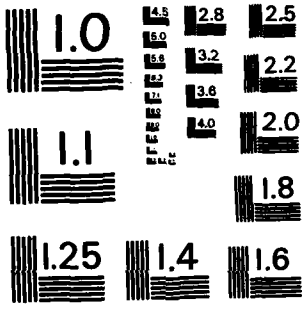
AD-A126 354 THE ANATOMY OF A SYSTEMIC CHOICE(U) UNIVERSITY OF SOUTHERN CALIFORNIA MARINA DEL REY INFORMATION SCIENCES INST W C MANN SEP 82 ISI/RR-82-104 AFOSR-TR-83-0135 1/1

UNCLASSIFIED F49620-79-C-0181 F/G 5/7 NL

END
DATE FILMED
5 83
DTIC



ISI/RR-82-104
October 1982

William C. Mann

[redacted] The Anatomy of a Systemic Choice

INFORMATION SCIENCES INSTITUTE
4676 Admiralty Way/ Marina del Rey/ California 90291
(213) 822-1511
UNIVERSITY OF SOUTHERN CALIFORNIA

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR-TR- 83-0135	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* The Anatomy of a Systemic Choice		5. TYPE OF REPORT & PERIOD COVERED<br~~Research Report~~ Interim 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) William C. Mann		8. CONTRACT OR GRANT NUMBER(s) F49620-79-C-0181
9. PERFORMING ORGANIZATION NAME AND ADDRESS USC/Information Sciences Institute 4676 Admiralty Way Marina del Rey, CA 90291		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2304/A2
11. CONTROLLING OFFICE NAME AND ADDRESS Mathematical and Information Sciences Air Force Office of Scientific Research Building 410 Bolling Air Force Base Washington, D.C. 20332		12. REPORT DATE September 1982 13. NUMBER OF PAGES 28
14. MONITORING AGENCY NAME & ADDRESS*(if different from Controlling Office)*		15. SECURITY CLASS. *(of this report)* Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT *(of this Report)*

This document is approved for public release and sale; distribution is unlimited.

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

.....

18. SUPPLEMENTARY NOTES

.....

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

artificial intelligence, choice expert, computational linguistics, computer generation of text, linguistics, natural language, systemic grammar, text generation

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

(OVER)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

20. ABSTRACT

Systemic grammar is one of the major varieties of syntactic theory in modern linguistics. It was originally defined by Michael A. K. Halliday around 1960 and has since been developed extensively by him and others. Unlike transformational grammar, systemic grammar is oriented to the ways that language functions for its users. Systemic grammars have been used in several well-known language-processing programs and have been found to be very advantageous for computer generation of text. This report presents a framework for expressing how choices are made in systemic grammars. Formalizing the description of choice processes enriches descriptions of the syntax and semantics of languages, and it contributes to constructive models of language use. There are applications in education and computation. The framework represents the grammar as a combination of systemic syntactic description and explicit choice processes, called "choice experts." Choice experts communicate across the boundary of the grammar to its environment, exploring an external intention to communicate. The environment's answers lead to choices and thereby to creation of sentences and other units, tending to satisfy the intention to communicate. The experts' communicative framework includes an extension to the systemic notion of a function, in the direction of a more explicit semantics. Choice expert processes are presented in two notations, one informal and the other formal. The informal notation yields a grammar-guided conversation in English between the grammar and its environment, while the formal notation yields complete accounts of what the grammar produces given a particular circumstance and intent.

ISI/RR-82-104
October 1982

William C. Mann

[redacted] The Anatomy of a Systemic Choice

INFORMATION SCIENCES INSTITUTE
4676 Admiralty Way/ Marina del Rey/ California 90291
(213) 822-1511
UNIVERSITY OF SOUTHERN CALIFORNIA

This research was supported by the Air Force Office of Scientific Research Contract No. F49620-79-C-0181. The views and conclusions contained in this document are those of the author and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Air Force Office of Scientific Research of the U.S. Government.

CONTENTS

ACKNOWLEDGMENTS

The work presented here would not have been possible without the active participation of Christian Matthiessen and Michael A. K. Halliday.

1. SCOPE AND PURPOSE

Choice is one of the most prominent organizing concepts in systemic linguistics. Languages are described in terms of the choices available to the speaker and the relationships of those choices to each other and to the utterance produced.

However, formal descriptions of choice are generally lacking. Within the framework of systemic linguistics there is relatively little description of the conditions under which particular choices are appropriate for a particular communicative purpose or, consequently, of how to choose in conformity with a communicative purpose, or intention to communicate. To create a more organized knowledge of how to choose we need both to formalize what we mean by "how to choose" and to express how to choose appropriately in particular grammars, a classical form-and-content development.

We expect that by exploring the nature of systemic choices, several kinds of interests in language will be served:

Grammar as language description

Describing choices can yield a richer understanding of the alternatives offered and the differences between one system of choices and another.

Describing choices leads to new argument forms about how grammars should be organized.

Choice descriptions greatly reduce the descriptive load carried by grammatical feature names and other labels.

Semantic descriptions

Choice descriptions can be a direct extension of present systemic frameworks in the direction of more explicit semantics.

Describing communication

Choice descriptions are helpful in creating descriptions of how particular intentions to communicate can be satisfied. They can be a component in a description of how speakers can use language in order to do things, to use the functionality of language.

Constructive models

Choice descriptions can be used as parts of a constructive model of the expressive communicative process, a model that can perform this process instead of describing it. Our current research is aimed at building such a model as a computer program.

Applications

Teaching English: Choice descriptions can be used to convey the ways that syntactic constructs are used.

Comprehensible computation: In making computers communicate effectively with people, choice processes can be part of a process that creates text for computer users.

We intend to convey primarily the skeletal framework for choice description, a set of well-motivated devices for describing choice processes. To be useful it will have to be complemented by a physiology, an account of how choice processes can interact to produce intended effects.

As part of an investigation of computer text production, this report addresses the problems of characterizing choice in a systemic framework and creating a corresponding notation for processes of choosing that fit both systemic grammars and some explicit model of having an intention to communicate. Focus on notation is necessary and timely, since development of notation must to some extent precede development of corresponding content.

Choices may be viewed from either of two perspectives:

1. as **actions** the speaker performs in preparation for speaking, or
2. as **classifications** of the speech produced, parts of a complex taxonomy of the potential available to the speaker.

For the linguist creating a grammar, the perspective chosen is often consequential. However, because the two perspectives are closely linked, it is common either to regard them as identical or to mix them freely. We can see that the taxonomic perspective is more thoroughly elaborated in the literature, since published systemic grammars always provide for expression of alternatives but seldom provide for expression of processes that lead to choosing. (However, see [6].)

In this report we focus on the perspective of choice as a speaker's action, without rejecting the other.

Although the notation and examples are drawn specifically from Halliday [2, 4, 5], the ideas are relevant in any grammatical framework organized around choice [1, 7, 8, 14].

2. CHOICE AND INTENT

People build grammars for different purposes, so it is appropriate to apply differing criteria to their products [3]. Although we believe our work serves a broad range of interests, we should still identify ours. We would like to create an account of **how one may generate text that is appropriate for satisfying particular intentions to communicate**. In other words, given something particular to say, we wish to be able to take an existing description of how to seek to satisfy intentions to communicate, apply it to the intention, and see at some level of detail that appropriate language is generated.

Descriptions need interpreters in order to apply them to cases. We could describe language in a way that requires a human interpreter, or we could describe it for some other interpreter such as a computer program. These two options lead to very different preferences for amounts of detail.

We want to do both, but we cannot develop both kinds of descriptions at the same time. A humanly comprehensible description must be created and refined before a computer program can be created.

By introducing a communicative purpose we immediately raise the issue of whether that purpose, or the potential of all purposes, should be regarded as within the grammar or outside of it. To include purposes as formal objects within systemic grammars would be a major departure from the dominant tradition.[1]

In the situation diagrammed in Figure 1, in which the grammar has a definite boundary, we consider communicative purposes as residing outside in its ***environment***. The purposes may range from very specific intentions about the scope and presentation of particular ideas to very general intentions about the text as a whole. They include the structural elements of plans, that is, sequence information and subgoals. (An example of a very specific intention would be the intention to give special contrastive emphasis to a particular object in presenting a particular proposition involving that object.)

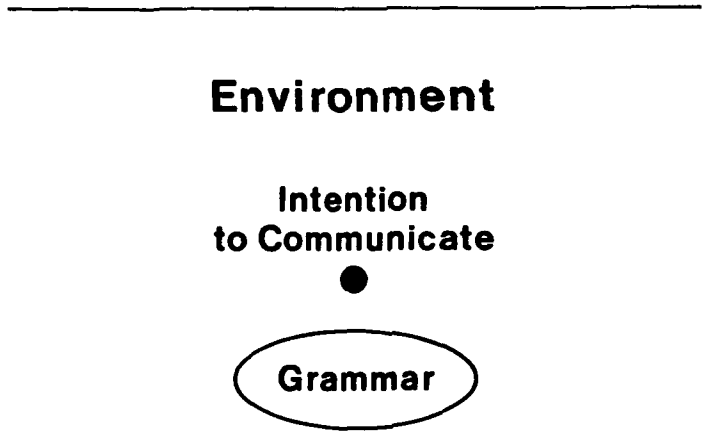


Figure 1: The boundary of the grammar

The grammar is composed of ***systems*** (hence the name systemic), which are collections of alternatives. Each alternative is named; the names are called ***grammatical features***. Each system has an ***entry condition***, which specifies when the alternatives are offered. For example, the Number system offers the alternatives Singular and Plural. Its entry condition is Count, that is, the alternative Count must have been chosen in some other system before the alternatives of Number are offered.

The act of offering the alternatives is called ***entering*** the system.

The grammar also includes the methods for choosing among alternatives, the "choice experts" to be developed below.

To keep the discussion simple we will make some assumptions about the framework:

[1] In contrast to Fawcett [1] (pp. 75-78), we treat the development of a particular intention to speak as outside of the grammar. We agree with Fawcett that the intention deserves theoretical treatment in conjunction with its grammatical realization, but we do not propose to derive it systemically.

- That if a choice is offered, it is relevant to the immediate communicative purpose, i.e., that entry conditions are strong enough to make choices sensible all of the time, rather than having a scheme in which throwaway choices are made.
- That there is no information available from systems that have not been entered.
- That once a system is entered the choice has to be made. (Alternative: One system could interrupt itself and ask another system to choose first.)

3. THINGS TO KNOW ABOUT A CHOICE

Because the grammar is to generate in conformity with a given intention to communicate, and the generated language will depend entirely on the grammar's choices (except for the possibility of underspecification of order, which we will ignore), the central problem is the following:

Globally: ***How can choices be made to conform to a given intention to communicate?***

We begin to answer this question by transforming the picture of the grammar in Figure 1 into the one below, Figure 2, in which each system addresses the environment individually.

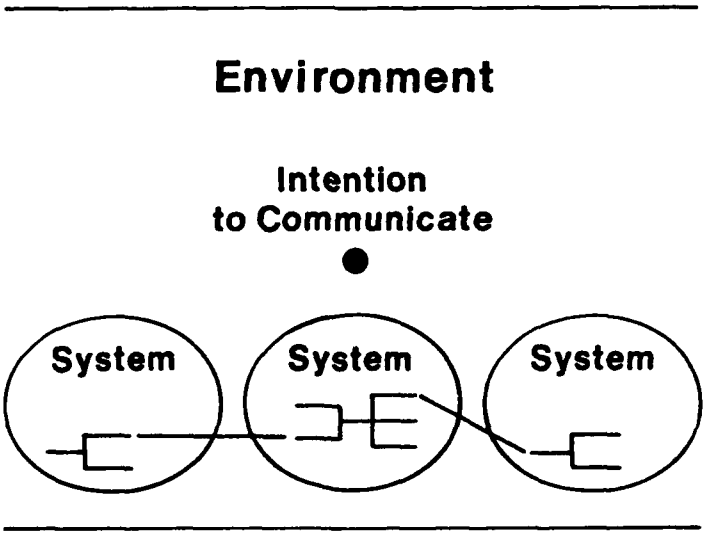


Figure 2: Isolated system boundaries

This yields a new problem, repeated for each choice point in the grammar:

Locally: ***How can this particular choice be made to conform to a given intention to communicate?***

It is helpful at this point to introduce the notion of a **choice expert** for a system. Following our intention to view choice as action, we define a choice expert as a process that interacts with the environment and determines what choice in the system conforms to the given intention.

Choice experts follow these conventions:[2]

- Each system has a distinct choice expert that operates independently of the other choice experts of the grammar.
- All of the interaction between choice experts is in the connectivity of the grammar, the entry conditions of systems.
- A choice expert does not function until the entry conditions of its system are fully satisfied.
- There are no constraints on the sequence in which choice experts become active beyond the constraints imposed by the entry conditions of the systems.

We give each choice expert the potential for two kinds of interaction with the environment: one where the expert may ask questions of the environment, and one where it may receive responses (Figure 3).

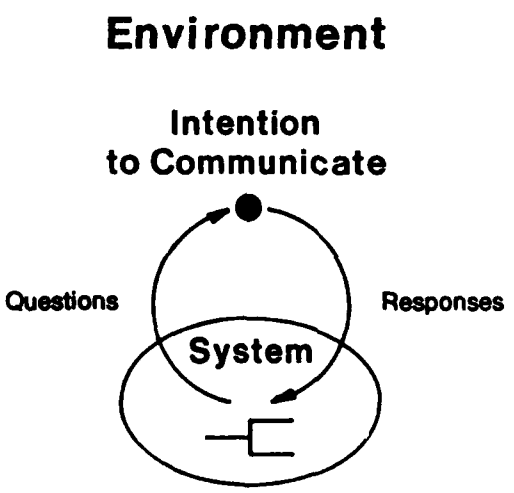


Figure 3: Two communication points for each choice expert

There are also constraints on the environment:

Initiative is in the grammar

The environment never volunteers any information; all of the information received by a choice expert is in response to questions asked.

The interaction is restricted to single responses to single questions.

The environment always responds.

[2] We overconstrain the choice experts slightly for simplicity. The latter part of the report identifies the substitute constraints that must replace some of these.

Invariant environment

> Except as explicitly noted, questions do not cause changes in the environment, and so do not affect the responses to other questions.
>
> The environment does not have to remember anything about the progress of its "conversation" with the expert.
>
> The environment is static and deterministic, so that if the expert were to ask a question repeatedly, it would always receive the same answer.

All of these restrictions are imposed for our convenience. They simplify the task of specifying the content of choices, and they do not restrict what the grammar can do.[3]

According to these conventions, there are two things a choice expert can do in the course of its interaction with the environment and the grammar:

1. It can ask a question of the environment.
2. It can make one of the choices offered by its system.

If it makes a choice, then its work is finished. If it asks a question, then by further convention it waits for the answer from the environment, inspects it, and proceeds in a way that is somehow conditioned by that answer.

The following become the central issues in describing each choice expert:

- Questions: What questions must be asked of the environment in order to choose according to the intent?
- Responses: What must the environment contain in order to answer the questions correctly?
- Continuation: How do answers condition asking further questions?
- Completion: How do answers condition making choices?

To explore these kinds of knowledge, we will examine the example of a choice between Singular and Plural in a particular systemic grammar.[4]

We select this particular pair of features because at first sight the way to make this choice seems to be obvious, in little need of explication. People feel that they understand it (i.e., as just a question of multiplicity). Any substantive issues raised in describing its meaning will support the conviction that all of the choices ought to be explicated.

For the moment the relevant fragments of this grammar are two:

---

[3] For the purposes here we also constrain the choice expert to be deterministic, so that faced with a particular intention to communicate it always does the same thing. This is not a necessary restriction; it simply clarifies the discussion. It is arguable whether there must be a source of uncontrolled variability, and if so, whether it should be in the choice experts. These issues are beyond the scope of this report.

[4] The grammar used for the examples is Nigel, a large grammar based on Halliday that is part of the Penman knowledge delivery system, currently under development. For related work see [9, 10, 11, 13].

1. The Number system, in which Singular and Plural are the only alternatives.
2. The systems that form the path from the entry of the grammar to Number.

These are shown in Figure 4. These systems are in the part of the grammar that forms NominalGroups.

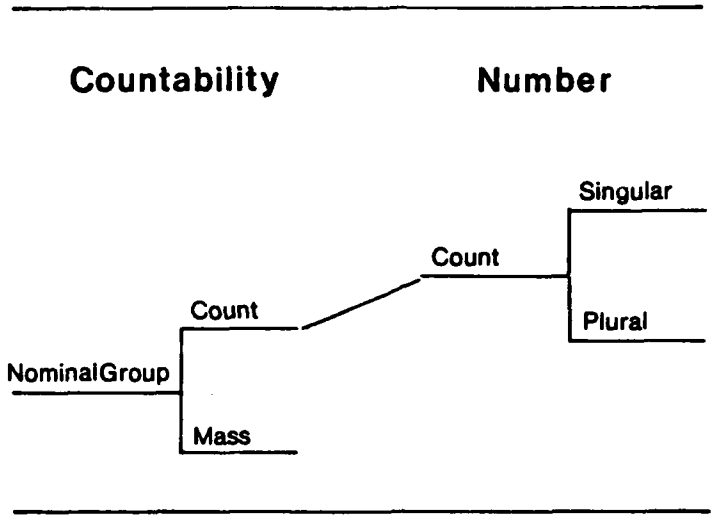


Figure 4: The systems leading to number in NominalGroup

The choice of Singular or Plural will govern the number attribute of the lexical item that realizes the head of the NominalGroup. We want it to control contrasts such as the distinction between *lion* and *lions* as used in the sentences below.

1. Of all of the kinds of animals in the world, I am most interested by the *lion*.
2. Of all of the kinds of animals in the world, I am most interested by *lions*.
3. Of all of the animals behind that bush, I am most afraid of the *lion*.
4. Of all of the animals behind that bush, I am most afraid of the *lions*.

Assume that the grammar is generating toward the first example. For the Number system, the first question from the list above becomes: **What questions must the Number system ask the environment to determine whether to choose Singular or Plural for this case?**

The Number system does not have any knowledge of the surrounding grammatical developments or what they represent. Its chooser must therefore be able to ask about everything it would need for at least the four cases above, and more for other cases. On the other hand, it will not need to ask all of its questions every time, and they may not all be relevant every time.

Our method below will be to develop the questions of the choice experts in English, postponing as long as possible any attempt to convert them to a formal notation. This has several benefits: Obscurity is avoided, issues of the adequacy of notations are taken up only after the material to be formalized is well developed, and, we can present the ideas in terms of progressive revisions without continually revising a formal version. Still, we recognize that the work of the choice expert needs to

be given formal treatment, and so each of the questions must eventually be converted into a more controlled notation.

For this case, the entity to be expressed is the intensional concept representing the species of animals called lions. The Number expert can begin by asking something roughly equivalent to the following:

Is the item unitary or multiple?

The environment must be able to make the determination and reply that the item is unitary. As we see from the first two examples, either a singular or plural form could express the concept, and the difference is in how the speaker wishes to convey this generic item in its context rather than in the speaker's notion of how many lions are behind the bush.

To continue, the expert could ask:

Is the item an intensional concept or an extensional one?

If the item is intensional, as in the example, then the expert could ask:

Does the plan of the text favor referring to the item as a species or as a collection?

The plan of the text, like the intention to communicate, resides outside of the grammar. If the plan included relating this concept to the title "King of the Jungle," then singular reference might be preferred. At any rate, the environment must be able to make this determination and respond accordingly. If species reference is favored, Singular is chosen; if collective reference is favored, Plural is chosen.

If at the first question the environment instead indicated that the referrent was multiple, the Number expert would choose Plural immediately.

If the item is extensional, then the expert can choose Singular after the second question. Note that this choice process covers other uses of English plurals not suggested by the "lions" sentences: a collection of kinds (e.g., three perfumes, a perfume), or a collection of quanta of a substance (e.g., three beers, a beer) are also correctly designated.

The questions above do not really cover the full complexity of this determination. Nevertheless, they are sufficient to raise issues.

For each question that we attribute to the choice expert, there is an issue of the question's appropriateness. Is the Singular/Plural expert really the right place to ask a question about intension/extension? Should there be alternatives beyond intension and extension, or should they be refined further? Should the grammar concentrate such questions at another point? Should the similarity of the two followup questions be captured as a generalization, somehow making them the same question?

There is also an issue of the completeness of any such account. Should there be additional questions? Do the questions cover the possibilities?

Responding to this last question by way of example, we know that Plural is involved in more complex ways with quantification and determination as in "every man," which is semantically multiple and grammatically singular. These phenomena are not allowed for under the choice process described above, but expansions of the choice processes could be proposed to allow them. (Note that such expansion is not like expanding the systemic syntactic account to offer alternatives beyond Singular and Plural. The syntactically motivated alternatives are still Singular and Plural, but we may have several distinct circumstances for choosing each.)

The fact that we can raise these questions indicates the utility of the choice expert concept as an aid to the development of choice-oriented theories.

The preceding example simply suggests some aspects of choosing that need to be explored. The sections below deal with three primary aspects: asking questions of the environment, answering, and choosing. These three are systematically incomplete because they do not deal with the subject matter of the questions, the entities represented by "the item" in the examples above. Later sections extend the discussion in order to make the subject matter explicit, and then partially formalize the result.

4. ASKING, ANSWERING, AND CHOOSING

Below we present a basic orientation to choice methods, first describing questions, then answers, and then choosing. Later sections supplement this orientation with other essential operations.

Varieties of Questions to Present to the Environment

The examples above are drawn from a grammar that contains over 200 systems, each raising relatively specific issues. Those examples do not suggest the diversity of questions that the grammar might address to the environment (see [12]).

We have attempted to draw up representative lists of questions for various systems. In one kind of experiment, we generate the unit structure for a particular sentence by projecting the questions that each choice expert would have to ask in order to make a well-informed correct choice. The resulting lists of questions are the basis of a representation of the content of the choices for those systems that are entered.

Some strong patterns emerge from the resulting sets of questions. Three kinds of questions are particularly influential in determining what is generated:

1. Some inquiries are used to determine whether information of a certain character is ***available,*** such as the location or duration of an event. These are generally used just before other inquiries that seek to characterize information.
2. Some questions try to categorize or characterize available information: The operators used for ***information characterization*** form the largest collection of inquiries. They are used to subcategorize, to discover relations of inclusion, identity, precedence, adjacency, and also to discover attributes of manner, number, completeness, intended emphasis, identifiability to the reader, decomposability, gender, hypotheticality, extensionality, and many other sorts.

3. Several inquiries about ***preference*** are concerned with whether available information should be expressed. Others seek a choice between lexical items that are syntactically and semantically undifferentiated for present purposes.

Note that for each of these kinds of inquiries, the set of possible answers is closed, since it is fixed by the inquiry and predictable in advance.

Answering

In this model there is a definite boundary between the grammar and its environment. Knowledge of the world and the intended communication belong to the environment. We could put a particular grammar in very different environments, and as long as the questions received the same answers, the generated units would be the same.

Two basic observations about answering can be made:

1. The method that the environment uses to determine its answer is not part of the grammar. A description of the grammar can therefore omit these methods.
2. The range of possible answers must not vary from one environment to another; rather it must be controlled (definitionally) by the grammar, and so a description of the grammar must include them.

Of course, there are many interesting questions about how the environment could possibly answer the grammar's questions. The question-answer interface simply provides a factoring of the descriptive problem. Suitably developed, it could also provide a grammar-independent way to specify a range of sentence-generation facilities, with varying degrees of capability, and correspondingly, a range of sentence-generation demands. These are particularly interesting options for those who are developing computer-based text-generation systems.

Choosing

Since the answers given to each choice expert have a predictable range, the response of the choice expert can be completely determined in advance. For each of the possible answers, there is a next action, either a question to ask or a choice to make.

The questions, answers, and choices therefore form a decision tree. We can describe a choice expert completely in terms of this decision tree. The questions are the only parts that present problems of formalization; the choices and answers have a natural but formal character.

5. A DECISION TREE OF ASKING, ANSWERING AND CHOOSING

The decision tree of the Number system's choice expert exemplified above is shown in Figure 5 below. It is an informal decision tree, since the questions are in English.

Since choice experts can be represented in decision trees, we have the obvious option of representing more of a choice expert's work in the systemic network. Then each choice could be made on the basis of a single question addressed to the environment. We do not do so, for several reasons:

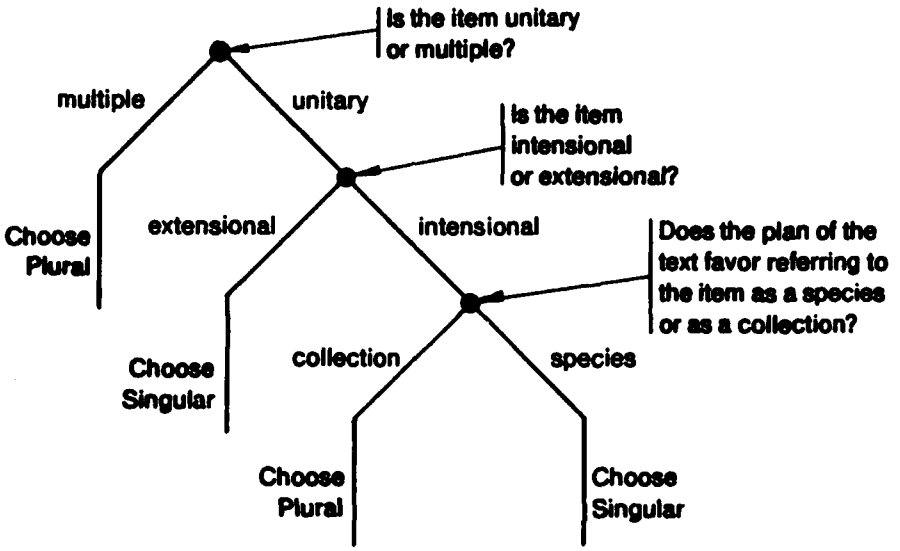


Figure 5: A decision tree for the choice expert of the Number system

- The grammars into which we might absorb such trees are much more mature and stable than the ideas being developed here about choice. The accomplishments that they embody should not be disturbed.
- Argument for particular content in the grammars is different from argument for particular content for a choice expert's decision tree. The two kinds of argument should develop as separate methods.
- It is premature to create such networks, since no extensive representation of choice experts in these terms has been done.

6. ASKING, ANSWERING, AND CHOOSING ARE NOT ENOUGH

The account of choice experts above is defective in that it does not adequately identify the subject matter of the choice expert's questions.[5] For example, in the decision tree for Number, each of the questions includes the phrase "the item." However, we have said that the environment is not required to remember anything about the ongoing "conversation" with the grammar, so asking about whether "the item" has a certain character is meaningless. "The item" stands for a hidden variable, an element that makes the question have different content in different instances of its use.

We need to expand the notion of a question so that it is interpretable by the environment. The expanded view should meet at least the following constraints:

[5] It also does not describe the three-way interaction between choice experts, the environment, and the lexicon. The extension of these ideas to the interaction with the lexicon is outside our scope here. However, see [13] for a description of how the lexicon, grammar, and knowledge representation of the environment might be suitably related.

1. The grammar does not need to know any of the literal symbology of the environment.
2. The grammar can ask an unlimited number of questions about any particular item.
3. More than one choice expert can ask about the same item.

We call an item in the environment a *hub*, by analogy to the hub of a spoked wheel, partly because in a network representation of knowledge such items tend to have a well-identified central structure with connections to surrounding structures.

Where does the grammar get its working knowledge of hubs? Nearly all of the systems are concerned with only a single hub, one identified as part of the work of systems that are "upstream" from the system in question. For the Number system, there is a NominalGroup being built, which will refer to an entity known in the environment. That entity is the hub for this system. In our example, it represents the intensional concept of a particular kind of animal (lions). The Number system can ask questions about lions provided that this hub has been identified to it as the one for which it should choose.

We will now again take up the example of lions, making the choice in the Number system for this hub. The method will be to associate hubs with function symbols, and to allow the choice experts to inquire about the hub currently associated with a function symbol they know.

Let us assume that the Number system has been entered and that a hub named LLL has been identified by being associated with the grammatical symbol THING by some choice expert that acts before Number is entered.

When Number is entered, it can start by asking

Is the hub associated with THING an intensional concept or an extensional one?

This question, presented to the environment, is effectively translated into

Is LLL an intensional concept or an extensional one?

The former version is in a fixed form that can be written as part of the fixed definition of the choice expert. The latter version does not contain any symbols unknown to the environment. A choice expert can ask an unlimited number of questions about any hub that has been associated with a grammatical function symbol. More than one choice expert can inquire about the hub associated with THING. Thus, all of the constraints listed beginning on page 12 are satisfied.

Because of the translation of inquiry forms, we can regard the interface between grammar and environment as a two-layer boundary, with a ***mediator*** process between the boundaries performing the inquiry translations, as depicted in Figure 6. It is a simple substitution process that uses a table of the existing associations between grammatical functions and hubs. The environment's responses are not translated.

The mediator isolates the grammar from the symbol system of the environment. The grammar is written in terms of grammatical functions; no symbols from the environment are written into the grammar. The environment does not encounter grammatical function symbols in inquiries. It sees only the question symbols of inquiries and hub names it has supplied itself.

The most important consequence of this arrangement is that ***the grammar can operate without any particular sensitivity to how knowledge is represented in the environment.***

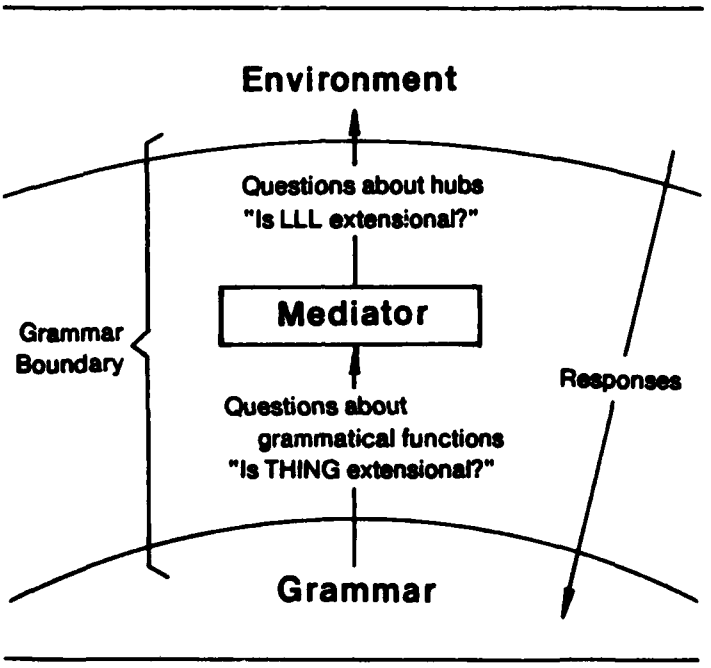


Figure 6: Two-layer interface

Up to this point we have not indicated how hubs are acquired by the grammar. For this purpose there is a second kind of inquiries--those for which the responses are drawn from open sets. These inquiries do not create branch points in the decision tree. Instead, they are used to create associations as described in the following section 7.

7. CREATING FUNCTION ASSOCIATIONS

Since some symbols come from the environment as responses to inquiries, and must be included in other later inquiries, there must be some sort of memory for remembering the symbols (and their significance) between inquiries. We have already identified this memory: associations between grammatical function symbols and hubs. This is an extension of the notion of a function symbol, since we can now ask of a function symbol what concept it represents and also what linguistic realization it has.

This extension seems particularly natural where reference is being performed. Function symbols such as ACTOR and BENEFICIARY are already in the grammar, and in satisfying intentions to communicate, ACTOR will be associated with hubs for actors, BENEFICIARY with hubs for beneficiaries, and so forth.

This use of function symbols is an extension in another way. Some function symbols will be associated with hubs but will not correspond to constituents in the generated structure. In Nigel the

function symbols EVENTTIME and RELEVANTTIME are used in the reasoning about tense, but do not have their own distinct constituents in clauses. The function symbols SPEAKER and HEARER are used in reasoning about pronouns, and the symbol SPEECHACT is used in reasoning about mood.

These various uses of function symbols are compatible: the way of identifying the hub to be associated with a function does not depend on whether that function will be inserted into the structure. Several functions are inserted in some instances but not others, and yet they carry the same hub information in each case. (For example, AGENT would carry the same hub symbol for either "Someone closed the door" or "The door was closed," but it would be inserted only in the second case.)

Associations between function symbols and hubs are created by the method used to present an open-set question to the environment. Part of the specification of such a question is the function symbol with which the environment's response will be associated. There is only one such symbol, and it must not have an existing association when the question is asked. Associations therefore cannot be changed, once made.

8. HUBS IN ANSWERS FROM THE ENVIRONMENT

We have seen examples of using hubs in questions, in the Number choice expert. This section exemplifies the opposite, hubs in answers from the environment.

Consider the construction of the clause structure for the previous example:

"Of all of the kinds of animals in the world, I am most interested by the *lion*."

The Agency system examines the process (ultimately represented by "interested") as to whether it is the kind of process that has an agent, and chooses Agentive. It does so by first asking the closed-set question

Does the process associated with PROCESS have an agent? (Answer: yes)

and the open-set question

What is the agent of the process associated with PROCESS? (Answer: LLL)

The second question yields the hub LLL. This is the first appearance of the hub identifier for "lions," and it is the only occasion in which LLL is conveyed to the grammar by the environment. LLL is then associated with the AGENT function by the process that presents the second question to the environment. This portion of the Agency choice expert can be expressed informally by

"Associate with AGENT the response from the inquiry 'What is the agent of the process associated with PROCESS?'"

9. HUBS AND RANK BOUNDARIES

Continuing with the lions, we note that the completed clause structure contains a function bundle that includes the AGENT function, associated with LLL. How does this lead to creation of the NominalGroup that expresses the concept, the one in which the choice of Singular is made?

The grammar treats the phrase "the lion" as part of the expression of AGENT, constructing "by the lion" in the clause. In the clause, it is under a function bundle containing the functions AGENT and PHENOMENON, which is realized in part by direct insertion of "by" in the clause structure, and in part by creation of a NominalGroup to realize the AGENT/PHENOMENON. In the NominalGroup the function THING is realized by the word "lion." To generate the word thus takes two passes through the grammar, the second of which includes the choice of Singular.

At the end of the first pass through the grammar, in the AGENT and PHENOMENON function bundle, both AGENT and PHENOMENON are associated with LLL. The bundle thus has a single, uniquely associated hub. (This is essential. The uniqueness of the hub association for a function bundle is a fundamental restriction on the form of a purpose-filling grammar. There must be one and only one hub associated with each bundle that is not to be immediately lexicalized.[6])

The grammar is a single network, always entered through the Rank system, where the choices are Clauses and GroupsPhrases. Tracing the hubs backward, we see that there must be a single hub representing the entire sentence (or other independent unit), and that it must be associated with ONUS at the moment the grammar is first entered, in order to generate the sentence. This hub represents the entire intention to communicate for that sentence. All of the other hubs representing aspects of that intention must be accessible from this hub. ONUS is, in effect, the root of every structure, the function that is always present when any choice expert is active.

So, when the grammar is reentered to generate the group "the lion," the previous association of AGENT with LLL is replaced by a new association of ONUS with LLL. The grammar is then ready to generate the group.

10. RELATING HUBS TO REALIZATION

Association of hubs with functions can be seen as a kind of realization of the hubs. It is entirely dependent on the function machinery. In contrast, the choice-expert processes themselves are not realizational. They have the character of mechanism rather than symbolic result. They operate before any realization takes place; some choice experts do not perform any association operations at all. However, choices can be seen as realizations of the intentions and knowledge that they discover in the environment.

If choice experts were to make all of their choices randomly, but still perform the related realizations (including associations), then the grammar could in principle generate its full range of units. However, if we restrict the generation so that the choice experts are operating and the hubs in the *environment* have randomly chosen attributes, then the grammar will generate a smaller range. The difference comes from the consistency of the environment's answers. If, for example, a certain hub is randomly selected to be masculine, then every time the grammar asks about the gender of that hub, the answer will be "masculine." Thus we could write a grammar in which agreement of gender between tags and subjects was not controlled grammatically, but which nevertheless did not make gender agreement errors. It would not be possible to generate "He washed, didn't she?" since every time the grammar asked gender of that hub, the same answer would be given by the environment. Gender agreement would prevail, even if the environment were constructed randomly. Our present

[6] The converse does not hold; a particular hub can be associated with several different functions at once.

grammar is in fact written this way; agreement of number and gender are achieved by having the environment make consistent answers to inquiries. The grammar writer has the option of representing such agreement properties in either of the two ways, by the consistency of the environment or by the symbolic operations of the grammar.

11. INQUIRY AND ANSWERING LANGUAGES

So far, we have presented all of the choice experts' questions in English. This seems right for developmental purposes, since a great deal of manual elaboration and refinement of these questions is needed. However, the questions should eventually have formal status and notation. This section presents a provisional formal notation for the inquiry and answer languages by which the grammar interacts with the interface. It is our hope that by developing these notations, the result will be stable enough so that it can be used in computer implementations of systemic text generators.

Four elements of notation are needed, each specifying some of the communicative capabilities of the choice expert. We will show formal correlates of the previous informal exchange:

- Grammar: ***Is the process PROCESS one that conceptually has some sort of entity that causes the process to occur?***
- Grammar (out of the Mediator): ***Is the process INTX one that conceptually has some sort of entity that causes the process to occur?***
- Environment: ***Agent.***
- Grammar: ***What is the entity that caused or causes PROCESS?***
- Grammar (out of the Mediator): ***What is the entity that caused or causes INTX?***
- Environment: ***LLL***

Question Symbols

For the question symbols we only need a fixed set of symbols, used by both the choice experts and the environment. The entire content of a question, exclusive of the hub identifications involved, is then represented by a single symbol. Any set of agreed-upon symbols would do, even the integers. For convenience, we will use perturbations of concatenated English words such as CausedProcessQ and CauserID, the Q-ending signifying closed-set responses and ID signifying open-set responses.

Identifying Hubs in Questions

A question may need to mention several hubs, as in "Is the activity PROCESS complete at time TIME?" The representation of the entire inquiry is a parenthesized ordered list containing the question symbol and zero or more hub identifiers, for example, (OngoingnessQ PROCESS TIME). The question symbol, say "OngoingnessQ," has an associated convention fixing the order of the hub identifiers. We will generally use the same left-to-right order in both the formal and English notations.

Answer Symbols

Answer symbols can also be an arbitrary collection of symbols. We will use mnemonic English words such as "multiple," "intensional," and "mental," recognizing that there is a small fixed set of answer symbols allowed to appear in response to each question symbol. Specifying this set is part of defining a question symbol.

Identifying Hubs in Answers

The open-set responses from the environment consist of single symbols, since only one hub need be mentioned in an answer.

Using these conventions, the formal version of the exchange above is as follows:

Grammar:	*(CausedProcessQ PROCESS)*
Mediator:	*(CausedProcessQ INTX)*
Environment:	*Agent*
Grammar:	*(CauserID PROCESS)*
Mediator:	*(CauserID INTX)*
Environment:	*LLL*

12. USING CHOICE DESCRIPTIONS TO INFLUENCE THE FORM OF THE GRAMMAR

Some systemicists, including Berry, Butler, Fawcett, and Matthiessen, have noted a need for a stronger tradition of selectivity in grammar writing: many local variants in a grammar seem equivalent, and often there is no good reason to select particular ones over others.

The presence of choice descriptions influences the form of a grammar, in that reasoning about choice representation leads to new preferences, but without limiting the generative power or representational perspicuity of the framework. Our experience in developing choice descriptions is that the preferences that arise are strong, they arise frequently, and they tend to help identify problem areas in the grammar; in addition, solving choice representation problems tends to suggest solutions that work well syntactically. The influences from the development of choice experts thus support grammar development rather than leading to compromises or approximations. It is as if the functional orientation of the grammar and the intention-fulfilling orientation of the choice experts tended toward a natural harmony.

In this section we exemplify a few of the argument forms by which reasoning about choice influences the grammar.

External Evidence for Metafunctional Organization

The ability to reason about the hubs makes possible new kinds of arguments about the overall organization of the grammar.

For example, we may reexamine Halliday's assignment of all systems as either Textual, Interpersonal or Ideational. In the PENMAN knowledge-delivery system for which the ideas here are being developed, the environment has three partitions: the knowledge base, the text plan, and the speaker state. To a first approximation, Textual systems ask questions about hubs in the text plan

portion, Ideational systems ask questions about hubs in the knowledge base, and Interpersonal systems ask questions about hubs in the speaker state. However, the correspondence is not complete, and the boundaries are somewhat uncertain. Nevertheless, the distribution of sites of the hubs that are asked about can be used as an independent basis for assigning metafunction.

The patterns of travel of hubs in the grammar are also informative. Some functions, for example, SUBJECT, get their hubs from other functions; others such as ACTOR do not. The set of functions without hub associations is also a distinguished set. And one may be suspicious of functions that carry hubs only part of the time.

Parallel Question Sets

If the decision trees of two choice experts contain questions in common, that suggests the possibility of a unification in which only one system asks the questions for both purposes.

Premature Offers of Choices

The grammar can offer choices prematurely. For example, it can offer the choice of subject number to be singular or plural before the subject has been identified. Under these circumstances, there is no reasonable way for the choice expert (e.g., of SubjectNumber) to choose. (Note that this is not a problem when the resulting unit must merely be grammatical; it is only problematic when pursuing an intention to communicate.)

Some **unreasonable** ways to respond would include the following:

1. The expert could ask all of the questions that (elsewhere) lead to choice of subject, figure out what the subject is going to be, and choose accordingly. This in effect requires the expert to simulate other experts. It is grossly unsatisfactory because it introduces covert requirements for rerepresentation of parts of the grammar.
2. The expert could stop and wait until the determination had been made. Our conventions do not allow this. If they did, the experts might all end up waiting.
3. The expert could compromise the intention.

Since the issue is the choice expert's access to the appropriate hub, one **reasonable** possibility is to relocate the choice to just beyond the choice-point where the function-bearing hub for the item (e.g., the ACTOR or GOAL) is conflated with the function representing the constituent being developed (e.g., SUBJECT). This requires some reorganization of the grammar, and so provides a reason to prefer the reorganized grammar over the previous (otherwise equivalent) one.

13. DEVELOPING THE CHOICE EXPERTS

This report has outlined a new framework for choice description. Obviously, the details are quite tentative, and a complete formalization of the area would be premature. Many kinds of refinement and development activity could be undertaken. However, we see that certain kinds of development activity now have special priority.

The most important kind of activity for developing choice descriptions is to **describe a large number of actual choice processes** relative to existing choice-oriented grammars. This activity involves repeatedly sitting down with a sentence from the middle of a text, assuming that one's favorite grammar of the language had actually generated that sentence, and identifying the questions and answers in the conversation that the grammar had with its environment. The description of a particular choice expert passes through various stages:

1. the choice as a labeled set of alternatives;
2. a single path of questions and answers, enough to cover a small number of examples;
3. an incomplete decision tree, with some answer options not identified or developed to the point of a choice;
4. a complete decision tree, in informal (English) notation;
5. a formal decision tree, including the related association operations.

After an initial time of elaboration, the description of each choice expert stabilizes as an independent descriptive unit. It would seem that the grammar is characterizing itself in a way that was never accessible by direct examination.

The highest priority activity is in the early stages of the sequence, up to the point at which the emerging descriptions are complete and stable, but informal, still in English. Why is this activity so important? Partly because it seems to effectively pursue many different goals at once. Whether one is interested in syntax, semantics, pragmatics, or in describing the social functions of language, the process extends and refines current knowledge. It documents itself in a form that is easily shared, but that also leads easily to formalization.

14. TOPICS FOR FURTHER DEVELOPMENT OF THE FRAMEWORK

The framework described here is seriously incomplete. In particular, the following areas must eventually be extended:

1. Lexicon: The environment knows relations between its knowledge and the lexicon. The inquiry and answering forms must be extended to cover issues of lexical availability and lexical choice.
2. Modification: Where other than single lexical items are used to refer, there must be a way to develop and control the content of referring phrases. We do not have the experience yet to tell whether the inquiry forms in this report are sufficient for control of modification.
3. Iteration: The mechanisms will need to be extended to handle unlimited iteration such as in "John's sisters are Amy, Barbara, Carolyn, Diana, ..." The syntactic mechanisms and choice processes need to be extended in a coordinated way.
4. Quality: As with any attempt at systematic, accurate description, there are quality considerations. Descriptions of choosers and descriptions of syntax interact in interesting ways. An approach to creating high-quality choice experts is needed. Our experience so far suggests that the same sorts of form criteria which lead to quality in a systemic description of syntax will apply, and that these criteria strongly resemble quality of form criteria for computer programs.
5. Understanding varieties of hubs: There is an interesting diversity to the hubs encountered, their uses and relations. Some of them represent new linguistic abstractions.

6. Understanding varieties of sources for hubs: Hubs arise from different kinds of external knowledge, leading to evidence for the structure of linguistic functionality.

Some of these extensions have already been accomplished or begun, but all are beyond the level of detail of this report.

15. CONCLUSIONS

We have presented a new way of thinking about choices, representing them, identifying their content, and progressively making the notion of grammatical choice more explicit. The key conceptual elements are the distinctness of the grammar and its environment, the metaphor of a choice expert that asks questions, closed sets of question and answer symbols, open sets of hubs and hub identifiers, association of hubs with grammatical function symbols, and choice expert processes as decision trees. This conception is compatible with the systemic framework and contributes to it. At the same time, it simplifies talking about how systemic grammars fit with various concepts of text and communication, and because it helps relate text to intentions to communicate, it contributes directly to the art of computer text generation.

REFERENCES

1. Fawcett, R. P., *Exeter Linguistic Studies.* Volume 3: *Cognitive Linguistics and Social Interaction,* Julius Groos Verlag Heidelberg and Exeter University, 1980.

2. Halliday, M. A. K., "Categories of the theory of grammar," *Word* 17, (3), 1961, 241-292.

3. Halliday, M. A. K., "Syntax and the consumer," in C.I.J.M. Stuart (ed.), *Monograph Series in Languages & Linguistics.* Volume 17: *Report of the Fifteenth Annual (First International) Round Table Meeting on Linguistics and Language Study*, pp. 11-24, Georgetown University Press, Washington, D.C., 1964.

4. Halliday, M. A. K., "Language structure and language function," in J. Lyons (ed.), *New Horizons in Linguistics*, Penguin, 1970.

5. Halliday, M. A. K., *System and Function in Language,* Oxford University Press, London, 1976.

6. Halliday, M.A.K., and J. R. Tastin (eds.), *Readings in Systemic Linguistics,* Batsford, London, 1981.

7. Hudson, R. A., *Arguments for a Non-Transformational Grammar,* University of Chicago Press, Chicago, 1976.

8. Hudson, R. A., DDG Working Papers. University College, London. Mimeo.

9. Mann, W. C., and J. A. Moore, Computer as Author--Results and Prospects, USC/Information Sciences Institute, RR-79-82, 1980.

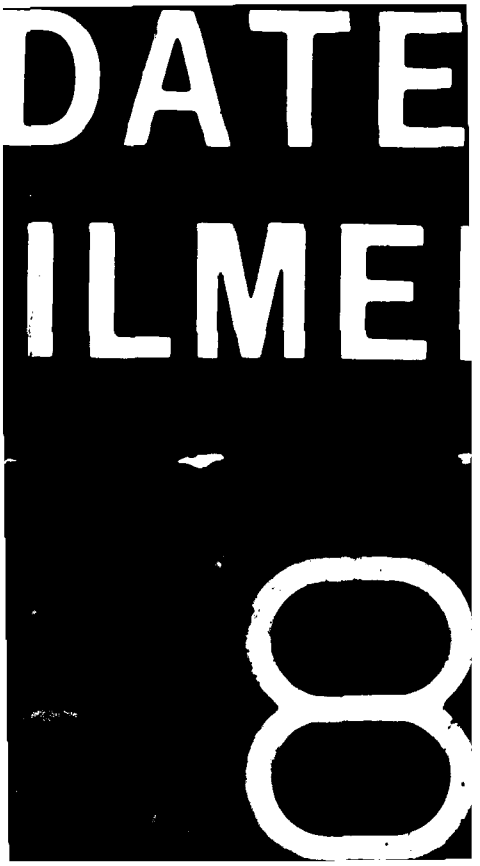
10. Mann, W. C., et al., *Text Generation: The State of the Art and the Literature*, USC/Information Sciences Institute, RR-81-101, December 1981. To appear in *American Journal of Computational Linguistics*.

11. Mann, W. C., "Two discourse generators," in *The Nineteenth Annual Meeting of the Association for Computational Linguistics*, Sperry Univac, 1981. Also available as Information Sciences Institute RR-82-102.

12. Mann, W. C., Generating text: The grammar's demands, 1982. Forthcoming as Information Sciences Institute report RR-82-106.

13. Matthiessen, C. M. I. M., "A grammar and a lexicon for a text-production system," in *The Nineteenth Annual Meeting of the Association for Computational Linguistics*, Sperry Univac, 1981. Also available as Information Sciences Institute RR-82-102.

14. Schachter, P., "Daughter-dependency grammar," in Moravcsik, E.A., and J. R. Wirth (eds.), *Syntax and Semantics*, pp. 267-299, Academic Press, 1980.


DATE
ILME
8