Expressing Rhetorical Relations in Instructional Text: A Case Study of the Purpose Relation*

Keith Vander Linden[†]
ITRI, University of Brighton

James H. Martin[‡]
University of Colorado

Natural language provides an extensive set of lexical and grammatical forms for expressing concepts, a set from which writers choose the particular form which they feel will produce the most effective expression given the communicative context. An important task of the text generation researcher is to specify both the range of these forms and the contexts in which they are used. This paper addresses this issue in the context of the expression of procedural relations between actions in instructional text. It employs the following four step approach to achieve this goal: (1) Collect a corpus of the relevant text type; (2) Perform a detailed linguistic study of a portion of this corpus, called the training set, and reserving the remainder as a testing set; (3) Implement the results of this study in a text generation system; (4) Compare the output of the system with the text found in the entire corpus. This has resulted in the construction of IMAGENE, *an instructional text generation system which embodies a model of the forms of expression consistently used by instructional text writers over a broad range of instruction types. The details of* IMAGENE*'s treatment of purpose expressions are given as representative of the coverage and form of the full system.*

1. Introduction

Natural language provides an extensive set of lexical and grammatical forms for expressing concepts, many of which may, taken out of context, appear to be interchangeable. They are not interchangeable. Writers systematically choose the particular form from this set which they feel will produce the most effective expression given the communicative context. An important task of the text generation researcher is to inform the text generation process with a specification of both the range of these forms and the contexts in which they are used.

The current study addresses this issue in the context of expressing procedural relations between actions in *instructional text*, that is in written, procedural directions. The complexity of procedural relationships typically expressed in such text has given rise to complex variations of expression in language. Consider, for example, the problem of expressing purpose relations. Such expressions could take many conceivable forms, all of which are perfectly grammatical:

(1a) Pull out sharply *in order to remove the phone*.

* This work was supported by the National Science Foundation under Contract No. IRI-9109859.
† Information Technology Research Institute, University of Brighton, Lewes Road, Brighton BN2 4AT, UK; knvl@itri.bton.ac.uk
‡ Department of Computer Science, University of Colorado, Boulder, CO 80309-0430, USA; martin@cs.colorado.edu

(1b) *To remove phone*, pull out sharply.
(1c) Pull out sharply *for phone removal*.
(1d) Pull out sharply *for removing the phone*.
(1e) *For the phone*, pull out sharply.
(1f) *Remove phone by* pulling out sharply.
(1g) *Remove the phone.* Pull out sharply.
(1h) The purpose of pulling out sharply is *to remove the phone*.
(1i) Pulling out sharply achieves the purpose of *removing the phone*.
(1j) *Removing the phone involves* pulling out sharply.
(1k) *The method for removing the phone* is to pull out sharply.

As can be seen, purpose expressions occur either before or after the expression of their related sub-actions (referred to here as the issue of *slot*), and are expressed in a number of grammatical forms (the issue of *form*). They may be linked with a variety of conjunctions or prepositions (the issue of *linker*), and may or may not be combined into a single sentence with the expression of their sub-actions (the issue of *clause combining*). The current study addresses these four issues of choice in the context of instructional text.

Text generation systems must know which forms to produce and when to produce them. Formal linguistic analyses are useful for weeding out grammatically unacceptable forms, but they do not provide a principled means of determining which of the grammatically acceptable forms should be used in any given communicative context. As an alternative, the current study has employed the following four step process for identifying both the relevant forms of expression and the contexts in which they are used:

1. Collect a corpus of text from the relevant genre, and encode a full range of the lexical and grammatical features of all of the text;
2. Perform a linguistic analysis of part of the corpus. This analysis involves determining the range of forms used in the corpus and then using an iterative cycle of hypothesis formation and testing to determine the communicative contexts in which each is used;
3. Implement the results of this analysis in the text generation system;
4. Compare, in detail, the output of the system with the text found in the corpus, differentiating between the predictions concerning text that was specifically used in the analysis (the training set), and text that was not (the testing set).

This process begins and ends with the corpus, providing an empirically based approach to identifying the range of lexical and grammatical forms that are used in real text and to determining the contextual issues that are relevant to choosing among them. Although the corpus study has become a common methodology in natural language generation, seldom are the representation and analysis techniques given in any detail, and neither are detailed evaluations of the resulting text provided. These details are provided, for our study, in this paper.

Our corpus is divided into training and testing portions. The training portion, used in step (2), constitutes approximately one-third of the full corpus and consists entirely of cordless telephone manuals. The methodology is successfully applied to this portion, showing that there are, in fact, patterns of expression in cordless telephone manuals that can be identified and implemented. The study is then extended by testing the system's predictions on a separate and more diverse portion of the corpus which includes instructions for different types of devices and processes. This additional testing serves both to disallow over-fitting of the data in the training portion, and to give a measure of how far beyond the telephone domain the predictions can legitimately be applied. No testing was done on non-instructional texts and no claims are made concerning the applicability of the system's predictions in those areas.

Following a review of relevant work in the area of natural language generation, this paper will discuss how these four steps have been applied to the generation of rhetorical relations in instructional text. It will detail what rhetorical relations in instructional text are, and how they were collected and represented. It will then discuss how the corpus analysis was performed and how the results were implemented in IMAGENE, an instructional text generation system. The details of IMAGENE's treatment of purpose expressions are given as representative of the coverage and form of the full system (more details concerning the other relations can be found elsewhere (Vander Linden, 1993c)). It will conclude with a discussion of how well IMAGENE's predictions match the text in the training and the testing portions of the corpus.

2. Related Work in Natural Language Generation

In pursuit of other issues, many studies have adopted a temporary solution to the problem of managing diverse forms of expression, namely that of choosing a single lexical and grammatical form to express each of the relevant types of information dealt with by the system. It is then a simple matter of allowing the type of information to determine the appropriate expressional form. Hovy's text structurer (Hovy, 1988b), for example, uses rhetorical relations as defined in Rhetorical Structure Theory (RST) (Mann and Thompson, 1987) to order a set of propositions to be expressed. It does not make any decision as to how a chosen relation is to be expressed, but rather leaves this task to the rudimentary implementation of the rhetorical relations provided by the Penman text generation system (Mann, 1985). In the case of a purpose, for example, Penman will produce a non-fronted *in order to* infinitive clause, as in the following output of the structurer (Hovy, 1988b, page 167): "Knox is en route *in order to rendezvous with CTG 070.10*." A similar approach to expressing rhetorical relations was taken in McKeown's TEXT system (McKeown, 1985).

Text generators specifically designed for instructional text, such as Mellish and Evans' generator (Mellish and Evans, 1989), EPICURE (Dale, 1992), COMET (McKeown et al., 1990), and TECHDOC (Rösner and Stede, 1992) display similar characteristics (See the analytical work of Delin et al (Delin et al., 1993) for a notable exception to this pattern).

Mellish and Evans' generator, for example, uses the output of NONLIN, a non-linear planner (Tate, 1976), as the preliminary rhetorical structure for the text. Because this often produced text which was monotonous or hard to understand, they included what was termed a *message optimization* phase which specifies rules for removing or modifying certain elements of the plan structure that are known to produce poor text. Although this greatly improves the text, it still tends towards text which is difficult to read. This problem is, in part, due to the fact that some of the plans they look at are quite complex and correspondingly difficult to express, but it is also attributable to the lack of a detailed corpus study of the linguistic tools used by technical writers in instructional text. The IMAGENE project can be seen as an extension of their work which employs such a study to help manage diversity of forms of expression. There are other natural language generation projects which have addressed similar issues. Two such examples are Hovy's Pauline (Hovy, 1988a) and Meteer's Spokesman (Meteer, 1992; Meteer, 1991), both of which are based, at least in part, on corpus studies.

Pauline produces an impressive range of expressional forms that are based on a list of pragmatic features of the communicative environment including information about the conversational atmosphere, the speaker, the hearer, the relationship between the two, and the interpersonal communicative goals of the speaker. Its construction required a considerable amount of analysis of sample texts, but, unfortunately, very little is said about how this analysis was actually performed, and how well the text produced by Pauline matches the text in the corpus. The concerns of the IMAGENE project are similar, except that both the text type, instructional text, and the linguistic phenomenon, expressing rhetorical relations, are much more focussed. The results of our study are therefore more detailed, but also more constrained (cf. Kittredge et al's concept of *domain communication knowledge*, 1991).

In the Spokesman project, Meteer proposed an architecture for addressing what she termed the problem of *expressibility* in text planning (Meteer, 1992). Her fundamental thesis is that an abstract linguistic representation is needed to provide the text planner with information on constraints of expression (cf. Huettner et al., 1987). Her constraints are taken, at least in part, from a study of text revisions made by expert editors. The IMAGENE project concurs with this concern for detailed forms of expression, but its methodology is geared toward determining the elements of the communicative context used to choose between equally acceptable alternative forms of expression. Meteer doesn't address what to do if, after using her constraints to remove unacceptable forms of expression, there are a number of remaining acceptable forms. This issue of choice is central to the current study.

3. Corpus Collection and Representation

The corpus developed for this study was taken from various types of instructional text, including instruction booklets, recipes, and auto-repair manuals. It contains approximately 1000 clauses (6000 words) of instructions, taken from 17 different sources representing a diverse array of process types. These sources include instructions for electronic devices like cordless telephones and clock radios, manipulative processes like auto repair and first aid, and creative processes like recipes and crafts. The one common feature is that they all involve the expression of actions and of the procedural relationships between them.[1] As an example of the nature of this text, consider the following excerpt from the instructions for the GTE Airfone (Airfone, 1991) which will be called the Remove-Phone text:

(2) When instructed (approx. 10 sec.) remove phone by firmly grasping top of handset and pulling out. Return to seat to place calls. (Airfone, 1991) [2]

This passage gives an example of the variation of expressional form that is common in instructional text. It contains, among other things, two expressions of purpose: "remove phone" and "to place calls." This notion of purpose, which will be detailed in the next section, is one of actions which are to be realized through the execution of expressed sub-actions. The first is stated as an imperative ("remove the phone"), with

1 It should be noted that this corpus is much smaller than the language corpora used in larger statistical studies (Church and Mercer, 1993). The deep semantic and pragmatic knowledge that was required for the current study has necessitated this.

2 This paper will add a reference to the end of all examples that have come directly from the corpus, indicating the manual from which they were taken. Examples of actual IMAGENE output will be fully italicized. All other examples are contrived for explanatory purposes.

the sub-actions expressed in participial form within a *by* prepositional phrase ("by firmly grasping top of handset and pulling out"). The second ("to place calls"), on the other hand, is expressed in final position as a *to* infinitive with its sub-action stated as an imperative ("return to seat"). The problem to be addressed by the corpus analysis in step 2, is to determine the contextual features used to choose these forms as opposed to the alternate forms which could have expressed the "same basic information."

A relational-style database is used to represent the rhetorical, grammatical, and lexical aspects of the corpus. The representation of the grammatical form of the clauses and phrases is based on traditional principles of syntax and semantics. Clauses and phrases are represented in separate tables. Links within the clause table are used to indicate subordinate relations, and links between the clause and phrase tables are used to represent relative clauses and predicate-argument relationships. It also includes semantic information such as the agent of a particular action (e.g., the reader or the device) and the semantic type of the predication (e.g., material or relational process). More detail on the database can be found elsewhere (Vander Linden, 1993c). The goal was to allow for the representation of any element of the pragmatic, semantic, or syntactic context that might be relevant in the analysis.

Mann and Thompson's Rhetorical Structure Theory (RST) has been used to encode the rhetorical relationships between expressions in the corpus (Mann and Thompson, 1987; Mann and Thompson, 1988). It was developed as a framework for describing text structure, viewed in terms of the semantic and pragmatic relations that hold between text spans at all levels. The current study has made use of five such relations: Purpose, Precondition, Result, Sequence, and Concurrent. This section will now make some general observations concerning RST, present an example RST analysis of the Remove-Phone text, and conclude with definitions of the relations used in the study.

RST distinguishes between what are called *nucleus-satellite* and *multi-nuclear* schemas. The nucleus-satellite schema relates two spans of text, designating a more central span, called the nucleus, and a more peripheral one, called the satellite. This relationship is represented graphically with a directed arrow from the satellite to the nucleus. Definitions of relations of this sort specify constraints that apply to the nucleus (N), the satellite (S), and the combination of the two, and specify the effects of the expression. Purpose, Precondition, and Result, are examples of such relations. The multi-nuclear schema relates one or more spans, designating no span as superordinate or subordinate to any other. Definitions of relations of this sort include specification of the constraints on the nuclei and the combination of nuclei, as well as a specification of the effect of the expression. The Sequence and Concurrent relations are such relations.

RST was attractive for the IMAGENE project because of its ability to represent the hierarchical structure of text with rhetorical structures that matched the level of analysis required for the study of expressions of procedural relationships. There is considerable debate in the field of discourse analysis concerning the relative importance of intentional structure and rhetorical relations (e.g., Moore and Pollack, 1992; Grosz and Sidner, 1986), most systems focusing on one or the other. The current study has conflated them, as the instructional texts have not tended to display the complex intentional structure common to persuasive texts and interactive discourses (Vander Linden, 1993b).

Finally, RST has been used by a large number of researchers for the purpose of text generation (e.g., Moore and Paris, 1988; Hovy and McCoy, 1989; Scott and Souza, 1990; Rösner and Stede, 1992). This testifies not only to RST's usefulness but also to the direct applicability of the results of the current study to the field of natural language generation. Because of the common use of RST, the results can be more easily applied to other work in this area. In particular, the focus on the precise forms of expression of rhetorical relations in instructional text fills an important gap in current work (cf. Scott

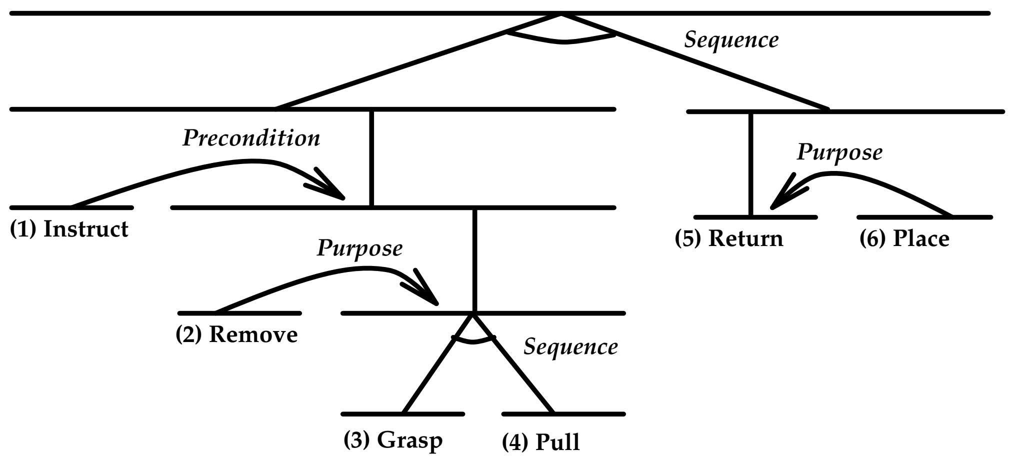

Figure 1
The RST Representation of the Remove-Phone Text

and Souza's work on expressing rhetorical relations (1990)).)

Consider the application of RST to the Remove-Phone Text. The first problem that must be addressed in any RST analysis is the segmentation of the text into spans that will serve as the atomic units of description. In RST, these spans have typically been clauses, as is the case in the Remove-Phone passage, but certain phrases with propositional content may be considered as well. The spans used in our analysis are propositional units that express single actions. In the Remove-Phone Text, there are six such action expressions, listed here in segmented form:

1. When instructed (approx. 10 sec.)
2. remove phone
3. by firmly grasping top of handset
4. and pulling out.
5. Return to seat
6. to place calls.

The second problem is one of relating these segments in the appropriate rhetorical structure. The current study has used the two aspects of the RST specification that can be mapped onto the procedural structure of the process being expressed, namely, the hierarchical structure of RST and the subset of RST relations that correspond to procedural relations. Each of these two aspects will be discussed with reference to the RST representation for the Remove-Phone Text, shown in Figure 1.[3]

The first aspect of this structure is its hierarchical nature. The procedural sequence schema at the top of the text hierarchy, for example, indicates that there are two spans

3 The actual RST and grammatical analyses of the text were carried out by one of the authors and the examples crucial to the formalization of the results were reviewed by both authors. This approach would be difficult in the analysis of certain more complex texts such as persuasive texts, but proved to be adequate in the study of local structure in instructional text.

of text that express a sequence of two actions. The spans themselves can be expressed as single propositional units or as more complex spans, the latter being the case with the two spans in this sequence. This is a representational manifestation of the hierarchical nature of the processes themselves, and is displayed graphically by extending the horizontal line of a span to cover all of its subordinate spans. The RST representation of the Remove-Phone Text is a small portion of the full hierarchy that represents the entire manual. The current study has focussed on the expression of just such local sub-trees; the problems of expression of macro-structure are beyond the scope of the analysis (cf. Mooney et al., 1991).

The second aspect of this structure is the nature of the rhetorical relations themselves. The representation makes use of five relations: Purpose, Precondition, Result, Sequence, and Concurrent, which are used as abstractions to identify the lexical and grammatical manifestations of the procedural relationships inherent in the process. They are termed *informational* by Moore and Pollack (1992) and *subject matter* by Mann and Thompson (1987) because they are based on semantic content, rather than on intentional or presentational content.[4] This section will now provide specific definitions of these relations.

PURPOSE (taken from Mann/Thompson 1987)

constraints on N:	presents an activity
constraints on S:	presents a situation that is unrealized
constraints on the N+S combination:	S presents a situation to be realized through the activity in N
the effect:	R recognizes that the activity in N is initiated in order to realize S

The Purpose relation is taken directly from the RST specification. In this paper, it refers to a situation where a higher-level activity is realized through the execution of lower-level sub-steps. An example can be found in the Remove-Phone text cited above: "... *remove phone* by firmly grasping top of handset and pulling out." Here the activity of removing the phone is realized by the execution of the sub-steps of grasping and pulling the phone.

PRECONDITION (taken from Rösner and Stede 1992)

constraints on N:	presents an action
constraints on S:	presents an unrealized situation
constraints on the N+S combination:	S must be realized in order to make it possible or sensible to carry out N
the effect:	R recognizes that situation S must be realized in order to successfully carry out action N

The Precondition relation is a simple amalgam of the standard RST relations Circumstance and Condition. It has been taken from Rösner and Stede's work on generating multilingual instructions (Rösner and Stede, 1992). This particular combination

4 There is some question as to whether these semantically-based relations should be termed *rhetorical* in the classical sense at all (Dale, 1993). Because of the prevalence of this use of the term, however, it will be retained in this paper.

has proven useful in analyzing various kinds of conditions and circumstances that frequently arise in instructions, such as the precondition found in the Remove-Phone text: "*When instructed* (approx. 10 sec.) remove phone ...." Here, the removal of the phone must not be attempted until after the device has instructed the user to do so.

RESULT (adapted from Mann/Thompson 1987)

constraints on N:	none
constraints on S:	presents either a volitional or non-volitional action or the situation that could have arisen from one
constraints on the N+S combination:	N presents a situation that could have caused the situation presented in S; presentation of N is more central to W's purposes in putting forth the N-S combination than is the presentation of S.
the effect:	R recognizes that the situation presented in N could be a cause for the action or situation presented in S

The Result relation is a simple amalgam of RST's volitional and non-volitional result. It was useful for analyzing expressions of actions or situations that are expressed as being the result of other actions, as in: "Place the handset in the base. *The BATTERY CHARGE INDICATOR will light.*" (Excursion, 1989). Here, the device's action of lighting the indicator is a result of the reader's action of placing the handset in the base.

SEQUENCE (taken from Mann/Thompson 1987)

constraints on N:	Multi-nuclear
constraints on the combination of nuclei:	A succession relationship between the situations is presented in the nuclei
the effect:	R recognizes the succession relationships among the nuclei

The Sequence relation is taken directly from the RST specification, and refers to actions that are in temporal sequence, as in the following excerpt from the Remove-Phone text: "... *by firmly grasping top of handset* and *pulling out*."

CONCURRENT (adapted from Mann/Thompson 1987)

constraints on N:	Multi-nuclear
constraints on the combination of nuclei:	A simultaneous relationship between distinct situations is presented in the nuclei
the effect:	R recognizes the simultaneous relationships among the nuclei

Finally, the Concurrent relation is a simple extension of Sequence, referring to actions which are distinct but simultaneous. An example can be found in "Press and hold the mouse button while you move the mouse." (Macintosh, 1988). Here, holding the mouse button and moving the mouse must be done simultaneously.

4. Corpus Analysis

Two related issues must be addressed in the corpus analysis:

1. Determining the range of expressional forms commonly used by instructional text writers;
2. Determining the precise communicative context in which each of these forms is used.

With a couple of minor exceptions, the study was performed exclusively on three instruction manuals for cordless telephones (approximately one third of our corpus) and the results were applied to the remainder of the corpus. The exceptions involved *so that* and *until* expressions, and expressions of concurrency, which were not well represented in the telephone manuals. Examples of these expressions were taken from the remainder of the corpus. The method employed, therefore, has much in common with the method proposed by Quinlan and implemented in ID3 (Quinlan, 1986) in that the training set is expanded in cases where there are insufficient examples on which to base a full analysis. So far, the testing set has not been expanded to include examples on which to test these *so that*, *until*, and concurrent expressions.

4.1 The Range of Expressions

The first task is that of determining the range of lexical and grammatical forms used to express each particular rhetorical relation. The full corpus contains 119 purpose expressions, all but four of which occur in one of the following seven forms (the purpose, i.e., the satellite span of the rhetorical relation, is italicized):

(3a) *To end a previous call*, hold down FLASH [6] for about two seconds, then release it. (Code-a-phone, 1989)
(3b) Follow the steps in the illustration below, *for desk installation*. (Code-a-phone, 1989)
(3c) The OFF position is primarily used *for charging the batteries*. (Code-a-phone, 1989)
(3d) *For frequently busy numbers*, you'll want to use REDIAL [7], and the pause will have to be in Redial memory. (Code-a-phone, 1989)
(3e) When instructed (approx. 10 sec.) *remove phone by* firmly grasping top of handset and pulling out. (Airfone, 1991)
(3f) *Return handset to wall unit from which it was taken*. Insert heel first as shown, then push top in firmly. (Airfone, 1991)
(3g) Tilt pan down slightly at the rear *so that the fluid drains out.* (Reader's Digest, 1981)

All four of the issues of lexical and grammatical choice we addressed are displayed here. The purpose expressions can be textually placed in the slot either before or after the expression of their sub-actions. Further, there are seven combinations of grammatical form, linker, and clause combining to choose from, the relative frequencies and percentages of which are given in Table 1 (where the letters in Example (3) correspond to the letters in the table). Example (3a) uses a *to* infinitive form (TNF). Example (3b) uses a *for* prepositional phrase with a nominalization ("installation") as the complement. Example (3c) uses a *for* preposition with a gerund phrase as the complement. Example (3d) uses a *for* preposition with a noun phrase that refers to the object (or goal) of the corresponding action as the complement. This is termed *Goal Metonymy*. Example (3e) uses a simple imperative for the purpose with *by* conjoining participial forms of the intended actions.

	Initial	Final	Total (count)	Total (percentage)
(a) To-Infinitive	38	33	71	59.6%
(b) For-Nominalization	2	7	9	7.5%
(c) For-Gerund	0	3	3	2.5%
(d) For-Goal-Metonymy	1	5	6	5.0%
(e) By-Purpose	11	1	12	10.0%
(f) Adjoined-Purpose	4	0	4	3.3%
(g) So-That-Purpose	0	10	10	8.4%
Other	4		4	3.3%

Table 1
The frequency of various form and slot combinations of purposes in instructional text

Example (3f) uses a simple imperative for the purpose, with the intended actions in a separate sentence following the purpose. Example (3g) uses a simple imperative for the intended actions, with a *so that* conjoining a present tense action form of the purpose.[5]

4.2 The Context of Expression

The second task, that of determining the functional context in which each of the forms is used, is more difficult. The IMAGENE project employs a hypothesis generation and test cycle, such as the one advocated by Cumming (1990) in an attempt to identify correlations between the contextual features of communicative context on the one hand, and the lexical and grammatical form on the other.

This methodology starts with the range of lexical and grammatical forms corresponding to each of the rhetorical relations considered. In the hypothesis phase, the analyst hypothesizes a feature of the communicative context that appears to correlate with the variation of some aspect of the lexical and grammatical forms. These hypotheses may come from an intuitive analysis of the texts, as well as from the current literature on the subject. The features themselves pertain to any of three aspects of the communicative context (termed *metafunctions* in Systemic Linguistics): Ideational — the propositional meaning of the material being expressed (associated with the traditional notion of semantics); Textual — the flow and structure of the text (associated with discourse analysis); and Interpersonal — the human relationships between interlocutors (associated with Socio-Linguistics) (Halliday, 1985). All of these types of features have proven relevant in the analysis. In the test phase, the analyst attempts to validate the hypothesis by querying the database for the relevant information. These two phases are repeated until a good match is achieved or until a relevant hypothesis cannot be found.

As an example of this methodology, consider the issue of slot, that is, the determination of which span in a rhetorical relation should be expressed first. The slot of purpose expressions in our corpus is split fairly evenly between initial and final purpose expressions. Of the 119 purpose expressions, approximately 48% are fronted and 52% are not fronted. These values are similar to the percentages of initial and final purpose clauses found by Thompson for procedural texts (1985). Although she reports just over 18% initial purpose clauses for English text in general, she reports 49% initial purpose clauses for a book of recipes and 34% for an auto-repair manual.

Thompson's study indicated that one common feature of fronted purpose clauses

5 This study has temporarily characterized these *so that* expressions as action/sub-action expressions. A more detailed analysis of how the situations that give rise to them differ from those for other purposes is yet to be performed.

is that their scope is global, that is, there is more than one expressed proposition that is directly related to the fulfillment of the purpose (e.g., "To achieve purpose A, do B and do C." where there are two sub-actions, B and C). Such a purpose clause is expressing a context in which the prescribed sub-actions are to be interpreted, and thus should be fronted. This provides a good starting hypothesis for determining the slot of purposes, namely that global purpose clauses are fronted. Upon inspection, we find that only 3 cases of global purpose expressions in our corpus (7.9%) are not fronted. This yields strong support for the hypothesis, and allows us to go on and discern what factors motivated the counter-examples. This process continues until either no distinctions can be found, or until there are not enough examples on which to base the distinctions.

The next question to be asked is if this is the *only* explanation of fronted purpose expressions, or if there are other fronted purpose expressions that are not global. This can be addressed by querying all the fronted, non-global purpose clauses in the corpus. In the current corpus, this set is non-empty, which leads to the conclusion that there are other factors at work in the question of purpose slot. The iterative process of hypothesis generation and testing can then be conducted on these other cases in a similar manner.

This analysis technique is designed to identify covariation between elements of the communicative context on the one hand and grammatical form on the other. The purpose slot analysis just discussed, for example, identifies the existence of a covariation between purpose scope and slot. This covariation, however, does not constitute proof that the technical writer actually considers the issue of slot during the generation process, nor that the prescribed form is actually more effective than any other. Proof of either of these issues would require psycholinguistic testing. The results of this study provide detailed prescriptions concerning how such testing should be performed, i.e., what forms should be tested and what contexts controlled for, but does not actually perform them. A discussion of this may be found elsewhere (Vander Linden, 1993a).

5. IMAGENE

This section will discuss the theoretical framework of the implementation and then detail the treatment for purpose expressions. The full IMAGENE architecture, as depicted in Figure 2, consists of a System Network and a Sentence Building routine, and is built on top of Penman. It transforms inputs (shown on the left) into instructional text (shown on the right).

5.1 IMAGENE's Architecture

Penman, a sentence-level generator developed at the USC Information Sciences Institute (Mann, 1985; Penman, 1989), was employed not only because of its broad coverage of English syntax, but also because it is based on a Systemic-Functional view of language (Halliday, 1976). The Systemic view is distinctly functional, that is, it is particularly interested in mapping elements of the communicative context onto the appropriate grammatical forms. As a by-product of this view of language, Penman contains a well-developed implementation of the *System Network*, the Systemic formalism for representing grammar.

The system network is basically a decision network where each choice point distinguishes between alternate features of the communicative context. It has been used extensively in Systemic Linguistics to address both sentence-level and text-level issues (e.g., Fawcett, 1990; Patten, 1988; Berry, 1981). Such networks are traversed based on the appropriate features of the communicative context, and as a side-effect of this traversal, linguistic structures are constructed by *realization statements* which are associated with each feature of the network. Penman's networks are specifically designed to construct

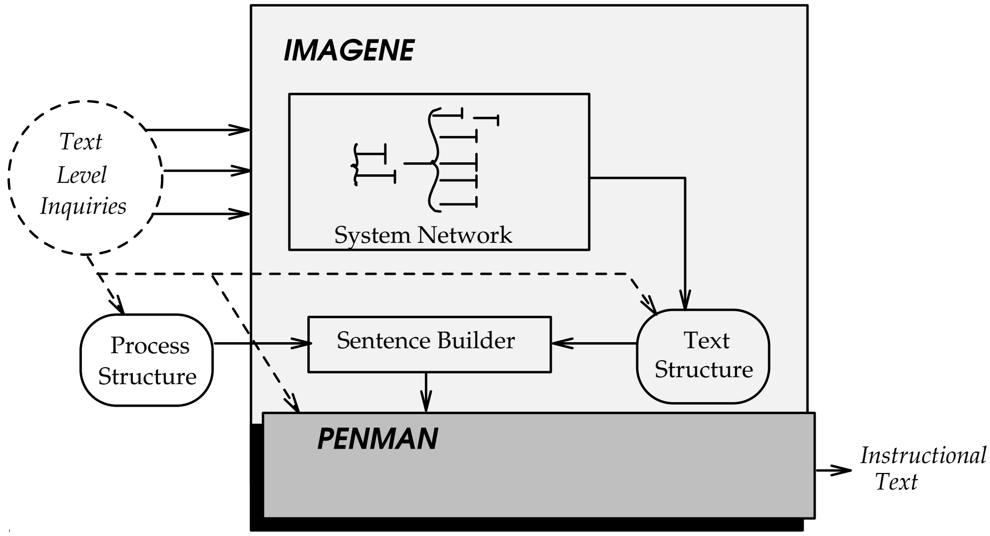

Figure 2
IMAGENE's Architecture

English sentences.

IMAGENE's system network is built in a similar manner, but because it constructs text structures rather than sentences, its realization statements have a significantly different flavor than their counterparts in the grammar developed for Penman.[6] We now give a short discussion of how IMAGENE's realization statements can manipulate the evolving text structure, making reference to the text structure produced by IMAGENE for the portion of the Remove-Phone text shown in Figure 3 (which corresponds to the text span "...remove phone by firmly grasping top of handset and pulling out"). The full analysis of the Remove-Phone text will be given later; here we intend only to illustrate the types of manipulations made by the realization statements.

- Inserting nodes into the text structure (**iterative-insert**, **insert**, **copy**) — IMAGENE starts with an empty text structure and uses these statements to insert action nodes as appropriate. In Figure 3, for example, the Remove-Action, Grasp-Action, and Pull-Action nodes refer to the actions of removing, grasping, and pulling the handset.
- Ordering the surface expression of the nodes (**order**, **reorder**, **insert-order**, **combine**) — IMAGENE uses these statements to specify the order of expression of the action nodes. The choice of clause combining strategies is made here as well. In Figure 3, the links are shown as light-faced directed arrows, marked with either *New-Sentence* or *Continue-Sentence* (in this segment, only the latter is used). **Reorder** is used to change an ordering made earlier in the processing, while the others establish the order of

6 Penman's sentence-level realization statements work with a single, pre-specified list of features of the sentence, called *grammatical functions*, such as ACTOR, PROCESS, GOAL, and THEME. At the text-level, there is no definitive list of what might be called *text functions*. Rather, IMAGENE's realization statements allow the insertion of subscripted elements which can correspond to lists of sequential commands, multiple preconditions, etc.

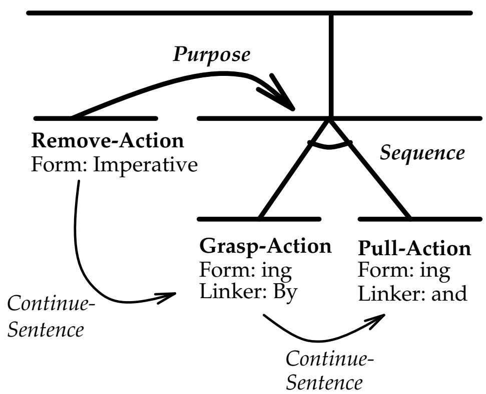

Figure 3
A Segment of the Text Structure for the Remove-Phone Example

newly inserted nodes.

- Building the RST structure between nodes (**structure**, **unlink**) — IMAGENE uses **structure** to create hierarchical text structure links between nodes, and **unlink** to remove them. Figure 3 contains a purpose relation and a sequential relation. As with **reorder**, **unlink** is used to "un-structure" a default structuring.
- Marking the lexical and grammatical form of expression of the nodes (**mark**, **iterative-mark**) — The realization statements also determine the grammatical form of the expression of each of the nodes in the structure. In Figure 3, the Remove-Phone node, for example, is marked as an imperative, and the Grasp-Action as an *ing* form with the linker *by*.

IMAGENE's network consists of approximately 70 systems. It maps those features of the communicative context deemed relevant in the corpus analysis performed in step 2 onto the appropriate lexical and grammatical forms for expressing each action. The network, having the basic high-level structure shown in Figure 4, performs the two basic functions of *Content and Rhetorical Status Selection*, and *Grammatical Form Selection*. We view Content Selection as the process of choosing the appropriate actions from the process plan to express, and Rhetorical Status Selection as the process of choosing the appropriate rhetorical relation to be used in expressing each of these actions. IMAGENE contains a sub-network implementing these two processes which is currently very preliminary (Vander Linden, 1993c).[7] This paper focuses on the Grammatical Form Selection portion of the network, that is, the choice, given an action to be expressed and its rhetorical status, of the appropriate lexical and grammatical form of expression. We will present a detailed discussion of the purpose sub-network below, which is representative of the other grammatical form sub-networks.

The input to IMAGENE is the set of features of the functional context that affect the form of expression of the plan, called the *text-level inquiries* in Figure 2. This input is implemented as a set of responses to the inquiries made by the IMAGENE system network pursuant to determining the appropriate path to be taken through the IMAGENE system

7 Paris (1988; 1993) discusses this issue in more detail, particularly as it pertains to user modeling.

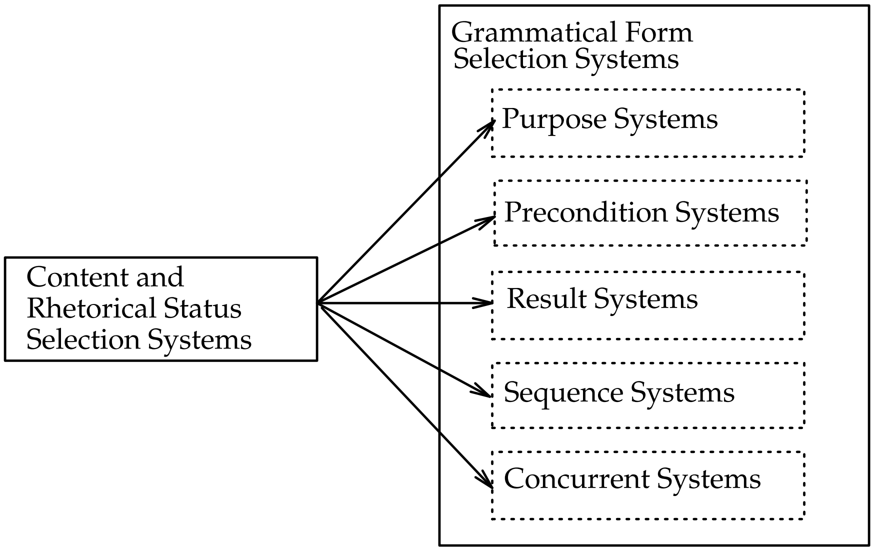

Figure 4
A High-Level View of the Systems in the Network.

network. They are analogous to Penman's sentence-level inquiries. Currently, the data structures and code necessary to respond to the inquiries automatically have not been implemented. Rather, the inquiries are answered manually, allowing us to focus on determining the appropriate set of inquiries and the precise lexical and grammatical consequences of the responses of these inquiries. The dashed lines in figure 2 indicate some of the information sources that the inquiry implementations will access (i.e., the Process Structure, the Penman lexicon and grammar, and the evolving text structure). As a side-effect of traversing the network, IMAGENE's realization statements automatically realize these consequences in a text structure, also shown in Figure 2. The Text Structure, to be discussed more fully in Section 5.3, is represented in IMAGENE's Text Representation Language (TRL). TRL itself is implemented in LOOM (MacGregor and Bates, 1987; Loom, 1993).

A second input shown in Figure 2, the *Process Structure*, is a representation of the process to be expressed. It is built in IMAGENE's Process Representation Language (PRL), which is also implemented in LOOM, and will also be discussed in Section 5.3. It is a representation like that produced by a procedural planner, containing the procedural hierarchy of the process being expressed as well as some information about the lexical items used to express each action and its arguments. It is currently built by hand which allows us to focus on the problem of expression rather than on planning.

As mentioned above, the Process Structure will eventually become a fundamental source of procedural information for the text inquiries. Currently, however, it is simply used by the final component of the architecture, the Sentence Builder, to specify the appropriate lexical items and case structures for the actions input to the text-level inquiries. The Sentence Builder uses the lexical information given in the Process Structure just described, to translate the Text Structure, described above, into the appropriate sentence specification to be passed to Penman for surface realization. This specification is coded in terms of Penman's Sentence Planning Language (SPL) (Kasper, 1989), a language which allows the specification the lexical items and grammatical structures to be generated by Penman. The translation process is performed by a recursive descent of the Text Structure hierarchy. One SPL command is produced for each sentence in the Text Structure.

5.2 Purpose Relations in Instructional Text

Purpose expressions arise in the context where actions are viewed as being related hierarchically, that is, where one higher level action is realized by the execution of a set of lower level actions.[8] As section 4 indicated, there are a large number of lexical and grammatical forms in which such procedural relationships are typically expressed, each used in a particular functional context. This section discusses the systems that have been included in IMAGENE to distinguish among these contexts.

There are other studies of purpose expressions from the point of view of representation and understanding which are of use here (Di Eugenio, 1992; Balkanski, 1992). Di Eugenio, for example, has worked with *by* purposes and *to* infinitive (TNF) purposes in the context of understanding, but doesn't appear to have distinguished the two forms in her analysis of the procedural relationships between actions. The current study is critically interested in discerning principled reasons for choosing between these sorts of expressions.

The issues of slot and form of purpose expressions are treated largely independently by IMAGENE. The slot is determined by the sub-network shown in Figure 5. The form is determined by the sub-networks shown later in Figures 7 and 8. This portion of the system network is capable of generating a greater range of purpose expressions than is typical in generation systems and of identifying the functional reasons for choosing one form over the other. It should be noted that while the determinations made by the systems are based solely on the results of the corpus analysis conducted in step 2, the following sections will include intuitive motivations for the realizations the systems make.

5.2.1 Purpose Slot. The slot determination portion of the purpose sub-network, shown in Figure 5, formalizes Thompson's notion of the "vastly different functions" for initial and final purpose clauses (Thompson, 1985) in the context of instructional text. It typically places the purpose expression in the final position. The exceptions to this are when the scope of the purpose is global, the purpose is considered optional, or the purpose is considered contrastive. These three exceptions are handled by the three systems depicted in the figure.[9]

The first exception, handled by the **Scope** system, concerns the number of actions the purpose pertains to. This correspondence between global purposes and fronted purpose expressions was already discussed in the corpus analysis section, but to give an intuitive feel for this empirical result, consider the awkwardness of restating (3a) as "?? Hold down FLASH [6] for about two seconds, then release it *to end a previous call*."[10] The restatement seems to imply, incorrectly, that the purpose applies to the last action alone rather than to the sequence of actions. The fronted form, in Example (3a), makes

8 Goldman has termed these hierarchical relationships *generation* relations (Goldman, 1970). A detailed discussion of them can be found elsewhere (Di Eugenio, 1993; Balkanski, 1993).

9 In the system network notation, vertical lines indicate decision points. The bold faced names are systems, the normal font names are features, and the italicized names are realization statements. The ordering realization statements are denoted with the operators $>$ and $|$, meaning order the clauses in the same sentence, and order the clauses in separate sentences, respectively. More detail on this notation can be found elsewhere (Winograd, 1983, chapter 6).

10 The "??" notation is used to denote a possible form of expression that is not typically found in our corpus; it does not indicate ungrammaticality.

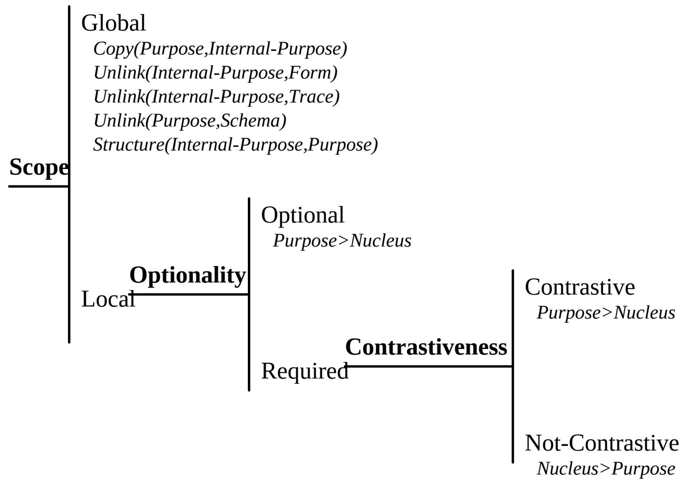

Figure 5
The Purpose Slot Selection Network

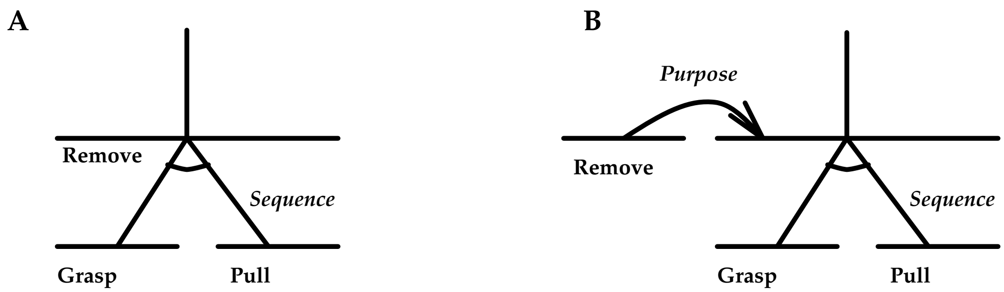

Figure 6
A Structural View of Purpose Demotion

no such implication.

As can be seen in Figure 5, the Global feature of the **Scope** system contains five realization statements, all making changes to the evolving text structure. In the Remove-Phone text, for example, these statements restructure or *demote* the **Remove** action in the hierarchical structure shown in Figure 6A, into a satellite node in the RST structure shown in Figure 6B. They do not, however, actually contain the realization statement to set the textual order of the purpose expression (which would read *Purpose>Nucleus*); later systems in this branch of the decision network execute this statement.

The remaining exceptions occur when the purpose is considered optional or contrastive and are handled by **Optionality** and **Contrastiveness** respectively. Here are examples of them from the corpus:

(4) *For more information and wall installation instructions*, see the Installation Notes on page 3. (Code-a-phone, 1989)

(5) To place call, dial AREA CODE and NUMBER. *To end call*, press red HANG UP button. (Airfone, 1991)

In (4), the action of getting more information is optional, that is, the reader may or may not want more information at this point in the text.[11] The purpose expression is, therefore, stated first to set the appropriate context for interpreting the prescribed sub-action. In (5), the purpose of ending a call is stated in contrast to placing a call in the previous sentence. It is thus fronted to set the appropriate context for the prescribed action. This fronting of contrastive purposes occurred in our corpus in the context of three oppositional semantic situations: (1) initiating/ending; (2) allowing/preventing; and (3) activating/deactivating.

The results of this study predict a number of cases where purposes should not be fronted, which is in contrast to the general claim made by Dixon (1987). He claimed that purposes should always be fronted because this facilitates the top-down construction of a procedural plan by readers as they progress through the text. Our results show cases where this rule is not followed by technical writers, i.e., when the purpose is neither global, optional, nor contrastive.

5.2.2 Purpose Form. The form selection sub-network, shown in Figures 7[12] and 8, determines the grammatical form of purpose expressions. The first element of the form selection sub-network is **Conditional-Status** which determines if the high-level purpose being expressed has special conditions pertaining to it, such as the expressed precondition in Example (6a) or other conditions that restrict the applicability of the purpose as in Example (7a) ("to wall unit *from which it was taken*"). If so, either a *by* purpose or an adjoined purpose expression are used, depending upon the complexity of the resulting sentence as determined by **Sentence-Complexity**. The slot of these forms is always initial and is determined here, rather than in the slot selection sub-network just discussed. As was the case in our corpus, IMAGENE expresses purposes that involve five or more propositions using the adjoined form, and otherwise with the *by* form. Consider the following examples of these situations:

(6a) When instructed (approx. 10 sec.) *remove phone by* firmly grasping top of handset and pulling out. (Airfone, 1991)
(6b) ?? *To remove phone* when instructed (approx. 10 sec.), firmly grasp top of handset and pull out.
(7a) Return handset to wall unit from which it was taken. Insert heel first as shown, then push top in firmly. (Airfone, 1991)
(7b) ?? Return handset to wall unit from which it was taken by inserting heel first as shown, then pushing top in firmly.

In example (6a), there is a precondition on the high-level purpose of removing the phone, a feature which correlates well with the use of the *by* form. Example (6b) seems to make the incorrect implication that the prescribed actions only work "when instructed." In the second example, the *by* form would similarly be prescribed by IMAGENE (because of the condition that the handset be returned "to wall unit *from which it was taken*"), but the number of propositions in the resulting sentence, (7b), appears to be too great (return, taken, insert, shown, push), forcing the use of the adjoined form, (7a). The adjoined purpose form is an example of a case where the rhetorical structure of a text need not

11 The distinction between conditions and optional purposes is under the purview of rhetorical status selection and is yet to be addressed.

12 Curly braces indicate that all sub-networks on the right should be entered. Square brackets indicate that all inputs must be true before entering the system on the right. The fact that the **Global-Purpose** feature is required for entry to the **Purpose-TNF** system, as well as the input conditions represented normally in the figure, is indicated with an arrow pointing to the additional input conditions. These determinations are made by the **Scope** system which is not repeated here.

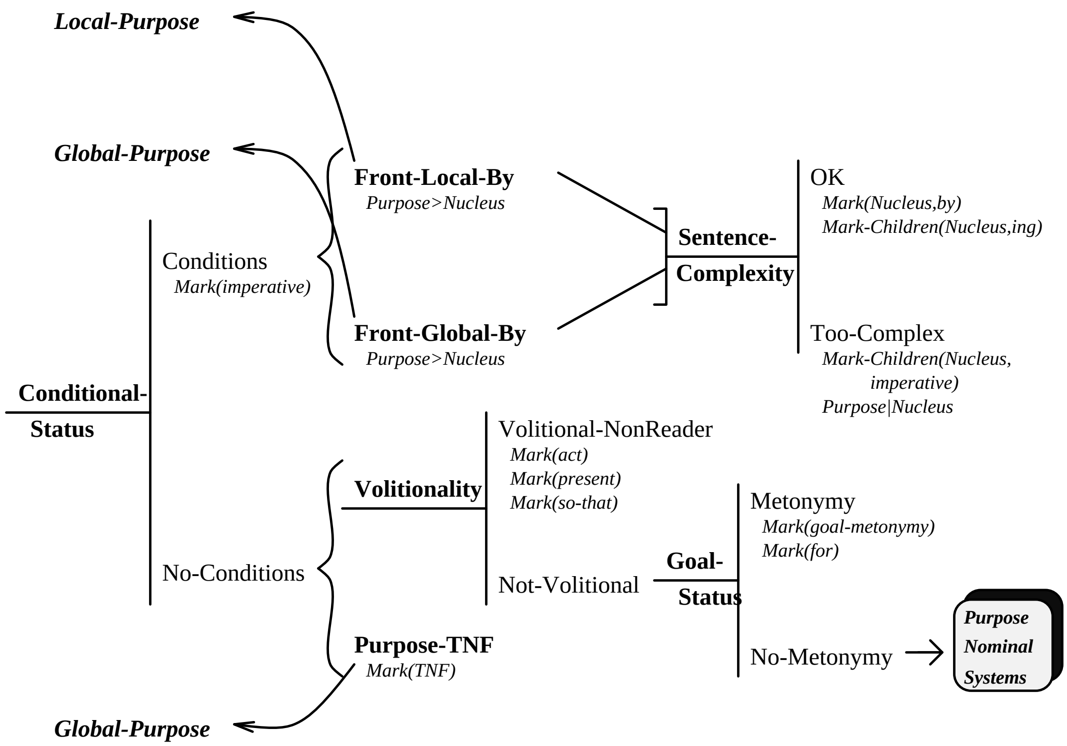

Figure 7
High-Level Systems for the Purpose Form System Network

be explicitly signaled with a lexical or grammatical cue (except textual order), called an “inferred connective” by Crothers (Crothers, 1979). RST allows the representation of this situation because its relations are not defined in terms of lexical and grammatical forms (Mann and Thompson, 1987).

When a purpose does not have conditions upon it and the scope is global, **Purpose-TNF** marks the purpose as a *to* infinitive (TNF). Example (3a) illustrated this. These sorts of context-setting purposes are not demoted to phrase status. This reflects that fact that global purposes are not expressed in phrasal form in our corpus.

The **Volitionality** system determines if the purpose expresses the desire of the reader to get some inanimate substance to perform in some volitional way. This context usually leads to the use of the *so that* purpose as shown in example (3g). Quite often these substances are liquids, but may also include other inanimates. What distinguishes liquids appears to be their ability to drip or drain over a period of time. Consider the following alternate forms for expressing purpose:

(8a) Sit the person up leaning slightly forward *so that blood and saliva can drain from his mouth.* (Rosenberg, 1985)
(8b) ?? Sit the person up leaning slightly forward in order to allow blood and saliva to drain from his mouth.

The form in example (8a) is more commonly used in our corpus in this context.

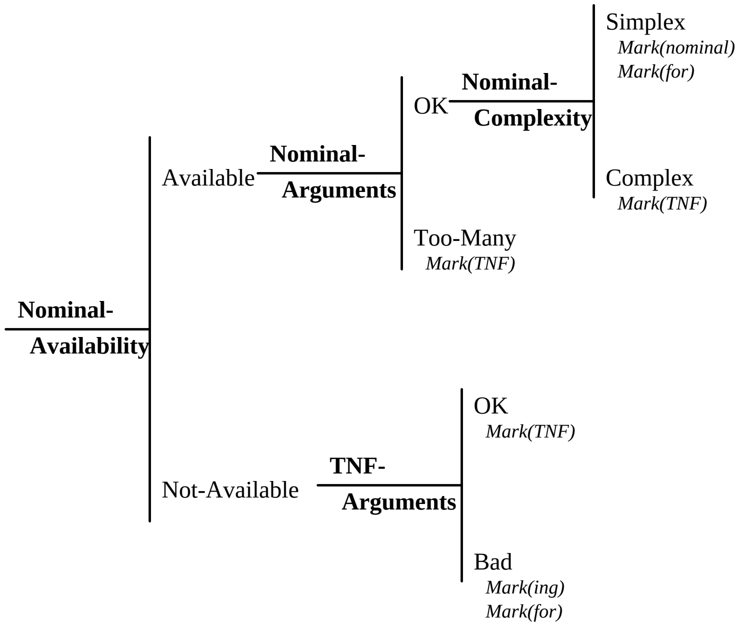

Figure 8
Nominalization Systems of the Purpose Form System Network

Goal-Status determines if the use of Goal Metonymy is warranted. The term *Goal* is used here as a case relation, corresponding to what is also called *theme* (Allen, 1987). This metonymy occurs in purposes where the direct object (or goal) of the purpose clause is more important than the action, as in:

(9) *For frequently busy numbers*, you'll want to use REDIAL [7], and the pause will have to be in Redial memory. (Code-a-phone, 1989)

The corpus study revealed that situations where the full purpose would be something like "to handle frequently busy numbers" or "for dealing with frequently busy numbers," tend to be expressed using this sort of ellipsis. The goal of the verb, in this case the busy numbers, metonymically refers to the action as a whole.

The remainder of the form selection sub-network, shown in Figure 8, is capable of generating three discrete points along the continuum from fully nominal to fully verbal forms (Quirk et al., 1985), namely the nominalization, the gerund, and *to* infinitive. These are the forms that were present in our corpus. **Nominal-Availability** will realize a prepositional phrase with a nominalization as the complement whenever the appropriate nominalization exists, as in example (9a).[13]

13 This analysis of nominalizations is an example of the descriptive nature of the current study of instructional text. The descriptive observation has been made that when nominalized forms of a verb exist in the lexicon, they tend to be used. A full explanatory account, in the spirit of current Discourse-Functional studies (e.g., Matthiessen and Thompson, 1987; Thompson, 1987), would attempt to identify the precise aspect of the action or the context of its expression which would dictate the use of a nominalization, thus resulting in the development of a nominalized form in the English Language. Such an account is beyond the scope of the current study.

(9a) Follow the steps in the illustration below, *for desk installation*. (Code-a-phone, 1989)

This use of phrases with nominalizations as propositional units is common in instructional text as well as in academic text (Cumming, 1991) and formal text in general (Hovy, 1987). IMAGENE's architecture implements a particular interpretation of Cumming's proposal (1991) that nominalizations be dealt with at two levels, one at which the actions are not specified for nominal or clausal expression, and another where they are. IMAGENE's Process Structure can be seen as the former level, its Text Structure as the latter.

Even if a nominalization exists, however, it still may not be used depending upon the determination of **Nominal-Arguments** and **Nominal-Complexity**. These systems, based on the examples in our corpus, restrict nominalizations to single, non-complex arguments. Consider the following examples:

(10a) Use the VOL LO/HI [2] switch *to adjust volume to your preferred listening level*. (Code-a-phone, 1989)
(10b) ?? Use the VOL LO/HI [2] switch *for volume adjustment to your preferred listening level*.
(11a) FLASH uses proper timing *to avoid an accidental hangup*. (Code-a-phone, 1989)
(11b) ?? FLASH uses proper timing *for accidental hangup avoidance*.

In cases (10a) and (11a), taken from our corpus, there were nominalizations available, namely "adjustment" and "avoidance," but neither was used. The "adjustment" nominalization in (10b) was apparently not used because it required more than one argument. The "avoidance" nominalization in (11b) appears to have been rejected because the argument "accidental hangup" was itself a nominalization and thus too complex. In both cases, the *to* infinitive form was preferred.

If no nominalization is available, **TNF-Arguments** will produce the *to* infinitive (TNF), unless the infinitive form requires the expression of redundant arguments. Here is an example of this case:

(12a) The BATT LOW Light [9] comes ON when the battery is weak. The handset must be returned to the base *for recharging*. (Code-a-phone, 1989)
(12b) ?? The BATT LOW Light [9] comes ON when the battery is weak. The handset must be returned to the base *to recharge (the battery?)*.

Examples similar to (12a) were found in the corpus, while those similar to the alternate *to* infinitive expression, (12b), were not.

5.3 The Remove-Phone Example

As an example of the data structures used by IMAGENE, consider the PRL representation of the actions from the Remove-Phone text, depicted graphically in Figure 9[14]. Note that this structure is currently built by hand. It is assumed that it could be constructed using AI planning methodologies. Note also that the PRL structure, in the various slots for each node, specifies the Penman lexical entries for most of the lexical choice issues, thus allowing IMAGENE to concentrate on expressing procedural relations.

14 **Return-Action** is a child of **Place-Action** because we have viewed it as the first of the sub-actions of "placing a call." In cases such as this one, the procedural distinction between child and sibling actions is a tricky one (cf. Di Eugenio, 1993). We have routinely classified actions expressed with *to* infinitive constructions as parent nodes rather than as sibling nodes. We leave a more complete treatment of this distinction to future work.

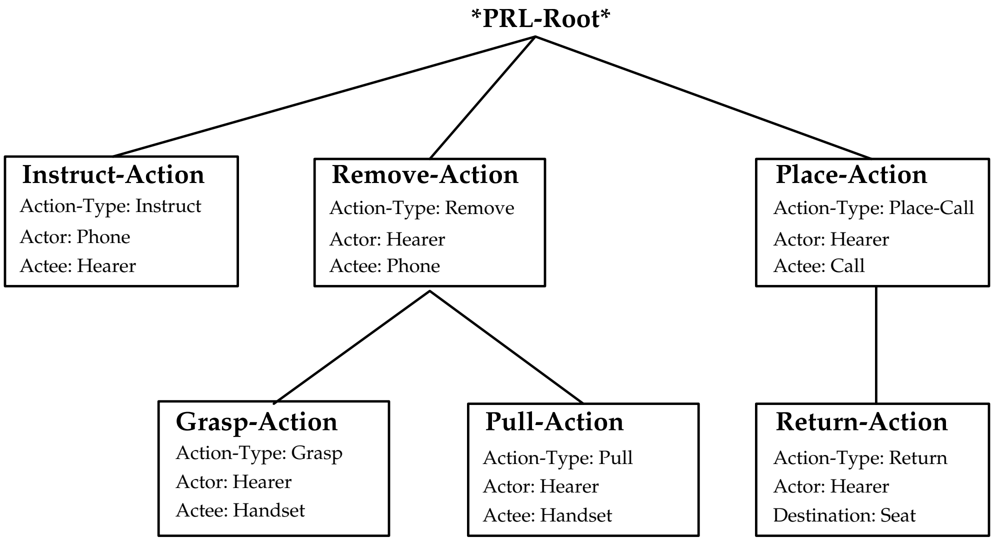

Figure 9
The Process Structure for the Remove-Phone Text

Given this structural representation of a sequence of actions, the Content and Rhetorical Status Selection system sub-network can be viewed as using the inquiry responses to produce the TRL structure shown in Figure 10.[15] Again, this process is not the subject of this paper, but is was mentioned to provide a more complete discussion of the data structures involved.

The Grammatical Form Selection sub-networks can then be seen as operating on the appropriate relations included in this representation, and producing the full TRL structure shown in Figure 11. TRL allows the Text Structure to include a representation of the hierarchical structure of the text in terms of RST, including both nucleus-satellite and multi-nuclear schemas. In addition, TRL specifies the textual order and clause combining using additional *New-Sentence* and *Continue-Sentence* links. For example, the **Instruct**, **Grasp**, **Remove**, and **Pull** nodes are all combined into one sentence in Figure 11. Finally, TRL specifies the grammatical form of each action expression using three features which may be attached to expressible nodes in the structure. The **Form** feature specifies the general grammatical form. For example, the **Instruct** is marked as Passive, indicating that the agentless passive should be used. The **Linker** and **Tense** markers are also used to mark the appropriate linker and tense of the expression.

The Sentence Builder then uses a straight-forward recursive descent algorithm to produce an SPL command for each of the sentences in the TRL structure. The generated text for this example is shown here:

(13) *When you are instructed, remove the phone by grasping the top of the handset and pulling it. Return to a seat to place a call.*

15 Because the execution of the Content and Rhetorical Status Selection sub-network is interleaved with the execution of the Grammatical Form Selection sub-networks, this structure alone would never exist at any point in the execution of the network. It is, rather, an illustrative view of what the Content and Rhetorical Status Selection sub-network would realize if it were executed in isolation.

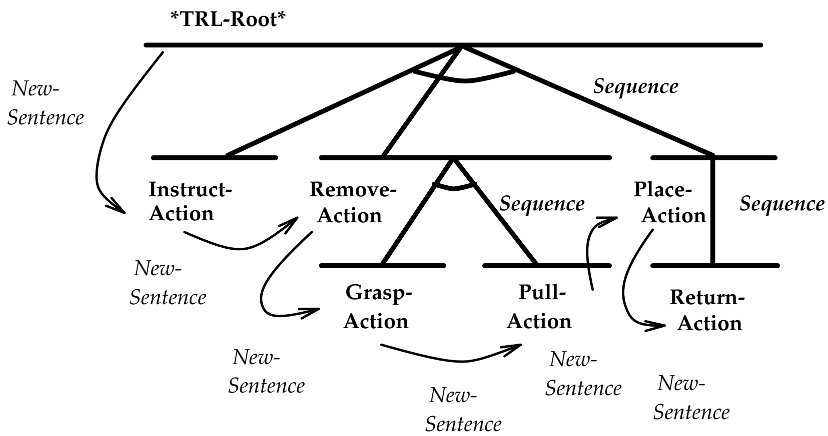

Figure 10
A Hypothetical View of the Output of the Content and Rhetorical Status Selection Sub-System for the Remove-Phone Text

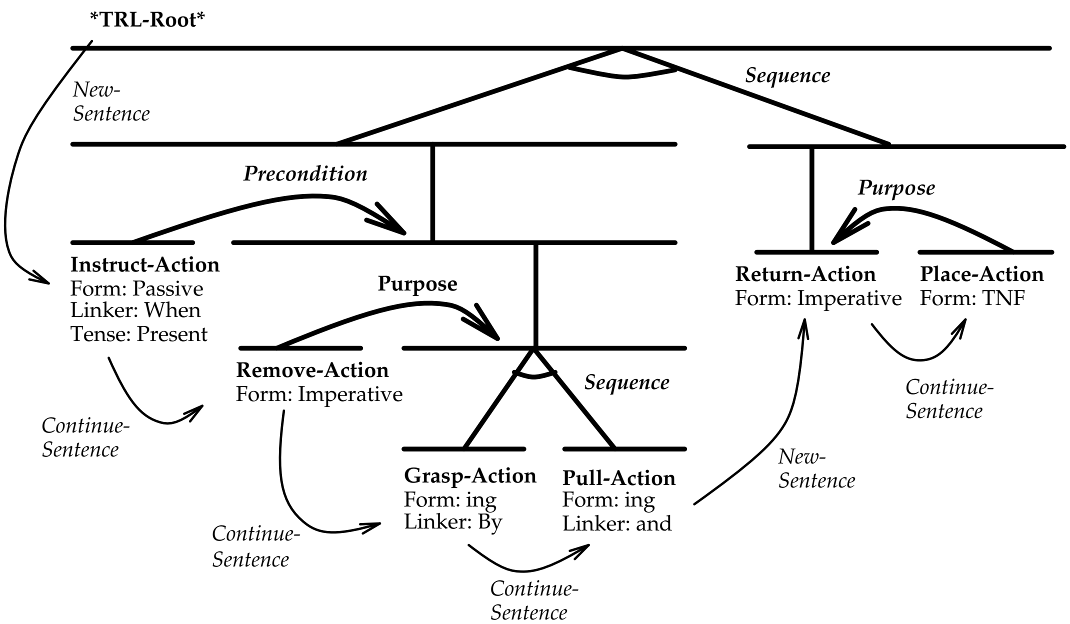

Figure 11
The Final Text Structure for the Remove-Phone Text

This text is identical to the original text with respect to the four lexical and grammatical issues addressed here. There are, however, a number of other lexical and phrasal differences, including the lexical items chosen for the object references and the use of determiners. These differences arise from the fact that the current study has not specifically addressed the issue of referring expressions. Currently, IMAGENE uses simple algorithms for pronominalization and the use of determiners, which are not based on a detailed corpus study of the forms and functions of the object reference domain. A study of referring expressions, similar to our work on expressing rhetorical relations, would allow the development of a more principled solution to this problem.

5.4 More Examples of IMAGENE's Output

This section includes examples of IMAGENE's output for the fundamental relations dealt with in the current study, i.e., purpose, precondition, result, and action sequence. It is intended to demonstrate IMAGENE's breadth of coverage and will not discuss the details of how the forms are motivated.

Given the choice to express an action, rhetorically, as a purpose, IMAGENE is capable of producing seven grammatical forms for its expression, most of which can either be fronted or not fronted. Here are the various forms, as generated by IMAGENE according to the distinctions discussed in the previous section:

(14a) *To end a call, hold down the FLASH button for two seconds, then release it.*
(14b) *Follow steps in the illustration for desk installation.*
(14c) *Use the OFF position for charging the batteries.*
(14d) *Use the REDIAL for frequently busy numbers.*
(14e) *When you are instructed, remove the phone by grasping the top of the handset and pulling it.*
(14f) *Remove the phone. Grasp the top of the handset, and pull it.*
(14g) *Tilt the pan so that the fluid drains out.*

Given the choice to express an action, rhetorically, as a precondition, IMAGENE is capable of producing four grammatical forms for its expression, all of which can either be fronted or not fronted and also linked with various lexical items. Here are some representative forms, as generated by IMAGENE:

(15a) *If light flashes, insert credit card.*
(15b) *The BATTERY LOW INDICATOR will light when the battery is low.*
(15c) *When the phone is installed, and the battery is charged, move the OFF/STBY/TALK switch to the STBY position.*
(15d) *Return the OFF/STBY/TALK switch to the STBY position after your call.*

There are two types of results that IMAGENE supports. The first type is non-reader actions that are not the result of an explicit command to monitor a particular device state. IMAGENE expresses this type of result as a future tense clause (see example (16a)). The second type is not based on an action in the process structure at all, but rather, is a span added by the system networks to signal a state resulting from an expressed action. IMAGENE expresses these as present tense relational expressions see example (16b)). Here are examples of these forms:

(16a) *The BATTERY LOW INDICATOR will light when the battery is low.*
(16b) *When the phone is installed, and the battery is charged, move the OFF/STBY/TALK switch to the STBY position. The phone is now ready to use.*

Simple sequential actions do not fit into the categories discussed above, and are marked as imperative commands. These commands are combined into clauses by the sentence tools system network using *and* when the concurrency that could be implied is impossible or inconsequential, as in example (14f), or *then* when there is possible unwanted concurrency, as in example (14a).

6. Verifying IMAGENE's Prescriptions

Finally, we compare the output of the text generator with the text in the corpus. For this purpose, IMAGENE's system network was re-run for all of the approximately 600 action expressions, both those from the training set and those from the testing set. Statistics were kept on how well its realizations matched the expressions in the corpus.[16] These tests were performed without the Penman realization component engaged, comparing the TRL output of the system network with the corpus text. This way, the extensive lexicon that would have been necessary for the surface realization was not required. IMAGENE currently includes a domain model and lexical entries for cordless telephones and a few other specific examples.

The match was judged on four separate lexical and grammatical issues: Linker, Form, Slot, and Clause Combining. The resulting TRL structure had to specify the identical linker (either preposition or conjunction), form (tense, aspect, mood, and voice, or non-finite verb or nominalization), slot (textual order), and combining (if the expression was combined with the following one). An example of this verification process can be found in Section 5.3, where the IMAGENE-produced Remove-Phone text is shown to match the original text on all four of these issues.

Note that the match must be exact. For example, if IMAGENE specifies the conjunction *and* for a sequence expression when *then* occurs in the text, the choice of linker would be counted as incorrect, in spite of the fact that the resulting text might be quite understandable. Note also that IMAGENE's realizations may even be better in some cases than the text in the corpus. Although the general philosophy of the approach taken in the current study is to assume that the choices made by the writers of the corpus are correct, there are isolated cases where the forms in the corpus are probably inappropriate. IMAGENE embodies choices which are consistently made over a range of instructions and thus does not reflect isolated examples.

The analysis conducted in step 2 has been based primarily on a small subset of the full corpus, namely on the instructions for a set of three cordless telephone manuals. This training set constitutes approximately 35% of our corpus. The results of this analysis were then implemented in IMAGENE and applied to the full corpus, providing a detailed characterization of the instructions found in the original telephone manuals, and a quantitative analysis of how well this characterization applies to the other forms of instructions. IMAGENE's realizations correctly match all four lexical and grammatical issues in 71% of the expressions in the training set, and 52% in the testing set. The specific levels of match for the four most common rhetorical relations are detailed in figures 12 and 13. There is one table for each of the major rhetorical relations, purpose, precondition, result, and sequence.[17] These tables show the percentage of IMAGENE's realizations for linker, form, slot, and clause combining that matched those in the corpus, differentiating between the training set and the testing set. As can be seen in all of the charts, the level of match is better for the training set, but still good for the testing set.

For purpose expressions, IMAGENE makes use of four different linkers (*by*, *for*, *so that*, and no linker) and six different forms (*to* infinitive, imperative, nominalization, gerund, goal metonymy, and simple present tense action) and produces a match on all

16 The training corpus included some non-procedural text which was included for a pilot study done before the focus on procedural text had been determined. It is not handled particularly well by IMAGENE and the results given in this section will not include it. The testing set is exclusively procedural and is included in full.

17 Because there are relatively few concurrent expressions in our corpus, only 33, those results are not included in this section.

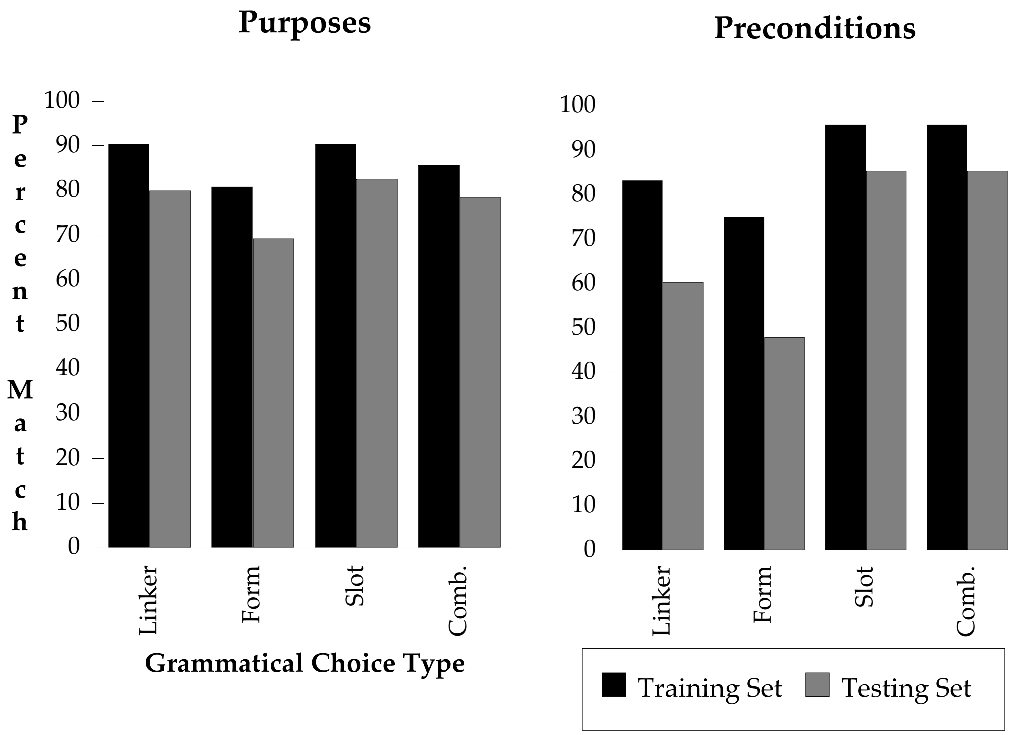

Figure 12
The Accuracy of IMAGENE's Realizations for Purpose and Precondition Expressions

four lexical and grammatical issues for 81% of the purpose expressions in the training set, and 59% in the testing set. Figure 12 gives a breakdown of IMAGENE's accuracy for the four lexical and grammatical issues. In order to more fully judge these results, consider an alternate system which always generates the single most common purpose form. In our corpus, this is the fronted *to* infinitive, which occurred in 34% of the purpose expressions.[18] Such a system would score 34% under the verification criteria used here.

For precondition expressions, the most common form in our corpus is the fronted *if* present tense clause, which occurred in 19% of the 98 precondition expressions in the corpus. IMAGENE, which produces five linkers and nine forms, produces a match for 67% of the precondition expressions in the training set, and 35% in the testing set. As can be seen in the precondition chart in Figure 12, IMAGENE's accuracy for preconditions is lower than for purposes, particularly in the testing set. This reflects the greater diversity of procedural contexts in which preconditions arise, and the corresponding diversity of the forms used to express them (cf. Vander Linden, 1994). Certainly, a larger training set is required here but it is not clear at this point how much larger it should be. IMAGENE's accuracy for results and sequence expressions is similar to that presented for purposes and preconditions. It is detailed in Figure 13.

18 Table 1 indicates 32%, but that would be for our corpus with the non-procedural portions of text included. They have been removed here to remain consistent with the statistics shown in this section.

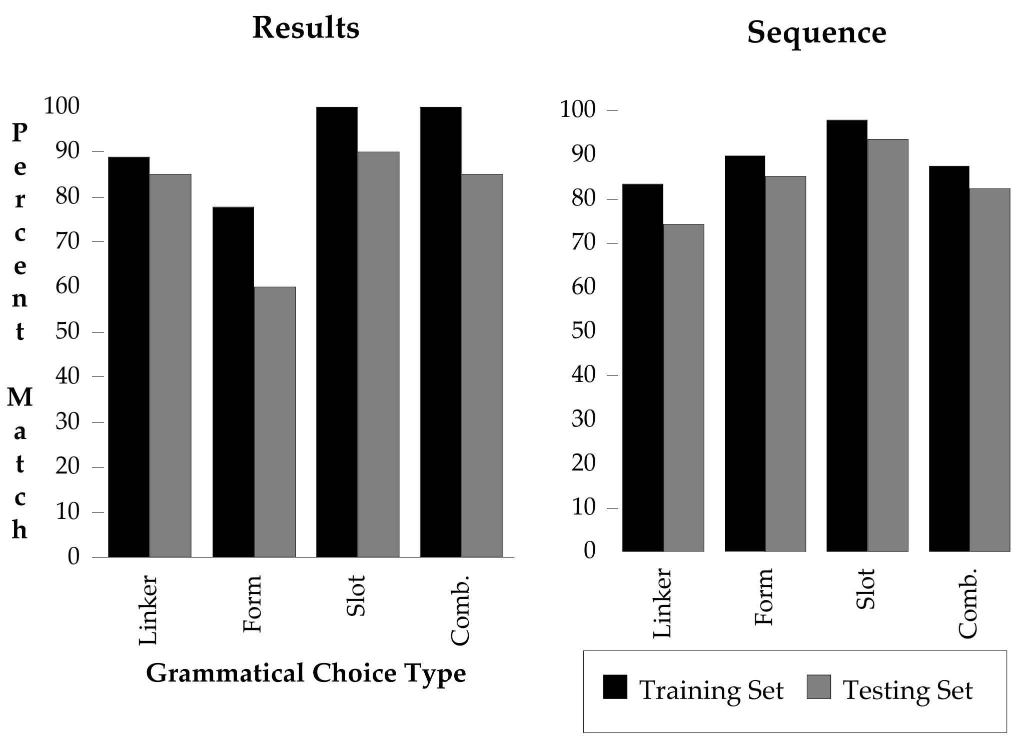

Figure 13
The Accuracy of IMAGENE's Realizations for Result and Sequence Expressions

7. Conclusions

This paper has addressed the problem of determining the precise lexical and grammatical forms for expressing procedural relations between actions in the context of instructional text generation. The corpus-based methodology employed is well suited for this problem, providing both a principled means for cataloging the lexical and grammatical forms that are consistently used in instructional text, and an environment for testing and confirming hypotheses concerning the contextual issues that co-vary with these forms.

The issues of procedural planning, user modeling, and content selection, although of unquestionable importance to the broad goal of generating instructions, were not specifically addressed here. The current study makes a number of prescriptions for the type of information that such techniques would need to provide to the text planner pursuant to the generation of instructional text, but says nothing about how they should be implemented in order to achieve this.

There are two fundamental contributions of the current study to the field of Computational Linguistics. The first is the analysis of instructional text itself. The current study has provided a characterization of certain aspects of instructional text that has been effectively applied to the generation of instructional text in general. This characterization is directly applicable to current work on instructional text, particularly in the context of natural language generation. The results can also serve as a source of preliminary hypotheses with respect to the analysis of other related genres.

The second is the detailed presentation of a methodology for managing diversity of expression at the textual level in the context of text generation. The approach involves collecting a suitable corpus of text, analyzing that text, implementing the results of the analysis in a text generator, and verifying the output of the generator. This approach is applicable, not just to the problem of expressing procedural relations in instructional text, but rather to any lexical or grammatical aspect of any linguistic genre.

Acknowledgments

This paper is based on earlier work done with Susanna Cumming whose approach to language study has inspired much of what we have done here. We also received valuable comments from the *Computational Linguistics* referees and from members of the ITRI Computational Linguistics group, including Tony Hartley, Cécile Paris, Richard Power, and Donia Scott.

References

Airfone (1991). *Inflight Entertainment & Information Guide*. United Airlines.

Allen, J. F. (1987). *Natural Language Understanding*. Benjamin/Cummings, Menlo Park, CA.

Balkanski, C. T. (1992). Actions, beliefs and intentions in rationale clauses and means clauses. In *Tenth National Conference on Artificial Intelligence,* July 12–16, San Jose, California.

Balkanski, C. T. (1993). *Actions, Beliefs and Intentions in Multi-Action Utterances*. PhD thesis, Harvard University.

Berry, M. (1981). Towards layers of exchange structure for directive exchanges. *Network*, 2:23–32.

Church, K. W. and Mercer, R. L. (1993). Introduction to the special issue on computational linguistics using large corpora. *Computational Linguistics*, 19(1):1–24.

Code-a-phone (1989). *Code-A-Phone Owner's Guide*. Code-A-Phone Corporation, P.O. Box 5678, Portland, OR 97228.

Crothers, E. (1979). *Paragraph Structure Inference*. Ablex, Norwood, NJ.

Cumming, S. (1990). Natural discourse hypothesis engine. In McKeown, K. R., Moore, J. D., and Nirenburg, S., editors, *Proceedings of the Fifth International Workshop on Natural Language Generation,* June 3–6, Dawson, PA.

Cumming, S. (1991). Nominalization in English and the organization of grammars. In *Proceedings of the IJCAI-91 Workshop on Decision Making Throughout the Generation Process,* August 24–25, Darling Harbor, Sydney, Australia.

Dale, R. (1992). *Generating Referring Expressions: Constructing Descriptions in a Domain of Objects and Processes*. MIT Press, Cambridge, MA.

Dale, R. (1993). Rhetoric and intentions in discourse. In Rambow, O., editor, *Proceedings of the Workshop on Intentionality and Structure in Discourse Relations,* June 21, Columbus, OH, pages 5–6.

Delin, J., Scott, D., and Hartley, T. (1993). Knowledge, intention, rhetoric: Levels of variation in multilingual instructions. In *Proceedings of the Workshop on Intentionality and Structure in Discourse Relations,* June 21, Columbus, OH, pages 7–10.

Di Eugenio, B. (1992). Understanding natural language instructions: The case of purpose clauses. In *Proceedings of the Annual Meeting of Association for Computational Linguistics,* Newark, DE, pages 120–127.

Di Eugenio, B. (1993). *Understanding Natural Language Instructions: A Computational Approach to Purpose Clauses*. PhD thesis, University of Pennsylvania. also available as IRCS Report 93-52.

Dixon, P. (1987). Actions and procedural directions. In Tomlin, R. S., editor, *Coherence and Grounding in Discourse*, pages 69–89. John Benjamins. Outcome of a symposium, Eugene, OR, June, 1984.

Excursion (1989). *Excursion 3100*. Northwestern Bell Phones, A USWest Company.

Fawcett, R. P. (1990). The COMMUNAL project: Two years old and going well. *Network*, 13/14:35–39.

Goldman, A. I. (1970). *A Theory of Human Action*. Prentice Hall, Englewood Cliffs, NJ.

Grosz, B. J. and Sidner, C. L. (1986). Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12(3).

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

Halliday, M. A. K. (1985). *An Introduction to Functional Grammar*. Edward Arnold, London.

Hovy, E. H. (1987). What makes language formal? In *Proceedings of the Ninth Annual Conference of the Cognitive Science Society,* Seattle, WA, pages 959–964. Poster Session.

Hovy, E. H. (1988a). *Generating Natural Language Under Pragmatic Constraints*. Lawrence Erlbaum Associates, Hillsdale, NJ.

Hovy, E. H. (1988b). Planning coherent multisentential text. In *Proceedings of the Twenty-Sixth Annual Meeting of the Association for Computational Linguistics*. State University of New York, Buffalo, NY, USA, June 7-10.

Hovy, E. H. and McCoy, K. F. (1989). Focusing your RST: A step toward generating coherent multisentential text. In *Program of the Eleventh Annual Conference of the Cognitive Science Society,* August 16–17, Ann Arbor, MI, pages 667–674. Lawrence Erlbaum Associates.

Huettner, A. K., Vaughan, M. M., and McDonald, D. D. (1987). Constraints on the generation of adjunct clauses. In *Proceedings of the 25th Annual Meeting of Association for Computational Linguistics,* Stanford, CA, pages 207–214.

Kasper, R. T. (1989). A flexible interface for linking applications to penman's sentence generator. In *Proceedings of the DARPA Speech and Natural Language Workshop* Philadelphia.

Kittredge, R., Korelsky, T., and Rambow, O. (1991). On the need for domain communication knowledge. *Computational Intelligence*, 7(4):305–314.

Loom (1993). *The LOOM Documentation*. USC Information Sciences Institute.

MacGregor, R. and Bates, R. (1987). The LOOM knowledge representation language. In *Proceedings of the Knowledge-Based Systems Workshop,* St. Louis, April 21–23. Also available as USC/ISI technical report RS-87-188.

Macintosh (1988). *Macintosh System Software User's Guide: Version 6.0*. Apple Computer, Inc.

Mann, W. C. (1985). An introduction to the Nigel text generation grammar. In Benson, J. D., Freedle, R. O., and Greaves, W. S., editors, *Systemic Perspectives on Discourse*, volume 1, pages 84–95. Ablex.

Mann, W. C. and Thompson, S. A. (1987). Rhetorical structure theory: A theory of text organization. Technical Report ISI/RS-87-190, USC/ISI.

Mann, W. C. and Thompson, S. A. (1988). Rhetorical structure theory: Toward a functional theory of text organization. *Text: An Interdisciplinary Journal for the Study of Text*, 8(2).

Matthiessen, C. M. I. M. and Thompson, S. A. (1987). The structure of discourse and 'subordination'. In Haiman, J. and Thompson, S., editors, *Clause Combining in Grammar and Discourse*, pages 275–329. Benjamins, Amsterdam.

McKeown, K. R. (1985). *Text Generation*. Cambridge University Press, New York.

McKeown, K. R., Elhadad, M., Fukumoto, Y., Lim, J., Lombardi, C., Robin, J., and Smadja, F. (1990). Natural language generation in COMET. In Dale, R., Mellish, C., and Zock, M., editors, *Current Research in Natural Language Generation*, chapter 5. Academic Press.

Mellish, C. and Evans, R. (1989). Natural language generation from plans. *Computational Linguistics*, 15(4):233–249.

Meteer, M. W. (1991). Bridging the generation gap between text planning and linguistic realization. *Computational Intelligence*, 7(4):296–304.

Meteer, M. W. (1992). *Expressibility and the problem of efficient text planning*. Pinter, London.

Mooney, D. J., Carberry, S., and McCoy, K. F. (1991). Capturing high-level structure of naturally occurring, extended explanation using bottom-up strategies. *Computational Intelligence*, 7(4):334–356.

Moore, J. D. and Paris, C. L. (1988). Constructing coherent text using rhetorical relations. In *Proceedings of the Tenth Annual Conference of the Cognitive Science Society*, August 17–19, Montreal, Quebec, pages 637–643.

Moore, J. D. and Pollack, M. E. (1992). A problem for RST: The need for multi-level discourse analysis. *Computational Linguistics*, 18(4):537–544. Squibs and Discussions.

Paris, C. (1993). *User Modelling in Text Generation*. Pinter Publishers, London.

Paris, C. L. (1988). Tailoring object descriptions to a user's level of expertise. *Computational Linguistics*, 14(3):64–78.

Patten, T. (1988). *Systemic Text Generation as Problem Solving*. Cambridge University Press.

Penman (1989). *The Penman Documentation*. USC Information Sciences Institute, Penman Natural Language Group.

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1:81–106.

Quirk, R., Greenbaum, S., Leech, G., and Svartvik, J. (1985). *A Comprehensive Grammar of the English Language*. Longman.

Reader's Digest (1981). *Reader's Digest Complete Car Care Manual*. The Reader's Digest Association, Inc., Pleasantville, NY.

Rosenberg, S. N. (1985). *The Johnson and Johnson First Aid Book*, pages 17,95,101,129. Johnson and Johnson.

Rösner, D. and Stede, M. (1992). Customizing RST for the automatic production of technical manuals. In Dale, R., Hovy, E., Röesner, D., and Stock, O., editors, *Aspects of Automated Natural Language Generation*, Lecture Notes in Artificial Intelligence 587, pages 199–214. Springer Verlag, Berlin.

Rösner, D. and Stede, M. (1992). TECHDOC: A system for the automatic production of multilingual technical documents. In *Proceedings of KONVENS-92*, Berlin. Springer. Also available as technical report FAW-TR-92021, FAW, Ulm, Germany.

Scott, D. R. and Souza, C. (1990). Getting the message across in RST-based text generation. In Dale, R., Mellish, C., and Zock, M., editors, *Current Research in Natural Language Generation*, chapter 3. Academic Press.

Tate, A. (1976). *Project Planning Using a Hierarchical Non-Linear Planner*. PhD thesis, University of Edinburgh. Dissertation Abstracts International Report 25.

Thompson, S. A. (1985). Grammar and written discourse: Initial and final purpose clauses in English. *Text*, 5(1–2):55–84.

Thompson, S. A. (1987). "Subordination" and narrative event structure. In Tomlin, R. S., editor, *Coherence and Grounding in Discourse*, pages 435–454. John Benjamins Publishing Co. Outcome of a Symposium, Eugene, Oregon, June, 1984, and published as vol. 11 of the series Typological Studies in Language.

Vander Linden, K. (1993a). Generating effective instructions. In *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society*, June 18–21, Boulder, CO, pages 1023–1028.

Vander Linden, K. (1993b). Rhetorical relations in instructional text generation. In Rambow, O., editor, *Proceedings of the Workshop on Intentionality and Structure in Discourse Relations*, June 21, Columbus, OH, pages 140–143.

Vander Linden, K. (1993c). *Speaking of Actions: Choosing Rhetorical Status and Grammatical Form in Instructional Text Generation*. PhD thesis, University of Colorado. Available as Technical Report CU-CS-654-93.

Vander Linden, K. (1994). Generating precondition expressions in instructional text. In *Proceedings of the 32nd Annual Meeting of Association for Computational Linguistics*, June 27–30, Las Cruces, NM, pages 42–49.

Winograd, T. (1983). *Language as a Cognitive Process, Volume I: Syntax*. Addison-Wesley, Reading, MA.