Generating Effective Instructions

Keith Vander Linden
Department of Computer Science
Institute of Cognitive Science
University of Colorado
Boulder, CO 80309-0430
linden@cs.colorado.edu

Abstract

This paper discusses a corpus-based approach to the generation of effective instructions. The approach advocated employs a detailed linguistic study of a corpus of a broad range of instructional texts to determine both the range of grammatical forms used in instructional text and the contexts in which they are used. The forms that are consistently used by technical writers are taken to be the most effective. The results of this study are implemented in an automated text generation system for instructional text. The primary focus of this study has been the use of rhetorical relations to effectively code actions and their procedural relationships in instructional text, but the approach can generally be applied to different linguistic issues and text genres.

Introduction

Text which prescribes the execution of actions to its reader, called *instructional text*, is a common part of everyday life. Because we repeatedly find ourselves dealing with it, the automation of its generation is greatly desirable, both from the point of view of the manufacturer who must produce the text and also the consumer who must read it. This automation, however, requires a deep knowledge of the rhetorical and grammatical forms of expression that most effectively convey actions and the procedural relationships among them. This paper discusses an approach for acquiring and implementing this knowledge.

What makes this task difficult is that the job of instructional text itself is extremely difficult. It must express, in the largely linear medium of text, an inherently non-linear process. Not surprisingly, we run into examples of text that we feel are poorly written. Consider, for example, the following excerpt from a instruction manual for a common household furnace:

(1) Depress knob and hold for 60 seconds after pilot has been lighted. Release knob and turn to ON position.

The point of interest here is the procedural status of the action of lighting of the pilot. Is the reader expected to manually light the pilot or not? Close inspection of the entire text reveals that the reader is indeed expected to manually light the pilot, making it clear that the rhetorical status and grammatical form used in the expression of the lighting action fail to effectively convey the procedural relationships involved.

How, then, are we to determine the most effective rhetorical and grammatical forms for expressing instructions. Although instructions themselves have a reputation as being a bad source for principles of effective expression, the author contends that a corpus-based study of a broad range of instructional texts, as produced by trained technical writers, will uncover the most effective forms of expression and also the precise contexts in which they should be used. The primary focus of this study has been the use of rhetorical relations to effectively code actions and their procedural relationships in instructional text, but the approach can generally be applied to different linguistic issues and text genres.

This paper will discuss the approaches taken to this problem in the fields of Artificial Intelligence, Psychology, and Linguistics, respectively, and then detail the approach advocated here. It will conclude with a discussion of IMAGENE, an instructional text generation system which embodies the results of this approach.

Other Approaches

Instructional text has been the focus of a number of studies and implementations in Cognitive Science,

and although some of this work has addressed other "instructional" texts such as complex object descriptions and textbook text rather than raw procedural text, they illustrate very well the general methodologies adopted in the sub-disciplines of Cognitive Science. All of these methodologies have their place certainly, but have failed to address the issues that are critical for the success of an automated text generation project.

Artificial Intelligence

In Artificial Intelligence, there are two bodies of work that are of interest, the work on understanding instructions, and the work on generating them. The understanding research, as is typical of understanding research in general, has paid a fair amount of attention to the expressional form of instructions; they must, after all, be capable of parsing the forms used in instructions. This concern has lead the researchers to catalogue the various forms found in instructional text corpora, such as those used for purposes (Di Eugenio, 1992), repetitions (Rock, 1992), and free adjuncts of various types (Webber and Di Eugenio, 1990). Although the results of this work have aided the IMAGENE project, their primary concern has been to determine the nature of the procedural relationships and how they are to be represented. The effect of the issues of textual and interpersonal context on expressional form are largely ignored. The furnace example, as will be seen later, cannot be explained without reference to precisely these issues.

One would expect the choice of expressional form to be the fundamental concern of the work on generating instructions, but this is not always the case. Instructional text generation systems (Dale, 1990; McKeown *et al.*, 1990; Mellish and Evans, 1989), dealing with other issues, have tended to hard code choices at the rhetorical level. This approach is at variance with the practice in actual instructional manuals, where the actions being specified are systematically expressed within any one of a number of rhetorical relations, depending upon the functional context, and, similarly, each rhetorical relation is expressed in any one of a number of grammatical forms, again, depending upon context.

Psychology

Psychological approaches typically involve experimental testing of pre-specified expressional forms. The general approach is to design an experimental setting where a set of instructional forms can be tested for rapidity understanding and accuracy of execution. Dixon (1988), for example, tested the relative efficacy of commands like:

(2a) Turn the left knob to set the top meter to 20.

(2b) To set the top meter to 20 turn the left knob.

The results indicated that example (2a) was better, but said nothing of the large number of other forms of expression that are commonly used in instructions to express this type of information, such as:

(3a) Set the top meter to 20 by turning the left knob.

(3b) For setting the top meter to 20, turn the left knob

(3c) Turn the left knob so that the top meter is set to 20.

(3d) Turn the left knob until the top meter is set to 20.

The analysis for the current study indicates that the form in example (3d) is most commonly used in contexts that could include this sentence. These psychological methods are well suited to test the relative efficacy of a set of expressional forms, but do not provide an effective means for determining the appropriate forms to test. What forms, for instance, should be tested for the furnace example?[1] Other psychological studies of more general instructional material (e.g., Britton and Gülgöz, 1991), using general reading comprehension models, largely ignore the grammatical form of the expressions altogether.

Linguistics

There has simply not been much work in Linguistics on instructional text. Most studies have focussed on narrative and expository text. Some of this work, particularly the work within Functional and Cognitive Linguistics, is still of considerable interest for this paper. These schools have attempted to specify the elements of the functional and cognitive context that are relevant in a writer's choice of one expressional form over another (e.g., Thompson, 1985; Thompson, 1987; Matthiessen and Thompson, 1987). This work, however, dealing with other genres, cannot always be directly applied to instructional text.

[1] The author came up with a list of alternative forms early in the IMAGENE project, none of which turned out to be commonly employed in the corpus, leading one to suspect our ability to invent relevant forms to test.

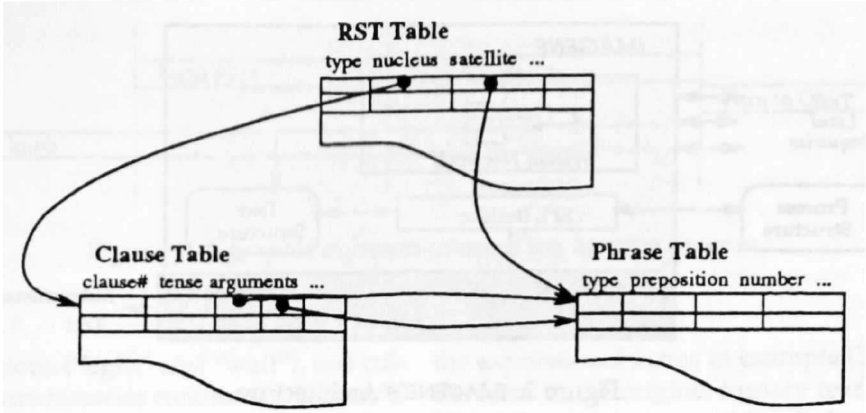

Figure 1: The Structure of the Database

The IMAGENE Approach

This paper now details the approach taken in the IMAGENE project to identify the relevant factors that weigh on the choice of rhetorical status and expressional form for various actions and their procedural relationships in instructional text and how these factors have been coded in IMAGENE.

Corpus Analysis

Because our intuitions concerning rhetorical status and grammatical form are not always accurate, we must refer to a corpus of real text to determine both the range of the rhetorical and grammatical forms that are used and the contexts in which they occur. The corpus currently consists of approximately 1000 clauses (6000 words) of instructional text taken from various sources including manuals for consumer electronic devices and auto-repair, craft instructions, recipes, and first-aid books. This corpus was loaded into a relational database for text, the structure of which is depicted in figure 1. Each occurrence of a set of rhetorical relations similar to those specified in Mann and Thompson's Rhetorical Structure Theory (Mann and Thompson, 1989) was then marked, along with a detailed representation of the grammatical form of its expression.

Determining the range of the lexical and grammatical forms used in the corpus for each rhetorical relation type simply involved querying all occurrences of that relation, and then noting the forms that were used. The process of distinguishing between these forms in the generation system involved forming and testing hypotheses concerning the co-occurrence of functional features and lexical and grammatical forms (similar to the methodology advocated in Cumming, 1990). For example, given the results of Thompson's study of purpose clauses (1985), one might hypothesize that all global purpose clauses are fronted. A simple database query shows that 90% of the 30 global purpose expressions in the database are fronted, providing good evidence in support of the hypothesized co-occurrence. The structural information that supported this query, as well as the other functional information that was used in other parts of the analysis had to be hand-coded in the database to be used.

This analysis has resulted in broad generalizations concerning the way actions are expressed in instructional text and has tended to be insensitive to isolated examples of ineffective text, such as the furnace example. The results specific to the furnace example are the focus of this paper and will be discussed below.

IMAGENE

The results of the corpus analysis just reported have been implemented in IMAGENE, an instructional text generation system whose architecture is shown in figure 2. IMAGENE takes as input a non-linguistic representation of the actions to be expressed called the *process structure*[2] and produces a *text structure* that specifies the rhetorical and grammatical form of the text. It uses a single system network of the form developed in Systemic-Functional Linguistics (Halliday, 1976) and implemented in the Penman text generation system (Mann, 1985) to create and manip-

[2]This representation is like that produced by a procedural planner, but is currently built by hand. This has allowed the current study to focus on the problem of expression rather than planning.

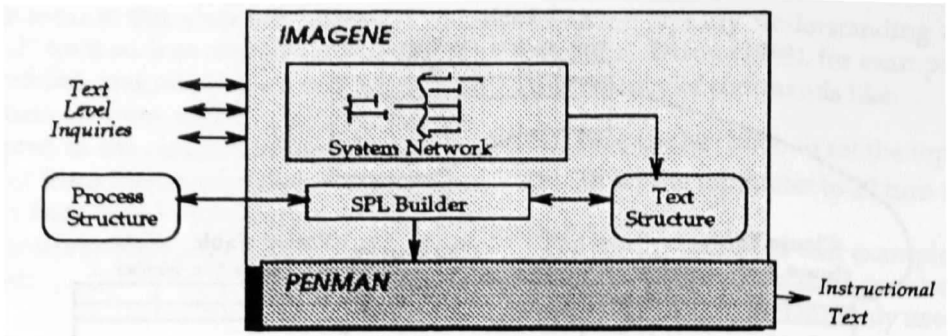

Figure 2: IMAGENE's Architecture

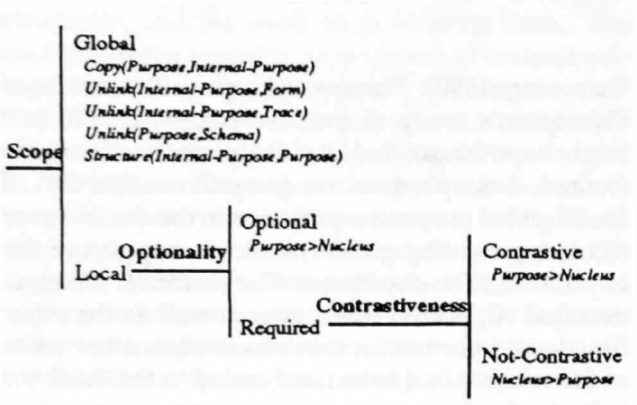

Figure 3: IMAGENE's Purpose Slot Systems

ulate the text structure. This network currently contains 67 systems (or decision points). The fragment of the system network dedicated to purpose slot determination is shown in figure 3. This network is to be read as a discrimination network inquiring about three conditions, scope, optionality, or contrastivity. The purpose is fronted by a *Purpose>Nucleus* realization statement if it is global, optional, or contrastive. The surface realization of the text is performed by Penman.

The system networks base their manipulation on the responses to a set of *text-level inquiries*, analogous to Penman's sentence-level inquiries.[3] These inquiries help to determine which of the paths through the network are to be taken. In figure 3, for example if the inquiries determine that the purpose is global the *Copy*, *Unlink*, and *Structure* realization statements are executed, building the appropriate rhetorical and grammatical structure for a global purpose expression. If the purpose is not global, then the Optionality system is entered and so forth. The details of this architecture and how it deals with purposes, preconditions, results, and action sequences are beyond the scope of this article (for more detail see Vander Linden, Cumming and Martin, 1992).

The Furnace Example

This section focuses on IMAGENE's approach to the expression of the pilot lighting action in the furnace text. There are clearly problems with the form of expression in example (1), but before we get too hard on the technical writer, consider the complexity of the process being expressed, depicted graphically in figure 4. There are four basic actions (depressing, lighting, waiting and turning), each of which would normally be expressed as imperatives were it not for the concurrency involved. The corpus study has shown that the use of the passive, perfective form, as was used in the furnace text ("after pilot has been lighted"), is not at all common in this context. IMAGENE, when given the relevant aspects of the context, produces the following more commonly used form:

(4) *Depress the knob. Light the pilot, and wait for sixty seconds. Release the knob, and turn it to the ON position.* [4]

An action, intended to be performed concurrently with other actions, is generally expressed, if possible, with a *durative* verb ("depress"), followed by

[3]Currently, these inquiries have not implemented, that is the data structures and code necessary to respond to the inquiry automatically have not been designed. Rather, the inquiries are answered manually allowing the focus of attention to be on the rhetorical and grammatical consequences of the results of the inquiries.

[4]In this paper, actual IMAGENE output will be typographically set apart as numbered examples in italic font.

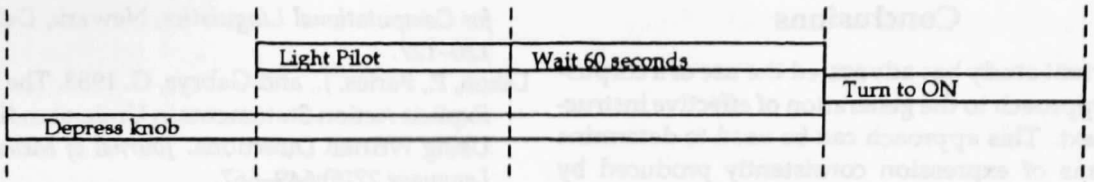

Figure 4: A graphic representation of the furnace process

the concurrent actions ("light" and "wait"), and culminated with a *discontinuation* command ("release"). A durative verb is one whose action is assumed to be continued until an explicit discontinuation verb is encountered. The semantics of the verbs themselves are sufficient to imply the concurrency. In this example, then, the principles of expression derived from a corpus-based analysis uncovered what appears to be a more effective expression for this procedure.

The nature of IMAGENE's implementation of the results also provides further insight to the writing process. The contextual information input to IMAGENE can be manipulated in an attempt to produce a form that more closely matches the one found in a particular text, such as the furnace text. This process sheds some light on the functional criteria the writer may have been using when they produced the form in the manual. For example, consider the following forms, one from the furnace domain and another from the telephone domain:

(5a) *Depress the knob. Wait for sixty seconds after lighting the pilot. Release the knob, and turn it to the ON position.*

(5b) *Return the OFF/STBY/TALK switch to the STBY position after your call.*

If IMAGENE is told that the action of lighting the pilot is well-known or obvious to the reader, it produces the form in example (5a). Earlier in the furnace text, the technical writer did, in fact, mention that the pilot must be manually lighted and may have assumed that this would render the action obvious to the reader. Unfortunately, such a mention doesn't make the lighting action obvious. IMAGENE requires both that the action be known to the reader prior to the reading of the manual and that it be the focus of the current paragraph, before it is considered obvious. This is the case with the action of making a call on a cordless telephone expressed in example (5b). Such manuals tend to assume that the reader already knows how to use a normal (non-cordless) phone, thus making the "your call" action obvious in the context of a paragraph on making a call. Notice that the expressional forms in example (5) closely match the form in the original furnace text in example (1). The actions that are considered obvious, the "lighting" and the "your call" actions, are expressed out of temporal order, grammatically marked with the preposition "after". This process is called *rhetorical demotion*, because an action is being demoted from a more prominent imperative status to either a gerund or nominalized form for rhetorical reasons. In the furnace text this demotion is inappropriate.

IMAGENE also varies the expressional form of an action that has already been mentioned in the text. Consider these examples:

(6a) *Depress the knob. When the pilot is lighted, wait for sixty seconds. Release the knob, and turn it to the ON position.*

(6b) *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.*

In example (6a), IMAGENE has rhetorically demoted the action of lighting the pilot to a precondition expression, presuming that the action has already been mentioned in the previous text. Although the furnace text did mention the fact that the pilot must be lighted manually, this previous mention did not actually specify this action as an imperative. IMAGENE requires that previous mentions cited in this context be explicit imperative commands. This pattern of repeated mention, although of questionable use in this furnace text, is common in instruction manuals. A preliminary section will tell the readers how to install a device using imperatives which quite often give detailed sub-steps for installation, and a subsequent section, on how to use the device, will begin with a rhetorically demoted action in precondition form such as the ones in example (6b). Here, both of the preconditions expressed ("the phone is installed" and "the battery is charged") have been prescribed to the reader in an earlier section as imperatives, thus warranting the rhetorical demotion. This form of rhetorical demotion is, again, not appropriate for the lighting action in the furnace example.

Conclusions

The current study has advocated the use of a corpus-based approach to the generation of effective instructional text. This approach can be used to determine the forms of expression consistently produced by technical writers in particular functional contexts. These results can be taken as indications of the most effective means of expression of instructions. This approach was contrasted with the approaches taken by other researchers in various sub-disciplines of Cognitive Science to the problem of generating instructions.

The paper concluded with a discussion of IMAGENE, an implementation of the results of the corpus study, and, in particular, a detailed discussion of how the results of this study allowed IMAGENE to produce a more effective form of expression for the furnace example, a process that is quite difficult to express. The broad generalizations concerning the grammatical form of instructional text brought out by this study and embodied in IMAGENE, are explicit enough to be subjected to psycholinguistic verification. Further, the approach taken is, in principle, applicable to other grammatical phenomena and text genres.

Future work on this project includes the use of the same methodology to address referring expressions in instructions, the incorporation of a procedural planner and user model to automatically produce the input to the text generator, and the psycholinguistic verification of the broad claims made by IMAGENE concerning the most effective forms of expression for actions in various functional contexts.

Acknowledgments: This work would not have been possible without the assistance of Susanna Cumming and James Martin.

References

Britton, B. K. and Gülgöz, S. 1991. Using Kintsch's Computational Model to Improve Instructional Text: Effects of Repairing Inference Calls on Recall and Cognitive Structures. *Journal of Educational Psychology* 83(3):329–345.

Cumming, S. 1990. Natural Discourse Hypothesis Engine. In McKeown, K.*et al.*(Eds.), *Proceedings of the Fifth International Workshop on Natural Language Generation.*

Dale, R. 1990. Generating Recipes: An Overview of Epicure. In Dale, R.*et al.*(Eds.), *Current Research in Natural Language Generation.* Academic Press.

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, Delaware, 120–127.

Dixon, P., Faries, J., and Gabrys, G. 1988. The Role of Explicit Action Statements in Understanding and Using Written Directions. *Journal of Memory and Language* 27(6):649–667.

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

Mann, W. C. and Thompson, S. A. 1989. Rhetorical Structure Theory: A Theory of Text Organization. In Polanyi, L.(Ed.), *The Structure of Discourse.* Norwood, NJ: Ablex. Also available as ISI tech. report ISI/RS-87-190.

Mann, W. C. 1985. An Introduction to the Nigel Text Generation Grammar. In Benson, J. D.*et al.*(Eds.), *Systemic Perspectives on Discourse*, volume 1, 84–95. Ablex.

Matthiessen, C. and Thompson, S. A. 1987. The structure of discourse and 'subordination'. In Haiman, J.*et al.*(Eds.), *Clause Combining in Grammar and Discourse*, 275–329. Amsterdam: Benjamins.

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.*et al.*(Eds.), *Current Research in Natural Language Generation.* Academic Press.

Mellish, C. and Evans, R. 1989. Natural Language Generation from Plans. *Computational Linguistics* 15(4):233–249.

Rock, S. 1992. Understanding Repetition in Natural Language Instructions — The Semantics of Extent. In *Proceedings of the Annual Meeting of Association for Computational Linguistics*, (Student Session), Newark, Delaware, 279–281.

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.(Ed.), *Coherence and Grounding in Discourse*, 435–454. John Benjamins Publishing Co.

Vander Linden, K., Cumming, S., and Martin, J. 1992. Using System Networks to Build Rhetorical Structures. In Dale, R.*et al.*(Eds.), *Aspects of Automated Natural Language Generation*, 183–198. Berlin: Springer Verlag.

Webber, B. L. and Di Eugenio, B. 1990. Free Adjuncts in Natural Language Instructions. In Karlgren, H.(Ed.), *Proceedings of the 13th International Conference on Computational Linguistics, Helsinki, Finland.*