of formality; rather, text seems to contain a number of little clues that cumulatively create a certain impression. What are these little clues? Where do they appear in language and how do we decide to use them?

As discussed in Section 2.4, handbooks of writing are of little use to answer this question. A more useful approach is to describe styles in terms of the decisions a generator has to make: decisions such as sentence content, sentence constituent order and content, and word selection.

4.2. Characteristics of formal text

All human language users have rules for making their text more or less formal. The desired level of textual formality comes into play along the whole range of generator decisions (from the initial sentence topic selection and organization down to the final word selection). The best way to illustrate these rules is to dissect a piece of text:

> Yesterday, December 7, 1941—a date which will live in infamy—the United States of America was suddenly and deliberately attacked by naval and air forces of the Empire of Japan.
>
> The United States was at peace with that nation and, at the solicitation of Japan, was still in conversation with its Government and its Emperor looking forward to the maintenance of peace in the Pacific.
>
> Indeed, one hour after Japanese air squadrons had commenced bombing Oahu, the Japanese Ambassador to the United States and his colleague delivered to the Secretary of State a formal reply to a recent American message. While this reply stated that it seemed useless to continue the existing diplomatic negotiations, it contained no threat or hint of war or armed attack.

It will be recorded that the distance of Hawaii from Japan makes it obvious that the attack was deliberately planned many days or even weeks ago. During the intervening time, the Japanese Government has deliberately sought to deceive the United States by false statements and expressions of hope for continued peace.

["We Will Gain the Inevitable Triumph—So Help Us God", war address by F.D. Roosevelt to joint session of Congress of the United States, December 8, 1941.]

What characteristics make this address formal? Certainly, one factor is the use of formal verbs and nouns instead of more common ones, such as "solicitation" instead of "request." Another factor is the use of full names and titles instead of their common abbreviations. Accordingly, one can replace words and phrases in the address by less formal equivalents ((a) below) and use the everyday names for entities (b).

The result, however, is definitely not informal. The sentences still seem long and involved. In order to simplify them, one can (c) remove conjunctions and multi-predicate phrases, and (d) remove adverbial groups, or place them toward the ends of sentences. Now, however, the text seems odd; for example, phrases such as "it will be recorded" do not blend with phrases such as "deliberately tried to cheat" (introduced by (a)). To improve this, (e) eliminate the use of passive voice, and (f) refer to the involved parties—speaker, hearer, and others—directly.

Now some phrases sound flowery and out of place. To simplify, some nominalized verbs can be converted to verbs (g); noun groups can be simplified by dropping redundant adjectives and nouns (h); pronominalization can be increased (i). Finally, a few finishing touches: simplified tenses (j); colloquial phrases (k); complete elision of redundant words where grammatical (l). The result, while not yet exactly street slang, is much less formal:

The $US_{(h)}$ was suddenly and deliberately attacked by naval and air forces of $Japan_{(b)}$ yesterday, December 7, $1941_{(d)}$. We'll never forget this $date_{(c,d,k,l)}$

We $were_{(b,f)}$ at peace with $them_{(i)}$. [and,]$_{(c)}$ At Japan's $request_{(a,h)}$ we $were_{(b,f)}$ still talking $to_{(a)}$ $their_{(e)}$ Government. [and its Emperor.]$_{(h)}$ We $were_{(b,f)}$ looking forward to $having_{(a,g)}$ peace in the Pacific.

[Indeed,]$_{(l)}$ One hour after Japanese air squadrons [had]$_{(j)}$ $started_{(a)}$ bombing Oahu, their $Ambassador_{(a)}$ [and his colleague]$_{(l)}$ $gave_{(a)}$ $our_{(f)}$ Secretary of State a formal reply to a recent message. [While]$_{(c)}$ [This reply $said_{(a)}$ that]$_{(l)}$ $They_{(b,f,i)}$ thought it $was_{(e)}$ useless to continue $negotiating_{(g)}$. [there $was_{(a)}$]$_{(c)}$ But $they_{(i)}$ $didn't_{(k)}$ [threaten or]$_{(h)}$ talk $about_{(a)}$ war. [or armed attack.]$_{(h)}$

[Note$_{(e,f)}$ that]$_{(l)}$ The distance of Hawaii from Japan makes it obvious that they$_{(f)}$ deliberately planned$_{(e)}$ the attack a while$_{(k)}$ [or even weeks]$_{(l)}$ ago. [In$_{(a)}$ the intervening time,]$_{(d)}$ The Japanese Government [has]$_{(j)}$ deliberately tried$_{(a)}$ to cheat$_{(a)}$ us$_{(b,f)}$ by [false statements and]$_{(h)}$ pretending$_{(a)}$ [expressions of hope for continued]$_{(l)}$ to hope for peace in the mean time$_{(k)}$.

4.3. Rules for creating formal text

A number of texts, ranging from politicians' speeches and writings to discussions with friends, were analyzed in the manner above. The transformation steps above were stated as rules that provide criteria by which PAULINE makes appropriate choices at decision points (more details appear in [51]). One of the program's rhetorical goals, the goal controlling formality, takes one of the values *highfalutin, normal, colloquial*. In order to make text more formal, the program examines its options at decision points and applies the strategies paraphrased as follows:

- *topic inclusion*: to make long sentences, select options that contain causal, temporal, or other relations to other sentence topics;
- *topic organization*: to make complex sentences, select options that are subordinated in relative clauses; that conjoin two or more sentence topics; that are juxtaposed into relations and multi-predicate enhancer and mitigator phrases;
- *sentence organization*: make sentence seem weighty by including many adverbial groups; by placing these groups toward the beginnings of sentences; by building parallel clauses within sentences; by using passive voice; by using more "complex" tenses such as the perfect tenses; by avoiding ellipsis, even though it may be grammatical (such as "Joe got more than Pete [did]," "when [I was] 20 years old, I got married");
- *sentence constituent organization*: make weighty, formal noun groups, by including many adjectives and adjectival groups in noun groups; by doubling nouns in noun groups ("Government and Emperor," "statements and expressions"); by including many adverbs and stress words in predicates; by using long, formal phrases; by nominalizing verbs and adverbs ("their flight circled the tree" instead of "they flew round the tree"); by prenominalizing where possible; by not referring directly to the interlocutors or the setting;
- *phrase/word choice*: select formal phrases and words; avoid doubtful grammar, slang, and contractions (say "man" rather than "guy" and "cannot" rather than "can't").

In contrast, by following inverted strategies, PAULINE makes its text less formal.

4.4 Determining appropriate levels of formality

Knowing how to make formal text is not enough. The generator must also know when it is appropriate. Since the level of formality is not actually measurable, it is most apparent only when the level is suddenly changed or is inappropriate. In order to determine the pragmatic effects of formality, then, the important question is: *what does the speaker achieve by altering the level of formality?*

First, since formality in language mirrors interpersonal distance between speaker and hearer, when your language becomes less formal you signal a perceived or desired decrease in distance. Which interpersonal distance corresponds to which level of formality and how the correspondence is signaled depends, of course, on social convention and on the interlocutors and their relationship. See [12] on the use of formal honorifics and [45, 63] on Japanese deictic honorifics.

Second, if you alter the level of textual formality, you may perturb the tone or atmosphere of the conversation. Since the conversational atmosphere is also mirrored by textual formality, a serious conversation (a burial speech or a conference talk) requires more formality than an everyday conversation (a report to the family of the day's events). An inappropriate level of formality can affect the hearer's emotion toward you: if you are too informal, you may seem cheeky or irreverent; if you are too distant, you may seem snooty or cold. A large amount of work by sociologists, anthropologists, and psycholinguists describes the characteristics of various settings and the appropriate levels of formality in various cultures (see, for example, [3, 54] on formal events; [37, 64] on politeness).

Based on these considerations, after PAULINE is given values for the parameters that characterize the conversational setting, the speaker, and the hearer (in sanserif), it uses the following rules to activate its rhetorical goal of formality:

(1) set the rhetorical goal of formality to
 − *colloquial* when the depth of acquaintance is marked *friends*, or when the *relative social status* is marked *equals* in an atmosphere (tone) marked *informal*,
 − *normal* when the depth of acquaintance is marked *acquaintances*,
 − *highfalutin* when the depth of acquaintance is marked *strangers*;
(2) then, reset the goal value one step toward *colloquial* if desired effect on interpersonal distance is marked *close* or if tone is marked *informal*;
(3) or reset the goal value one step toward *highfalutin* if desired effect on interpersonal distance is marked *distant* or if tone is marked *formal*;
(4) and invert the value if desired effect on hearer's emotion toward speaker is marked *dislike* or if desired effect on hearer's emotional state is marked *anger*.

4.5. The rules at work

PAULINE uses these rules to produce the two shantytown texts in Table 1 when it is being *highfalutin* (say, writing for a newspaper) and *colloquial* (say, talking to a friend).

Table 1
Formality

Highfalutin	Colloquial	Decision type
[IN EARLY APRIL],	[ ]	adjunct position
	STUDENTS [PUT]	verb formality
A SHANTYTOWN—[NAMED]	A SHANTYTOWN, [ ]	ellipsis
WINNIE MANDELA CITY—	WINNIE MANDELA CITY, UP	
[WAS ERECTED] BY]		mode, verb formality
[SEVERAL] STUDENTS		adjective inclusion
ON BEINECKE PLAZA,	ON BEINECKE PLAZA	
	[IN EARLY APRIL].	
[SO THAT]		conjunction
	THE STUDENTS WANTED	
YALE UNIVERSITY WOULD	YALE UNIVERSITY TO	
[DIVEST FROM]	[PULL THEIR MONEY OUT OF]	verb formality
COMPANIES DOING	COMPANIES DOING	
BUSINESS IN SOUTH AFRICA.	BUSINESS IN SOUTH AFRICA.	
[LATER, AT 5:30 AM ON	[ ]	adverb group, position
APRIL 14],		
THE SHANTYTOWN	OFFICIALS	
[WAS DESTROYED]	[TORE IT DOWN]	mode, verb formality
BY OFFICIALS;	AT 5:30 AM ON APRIL 14,	
[ALSO, AT THAT TIME,] THE	[AND] THE	conjunction, adjunct
POLICE ARRESTED 76	POLICE ARRESTED 76	
STUDENTS.	STUDENTS.	
SEVERAL LOCAL POLITICIANS	SEVERAL LOCAL POLITICIANS	
AND FACULTY MEMBERS	AND FACULTY MEMBERS	
[EXPRESSED CRITICISM]	[CRITICIZED]	verb formality
OF [YALE'S] ACTION.	THE [ ] ACTION.	adjective inclusion
[FINALLY],	[LATER,]	word formality
YALE [GAVE]	YALE [ALLOWED]	verb formality
THE STUDENTS [PERMISSION]	THE STUDENTS	
TO [REASSEMBLE]	TO [PUT IT UP]	verb formality
THE SHANTYTOWN THERE	THERE AGAIN	
[AND, CONCURRENTLY],	[ ]	conjunction
THE UNIVERSITY [ANNOUNCED]	THE UNIVERSITY [SAID]	verb formality
THAT A COMMISSION WOULD	THAT A COMMISSION WOULD	
GO TO SOUTH AFRICA IN JULY	GO TO SOUTH AFRICA IN JULY	
TO [INVESTIGATE] THE	TO [STUDY] THE	verb formality
SYSTEM OF APARTHEID.	SYSTEM OF APARTHEID.	

5. Detail

5.1. The problem

Simply put, given sentence topic, the generator's task is to find a form of expression—either a syntactic rule or a phrase—that will enable it to select and to order aspects of the topic in order to build a sentence. The straightforward approach is to define a fixed correspondence between topic representation types on the one hand and grammatical rules and lexical elements on the other. This approach has a flaw: the results are invariably bad or boring. How bad, of course, depends on the representation, but anything detailed enough to be useful for other purposes, such as learning or diagnosing, simply does not make great prose in practice. A good example is furnished by the following text, in which the generator's input consists of a list of topics, where each topic describes some episode in a fight between two people. Straightforward generation (by PAULINE, using a simplified grammar) produces:

(a)   FIRST, JIM BUMPED MIKE ONCE, HURTING HIM. THEN MIKE HIT JIM, HURTING HIM. THEN JIM HIT MIKE ONCE, KNOCKING HIM DOWN. THEN MIKE HIT JIM SEVERAL TIMES, KNOCKING HIM DOWN. THEN JIM SLAPPED MIKE SEVERAL TIMES, HURTING HIM. THEN MIKE STABBED JIM. AS A RESULT, JIM DIED.

(The input was produced by Bain's JUDGE program [4, 5], a case-based expert system that models the sentencing behavior of a judge. As input, JUDGE accepts the representation of a fight—a set of actions and resulting states—and as output it produces a set of interpretations of each action.) This example is an extreme case because it contains only two main representation types, ACTION and STATE, which can relate in only one way, RESULT. When the generator knows only one way to express this combination, what more can one hope for?

Correcting this inflexibility seems straightforward. Though there is nothing wrong with the sentence form used above, namely,

$$[[\textit{say-time} \ \#\text{TIME}] \ [\textit{say-sentence} \ \#\text{ACTION}], \ [\textit{say-participle} \ \#\text{STATE}]] \ ,$$

one can add to the grammar a few more sentence forms expressing actions and their results, more time words, and more verbs, and then make the generator cycle through its options whenever it encounters a choice point:

(b)   FIRST, JIM BUMPED MIKE ONCE AND HURT HIM. THEN MIKE SMACKED JIM, HURTING HIM. NEXT, JIM HIT MIKE ONCE. THE RESULT WAS THAT HE KNOCKED HIM DOWN. AFTER THAT, MIKE SMACKED JIM SEVERAL TIMES AND KNOCKED HIM DOWN. JIM SLAPPED MIKE SEVERAL TIMES, HURTING HIM. AFTER THAT, MIKE STABBED JIM. AS A RESULT, JIM DIED.

Yet this produces no real improvement! Clearly, simply extending the number of phrase patterns for each representation type does not solve the problem. When we speak, we do a lot more than simply cast input topics in different forms; for example, we might say:

(c)   JIM DIED IN A FIGHT WITH MIKE.

(d)   AFTER JIM BUMPED MIKE ONCE, THEY FOUGHT, AND EVENTUALLY MIKE KILLED JIM.

(e)   AFTER JIM BUMPED MIKE ONCE, THEY FOUGHT, AND EVENTUALLY HE WAS KNOCKED TO THE GROUND BY MIKE. HE SLAPPED MIKE A FEW TIMES. THEN MIKE STABBED JIM AND JIM DIED.

Illustrated this way, the problem seems rather simple. Obviously, the solution is to group together similar enough topics, where the similarity criterion can be varied depending on external factors, and then to generate the groupings instead of the individual actions. Grouping together contiguous actions of similar force, PAULINE produced variants (c), (d), and (e). (In the first variant, all actions were grouped together; in the second, all actions more violent than bumping but less violent than killing were accepted; and in the third, the grouping resulted from defining four levels of violence: bumping, hitting and slapping, knocking to the ground, and killing.)

Clearly, though it improves the JUDGE examples, the technique of grouping actions by levels of force is very specific and not very useful. However, when "group" is used in a wider sense to mean "interpret," this technique becomes both difficult and interesting, and provides a very powerful way to increase the expressive flexibility and text quality of a generator. So the questions are: what interpretation/grouping criteria are general and still useful? When and how should the generator interpret input topics? How should it find appropriate grouping criteria?

5.2. An example of interpretation

In a second example, PAULINE produces a number of versions describing a hypothetical primary election between Carter and Kennedy during the 1980 Democratic Presidential nomination race. In the election, Kennedy narrows Carter's lead. The underlying representation comprises about 80 distinct units. When PAULINE is given as input the outcome for each candidate, straightforward generation produces:

(f)   IN THE PRIMARY ON 20 FEBRUARY CARTER GOT 1850 VOTES. KENNEDY GOT 2185.

However, PAULINE can notice that both outcomes relate to the same primary, and can say instead:

  (g)   IN THE PRIMARY ON 20 FEBRUARY, KENNEDY BEAT CARTER BY 335 VOTES.

(or any of a number of similar sentences using "beat," "win," and "lose"). But why stop there? If PAULINE examines the input further, it can notice that Carter's current delegate count is greater than Kennedy's, that this was also the case before the primary, and that this primary is part of a series that culminates in the final election, the nomination. In other words, PAULINE can recognize that what happened in this primary was:

  (h)   IN THE PRIMARY ON 20 FEBRUARY, KENNEDY NARROWED CARTER'S LEAD BY GETTING 2185 VOTES TO HIS 1850.

If we want good text from our generators, we have to give them the ability to recognize that "beat" or "lose" or "narrow lead" can be used instead of only the straightforward sentences (f).

This ability is more than a simple grouping of the two outcomes. It is an act of generator-directed inference, of interpretation, forming out of the two topics a new topic, perhaps one that does not even exist in memory yet. And the new topic is not simply a generator construct, but is a valid concept in memory. The act of determining that "beat" is appropriate *is* the act of interpreting the input as an instance of BEAT—denying this is to imply that "beat" can logically be used where BEAT is not appropriate, which is a contradiction. This is not an obvious point; one could hold that the task of finding "beat" to satisfy a syntactic or pragmatic goal is a legitimate generator function, whereas the task of instantiating it and incorporating it into memory is not. However, it is clearly inefficient for a generator to interpret its input, say it, and then simply forget it again!—especially when there is no principled reason why generator inferences should be distinct from other memory processes.

Thus, after interpretation, the newly built instance of the concept should be added to the story representation, where it can also be used by other processes, or by the generator the next time it tells the story. In this way the content of memory can change as a result of generation. This is consistent with the fact that you often understand a topic better after you have told someone about it: the act of generating has caused you to make explicit and to remember some information you didn't have before.

Immediately, this view poses the question: *which process is responsible for making these inferences?* The two possible positions on this issue reflect the amount of work one expects the generator to do. According to the strict minimalist position—a position held by most, if not all, generator builders today—the generator's responsibility is to produce text that faithfully mirrors

the input topics with minimal deviation: each sentence-level input topic produces a distinct output sentence (though perhaps conjoined with or subordinated to another). This inflexible attitude gave rise to the JUDGE texts (a) and (b). To circumvent this problem, in practice, most generator builders employ in their programs a number of special-purpose techniques, such as sophisticated sentence specialists that are sensitive to the subsequent input topics. This is a tacit acknowledgment that the strict position does not hold. However, on renouncing the hard-line position, one must face the question *how much generator-directed inference are you prepared to do*?

I do not believe that a simple answer can be given to this question. The issue here is economic: a tradeoff exists between the time and effort required to do interpretation (which includes finding candidate interpretations, making them, and deciding on one) on the one hand, and the importance of flowing, good text on the other. Greater expense in time and effort produces better text. Thus pragmatic criteria are appropriate for treating this question. Hence a reasonable answer is *I'll do as much inference as I can do, given the available time, the pragmatic constraints on what I want the hearer to know, and the richness of my memory and my lexicon*. Of these three factors, the most difficult is clearly the pragmatic constraints on what the hearer is to be told. When does the hearer need to know the details of the topic? What is the effect of saying only interpretations? Or of saying both? The answer can be summarized as: if you can trust the hearer to make the interpretations himself, then all you need say are the details. Thus, if the hearer is a political pundit who is following the nomination race with interest, then clearly (f) is better, since he or she can draw the conclusion without difficulty, and, in addition, now has precise numerical information. If, in contrast, the hearer has only minimal knowledge about or interest in the nomination procedure, then (h) is better, since it removes the burden of details and the task of doing the interpretation. What must you say, however, if the hearer is interested and has a limited amount of knowledge—say, he or she is a student of the political process—or is knowledgeable but unlikely to make the right interpretation—say, he or she is a strong Kennedy supporter, whereas you are pro-Carter? In both these cases you must ensure that the hearer understands how you expect him or her to interpret the facts. So you give the details *and* the interpretations:

(i)   KENNEDY NARROWED CARTER'S LEAD IN THE PRIMARY ON 20
       FEBRUARY. HE GOT 2185 VOTES AND CARTER GOT 1850.

These considerations can be stated as rules, using the terms defined above to characterize the pragmatic aspects of conversations and the goals of speakers). PAULINE uses these rules to activate the rhetorical goal detail that controls the level of detail of topics generated. The goal takes one of the values *details, interpretations, all* (both details and interpretations):

- Set the goal's value to *details* if the hearer is likely to understand the details or wants to hear the details. This rule bears on information about the hearer: Is the hearer's knowledge level marked *expert*; or is the hearer's interest level marked *high*?
- Otherwise, set it to *all* if the hearer is likely to make the wrong interpretations of the details, that is, when the hearer's knowledge level is marked *student* or *novice*; the atmosphere (time) is not marked *little*; and the hearer's sympathies and antipathies for the central topic of the conversation are not the opposite of the speaker's.
- Otherwise, set it to *interpretations*.

In addition to these considerations, the value of the goal can be affected by the desire not to upset the hearer:

- Then, set the value to *interpretations* if it is better to avoid painful topics, to ensure that painful aspects (the details, the interpretation, or the inferences used to make it) can simply be left out. This rule translates as follows: Is speaker-hearer depth of acquaintance marked *strangers*, or is speaker-hearer relative social status marked *subordinate*, or is desired effect on hearer's emotion toward speaker marked *like*, or is desired effect on interpersonal distance marked *close*, or is desired effect on hearer's emotional state marked *calm*?

In summary, you must be as specific as the hearer's knowledge of the topic allows: if you are too specific he or she won't understand, and if you are too general you run the risk of seeming to hide things, or of being uncooperative. In the first case, you violate the goal to be intelligible, and in the second, you violate the goal to avoid unacceptable implications. In either case, you violate Grice's maxim of quantity to say neither more nor less than is required [39].

5.3. Finding candidates and making interpretations

In order to interpret the input topics as instances of some concept, the interpretation process must recognize when the topics (or some of them) conform to the definition (or part of the definition) of the concept. Thus, either concepts must be defined in such a way as to allow their definitions to be read, or inferences must exist that fire when a definition is matched—in other words, where the antecedent of an inference is the definition and the consequent asserts the existence of the new concept.

PAULINE was implemented with the second approach, using patterns called configurations (more details can be found in [50]). A configuration is the description of the way in which a collection of concepts must relate to one

other to form a legitimate instance of a high-level concept. For example, the
configuration matching the concept BEAT is

> ; ?X is someone's CONTEST-OUTCOME in some contest ?Y,
>
> (CONTEST-OUTCOME ?X
>
>   (INSTANCE (ELECTION ?Y))
>
>   ; and it is greater than another CONTEST-OUTCOME in ?Y
>
>   (RELATIONS (REL-GREATER ()
>
>   (CONC1 (?X))
>
>   (CONC2 (CONTEST-OUTCOME ()
>
>   (INSTANCE (?Y)))))))))

which means: some concept is a CONTEST-OUTCOME; its aspect RELATIONS
contains a GREATER relation, of which the greater part is that same concept
and the smaller part is another CONTEST-OUTCOME in the same primary. Thus,
since Kennedy's outcome resulted from a primary and it is greater than
Carter's outcome, the two form an instance of BEATing.

For the Carter–Kennedy and shantytown examples, PAULINE has 14 configu-
rations beside BEAT, all considerably more complex. These configuration
patterns obviously depend on the exact representations used. A more sophisti-
cated implementation of this idea would use an automatic classifier system such
as the KL-ONE classifier (see [89]).

The problem in interpretation is to find valid interpretations easily and
quickly. This can be done in two ways.

One solution to this problem is to run interpretation inferences directly on
the input topics. In this *bottom-up* method of interpretation, inferences reside
in memory and the lexicon, linked to definitions of concept types. (This
scheme forms a concept representation network slightly different from the
usual multi-parent schemes used in, say, [10, 17, 91]). From the type of each
input topic, the program collects and runs inferences, collects the results. This
is not a wonderful solution—it depends on the right links being defined
beforehand—but it is practical in limited domains.

Another way to find interpretations is *top-down*: potentially useful infer-
ences can be explicitly included in plans that serve the generator's pragmatic
goals, and can be tried on candidate sentence topics. Since interpretation is a
powerful way of slanting the text, the pragmatic goals to communicate opinions
(see Section 6) are an eminently suitable source of guidance. Indeed, many of
these goals can *only* be achieved through interpreting the input topics appropri-
ately.

During its planning stage, PAULINE gathers likely interpretation inferences,
both top-down and bottom-up, and then, using a simple pattern-matcher,
applies their configurations to the candidate topics and collects all the matches.
Its strategies for selecting configurations are based upon the pragmatic factors

knowledge, slant, and time, described above. If an instance of a newly made interpretation does not yet exist in memory, PAULINE creates one and indexes it following the memory organization principles described in [86], so that it can be found again and used in future.

For example, when generating the shantytown text with the goal to slant the input, PAULINE uses top-down inferences to interpret the input topics appropriately (see the section on partiality below for more details):

(k)   IN EARLY APRIL, A SMALL NUMBER OF STUDENTS [*WERE IN-VOLVED IN A CONFRONTATION*]$_{(a)}$ WITH YALE UNIVERSITY OVER YALE'S INVESTMENT IN COMPANIES DOING BUSINESS IN SOUTH AFRICA. THE STUDENTS [*TOOK OVER*]$_{(b)}$ BEINECKE PLAZA AND CONSTRUCTED A SHANTYTOWN NAMED WINNIE MANDELA CITY [*IN ORDER TO FORCE*]$_{(c)}$ THE UNIVERSITY TO DIVEST FROM THOSE COMPANIES. YALE REQUESTED THAT THE STUDENTS ERECT IT ELSEWHERE, BUT THEY REFUSED TO LEAVE. LATER, AT 5:30 AM ON APRIL 14, OFFICIALS HAD TO DISASSEMBLE THE SHANTYTOWN. FINALLY, YALE, [*BEING CONCILIATORY*]$_{(d)}$ TOWARD THE STU-DENTS, NOT ONLY PERMITTED THEM TO RECONSTRUCT IT, BUT ALSO ANNOUNCED THAT A COMMISSION WOULD GO TO SOUTH AFRICA IN JULY TO EXAMINE THE SYSTEM OF APARTHEID.

PAULINE made the interpretations *confrontation* (a), *appropriation* (b), *coercion* (c), and *conciliation* (d), none of which were contained in the original input story.

As generators become larger and more complex, and as they are increasingly used together with other programs, they should use the capabilities of those programs to further their own ends. Therefore, we should study the kinds of tasks that generators share with other processes and the purposes generators require them to fulfill. The strategies described here embody some of the kinds of demands a generator can be expected to place on a general-purpose inference engine. And even with PAULINE's limited inferential capability, the program can greatly enhance the quality of its text and the efficiency of its communication of nonliteral pragmatic information.

6. Partiality

6.1. Introduction

Much language is devoted to the communication of opinions. Such communication is often implicit; our biases sneak into what we say so easily and so often that producing genuinely unslanted text can be quite difficult! For example, compare the first two sentences from the protester and university texts in Table 2.

Table 2

Protester	University
On April 4, *concerned* Yale students	On April 4, *a small group of* students *took over*
constructed a shantytown on *Boesak Plaza*	*Beinecke Plaza* and built some shanties;
as a *reminder* to *those in Woodbridge Hall* that Yale is complicit. . .	they wanted to *force* *Yale* to sell its stocks. . .

Clearly, the protester had a reason for saying "concerned" where the university speaker said "a small group of"; also, the latter wanted to imply something specific with "took over"; and even the different names used for the location (Boesak Plaza and Beinecke Plaza, the official name) and for the university ("Yale" and "those in Woodbridge Hall," which is the president's office) have different connotations. These differences are not haphazard; each speaker made the decisions that slanted the text in his or her favor. What are these decisions? When and how are they made?

6.2. Affect in PAULINE

In order to slant the text to fit the hearer's opinions, the speaker must be able to determine what the hearer is likely to find sympathetic, what he or she is likely to dislike, and what he or she is likely not to care about much. PAULINE uses three values of affect: GOOD, BAD, and NEUTRAL. (Of course, *affect* here simply denotes something akin to "like." But even with this limited denotation, three values are sufficient to give the program interesting behavior. In this regard it is similar to the work on narrative summarization in [65].)

PAULINE's affects derive from two sources: provided by the user and defined as intrinsic to certain representation elements. To give PAULINE opinions, the user must specify one or more representation elements as *sympathies* or as *antipathies*. (In PAULINE, this is simply implemented by having a sympathy and an antipathy list. Elements on these lists will be characterized as GOOD and BAD respectively.) The second source of affect is defined for those generic representation elements that carry some intrinsic affect in the example domain. For example, in neutral context, the concept ARREST is BAD, the university's goal to be reasonable and fair is GOOD, and all other concepts, such as STUDENTS and CONSTRUCTION, are NEUTRAL.

In order to compute an opinion about any arbitrary piece of input representation, PAULINE has the ability to combine its given affects and concepts' intrinsic affects and to propagate affect along relations to other concepts. Though their exact form obviously depends on the design of the representa-

tion, the basic rules are:

(1) affect is preserved when combined with NEUTRAL;
(2) like affects combine to GOOD;
(3) unlike affects combine to BAD;
(4) affect inverts when propagated along certain relations (e.g., the *patient* of a BAD act is GOOD). A special rule for affect propagation is defined for each such relation.

6.3. Determining partiality

When should the speaker exhibit partiality? In general, since his or her sympathies and antipathies reflect so accurately the speaker's disposition toward the world, any opinion with which the hearer disagrees implies distance between them—perhaps even censure on the part of the speaker. Thus, to simplify, when the speaker's opinion agrees with the hearer's, expressing it will tend to make them closer; when it disagrees, expressing it may cause problems. Furthermore, partiality can be expressed explicitly, using clauses that state the speaker's opinion, or implicitly, using techniques such as phrasal juxtaposition and stress words. The rules PAULINE uses to activate its rhetorical goal of partiality are:

(1) Set the value of the goal to *explicit* if the speaker's and hearer's affects for the topic agree and desired effect on hearer's emotion toward speaker is marked *like*; or desired effect on interpersonal distance is marked *close*; or tone is marked *informal*.
(2) Set it to *implicit* if the speaker's and hearer's affects for the topic agree and desired effect on interpersonal distance is marked *distant*, since being lukewarm about the agreement with the hearer separates them; or speaker-hearer relative social status is marked *dominant*, for the same reason; or desire to involve hearer is marked *repel*, that is, if the speaker does not want to make the hearer too involved in the conversation.
(3) Otherwise, set it to *impartial* if their affects agree, or if their affects disagree and desired effect on hearer's opinion is marked *none*, hearer's knowledge level is marked *expert*, and speaker's knowledge level is marked *student* or *novice*, and desired effect on hearer's emotion toward speaker is marked *respect* or *like*, since when the speaker cares about an expert hearer's opinion, he or she will not want to exhibit partiality and lack of knowledge.
(4) Set the value of the goal to *explicit* if the speaker's and hearer's affects for the topic disagree and desired effect on hearer's opinion is marked *switch*; or desired effect on hearer's emotional state is marked *anger*; or desired effect on hearer's emotion toward speaker is marked *dislike*; or desired effect on interpersonal distance is marked *distant*.

(5) Otherwise, set it to *implicit* if their affects disagree and desired effect on hearer's opinion is marked *switch*; or desire to involve hearer is marked *involve*; or relative social status is marked *subordinate* (that is, when the hearer is subordinate to the speaker).

Having determined a value for this goal of partiality, PAULINE uses the following strategies of style that act as criteria at decision points to make text partial (both *explicit* and *implicit*):

(1) *topic inclusion*: include explicit expressions of opinion (if *explicit*);
(2) *topic organization*: make appropriate interpretations of topics, as discussed below (if *implicit*);
(3) *topic/phrase organization*: juxtapose topics in affect-imputing phrases (*explicit* and *implicit*);
(4) *sentence inclusion and organization*: include appropriate descriptive adjunct groups, adverbial and adjectival (*explicit*);
(5) *sentence constituent inclusion*: include appropriate affect-laden adjectives and adverbs; and include stress words (*explicit* and *implicit*);
(6) *word choice*: select nouns and verbs that carry affect (*explicit* and *implicit*).

In contrast, in order to make its text as *impartial* as possible, the program uses inverse strategies.

6.4. Slanting techniques

As suggested by the abovementioned list, affect can be injected into text by a number of different slanting techniques. These techniques are either *content-related*, involving judicious topic collection and the biased interpretation of topics, or *form-related*, involving the appropriate juxtaposition of topics in phrases, the use of enhancing and mitigating stress words, and prejudiced word choice.

6.4.1. *Biased topic collection*

As input, PAULINE is given one or more initial sentence topics. It uses one of three topic collection plans (the CONVINCE plan, if the goal partiality is not *impartial*; the DESCRIBE plan (similar to the schemas described in [74]) for objects; and otherwise the RELATE plan) to collect additional topics from the concept representation network. Each step of a topic collection plan tells the generator where in the representation network to find other potentially useful sentence topics. Applying each step to the input topics, PAULINE gathers additional topics, and, for each of them, activates and runs an appropriate collection plan in turn, thereby doing a depth-first topic search and collection.

When the hearer shares PAULINE's affect for a topic, there is no problem; the topic can simply be said directly. But when they disagree, the program has to

be more careful. The strategy tried in the Carter–Kennedy examples (and described more fully in [48]) is *evasion*:

The wishful suppression and mitigation plan:
– Say GOOD topics.
– Juxtapose NEUTRAL topics with GOOD ones in enhancer phrases (explained below).
– Leave out BAD topics altogether, unless they can be mitigated using mitigator phrases and words, or unless they are central to the story.

Sometimes, however, the program is explicitly given the goal to discuss a sensitive topic. In such cases PAULINE uses the strategy of *selectivity*: saying only aspects of topics that support its opinion. The CONVINCE plan, described more fully in [47], contains the following suggestions:

The convince plan:
– Consider *worse examples* of the topic with the sensitive aspect—from the concept(s) immediately superior to the topic in the memory network, compute the affects of other, similar instances, and collect those with BAD affect.
– Consider *good results* of the topic with the sensitive aspect—collect all the results and outcomes of the topic (if it is (part of) a goal, a plan, or a script, examine the final outcomes too) if they are GOOD.
– Consider *good intergoal relations* of the topic with the sensitive aspect— collect those goals the topic supports, opposes, is a side-effect of if they are GOOD.
– Consider *good side-effects* of the topic with the sensitive aspect—collect all the side-effects of the topic (if it is a goal or a plan) as far as they are known to be GOOD.
– *Appeal to authority*—if any of the immediate aspects of the topic refer to people or organizations who share in, have, or support the sensitive aspect, and if the hearer's affect for these authorities is GOOD, collect them (not implemented in PAULINE).
– Simply *enhance or mitigate the topic* with the sensitive aspect—just say it and allow subsequent realization decisions to give it the appropriate slant.

In the shantytown example of Table 3, PAULINE is given three input topics (the building of the shanties, their being taken down, and Yale's permission for them to be rebuilt). When the program has the goal to switch the hearer's opinions to correspond to its own, it activates the CONVINCE topic collection plan. When PAULINE is speaking as a university supporter, the *good results* step of the plan causes it to collect, as additional topics, the university's offer of an alternative site for the shanties and the protesters' refusal to move, since these topics are (a) direct results of the building of the shanties and are (b) GOOD from the program's point of view, for they serve the university's goal to be

Table 3
Partiality

For protesters	For university	Decision strategy
[AS A REMINDER TO] YALE UNIVERSITY TO DIVEST FROM COMPANIES DOING BUSINESS IN SOUTH AFRICA,	IN APRIL,	interp: *peaceful*
[A LARGE NUMBER OF] [CONCERNED] STUDENTS	[A SMALL NUMBER OF] [ ] STUDENTS [TOOK OVER] BEINECKE PLAZA	interp: *support* adj choice: enhancer interp: *tactics*
ERECTED A SHANTYTOWN NAMED WINNIE MANDELA CITY ON BEINECKE PLAZA IN APRIL.	AND ERECTED A SHANTYTOWN NAMED WINNIE MANDELA CITY	topic: given in input
	[IN ORDER TO FORCE] YALE UNIVERSITY TO DIVEST FROM COMPANIES DOING BUSINESS IN SOUTH AFRICA.	interp: *coercion*
	YALE [REQUESTED] THAT THE STUDENTS BUILD IT ELSEWHERE, BUT THEY REFUSED TO LEAVE.	verb choice: leniency topic: pro-university
	SO THE UNIVERSITY GAVE IT PERMISSION TO EXIST UNTIL THE MEETING OF THE YALE CORPORATION, BUT [EVEN] AFTER THAT THEY [STILL] REFUSED TO MOVE.	topic: pro-university adv choice: enhancer adv choice: enhancer topic: pro-university
[AT 5:30 AM ON APRIL 14,] [YALE HAD] OFFICIALS [DESTROY] IT; ALSO, AT THAT TIME, THE UNIVERSITY [HAD] THE POLICE ARREST 76 STUDENTS AFTER THE LOCAL COMMUNITY'S [HUGE] [OUTCRY],	OFFICIALS [HAD TO] [DISASSEMBLE] THE SHANTYTOWN [ ]. FINALLY,	interp: *abnormal-circ* interp: *coercion* verb choice: force topic: given in input interp: *coercion* topic: pro-protesters topic: pro-protesters interp: *support*
	YALE, [BEING CONCILIATORY] TOWARD THE STUDENTS,	interp: *conciliation*
YALE PERMITTED THE STUDENTS TO RECONSTRUCT THE SHANTYTOWN.	[NOT ONLY] PERMITTED THEM TO, RECONSTRUCT IT [BUT ALSO] ANNOUNCED THAT A COMMISSION WOULD GO TO SOUTH AFRICA IN JULY TO STUDY THE SYSTEM OF APARTHEID.	phrase juxta: enhancer topic: given in input phrase juxta: enhancer topic: pro-university

lenient and show the protesters' intransigence. When on the other hand it is speaking as protester, the same step causes it to collect the item representing the students' arrest. Other steps of the plan provide other topics. Eventually,

having performed the collection, PAULINE begins topic organization with the initial input and the topics it has collected.

6.4.2. *Biased interpretation under the rhetorical goals of opinion*

Having selected topics to be said, the next problem is to find suitable forms of expression for them. In certain cases, selections with appropriate biases are easy to make; in the same way that "erect" was defined as a formal version of "build," "tear down" can be defined as a slanted version of "disassemble." However, most slanting techniques require much more information. For example, the phrases "many students" and "a small group of students" are not simply two different lexicalizations of an underlying concept. In order to find when they can be used and which phrases are appropriate the generator requires, at least, limited inferential capability together with rules (special-purpose inferences) that suggest making specific slanting implications.

As described in Section 5, PAULINE has limited inferential capability. It also has a list of rules that prescribe how the generator should proceed to find forms of expression for input topics with certain characteristics, and that indicate what aspects of these topics can be used to create an appropriate slant. The goals that activate these rules are called the *rhetorical goals of opinion*. When the program is given sympathies that oppose the hearer's sympathies, and when the pragmatic value for effect on hearer's opinion of the topic is *switch*, PAULINE activates these goals, which can be paraphrased as:

- *State outright* that our side is good and theirs is bad.
- Show how our side has *good goals*, by describing how (a) we help other people; (b) we want a solution to the conflict; and (c) our goals are good according to accepted standards.
- Explain how our side does *good actions* to achieve the goals: (a) the actions are not unreasonable or nasty; (b) they are good according to accepted standards; and (c) they are performed in the open.
- Specifically, describe our side's *response to the opponent*: (a) negotiations that have taken place and (b) how we have moderated our demands.
- Finally, show how *other people* believe that we are good, by describing (a) their active support and (b) their statements and recommendations to that effect.

A similar list exists for the inverse goal, to show how bad the opponents' side is. Both lists contain a large number of specific inferences and explicit suggestions for sentences. For example, a strategy to make the opponents look bad is:

- Show how *they are unreasonable*: (a) they started the whole affair; (b) they coerce others into doing things; (c) they have little support; (d) they don't seem to want a solution; (e) their demands/goals are beyond reasonable

expectations; (f) they are only in it for their own good; (g) they are immoral, unfair; (h) they use distasteful/ugly tactics, misuse their rights, or overstep the bounds of propriety; (i) they disseminate false or misleading information about the dispute; (j) they have a hidden agenda; (k) they won't discuss/negotiate the issue; (l) they won't moderate their stance, are unconciliatory, intransigent.

These strategies are encoded as top-down interpretation inferences. They fire when the input sentence topics have characteristics that match their activation conditions; the left-hand sides are patterns of representation element types and their right-hand sides cause the generator to select the appropriate forms of expression. In somewhat more detail, the inferences *coercion* and *limited support* (see Table 3) can be paraphrased as (the terms in small capitals are elements of the representation language; MTRANS denotes the act of transferring information; PTRANS the act of transferring physical objects; and ATRANS the act of transferring control over something; see [84, 87]):

– *Coercion*: they force their will on others (corresponding to the university speaker's "in order to force"):

> IF the current topic is an ACTION,
>> AND its affect is BAD,
>> AND the action serves one of the opponent's goals,
>> AND the goal's desire is to have some other party do some act,
>> AND the other party's affect is GOOD,
> THEN imply that the opponents force their will on them (using verbs and phrases such as "force," "make them do").

– *Limited support*: they claim to have more support than they have (corresponding to the university speaker's adjective "a small number"):

> IF the current topic claims support (an MTRANS of a SUPPORT),
>> AND the ACTOR's affect is BAD,
>> AND the SUPPORT contains a number of people,
> THEN minimize that number,
>> by using adjectives such as "a small number," "a few."

Different inferences are applied at different times in the generation process. This depends on the kinds of effect they have on the processing and is controlled by the grammar. Inferences that call for the candidate topic(s) to be interpreted and completely replaced by other topics (such as interpreting a request as a coercion) are run during the topic organization phase; inferences

that suggest appropriate adjectives ("a large number," "a small group") are run when noun phrases are built; those that prescribe specific verbs when predicates are constructed.

6.4.3. *Slanted topic juxtaposition*

After it has collected candidate topics and before it says them, given the time and the inclination (controlled by the rhetorical goals of haste and simplicity), PAULINE can perform a number of topic organization tasks, one of which (conjunction) was described above under formality.

Sentence topics can also be juxtaposed in ways that carry affective implications. Conjunctive patterns such as "not only X but Y" can be used to imply that X and Y carry the same affective value, and in fact that the value is to be strengthened due to their juxtaposition. Compare (a) below to (b) and (c), which hold no such cumulative affective import:

(a)   "Not only did he play baseball, but *he hit five home runs!*"

(b)   "Pete played the game and he hit five home runs."

(c)   "When Pete played the game he hit five home runs."

The "not only X but Y" form can be called an *enhancer*. Other enhancing phrases are:

(d)   "Pete played the game; *also*, he hit five home runs."

(e)   "Pete played the game, *what's more*, he hit five home runs."

When an enhancer phrase juxtaposes two affect-laden sentences, the affect is strengthened; when it juxtaposes an affect-laden sentence with a neutral one, the affect is imputed to the latter. Thus, in addition to stressing affective concepts, a speaker can strengthen his or her case by imputing affect to neutral concepts too! This is, for example, what PAULINE does to produce

> NOT ONLY DID YALE UNIVERSITY PERMIT THE STUDENTS TO REBUILD THE SHANTYTOWN, BUT YALE ANNOUNCED THAT A COMMISSION WOULD GO TO SOUTH AFRICA TO STUDY THE SYSTEM OF APARTHEID.

when defending the university (see Table 3). For PAULINE, the commission visit topic is simply NEUTRAL, whereas permission to rebuild, because it serves the goal to be reasonable (which is intrinsically GOOD) is GOOD. When juxtaposed in this way, *both* sentences seem GOOD for Yale—exactly what PAULINE wants.

Similarly, phrases with weakening effect are *mitigators*. When a mitigator phrase juxtaposes two sentences carrying opposite affect, the resulting affect is that of the first sentence, weakened; when it juxtaposes an affect-laden sentence with a neutral one, the opposite affect is imputed to the latter. In the

following sentences, if "John whipped the dog" carries BAD affect, then, if we know nothing more, "he remembered the cat" becomes GOOD:

(f)   "*Although* John remembered the cat, he whipped the dog."

(g)   "John remembered the cat. *However*, he whipped the dog."

The first part, taken by itself, is neutral; it could just as well have been made BAD by using an enhancer:

(h)   "*Not only* did John remember the cat, he whipped the dog."

In a two-predicate mitigator, the sentence with the desired affect usually appears last.

A number of other constraints must be met before two topics can be juxtaposed in an enhancer or mitigator phrase. These are described in [48].

The juxtaposition of topics is controlled by the active rhetorical goals of opinion. In the shantytown example, for instance, the program's first goal is to introduce the topic. Its topic collection strategies provide it with two topics (the shanty construction and the protesters' intention) that are related by a SUBGOAL-TO relation. As at any decision point, the active rhetorical strategies of style are queried: should the relation between the two topics be used to conjoin them into a compound sentence? The answer is *yes*, since the relevant topic organization strategy, activated for both *explicit* and *implicit* values of partiality, calls for the use of affect-imputing enhancer and mitigator phrases. What is an appropriate way to express a SUBGOAL-TO relation? Here the inferences of opinion come into play, making decisions about the appropriateness of various interpretations of the two topics and their relationship. When sympathetic toward the university, one inference that matches the construction and its goal, which has the desired state that Yale divest from the companies, is that of *coercion*, described above. This strategy spawns the instruction to say a newly-formed interpretation, CAUSE-TO-DO, with the protesters' intent as attached topic, and the conjunction "in order to force." In contrast, when PAULINE is speaking as a protester, the strategy *we are lenient, offer passive resistance* causes it to join the topics using the phrase "as a reminder to." (When the program has no opinions, it would simply use a neutral phrase such as "in order to" or "so as to.") All these phrases are in the lexicon, indexed in a discrimination net linked to the relation SUBGOAL-TO.

6.4.4. *Biased word choice*

Nouns and verbs often carry affective value themselves. The words in PAULINE's lexicon are organized in discrimination nets to provide enhancing and mitigating alternatives when required. For example, the representation primitive MTRANS indexes to, amongst others, the verbs "order," "tell," "ask," and "request"; and DECONSTRUCT to "tear down," "disassemble," and "remove." See Table 3.

6.5. A final example

In summary, compare PAULINE's generation of the Carter–Kennedy example under three pragmatically different scenarios. In all three cases, the input is the same; the differences in the text result from the different values for the active rhetorical goals, which result from the different initial pragmatic settings. Only the effects of the rhetorical goals of formality, detail, partiality and haste will be discussed here (see [49] for more details).

Case 1. Neither interlocutor has opinions about the topic (causing partiality to be set to *impartial*); both have the usual knowledge of the electoral process (making detail be *details*); the level of formality is *colloquial*; and when the program is given enough time, haste is activated with the value *somewhat planned*. The result appears in Table 4.

Case 2. The hearer is a *friend* and social *equal* (therefore again *colloquial* formality) who is not as expert as the sibling (i.e., knowledge level is *student*, which makes detail be *details and interpretations*). But now both interlocutors have opinions: PAULINE's sympathy is for Kennedy and the hearer's is for Carter (so that partiality is *implicit*). The program is given as much time (mainly to make interpretations) as it needs: haste is *planned*. The result appears in Table 5.

Case 3. PAULINE is a Carter supporter and is speaking to its boss, an irascible Kennedy man. They are making a long-distance telephone call, which

Table 4
Case 1 (to an acquaintance): *colloquial, impartial, details, somewhat planned*

Text	Decision	Rhetorical goal value
Topic: central topic	RELATE plan	
[ ] CARTER AND KENNEDY WERE	no adjuncts before	*colloquial*
THE CANDIDATES IN A PRIMARY		
[IN MICHIGAN] [ON 20 FEBRUARY].	adjuncts after subject	*colloquial, planned*
Topic: result	RELATE plan	
CARTER [LOST]	neutral verb	*impartial*
TO KENNEDY BY [1335] VOTES.	neutral details	*impartial, details*
Topic: outcome with good affect for Kennedy	RELATE plan	*impartial*
AT PRESENT, KENNEDY		
HAS A BETTER CHANCE		
OF [GETTING] THE NOMINATION	informal word	*colloquial*
THAN [ ] BEFORE.	elide *he had*	*colloquial*
Topic: outcome with good affect for Carter	RELATE plan	*impartial*
CARTER IS ALSO CLOSER	separate sentence	*colloquial*
TO [GETTING] THE	informal world	*colloquial*
NOMINATION THAN [ ] BEFORE.	elide *he was*	*colloquial*
Topic: actor's goals (twice)	RELATE plan	
BOTH CARTER AND KENNEDY [WANT]	informal verb	*colloquial*
TO [GET] THE NOMINATION.	informal verb	*colloquial*

Table 5
Case 2 (to a friend): *colloquial, implicit, all* (details and interpretations), *planned*

Text	Decision	Rhetorical goal values
Topic: results with good affect for Kennedy	CONVINCE plan	*implicit*
[ ] KENNEDY	no adjuncts before	*colloquial*
[DIMINISHED] CARTER'S [LEAD]	interpretation	*all, planned*
BY [GETTING]	informal verb	*colloquial*
[ALL OF]	enhancer adj	*implicit*
[21850] VOTES	details	*all*
[IN THE PRIMARY] [IN MICHIGAN].	adjuncts after subject	*colloquial*
Topic: reminding	indexed off interp	*planned*
IN A SIMILAR CASE, CARTER DECREASED	reminding	*implicit, planned*
UDALL'S LEAD IN A PRIMARY		
IN 1976, AND HE [EASILY]	enhancer adv	*implicit*
[TROUNCED] UDALL TO BE NOMINATED	enhancer verb	*implicit*
BY [2600] DELEGATES.	details	*all*
Topic: outcome with good affect for Kennedy	CONVINCE plan	*implicit*
[I AM REAL GLAD THAT]	informal opinion	*colloquial, explicit*
KENNEDY IS [NOW] CLOSER TO	adjunct after, informal	*colloquial*
[GETTING] THE NOMINATION THAN	informal verb	*colloquial*
[ ] BEFORE.	elide *he was*	*colloquial*

Table 6
Case 3 (to the boss): *colloquial, implicit, interpretations, pressured*

Text	Decision	Rhetorical goal values
Topic results and outcomes for Carter	CONVINCE plan	*implicit*
. . . .	no time for mitigation	*pressured*

gives the program *little* time and makes conversational conditions *noisy* (activating the haste goal with the value *pressured*.) Furthermore, the program is *distant* from its boss, does not wish to anger him (desired emotional effect is *calm down*), and still wants to make him feel socially *dominant* (resulting in *implicit* partiality and *interpretations* for detail). But to its boss (Table 6), the program says nothing!

This text came as a surprise. Investigation showed that the lack of time prevented any of the strategies for implicitly stating opinions from being applied: no topic collection plan was activated; no search for mitigating interpretations took place; the lack of a second topic meant no topic juxtaposition was possible; no rhetorical goals of opinion were present to guide mitigating adverb and adjective selection and appropriate word choice. Therefore, the goal to present the topic only in mitigated (implicit opinion) form couldn't be satisfied, and no sentence could be generated.

7. Conclusion

The question "why and how is it that we say the same thing in different ways to different people, or even to the same person in different circumstances?" is interesting from a number of perspectives. From a cognitive perspective, it highlights speakers' goals and personal interrelationships in communication; from a linguistic perspective, it raises interesting questions about the information content of language; and from an engineering-AI perspective, it illustrates the need for principled reasons by which a program that can realize the same input in various ways can make its selections.

As described in this paper, the answer deals with the pragmatic nature of communication—a big and complex field of study. In order to begin to study how pragmatics is used in generation, a number of assumptions about plausible types of speaker goals and the relevant characteristics of hearers and of conversational settings must be made. The specific pragmatic features used by PAULINE are but a first step. They are the types of factors that play a role in conversation; no claims are made about their literal veracity. Similarly, the strategies PAULINE uses to link its pragmatic features to the actual generator decisions, being dependent on the definitions of the features are equally primitive; again, no strong claims are made about their existence in people in exactly the form shown. However, in even such a simple theory as this, certain lessons emerge, and these lessons, I believe, hold true no matter how sophisticated the eventual theory is. The lessons pertain primarily to the organization of pragmatic information in generation: the fact that interpersonal and situational information and goals are too general to be of immediate use; the resulting fact that intermediate strategies, here called rhetorical strategies, are required to guide generation; the fact that, in a model of generation that incorporates these goals, rhetorical planning and realization must be interleaved processes, where the interleaving takes place at the choice points.

The study of language generation by computer has traditionally been divided into two questions: *what shall I say?* and *how shall I say it?* The aim of this work is to illustrate the importance of a third question: *why should I say it?* If generators do not face up to this question, they will never be able to address the other two satisfactorily.

REFERENCES

1. D.E. Appelt, TELEGRAM: A grammar formalism for language planning, in: *Proceedings IJCAI-83*, Karlsruhe, FRG (1983) 595–599.
2. D.E. Appelt, *Planning English Sentences* (Cambridge University Press, Cambridge, MA, 1985).
3. J.M. Atkinson, Understanding formality: The categorization and production of 'formal' interaction, *British J. Sociol.* **33** (1) (1982) 86–117.
4. W.M. Bain, Case-based reasoning: A computer model of subjective assessment, Ph.D. Dissertation, Yale University, New Haven, CT (1985).
5. W.M. Bain, A case-based reasoning system for subjective assessment, in: *Proceedings AAAI-86*, Philadelphia, PA (1986) 523–526.
6. S. Baker, *The Complete Stylist* (Thomas Crowell, New York, 1966).
7. M.A. Bienkowski, A computational model for extemporaneous elaborations, Tech. Rept. 1, Princeton University Cognitive Science Laboratory, Princeton, NJ (1986).
8. N.P. Birk and G.B. Birk, *Understanding and Using English* (Odyssey Press, New York, 4th ed., 1965).
9. L. Bloomfield, *An Introduction to the Study of Language* (Holt, New York, 1914).
10. D.G. Bobrow and T. Winograd, An overview of KRL: A knowledge-representation language, *Cognitive Sci.* **1** (1) (1977) 3–46.
11. C.A. Broverman and W.B. Croft, Reasoning about exceptions during plan execution monitoring, *Proceedings AAAI-87*, Seattle, WA (1987) 190–193.
12. P. Brown and S.C. Levinson, Universals in language usage: Politeness phenomena, in: E. Goody, ed., *Questions and Politeness: Strategies in Social Interaction* (Cambridge University Press, Cambridge, MA, 1978) 56–311.
13. B. Bruce, A. Collins, A.D. Rubin and D. Gentner, A cognitive science approach to writing, Tech. Rept. 89, Bolt, Beranek and Newman, Cambridge, MA (1978).
14. K. Bühler, *Sprachtheorie* (Fischer, Jena, 1934).
15. J.G. Carbonell, Intentionality and human conversations, in: *Proceedings TINLAP2*, Urbana, IL (1978) 141–148.
16. R. Carnap, Foundations of logic and mathematics, in: O. Neurath, R. Carnap and C.W. Morris, eds., *International Encyclopedia of Unified Science* **1** (Chicago University Press, Chicago, IL 1938) 139–214.
17. E. Charniak, C.K. Riesbeck and D.V. McDermott, *Artificial Intelligence Programming* (Erlbaum, Hillsdale, NJ, 1980).
18. H.H. Clark, and T.B. Carlson, Context for comprehension, in: J. Long and A. Baddeley, eds., *Attention and Performance* **9** (Erlbaum, Hillsdale, NJ, 1981).
19. H.H. Clark and G.L. Murphy, Audience design in meaning and reference, in: J.F. Le Ny and W. Kintsch, eds., *Language and Comprehension* (North-Holland, Amsterdam, 1982).
20. H.H. Clark and D. Schunk, Polite responses to polite requests, *Cognition* **8** (1980).
21. J.H. Clippinger, A discourse speaking program as a preliminary theory of discourse behavior and a limited theory of psychoanalytic discourse, Ph.D. Dissertation, University of Pennsylvania, Philadelphia, PA (1974).
22. P.R. Cohen, On knowing what to say: Planning speech acts, Ph.D. Dissertation, University of Toronto, Toronto, Ont. (1978).
23. G. Cowan and E. McPherson, *Plain English Rhetoric and Reader* (Random House, New York, 1977).
24. J.H. Danks, Producing ideas and sentences, in: S. Rosenberg, ed., *Sentence Production: Developments in Research and Theory* (Erlbaum, Hillsdale, NJ, 1977).
25. L. Danlos, *Génération Automatique de Textes en Langues Naturelles* (Masson, Paris, 1985).
26. R. de Beaugrande, *Text Production*, Advances in Discourse Processes **11** (Ablex, Norwood, NJ, 1984).

27. K. de Smedt and G. Kempen, Incremental sentence production, in: G. Kempen, ed., *Natural Language Generation: New Results in Artificial Intelligence, Psychology, and Linguistics* (Kluwer Academic Publishers, Dordrecht, Netherlands, 1987) 365–376.

28. R.J. Doyle, D.J. Atkinson and R.S. Doshi, Generating perception requests and expectations to verify the execution of plans, in: *Proceedings AAAI-86*, Philadelphia, PA (1986) 81–88.

29. E.H. Durfee and V.R. Lesser, Incremental planning to control a blackboard-based problem solver, in: *Proceedings AAAI-86*, Philadelphia, PA (1986) 58–64.

30. R.P. Fawcett, *Cognitive Linguistics and Social Interaction* (Julius Groos, Heidelberg, 1980).

31. R.E. Fikes, P.E. Hart and N.J. Nilsson, New directions in robot problem solving, in: B. Meltzer and D. Michie, eds., *Machine Intelligence* 7 (Edinburgh University Press, Edinburgh, Scotland, 1972).

32. G. Gazdar, *Pragmatics: Implicature, Presupposition, and Logical Form* (Academic Press, New York, 1979).

33. G. Gazdar, Pragmatic constraints on linguistic production, in: B. Butterworth, ed., *Language Production* 1 (Academic Press, New York, 1980) 49–68.

34. R. Gibbs, Contextual effects in understanding indirect requests, *Discourse Process.* 2 (1979) 1–10.

35. R. Gibbs, Your wish is my command: Convention and context in interpreting indirect requests, *J. Verbal Learning and Verbal Behav.* 20 (1981) 431–444.

36. N.M. Goldman, Conceptual generation, in: R.C. Schank, ed., *Conceptual Information Processing* (North-Holland, Amsterdam, 1975) 289–358.

37. E. Goody, ed., *Questions and Politeness: Strategies in Social Interaction* (Cambridge University Press, Cambridge, 1978).

38. M. Gregory, Towards 'communication' linguistics: A framework, in: J.D. Benson and W.S. Greaves, eds., *Systemic Perspectives on Discourse* (Edward Arnold Press, London, 1982).

39. H.P. Grice, Logic and conversation, in: D. Davidson and G. Harman, eds., *The Logic of Grammar* (Dickinson, 1975) 41–58.

40. J.E. Grimes, *The Thread of Discourse* (Mouton, The Hague, Netherlands, 1975).

41. B.J. Grosz, A theory of discourse structure, in: *Proceedings Eighth Annual Conference of the Cognitive Science Society*, Amherst, MA (1986) 432–436.

42. B.J. Grosz and C.L. Sidner, Discourse structure and the proper treatment of interruptions, in: *Proceedings IJCAI-85*, Los Angeles, CA (1985) 832–836.

43. M.A.K. Halliday, *Halliday: System and Function in Language, Selected Papers*, edited by G.R. Kress (Oxford University Press, Oxford, 1976).

44. M.A.K. Halliday, *Language as Social Semiotic* (Edward Arnold Press, London, 1978).

45. S.I. Harada, Honorifics, in: M. Shibatani, ed., *Syntax and Semantics* 5: *Japanese Generative Grammar* (Academic Press, New York, 1976) 499–561.

46. A.S. Hill, *The Foundations of Rhetoric* (Harper, New York, 1892).

47. E.H. Hovy, Integrating text planning and production in generation, in: *Proceedings of IJCAI-85*, Los Angeles, CA (1985) 848–852.

48. E.H. Hovy, Putting affect into text, in: *Proceedings Eighth Annual Conference of the Cognitive Science Society*, Amherst, MA (1986) 669–671.

49. E.H. Hovy, Some pragmatic decision criteria in generation, in: G. Kempen, ed., *Natural Language Generation: New Results in Artificial Intelligence, Psychology, and Linguistics* (Kluwer Academic Publishers, Dordrecht, Netherlands 1987) 3–19.

50. E.H. Hovy, Interpretation in generation, in: *Proceedings AAAI-87*, Seattle, WA (1987) 545–548.

51. E.H. Hovy, What makes language formal?, in: *Proceedings Ninth Annual Conference of the Cognitive Science Society*, Seattle, WA (1987) 959–962.

52. E.H. Hovy, *Generating Natural Language under Pragmatic Constraints* (Erlbaum, Hillsdale, NJ, 1988).

53. E.H. Hovy, Generating language with a phrasal lexicon, in: D.D. McDonald and L. Bolc, eds., *Natural Language Generation Systems* (Springer, New York, 1988) 353–384.
54. J.T. Irvine, Formality and informality of speech events, *Am. Anthropol.* **81** (1979) 773–790.
55. R. Jackendoff, On Katz's autonomous semantics, *Language* **57** (1981) 425–435.
56. R. Jackendoff, *Semantics and Cognition* (MIT Press, Cambridge, MA, 1985).
57. P.S. Jacobs, A knowledge-based approach to language production, Ph.D. Dissertation, University of California, Berkeley, CA (1985).
58. R. Jakobson, Linguistics and poetics, in: T. Sebeok, ed., *Style in Language* (MIT Press, Cambridge, MA, 1960) 350–77.
59. A. Jameson, How to appear to be conforming to the 'maxims' even if you prefer to violate them, in: G. Kempen, ed., *Natural Language Generation: New Results in Artificial Intelligence, Psychology, and Linguistics* (Kluwer Academic Publishers, Dordrecht, Netherlands, 1987) 19–42.
60. P.N. Johnson and S.P. Robertson, MAGPIE: A goal-based model of conversation, Tech. Rept. 206, Yale University, New Haven, CT (1981).
61. J.J. Katz, *Propositional Structure and Illocutionary Force* (Crowell, New York, 1977).
62. G. Kempen, Conceptualizing and formulating in sentence production, in: S. Rosenberg, ed., *Sentence Production: Developments in Research and Theory* (Erlbaum, Hillsdale, NJ, 1977).
63. S. Kuno, *The Structure of the Japanese Language* (Harvard University Press, Cambridge, MA, 1973).
64. R. Lakoff, Politeness, pragmatics, and performatives, in: *Proceedings Conference on Performatives, Presuppositions, and Implicatures*, TX (1977) 79–106.
65. W.G. Lehnert, Plot units: A narrative summarization strategy, in: W.G. Lehnert and M.H. Ringle, eds., *Strategies for Natural Language Processing* (Erlbaum, Hillsdale, NJ, 1982) 375–414.
66. W.J.M. Levelt and H. Schriefers, Stages of lexical access, in: G. Kempen, ed., *Natural Language Generation: New Results in Artificial Intelligence, Psychology, and Linguistics* (Kluwer Academic Publishers, Dordrecht, Netherlands, 1987) 395–404.
67. S.C. Levinson, *Pragmatics* (Cambridge University Press, Cambridge, 1983).
68. R.S. Loomis, H.R. Hull and M.L. Robinson, *The Art of Writing Prose* (Farrar and Rinehart, New York, 1936).
69. W.C. Mann, An overview of the Nigel text generation grammar, Tech. Rept. RR-83-113, USC/Information Sciences Institute, Marina del Rey, CA (1983).
70. W.C. Mann, An overview of the Penman text generation system, Tech. Rept. RR-83-114, USC/Information Sciences Institute, Marina del Rey, CA (1983).
71. K.F. McCoy, Correcting object-related misconceptions, Ph.D. Dissertation, University of Pennsylvania, Philadelphia, PA (1985).
72. D.D. McDonald, Natural language production as a process of decision making under constraint, Ph.D. Dissertation, MIT, Cambridge, MA (1980).
73. D.D. McDonald and J.D. Pustejovsky, Description-directed natural language generation, in: *Proceedings IJCAI-85*, Los Angeles, CA (1985) 799–804.
74. K.R. McKeown, *Text Generation: Using Discourse Strategies and Focus Constraints to Generate Natural Language Text* (Cambridge University Press, Cambridge, 1985).
75. D.P. Miller, Planning by search through simulations, Ph.D. Dissertation, Yale University, New Haven, CT (1985).
76. C.E. Osgood, A behavioristic analysis of perception and language as cognitive phenomena, in: J. Bruner, ed., *Contemporary Approaches to Cognition* (Harvard University Press, Cambridge, MA, 1957).
77. C.E. Osgood and J.K. Bock, Salience and sentencing: Some production principles, in: J. Rosenberg, ed., *Sentence Production: Developments in Research and Theory* (Erlbaum, Hillsdale, NJ, 1977).

78. C.E. Osgood, W.H. May and M. Miron, *Cross-Cultural Universals of Affective Meaning* (University of Illinois Press, Urbana, IL, 1975).
79. C.L. Paris, Discourse strategies for descriptions of complex physical objects, Ph.D. Dissertation, Columbia University, New York (1987).
80. T. Patten, *Systematic Generation as Problem Solving* (Cambridge University Press, Cambridge, 1988).
81. L.V. Payne, *The Lively Art of Writing* (Follet (Mentor Books), New York, 1969).
82. S. Rosenberg, Semantic constraints on sentence production: An experimental approach, in: S. Rosenberg, ed., *Sentence Production: Developments in Research and Theory* (Erlbaum, Hillsdale, NJ, 1977).
83. E. Sacerdoti, *A Structure for Plans and Behavior* (North-Holland, Amsterdam, 1977).
84. R.C. Schank, 'Semantics' in conceptual analysis, *Lingua* **30** (2) (1972).
85. R.C. Schank, *Conceptual Information Processing* (North-Holland, Amsterdam, 1975).
86. R.C. Schank *Dynamic Memory: A Theory of Reminding and Learning in Computers and People* (Cambridge University Press, Cambridge, 1982).
87. R.C. Schank and R.P. Abelson, *Scripts, Plans, Goals and Understanding* (Erlbaum, Hillsdale, NJ, 1977).
88. R.C. Schank, G.C. Collins, E. Davis, P.N. Johnson, S.L. Lytinen and B.J. Reiser, What's the point?, Tech. Rept. 205, Yale University, New Haven, CT (1981).
89. J.G. Schmolze and T.A. Lipkis, Classification in the KL-ONE knowledge representation system, in: *Proceedings IJCAI-83*, Karlsruhe, FRG (1983) 330–334.
90. R.F. Simmons and J. Slocum, Generating English discourse from semantic networks, *Commun. ACM* **15** (10) (1972).
91. M. Stefik and D.G. Bobrow, Object-oriented programming: Themes and variations, *AI Mag.* **6** (4) (1986).
92. W. Strunk, Jr and E.B. White, *The Elements of Style* (Macmillan, New York, 1959).
93. W.R. Swartout, Producing explanations and justifications of expert consulting programs, Tech. Rept. LCS-TR-251, MIT, Cambridge, MA (1981).
94. T. van Dijk, *Studies in the Pragmatics of Discourse* (Mouton, The Hague, Netherlands, 1985).
95. W. Weathers and O. Winchester, *The New Strategy of Style* (McGraw-Hill, New York, 1978).
96. R. Wilensky, *Planning and Understanding* (Addison-Wesley, Reading, MA, 1983).
97. H. Willis, *Structure, Style, and Usage: The Rhetoric of Exposition* (Holt, Rinehart and Winston, New York, 1969).
98. B. Woolf and D.D. McDonald, Context-dependent transitions in tutoring discourse, in: *Proceedings AAAI-84*, Austin, TX (1984) 355–359.