A Goal-Directed Functionally-Based Stylistic Analyzer

Pat Hoyt

September 23, 1993

Abstract

If sophisticated natural language systems are to handle the full range of communication, then they must be able to account for the nuances and subtleties of linguistic style. A computational treatment of style would be highly advantageous to natural language understanding and generation, with particular relevance to intelligent computer-assisted language instruction and machine translation. These systems would be able to understand more complex and expressive language, produce text suitable for a specific occasion, help a second-language learner develop a more natural and appropriate style, and produce higher quality translations of text.

A foundation for AI-based computational style has been laid by DiMarco, with extensions by Green, Makuta-Giluk, Mah, and Payette in generation, rhetoric, comparative stylistics, and intelligent computer-aided language instruction, respectively. These researchers found that DiMarco's work, while an important step in computational stylistics, was limited due to the lack of a theoretical foundation. DiMarco and Hirst provided a preliminary theoretical foundation and Green extended their work. This thesis unifies these complementary, and sometimes contradictory, theories of syntactic style. A definitive grammar of style, based on this revised theory, is developed and used to implement a stylistic analyzer, ASSET. The revised theory of syntactic style and its implementation show that human-independent computer analysis of style is a feasible goal for computational linguistics.

Acknowledgements

There are many people who, during the time I was working on this thesis, made the task easier. They either distracted me, expressed their faith in me, went out of their way to be helpful, or gave me their support. I am grateful to all. There are four people, however, whom I want to publicly acknowledge.

First, many thanks to my supervisor, Chrysanne DiMarco. Her unfailing faith in my ability, her encouragement when I needed it, and her generosity let me believe that I *could* get this thesis done.

Devin and Logan, just by being there, were a constant reminder that there was more to life than writing a thesis. As a result, there were many times I returned, refreshed, to the task of writing.

And finally there is John, without whom keeping things in perspective would have been almost impossible. I thank you.

Contents

List of Figures

Chapter 1

Introduction

1.1 Style: Why is it Important?

For most people, writing effectively is a difficult, even excruciating, task. Why is this so, even when a person knows exactly what it is that they want to say?

Writing involves much more than simply putting words down in correct grammatical order. Although we usually have the primary purpose of communication in mind when writing, we also have other, often implicit, reasons. We are communicating more than the facts and ideas expressed by the words. For example, one could say that we are only expressing our views and opinions when writing on a controversial topic, such as environmental concerns rising from particular logging practices. We are, however, also trying to achieve the secondary purpose of influencing the reader's opinions and beliefs. Thus, how we write, or our *style*, conveys information about our interpersonal relationships, our views, and our attitudes. It is the style of our writing that determines its effectiveness. And the lack of knowledge needed for producing an effective style is what makes writing so difficult.

When we write, we make stylistic choices, often unconsciously, in order to achieve a particular purpose. These choices are made at all levels: lexical, syntactic, and semantic. A seemingly small linguistic difference can result in a dramatic stylistic difference. An example at the lexical level is shown below.

(1-1) I am going to my home.

(1-2) I am going to my domicile.

The only difference between the two sentences is the replacement of the word *home* with *domicile*. The former sounds perfectly natural and informal. The latter sounds unusual and stiff. One is left with the impression that the writer is a snob.

Syntactical choices also influence the style of our writing and Sentences (1-3) and (1-4) provide an example.

(1-3) The very timid man walked slowly.

(1-4) The man, very timid, walked slowly.

Both sentences impart the same information, but with different emphases. In the former the man's timidness is treated almost inconsequentially. In the latter, the postposed adjective places the emphasis on the timidness of the man and so a very different stylistic effect results.

We make these types of stylistic choices to achieve the purpose behind our written texts. How we choose to express our thoughts is defined by a particular set of stylistic goals. Style, thus, is *goal-directed*, a view supported, even if implicitly, by stylists such as Vinay and Darbelnet (1958), Crystal and Davy (1969), Kane (1983), and, especially, Lanham (1974).

The dramatic effect of a text that can result from particular stylistic choices indicates that style *must* be taken into consideration during any study of language. And this is as true from the viewpoint of Natural Language Processing (NLP) as from that of formal linguistics.

1.2 Style and Natural Language Processing

How could style be incorporated into a NLP system and how would that be helpful? Aren't the complexities of the many syntactic and semantic problems already enough to deal with? Let's examine these questions by looking at NLP from two aspects, Natural Language Generation (NLG) and Natural Language Understanding (NLU).

As pointed out in Section 1.1, style determines how we communicate. And one of the problems facing researchers in NLG is the issue of how to control the form of the output structure when there is more than one choice. Style, if incorporated into a NLG system, would provide this control. Work on incorporating style into generation systems was done by Hovy (1990) in his system PAULINE. The underlying objective of his system was the production of texts, different in form but similar in content, from a single knowledge base. And although his work—built on intuition-based heuristics with no formal theory of style—has limited reproducibility, it underlines the importance of computational style to generation systems.

An application that encompasses both NLG and NLU is Machine Translation (MT). Good quality translations should preserve the 'feel' of the original text across the translation. In other words, the style also must be translated. Incorporating style into MT systems would require adapting research from computational stylistics. This would include a stylistic analysis of the source text, the application of rules of comparative stylistics, and the stylistic control of the generation of the target text. Work by DiMarco (1990) began this process with the development of a computational theory of style. Mah (1991) extended her work to develop a codification of French—English comparative stylistics and Green (1992) incorporated a portion of her theory into the stylistic control mechanism of a NLG system. It is hoped that this work will ultimately improve the quality of machine-driven translation and simplify the human post-editing stage.

DiMarco's work also has applications in another area of NLP, Intelligent Computer-Aided Language Instruction (ICALI)—especially for second-language instruction. A student would benefit from the ability of an ICALI system to give feedback on the inappropriate use of style. It is much harder for a second-language learner to assimilate the appropriate style of a new language than its words and grammar. The feasibility of incorporating style into ICALI systems has been shown by (Payette, 1990) and (Payette and Hirst, 1992).

Although DiMarco's work is a significant contribution to AI-based computational stylistics, it is not without its shortcomings. The one most relevant to this thesis is the lack of a well-defined theoretical basis, or formal linguistic justification. This has made her work difficult to extend or reproduce. DiMarco and Hirst (1993a) have provided a formal justification based on the linguistic theory of Functional Grammar and Green (1992) has refined and extended their work. The objective of this thesis is to produce a 'definitive' computational theory of syntactic style by combining (DiMarco, 1990), (DiMarco and Hirst, 1993a), and (Green, 1992).

1.3 A Revised Theory of Style

The work of DiMarco (1990), DiMarco and Hirst (1993a), and Green (1992) has produced several versions of a theory of style, drawing on different sources for theoretical justification. The objective of the thesis is to integrate their complementary, and sometimes contradictory, work into a single unified theory of style, build a definitive grammar of style from this revised theory, and then use this grammar to implement a style analyzer.

The thesis is divided into four main sections. Chapter 2 addresses the confusion associated with the term *style*. It summarizes how the world, in general, views style and then suggests a working definition for the purposes of this thesis. The seminal work of Crystal and Davy (1969), who developed a methodology for the systematic study of style, is examined. And, finally, a brief overview of computational stylistics is presented.

Chapter 3 presents, in detail, the newly consolidated theory of style with its linguistic foundation. Before doing so, however, a brief discussion of DiMarco's original work and the linguistic theory behind the justifications of the revised theory is presented as background to the theory itself.

More than a theory of style is needed for an implementation of a style analyzer. A systematic way of representing the theory's concepts is required. Chapter 4 presents the rules that are contained in the definitive grammar of style and that codify the revised theory. It describes each of the rules with explanatory text and examples.[1]

The fourth section, Chapter 5, describes the stylistic analyzer that implements the definitive grammar of style. ASSET, named for Analyzing the Style of SEnTences, analyzes the style of single sentences. Both the design criteria used during its development and its architecture are described. ASSET has shown that human-independent computer-analysis of

[1]Appendix B presents the grammar stripped of all text as a convenience to the reader.

style is indeed a feasible goal for computational linguists.

Chapter 2

Style: An Overview

2.1 Introduction

Before discussing the implementation of the stylistic analyzer and the theory behind it in any detail, we must have an idea of what style is and what it does. This is not a simple question to address, for style is not a simple concept.

Style covers a broad range of language phenomena. Crystal and Davy (1969), when trying to show their readers the complexities involved in delineating the meaning of the word *style*, write "... at least four commonly occurring senses of the term 'style' need to be distinguished."[1] And Crystal (1987) also writes the following about style.

> Style is one of the thorniest concepts to be dealt with in this encyclopedia. To Samuel Wesley, it was 'the dress of thought'; to Jonathan Swift, it was 'proper words in proper places'; to W. B. Yeats, it was 'high breeding in words and in argument'. And so we could continue, through several hundred definitions and characterizations. It is a remarkable career for a word that originally meant no more than a 'writing-implement - a pointed object, or *stilus*, for inscribing wax.' (Crystal, 1987, p. 66)

The lack of a standard definition for style, even among linguists, is problematic for researchers investigating any aspect of style. They must define, for themselves, what is encompassed by the term.

A useful starting point in defining style is to look at the function of language. Obviously, we communicate ideas and facts through language. We are also, however, communicating, for example, about our interpersonal relationships, our views and attitudes, and the physical environment. Two different letters to the same person can have two very different salutations. A business letter will begin *Dear Ms. Creant* and a personal letter could begin *Hi there Liz!*. The former is very formal, as the writer wants to be taken seriously. In the latter the writer,

[1](Crystal and Davy, 1969, p. 9).

wanting to be friendly and personable, is informal. We also get the impression that the writer knows *Liz* fairly well. It is *style* that makes these two different letter salutations very different and it is *style* that fulfils the communicative functions of language.

The study of style has a long and ancient history dating back to ancient Greece and Rome. At that time, the study of Classical Rhetoric—a field in which the choice and organization of words and syntactic structure for effective use was the mainstay—became very important. It is from this classical background that the concept of style arose and this approach became part of a standard university education. In the nineteenth century, however, the teaching of Classical Rhetoric gave way to the teaching of a set of prescriptive rules for composition, a trend which has continued through until the modern era.

As a result of this history, there are generally two different ways of viewing style.[2] The first is *prescriptive style*, an evaluative style in which there is either a right way or wrong way of expressing oneself and which is rather entrenched in our society. This is the style taught in schools and in books on writing. Although there have to be some stylistic conventions so that communication is possible, the authoritarian approach does not make allowances for individual expression that defies convention. This is problematic for computational linguistic applications that aim to cover a variety of situations. Uses envisaged for style in computational linguistics include machine translation, second language instruction, and natural language generation. Thus, a prescriptive viewpoint is not useful for developing a computational theory of style.

The other viewpoint is *descriptive style*. Here, there is no sense of evaluation. Rather, the characteristics that distinguish text are used in characterizing style. There is no best or proper style. Instead, style is seen as the sum of linguistic choices used to achieve the purpose behind the written text. This is the approach behind the implementation of our theory of style in which style is *goal-directed* so that choices are made at semantic, syntactic, and lexical levels to fulfil specific *stylistic goals*.

In addition to these two modes, style can be divided along another dimension. First, there is *literary style*. Here, the idiosyncratic style of a single author, compared to others writing around the same time, is identified. Researchers are interested in the comparisons of individual writers in order to determine just *what* gives an author a characteristic 'voice'. They are also interested in determining reliable ways to prove the authenticity of an author of some particular piece of literature.

The other type of style is non-literary. Here, the commonality of a group of writers—hence the term *group style*—is investigated. This introduces the notion of genre, exemplified by the stylistic sublanguages used by the legal community and by the scientific community. The theory of style that lies behind my implementation is based on the characteristics of non-literary group style, specifically in high-quality magazine writing.

The rest of this chapter will discuss the background of style for chapter 3, where the theory behind the implementation is presented. First, to discover the general viewpoint

[2]Actually, this is true of language in general. This point is made very well by Crystal (1987).

of style, I will look at how modern composition is taught. I will then discuss an attempt to codify style—from a linguist's point of view. The field of computational stylistics will next be examined. I will discuss the so-called style checkers, which are becoming quite common; stylostatistics, which has produced important linguistic data; and, finally, Artificial Intelligence (AI)-based computational stylistics.

2.2 Style in the Modern World

An appropriate place to begin an examination of style is to look at the general viewpoint of the modern world. This can be done by looking at the perspective of style in a 'how-to-write' book aimed at college students. The one used here is Kane's (1983) *The Oxford Guide to Writing*.

The attitude towards style in this book is very interesting—there seems to be a contrast between the two major viewpoints of style. Implicitly, there is an emphasis on a single correct way to write. Additionally, the point is made that most errors in writing are those of style.[3] At first glance, the book seems to be espousing a prescriptive view of style. There is also, however, an implicit appreciation of the value of goal-directed style. For instance, Kane discusses how to structure a sentence to achieve conciseness and also to achieve emphasis. This is supportive of a descriptive view of style, as is an explicit discussion on style that emphasizes that there is no bad style and that style is the total sum of all choices of words and their arrangements. Overall, the attitude towards style is somewhat ambiguous.

This ambiguity is not directly useful in the construction of a theory of style as there is no systematic approach to describing or organizing the constituents of style. However, since there is a prevailing descriptive point of view, a consideration of the syntactic coverage in Kane's book is quite useful. He includes paragraph development, such as the patterns of sentences found in paragraphs. Additionally, Kane distinguishes between different grammatical types (*complete, simple, complex*, and *fragments*) and also between different stylistic forms (*serial, parallel* and *balanced*, and the *hierarchical structure*) of individual sentences. The stylistic effects associated with paragraphs and sentences are important to consider in the development of a syntactic theory of style.

Though Kane's book proves useful to the development of a codified view of style by providing justification for the syntactic coverage,[4] there is no systematic approach to the description or organization of style. Before looking at what computational stylistics has to offer, a methodology developed for the objective study of style will be discussed.

[3] See (Kane, 1983, p. 18).

[4] See (DiMarco and Hirst, 1993a).

2.3 A Systematic Approach to the Investigation of Style

Crystal and Davy (1969) proposed a methodology, from a linguistic perspective, that would allow for an objective study of style characteristics. They felt that the different 'varieties'[5] of English should be "put on a more systematic basis and extended".[6] Their reasoning was that this would result in more successful oral and written communications, thus enhancing social interactions. They stressed that this stylistic awareness is as important for the native speaker as for someone learning English as a second language. If style were codified then the teaching of these different types of English could be done methodically. Another motivation for the codification of style was to resolve disputes involving linguistic interpretations of texts. Arguing about different interpretations could be done more rationally if one could rely on more than an intuitive reading of the text in question.[7] In other words, one needs "analytical methods ... to allow evaluation and comparative study."[8]

Crystal and Davy proposed a theory that made use of a hierarchy of five interrelated language levels: the phonetic/graphetic; the phonological/graphological; the grammatical; the lexical; and the semantic.[9] They analyze a text, taken from a chosen variety, on each of these levels, picking out the important stylistic features. To do so, they developed two criteria for discerning just what the important features of a particular genre are. These features were defined as those that occur more frequently within a variety and also those that are shared less by other varieties. For instance, the use of passive constructions in scientific English is considered important for both of these criteria. It is the most frequent feature and so is important to include in a corresponding codification. Also, because of the more infrequent use of the passive in other groups, the use of the passive serves to distinguish the scientific genre. Another example involves the use of *hereinbefore* in legal text. It is distinctive as it is used nowhere else in English. It also has great stylistic importance when it is used as it rarely occurs.[10]

A more-indepth examination of the grammatical level is useful to researchers interested in syntactic style. Crystal and Davy developed five different components at this level, which I list below:

- **Inter-sentence Relationships:** Sentence-linking features, such as ellipsis; anaphora; use of concord; and lexical repetition, define these relationships.
- **Sentence Typology:** Sentences are divided into the two classes of *complete* and *incomplete* with the completes being further divided into *majors* (simple, compound, complex, and mixed) and *minors*.

[5]The term *varieties*, as used here, has some equivalence to the more familiar notion of *genre*.

[6](Crystal and Davy, 1969, p. 7).

[7](Crystal and Davy, 1969, p. 6).

[8](Crystal and Davy, 1969, p. vii).

[9]The term *semantic* is used here to refer to the "linguistic meaning of a text over and above the meaning of the lexical items taken singly" (Crystal and Davy, 1969, p. 19).

[10]See (Crystal and Davy, 1969, p. 21).

- **Clause Typology:** Five elements of clause structure are used in the analysis: *subject, predicator, complement, adverbial*, and *vocative*.

- **Group Typology:** The composition of the nominal group—presence and type of, or absence of, premodification and postmodification—and the verbal group—tense used and finiteness of lexical verb—are examined.

- **Word Typology:** The morphological considerations of *root, prefix*, and *suffix* are considered.

These different components are then used as a basis when deciding which of the syntactic features are stylistically important.

The result of this work was the concept of stylistic sublanguages and a proposed rudimentary grammar for style. The latter made the objective study of style possible which in turn allowed the investigation of the stylistic characteristics of different genres. Although they did not deal specifically with goal-driven style, the philosophy and methodology behind their work can be adopted. Their philosophy is apparent in their determination to minimize the intuitive element. In addition, their methodology, in which one has a well-defined, coherent, and systematic method for language examination, can be adapted when defining a theory of style from a goal-directed point of view.

Having looked at style from a general point of view and from the perspective of theoretical linguists, we now turn to computational stylistics.

2.4 Computational Stylistics

2.4.1 Style Checkers

The best known, but least important from an Artificial Intelligence point of view, computer programs that deal with style are the so-called *style checkers*. These include such programs as Correct Grammar for Windows, Right Writer for Windows, Grammatik for Windows, CorrectText, Reader, and PowerEdit[11], which 'proofread' a writer's text checking for grammatical errors such as subject-verb agreement, pronoun-case problems, split infinitives, deictic usage (*this/that/these/those*), article usage, incomplete sentences, sentence fragments, run-on sentences, excessive sentence length, and excessive sentence complexity. In addition, they check for such stylistic variables as jargon, clichés, redundancy, archaic or colloquial use, weak wording, and long-windedness. The result is a tendency towards the enforcement of prescriptive style, looking only for clearness, simplicity, and preciseness.

Although these programs are quite popular—and their usefulness cannot be denied—they do not account for how style is used. They are, instead, mindless enforcers of a particular style no matter what the context. There is no methodical treatment of style from an objective

[11]See (Bolt, 1993) for a detailed performance review of these programs.

viewpoint. There is no systematic representation of stylistic knowledge. These type of programs provide no useful base or background from which to develop a computational theory of style.

2.4.2 Stylostatistics

Crystal (1987) describes stylostatistics as follows:

> Such studies comprise a major part of the field of statistical linguistics — a field which investigates not only the differences between samples or texts, but also the properties that samples (and, ultimately, whole languages, and all languages) have in common, as part of the search for linguistic universals. (Crystal, 1987, p. 67)

In this field, collected data on syntactic text patterns are subjected to statistical analysis. The object is to discover either universal or characteristic patterns among selected samples of texts. Enlisting the aid of the computer greatly reduces the tedious part of this process while leaving the interpretation part of the analysis to the researchers.

A Pioneer—Milic

One of the pioneers in computational stylostatistics is Louis T. Milic, who was interested in showing that style could be described objectively and quantitatively. His thesis[12] was devoted to proving this by comparing Swift's writing to that of other eighteenth century writers. He began his work by manually and painstakingly counting such things as words, phrases, sentences, pronouns, nouns, verbs, adjectives, adverbs, classifications of the first and last words of sentences, and also by categorizing groups of two and three and four words. During this process, he was introduced to computers and immediately appreciated how they could be used to reduce the tediousness and time-consuming part of his work.

In order to incorporate a computer into his task, Milic set up a grammatical word-class classification scheme based on that of Fries. He then developed a two-digit coding system for each of the word classes in the grammar scheme. Texts to be analyzed were first translated into corresponding codes, the input data for Milic's system. Frequency distributions of word classes, totals of first and last sentence word classes, and totals of overlapping three-word combinations were the statistics collected. A compilation of the total number of different three-word classes was also included as an afterthought.[13]

Milic wanted to discover syntactic patterns that would show distinctions between the writers that he was investigating. He was especially interested in the results of the totals of the three-word combinations. Interestingly, the latter showed linguistic constancy among

[12] See (Milic, 1982).

[13] See (Milic, 1982) for a more complete description of his methodology.

the writers instead of the style distinctions that were anticipated. Moreover, it was the afterthought, or "D-statistic", the number of *different* three-word combinations, that indicated stylistic uniqueness among different authors. His pioneering work in computational stylistics was adapted by the York Computer Inventory of Prose Style, the objectives, methodology, and results of which are summarized below.

The York Computer Inventory of Prose Style—Cluett

Cluett (1976) describes a large project, the York Computer Inventory of Prose Style.[14] He states the objectives of this project:

> ... to develop hitherto undiscovered data about the syntactic habits of individual writers, to collect those data and the texts on which they are based in a single convenient repository, and to maintain both this material and the programs that manipulate it ... (Cluett, 1976, p. 15)

Statistical analyses, on syntactic data from many different literary text sources, were compiled by a computer. This was an ambitious and challenging undertaking.

The York Project was based on some of the work done by Milic. The classification scheme used in this project was based on the Fries-Milic Syntactic Code for computers where the syntax was divided into thirty-one separate categories, each assigned a two-digit code. This code was enhanced by adding a third descriptor digit to the existing two-digit code. For example the two-digit code for a noun is **01**, no matter how the noun is used in a sentence. Increasing the descriptive power of the grammar now assigns **011** to ordinary uses of the noun and a different third digit to specialized uses. For example a noun used as a possessive, as in *railroad's president*, is assigned the code **013**. The resulting grammar contains ninety-seven different categories.[15]

The preparation of the text involved the development of a methodology for text selection and then translating, or parsing, the text into its equivalent three-digit code. The selection of authors and then the selection of texts from their writings was done using three different criteria: the judgement of the researchers involved in the Inventory, other ongoing research, and the demands placed upon the Inventory by other researchers. Once a particular work was selected, several sections, or cuts, each consisting of the text from the first terminal punctuation on the left-hand page to the last terminal punctuation on the right-hand page, were selected using a random number table. Enough cuts were taken to obtain 3400 to 3500 words of natural language, resulting in a total of between five and ten cuts, as each cut contained between 350 and 700 continuous words. The selected material, once translated and encoded, was entered into the computer as three-digit codes for analysis.

The computer analysis consisted of a number of different manipulations of the data, such as the frequency-ordered distributions of both the three-digit and the two-digit word classes

[14] From now on referred to as the York Project or the Inventory.

[15] For a full description of the grammar see (Cluett, 1976, p. 20).

and Milic's "D" Statistic—the number of different consecutive three-word sequences. Cluett gives a detailed account of the interpretation of these analyses on selected literary prose. Though a non-AI method of investigating style was used, the Inventory showed that the characteristics of literary style could be correlated to syntactic constructions. In addition to validating a formalization of style based on syntax, the York Project also itemized syntactic features important in producing stylistic effects.

Importance of Stylostatistics

Stylostatistics, in conjunction with the use of computers, has evolved into a very sophisticated field. And the definitive results that have come out of these investigations into syntactic style give impetus to further research in computational stylistics. For though the computer has no part in the interpretation of the results of the analysis and cannot deal with natural language, stylostatistics has pointed out the validity of developing an objective theory of style based on the syntax of the language. We turn now to attempts to incorporate style into AI-based systems dealing with natural language.

2.4.3 Style and Machine Translation

An important area of computational linguistics, Machine Translation (MT), is starting to develop an awareness of the importance of incorporating style into MT systems. The rationale for doing so is that the quality of the translations would improve and the human post-editing stage would be simpler. Though much work is needed, investigators have begun research in this direction.

One of these, Loffler-Laurian (1987), has attempted to describe group style, the set of characteristics common to the stylistic conventions of a group of writers, and has also investigated the stylistic aspects of the post-editing stage of MT. The important contribution that Loffler-Laurian makes is to emphasize that, although group styles might exist across languages, their realizations might be different.

Tsutsumi (1990) is interested in stylistic differences that appear across languages and has developed a methodology for dealing with this problem in MT. He identified categories of 'stylistic gaps', for example, the inability of certain syntactic constructions in the source language to be translated directly into the target one, which causes problems during the translation process. These gaps are dealt with by a wide-range restructuring of intermediate representations, where the sentence is rewritten before translation using an augmented context-free grammar. Tsutsumi's work has shown that computational stylistics is useful for the translation of pairs of languages *not* in the same language group as well as those that are.

2.4.4 Style and Natural Language

The computer processing of natural language has spawned a great deal of research. The use of linguistic style, however, as an aid to either the comprehension or generation of natural language is almost non-existent. PAULINE, a text generation system, and STYLISTIQUE, the first implementation to use a formal representation of sentence style, are two systems that are centred around the importance of style. I now describe each of these systems in more detail.

Generation and Style

Hovy's (1990) text generation system PAULINE (Planning And Uttering Language In Natural Environments) was among the first that was able to deliver 'stylistically-appropriate' text. Until then, generators produced the same text, no matter the hearer or the circumstance in which the interaction took place. This was not sufficient as "we tailor our text to the hearer and to the situation."[16] Doing so, we convey more information in our communications than is contained by the literal meaning of the text and this information has an important effect on the participants involved in the exchange.

The objective underlying PAULINE is the production of texts, different in form but similar in content, from a single knowledge base. The generated output depends on situational settings that portray pragmatic concerns associated with each text. The pragmatic concerns used by PAULINE are categorized based on the conversational setting, the interlocuters' personal characteristics, and the speaker's goals with respect to the hearer. Each of these categories has a number of features, each with a fixed set of values. For instance, the speaker (one of the pragmatic interlocutor categories) is given the following features: knowledge of the topic (*expert, student, novice*), interest in the topic (*high, low*), opinions of the topic (*good, neutral, bad*), and emotional state (*happy, angry, calm*).

PAULINE showed, however, that these pragmatic goals were too general to guide the decision process during text realization. This generality is apparent, for example, when an attempt is made to correlate a speaker's goal to be friendly with active versus passive sentence constructions.[17] To enable pragmatic goals to influence the realization process, a system of *rhetorical* goals of style was developed to act as an intermediary between the overall communicative goals and the realization module making syntactic decisions. The following twelve goals of style are defined in PAULINE: *formality, simplicity, timidity, partiality, detail, haste, force, floridity, colour, personal reference, openmindedness*, and *respect*. As do the pragmatic features, the stylistic goals all have a fixed set of values, usually within some range. The values of the stylistic goals are set by the pragmatic goals as specified by the user and have a direct effect on the style of the text produced by PAULINE.

Hovy's success with PAULINE underlines the importance of goal-directed style during the

[16](Hovy, 1990, p. 154).
[17](Hovy, 1990, p. 162).

processing of natural language and as such his work is pioneering. The ability of PAULINE to produce radically different texts using the same knowledge base but different pragmatic settings illustrates the importance of style. The difficulty with this work is that no formal theory was developed to underline the interaction of pragmatics with stylistic goals or the effect stylistic goals have on syntax. PAULINE uses heuristics when determining the relationship between the pragmatic categories and the rhetorical goals of style. In addition, the effect of the individual stylistic goals on syntax has also been determined by the use of heuristics based on intuition. What is needed is a theoretical basis of style and rhetoric which could be used by Natural Language Processing systems in any domain or situation. This leads naturally to the next topic—formalizing syntactic style in a computational theory.

A Formalization of Style

STYLISTIQUE is a Prolog system, created by DiMarco (1990), that analyzes the style of single sentences. The basis of this implementation is the first formalized theory of style developed for computational linguistics. This theory is based on non-literary group style[18] and takes a view of style that is goal-directed. STYLISTIQUE uses the syntactic features of the input sentence to produce a detailed analysis of its resulting style. One of the more important contributions of this work is the separation of the computational theory from the actual implementation. As Stylistique is the springboard for the theory of style that underlies my implementation, I will discuss it in more detail in chapter 3.

2.5 Summary

To summarize, I have presented the background for a formal computational theory of style. First, an attempt was made to define what is meant by the term *style*; next, a formal and objective theory of style from a linguistic point of view was described; and finally, the field of computational stylistics was reviewed. We saw that there has been little work at integrating style into natural language processing. Rather, the main computational developments have been along two fronts: the proofreading of documents using style superficially and the statistical processing of syntactic data that is then used by researchers in their investigations into style.

The next chapter will introduce our syntactic theory of style. There I will discuss STYLISTIQUE, the linguistic background underlying the theory, and finally, the theory itself.

[18] An example of non-literary group style would be high-quality magazine writing.

Chapter 3

A Theory Of Syntactic Style

The main focus of this chapter is the theory of syntactic style that underlies my sentence-style analyzer. Before giving the details of this theory, however, I examine the theory behind DiMarco's (1990) STYLISTIQUE system and describe the parts of it that directly contribute to ours.[1] I then outline the linguistic theory that provides a formal justification for the correlation of stylistic and syntactic features in our theory. And finally, I present the theory itself.

3.1 The Theory of Style Underlying STYLISTIQUE

STYLISTIQUE, a Prolog system developed by DiMarco (1990), analyzes the syntactic style of a single sentence. It was one of the first systems in which the computer performed stylistic analysis without human intervention. One of DiMarco's contributions was the separation of stylistic theory and implementation.[2] As DiMarco's computational theory of style has been the springboard for our present one, I now present the parts directly relevant to it. Full definitions of concepts introduced here are given in Section 3.3.

3.1.1 Goal-Directed Style

An essential element for developing any theory of style is a decision about what the term *style* will encompass. DiMarco considers style to be near the descriptive end of the descriptive-prescriptive continuum that encompasses the many attempts to define style. The resulting approach to style is a *goal-directed* one in which lexical choice, syntactic structure, and semantic organization are correlated with particular stylistic goals.[3] In her syntactic, goal-

[1] I have used the plural form of the first person pronoun to refer to the theory of style discussed in this chapter.

[2] To make this task more tractable, DiMarco concentrated on non-literary group style, specifically, high-quality magazine writing.

[3] See (DiMarco, 1990, p. 9).

driven theory of style, she has chosen three sets of goals, which she took from (Vinay and Darbelnet, 1958). The goals in each set represent opposite ends of a continuum and are presented here with simplified descriptions.

Clarity and Obscurity: Clear sentences tend to be easily understood, while obscure ones are difficult to interpret.

Abstraction and Concreteness: Abstract sentences contain a general lack of modification and are very ordered. Concrete sentences suggest an effect of immediacy.

Staticness and Dynamism: Static sentences are uniform and predictable while dynamic sentences deviate from the norm.

3.1.2 A Fundamental Concept

DiMarco, in her theory, has incorporated the idea that style is produced by deviations from a stylistic norm, for it is her contention that "style is created by patterns of concord and discord giving an overall integrated arrangement".[4] Consequently, the concepts of concord and discord are fundamental principles in DiMarco's formalization. The term *concord* is defined as a conformity with the norm and *discord*, a deviation from the norm. The fundamental nature of these terms is shown by the following one-sentence description, albeit oversimplified, of DiMarco's theory. Briefly, it is a formalized description of how the syntactic elements of a sentence either conform to or deviate from a stylistic norm. The theory revolves around this concept.

3.1.3 A Theory of Syntactic Style—in Brief

DiMarco's theory formally correlates the syntax of a sentence with stylistic goals. To do so, a stylistic vocabulary was developed in order to express stylistic concepts precisely. I presented, in Section 3.1.1, the stylistic goals, which are the terms developed to introduce the more general stylistic concepts. I will now describe the rest of the stylistic vocabulary.

The Abstract Elements

The *abstract elements* were developed to describe groups of stylistically similar sentences, a description based on syntactic properties. When DiMarco constructed these elements, she looked at three different syntactic categories of properties found in sentences: *balance*, *dominance*, and *position*. It is from these properties that the abstract elements were derived. For example, the *homopoise* abstract element is in the balance group and is exemplified by Sentence (3-1).

[4](DiMarco, 1990, p. 40).

(3-1) The style was formed and the principles were acquired.[5]

The abstract elements provide a way of linking sentences to particular stylistic goals as each goal is *defined* by these elements. For example, the goal of clarity can be achieved with a homopoisal sentence.

The abstract elements and the syntactic properties on which they are based have been introduced very briefly here. A detailed description is given Section 3.3.2.

The Primitive Elements

The *primitive elements* link actual syntactic constructs, such as the premodification in a noun phrase, with the abstract elements. To begin with, two different viewpoints of the sentence are important—a connective one and a hierarchical one. The stylistic effect of linear cohesive bonds of sentence components are described by the *conjunct* and *antijunct* primitive elements in the former view. Similarly, nested bonds of sentence components are described by the *subjunct* and the *superjunct* primitive elements in the latter one. The *degree* of strength of these bonds is incorporated into this primitive-element description, thus increasing the subtlety of the stylistic theory. The primitive elements are detailed fully in Section 3.3.3.

3.1.4 Conclusion

I have very briefly summarized the development of a stylistic vocabulary, which is used to describe stylistic attributes of sentences and which is ultimately based on underlying syntax. The abstract elements connect the abstract stylistic goals with the concrete primitive elements. This theory, thus, relates fairly general stylistic concepts—stylistic goals—to the syntactic characterizations of a sentence. The use of three different classes of stylistic terminology and constructs, each more abstract than the preceding, has resulted in a hierarchical theory of style that allows the mapping of specific syntactic constructs to the general stylistic goals to be done in a intuitive, yet precise way.

I have given a brief summary of DiMarco's theory of style with the purpose of providing a background to ours, described later in this chapter. Our theory has been built on the ideas described in this section. To summarize, these include the view of style as goal-directed, the concept of stylistic norm, the concepts of concord and discord, the hierarchical shape of the theory, and a stylistic vocabulary to describe sentences. One of the things missing, however, from DiMarco's original theory is the lack of formal linguistic justification for the classification of syntactic constructs as primitive elements. To a large extent, this has been remedied and in the next section I describe the linguistic theory that now provides the underpinnings of the theory of style.

[5]Adapted from the *Manchester Guardian Weekly*, 14 February 1988, p. 15.

3.2 Linguistic Foundations

In this section, I outline the linguistic theory on which our syntactic theory of style is now based. Our main source is *Functional Grammar*[6] a formalism chosen because of its view of language as a network of choices. This allows us to also express style as choice. I will begin with a section outlining the main concepts, especially those relevant to our theory of style, of Functional Linguistics from (Halliday, 1985). I then review the work of Halliday and Hasan (1976), who describe how text cohesion is accomplished in English. Finally, I end by describing two other linguistic contributions to the foundation of our theory of style.

3.2.1 Functional Linguistics—Halliday

Due to space considerations and the complexity of the subject, I will give only a very brief overview of *Functional Grammar*. I have tried to glean those concepts necessary for an understanding of our theory. For more detail, the reader is referred to (Green, 1992), (Green and DiMarco, 1993), or (Halliday, 1985), the latter being a definitive work on Functional Grammar.

Functional Grammar is "functional in the sense that it is designed to account for how language is used."[7] Halliday uses a constituent analysis, with minimal bracketing, and then augments the resulting nodes with functional labels. See Figure 3.1 for an example of a functionally labelled constituent analysis in the style of Halliday.

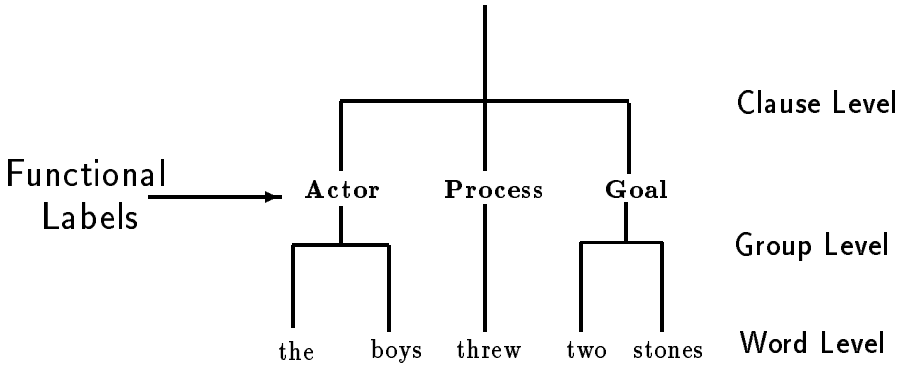

Figure 3.1: An example of minimal bracketing with functional labels.

Minimal bracketing is a way of combining linguistic sequences that perform a function in language. A hierarchy of constituents, where each constituent is related to the next (in

[6]The reader should be aware that *Systemic Functional Grammar* is the linguistic basis of this work. However, as only the *Functional* part is directly relevant to my work, I have chosen to use the term *Functional Grammar* throughout this thesis.

[7](Halliday, 1985, p. xiii).

the hierarchy) by function, is used. The resulting language structure consists of sentences composed of clauses composed of groups and/or phrases composed of words composed of morphemes. This classification leads to the concept of *rank* in Functional Grammar. The hierarchy used in the constituent analysis provides a scale where each level is considered to be a unit. Hence, the use of minimal bracketing is also termed *ranked constituent analysis.* Usually, sentence structure follows this ranking scale strictly. However, the phenomenon of *rank-shifting*, where a usually higher-ranked constituent realizes some function at a lower rank, is not an unusual occurrence.

While constituent analysis shows how sentence components fit together, there is not much information present about linguistic structure and function. To present this information, nodes are marked by functional labels. A constituent, moreover, nearly always realizes more than one function at a time. This concept of function is central to Functional Grammar and we will explore it looking at the clause itself, below the clause, and above the clause.

At the Clause Level

Halliday uses three different ways of exploring or analyzing a clause: *clause as message, clause as exchange*, and *clause as representation.* These ways of examining the clause refer to the three principal kinds of meaning embodied in clause structure. Halliday states it best: "Three distinct structures, each expressing one kind of semantic organization, are mapped on to one another to produce a single wording."[8] In the following paragraphs, we will explore each one in turn.

Clause as Message Here, we are concerned with the *thematic* structure, that which characterizes the clause as a message. The notion of theme relates the clause to the discourse and context in which it is found. The *Theme* and *Rheme* are the two functions underlying the thematic structure. The Theme[9] is realized by the constituent in the first position in the clause and is the point of departure for the message. The remainder of the message is developed in the Rheme, the rest of the clause. I give two examples of the Theme-Rheme structure of sentences in Figure 3.2.

Clause as Exchange In addition to conveying information, the clause is an interaction between two parties. Thus, Halliday does an analysis of mood where the two top functions realized by sentence constituents are Mood and Residue. The Mood, consisting of a Subject and a Finite, determines whether the the clause is a statement, question, or command. The Residue is realized by the rest of the clause and is composed of the Predicator, Complements, and Adjuncts. An example of an analysis of a sentence as exchange is shown in Figure 3.3.

[8](Halliday, 1985, p. 38).

[9]I am following Halliday's use of first-letter-capitalization of function labels throughout this section.

Theme	Rheme
my aunt	has been given that teapot by the duke
on Friday night	I go backwards to bed

Figure 3.2: The Theme-Rheme Structure of a Clause

Mood		Residue	
Subject	Finite	Predicator	Complement
Mary	has	had	a cold
John	is	being	naughty

Figure 3.3: The Mood-Residue Structure of a Clause

Clause as Representation Here we are concerned with how the clause expresses the "reflective, experiential aspect of meaning".[10] *Transitivity* structure expresses the representational meaning and centres on what *processes* are being represented and what structures support them. The components of a transitivity analysis consist of the process itself, the participants in the process, and the circumstances under which the process happens. There are four types of processes: material, mental, verbal, and relational. The functions of the other constituents in the analysis depend on the type of process present. I will use a material process, which is one of creating or doing, as an example. There are three participants that may be present: Actor, the entity creating or doing; Goal; the object of the creating or doing; and Beneficiary, the entity that benefits from the creating or doing. In addition, there may be constituents in the clause carrying out circumstantial functions such as Extent, Location, Manner, or Cause. See Figure 3.4 for an example of a transitivity analysis.

Below the Clause

Halliday maintains that the three ways of looking at the clause are also valid at the group level. Here, however, distinct structures do not exist for each viewpoint. Rather, all three

[10](Halliday, 1985, p. 101).

Actor	Process	Goal
the lion	caught	the tourist

Figure 3.4: The Process (Material) Structure of a Clause

those	two	splendid	old	electric	trains	with pantographs
Deictic	Numerative	Epithet	Epithet	Classifier	Thing	Qualifier

Figure 3.5: The Experiential Structure of the Nominal Group

views contribute to one structure. Halliday analyzes this one structure from an *experiential* viewpoint, where "meaning is organized as an experience"[11], and from a *logical* structure, where language is expressed in logical relations. I discuss three groups found within the clause: the *nominal* group, the *verbal* group, and the *adverbial* group.

The Nominal Group The experiential structure of the nominal group is separated into different types of modifiers and the noun itself. The noun realizes the function of Thing—what is being discussed. The modifiers fill the following functions:

- **Deictic:** Indicates a specific subset, if any, of the Thing;
- **Numerative:** Expresses some numerical quality of the subset;
- **Epithet:** Indicates some quality, either an objective property or the speaker's subjective attitude, of the subset;
- **Classifier:** Indicates a particular subclass of the Thing.

Usually, the Thing is the Head of the nominal group. However, this function can also be filled by a Numerative or Deictic. In addition, the function of Qualifier is realized by the constituent which follows the Head. The Qualifier is *embedded* or *rank-shifted* as it is usually realized by a prepositional phrase or a clause functioning at a lower rank. An example of the experiential structure of a nominal group is shown in Figure 3.5.

[11](Halliday, 1985, p. 158).

The logical structure of the nominal group is shown by the specification of a subset of things. This is done by the elements that form the premodification. Each step in the progression from right to left further specifies the Head of the nominal group.

The Verbal and Adverbial Groups The experiential structure of the verbal group is Finite with Event and optional Auxiliaries. The Event is the verbal equivalent of the nominal group's Thing, expressing some process, and the Finite relates the Event to the speaker and to the present time. The Auxiliaries, in turn, relate the time of the process to the time of the utterance. In the verbal group *has been eating*, the Finite is realized by *has*, the Event by *eating*, and an Auxiliary by *been*. The logical structure of the verbal group realizes tense, which in English is a complex system.

The adverb group has an adverb as a Head with optional premodifiers and an optional postmodifier. The logical structure parallels that of the nominal group with the premodification composed of submodifiers modifying adjectives which, in turn, modify the adverb as in *much more easily*. Postmodification, when present, is a comparison relationship and the postmodifiers are either embedded clauses or embedded prepositional phrases as in *faster than fifteen knots*.

Above the Clause

The system of interdependencies between clauses is *taxis*, divided into *parataxis* and *hypotaxis*. Halliday defines them as below:

> Parataxis is the linking of elements of equal status. Both the initiating and the continuing element are free, in the sense that each could stand as a functioning whole.
>
> Hypotaxis is the binding of elements of unequal status. The dominant element is free, but the dependent element is not.[12]

Taxis *expands* a clause, creating a type of clause complex.[13] There are three types of expansion: *elaboration, extension*, and *enhancement*. I now describe the hypotactic expansion of clauses.

Elaboration: The meaning of one clause is *elaborated* upon by further description. The finite and non-finite non-restrictive clauses below show elaboration.

(3-2) She was hard at work on the white kitten, which was lying quite still.[14]

(3-3) I worked for a local firm at that time, selling office equipment.[15]

[12](Halliday, 1985, p. 198).

[13]"A sentence can be interpreted as a CLAUSE COMPLEX: a Head [or dominant] clause together with other clauses that modify it." (Halliday, 1985, p. 192).

[14](Halliday, 1985, p. 205).

[15](Halliday, 1985, p. 206).

Extension: When the meaning of a clause is extended, something new is added as an addition, replacement, or alternative. Halliday lists two different categories of extension. Sentence (3-4) illustrates an *additive* extension while a *variative* extension is shown by sentence (3-5).

(3-4) The executioner, the King and the Queen were all talking at once, while all the rest were quite silent.[16]

(3-5) Instead of just working for a living you could be sitting on your backside all day.[17]

Enhancement: By using one of a number of methods—reference to time, place, manner, cause, or condition—one clause *enhances* the meaning of another. Two examples follow:

(3-6) He lives there while he's on the job.[18]—*temporal* enhancement

(3-7) I carry it upside down, so that the rain can't get in.[19]—*causal* enhancement

3.2.2 Cohesion—Halliday and Hasan

Halliday and Hasan (1976) investigated how texts that appeared *connected* differed from those that did not. They classified cohesive phenomena with the relations of *reference, substitution, ellipsis*, and *conjunction.* In addition, they ranked the degree of cohesiveness exhibited by the relations. Substitution and ellipsis are the most cohesive, followed by reference, and then conjunction.

Reference

By *reference*, Halliday and Hasan are alluding to items that cannot be interpreted on their own. Instead, reference is made to a specific item recoverable in the text. The resulting meaning is identical to that of the referred item. There are three types of reference relations.

Personal Reference The role of *personal reference* is assumed by the personal pronouns, the possessive determiners, and the possessive pronouns. Halliday and Hasan discuss how these entities exhibit exophoric (outside of the text) and anaphoric (inside the text) reference. It is the latter that is cohesive. An example of personal reference is shown below in Sentence (3-8), in which the personal pronoun *they* refers back to *the boys.*

(3-8) The boys said that they were going to the store.[20]

[16](Halliday, 1985, p. 208).
[17](Halliday, 1985, p. 208).
[18](Halliday, 1985, p. 215).
[19](Halliday, 1985, p. 215).
[20]Constructed example.

Demonstrative Reference Halliday and Hasan describe *demonstrative reference* as a "form of verbal pointing"[21] by location. This type of reference is realized by the deictic determiners (*this, that, these*, and *those*) and time/space adverbs (*here, now, there*, and *then*).

Comparatives *Comparative reference* is an indirect form of reference accomplished by comparison through an identity or similarity relationship. It is separated into *general comparison*—realized by adjectives (*e.g., same, equal, identical*, and *similar*) and adverbs (*e.g., identically* and *similarly*)—and into *particular comparison*, realized by the adjectives *better* and *more* and the adverbs *so, more, less,* and *equally.*

Substitution

Substitution is the replacement of one item by another in a text. Unlike reference, where both the referred and referring item are interpreted identically, part of what is being replaced can be repudiated in substitution. There are three types of substitution.

Nominal Substitution The lexical items substituting for nominal groups are *one*, *ones*, and *same.*

(3-9) My old watch worked all right, but this <u>one</u> is hopeless.[22]

Verbal Substitution *Do* is the lexical entity that often substitutes for a verb.[23] This use of *do* should not be confused with its use as a lexical verb in its own right.

(3-10) John is smoking more now that Mary is <u>doing</u>.[24]

Clausal Substitution In this type of substitution, an entire clause is replaced by either *so* or *not.*

(3-11) He may come, but he didn't say <u>so</u>.[25]

---

[21](Halliday and Hasan, 1976, p. 59).

[22]Adapted from (Halliday and Hasan, 1976, p. 104).

[23]Halliday and Hasan discuss the effect of dialect on verbal substitution and verbal ellipsis. This type of verbal substitution is more common in British English. This is why example (3-10) may not sound grammatical to all readers.

[24](Halliday and Hasan, 1976, p. 115).

[25](Halliday and Hasan, 1976, p. 139).

Ellipsis

Ellipsis is the absence of some entity in the sentence. The resulting unfilled grammatical spot is responsible for the presupposition of some preceding text. For example, sentence (3-12) has the verb *brought* missing in the second clause.

(3-12) Joan brought some carnations, and Catherine ⊘ some sweet peas.[26]

The same three kinds of ellipsis, as in substitution, are present: *nominal*, *verbal*, and *clausal*. Sentence (3-12) is an example of verbal ellipsis. The next two sentences show nominal and clausal ellipsis, respectively.

(3-13) Four other Oysters followed them, and yet another four ⊘.[27]

(3-14) You can borrow my pen if you want ⊘.[28]

Conjunction

Conjunction is a semantic relationship "express[ing] certain meanings which presuppose the presence of other components in the discourse."[29] The conjunctive relationship is filled by three kinds of sentence adjuncts:

- Simple adverbs (*but, so, then, ...*), compound adverbs ending in *-ly* (*accordingly, actually, ...*), and compound adverbs using *there* and *where* (*therefore, whereat, ...*);
- Other compound adverbs (*nevertheless, anyway, besides*) and prepositional phrases (*on the contrary, as a result, ...*);
- Prepositional phrases with a reference item that may be optional (*as a result of that, as a result of, instead of that, instead of, ...*) or requisite (*in spite of that, because of that, ...*).

The adjuncts are organized by Halliday and Hasan into four different semantic categories—*additive, adversative, causal*, and *temporal*—illustrated by the following example:

(3-15) For the whole day he climbed up the steep mountainside, almost without stopping.[30]
a. And in all this time he met no one. (additive)
b. Yet he was hardly aware of being tired. (adversative)
c. So by night time the valley was far below him. (causal)
d. Then, as dusk fell, he sat down to rest. (temporal)

[26](Halliday and Hasan, 1976, p.143). The symbol ⊘ shows the position of the elided component.
[27](Halliday and Hasan, 1976, p. 148). It is interesting to note that verbal ellipsis (the verb *followed* has been elided) is present as well as nominal ellipsis.
[28](Quirk *et al.*, 1985, p. 909).
[29](Halliday and Hasan, 1976, p. 226).
[30](Halliday and Hasan, 1976, p. 238).

As noted by Green (1992, p. 32), the conjunctive relation is a semantic one. Thus, a syntactic-based style analyzer will be unable to fully exploit the conjunctive relationship.

3.2.3 Contributions From Other Sources

Quirk et al.

We now consider the work of Quirk *et al.*(1985) that is relevant to this thesis.. They distinguish adverbials from other elements by the fact that their position (in a clause) can be varied, that they can be omitted from the clause, and that they cover a spectrum from those that cannot be omitted and are relatively immobile[31] to those which tend to qualify a sentence or clause. Two main groups of adverbials are distinguished: the ones which more resemble complements, the *adjuncts* and *subjuncts*, and the ones which some grammarians call *sentence adverbials*, the *disjuncts* and *conjuncts*. It is characteristic of the latter to be separated from the rest of the clause or sentence by commas.

- **Adjuncts:** These adverbials closely resemble other sentence elements, such as the subject or object, because they can be the focus of cleft sentences, be the focus of the contrast in alternative interrogation or negation, come within the scope of verbal ellipsis, and be elicited by question forms. The "adjuncts are similar in the weight and balance of their sentence role to other sentence elements object."[32]

- **Subjuncts:** These adverbials cannot be treated the same as the adjuncts as they do not resemble other sentence elements. They thus cannot fill the same grammatical functions that were described above. They generally fill a role subordinate to other sentence elements.

- **Disjuncts:** These adverbials differ from the adjuncts in the same manner as the subjuncts. As they modify the clause itself by "comment[ing] on the form or content",[33] they are considered to be syntactically detached from, and superordinate to, the rest of the sentence.

- **Conjuncts:** Like disjuncts, these adverbials are syntactically outside the sentence or clause. Unlike disjuncts, conjuncts comment on the speaker's view of how two linguistic units are connected. They "thus *both* indicate the [connective] relation *and* are demonstratively outside the syntactically integrated clause structure that admits adjuncts."[34]

[31] Quirk *et al.* mention that these type of adverbials resemble complements and are classified as such by some grammarians.
[32] (Quirk *et al.*, 1985, p. 613).
[33] (Quirk *et al.*, 1985, p. 52).
[34] (Quirk *et al.*, 1985, p. 633).

It should be noted that the boundaries between the four classes of adverbials is not always distinct. Quirk *et al.* acknowledge the indeterminacy of grammar by discussing the *gradience* along a grammar scale of some syntactic categories. They point out that the differences between the different types of adverbials and between adverbials and complements display this indeterminacy.

Interpolation

DiMarco and Hirst (1993a) have identified *interpolation* as a syntactic organization that works against sentence cohesion. They define interpolation as "certain instances of parenthetical constructions, those that display none of the forms of cohesion listed earlier [by Halliday and Hasan (1976)],"[35] as in sentence (3-16).

(3-16) Stephen and Jennifer are not going to buy the house, according to a spokesman.[36]

3.3 The Theory

In this section, I present our computational theory of syntactic style; grounded in, and now superceding, the one developed by DiMarco (1990). I will first list the concepts that have been retained from the original theory and then go on to give an overview of the more important differences. The similarities are listed below:

- Style is goal-directed. Ultimately, specific stylistic goals are correlated with specific syntactic constructs in a sentence.
- The concepts of concord and discord have retained their fundamental natures. The syntactic elements either conform to or deviate from a stylistic norm. Conforming elements are concordant and deviating ones are discordant.
- We deal only with the genre of non-literary group style. An example would be high-quality magazine-writing.
- The hierarchical shape of the theory has been retained. There is the primitive-element layer, the abstract-element layer, and the stylistic-goal layer.

The items in the above list were introduced in Section 3.1. Before presenting our theory in more detail, I will now outline how it overcomes some of the shortcomings in DiMarco's (1990) original work.

One of the major problems with the original theory was the lack of formal justification when syntactic constructs were correlated with primitive elements. Work by DiMarco and Hirst (1993a) and by Green (1992) has remedied this and will be discussed in Section 3.3.3.

[35](DiMarco and Hirst, 1993a, p. 37).
[36](DiMarco and Hirst, 1993a, p. 26).

Another shortcoming was the lack of any discussion on the concept of *stylistic norm*, which underlies the definitions of concord and discord. Green (1992) has investigated this concept and the reader is encouraged to consult his work for full details. Here I present a summary.

Two level of norms have been defined: the *primary* norm and the *secondary* norm. Green considers *clarity* to be primary across the entire language. Thus, no matter the genre, the goal of language is to be clear—this is the primary norm. But, as he also points out, this is insufficient to describe all writing. What is clear in one genre can be obscure in another. The secondary norms allow for this variation across genres and are realized through the other stylistic goals.

I now present details of the theory by describing each of its layers in turn.

3.3.1 The Stylistic Goals

As style is goal-directed—that is, it is used for some purpose—sentences realize certain *stylistic goals*. According to our theory of syntactic style, these goals are ultimately defined by syntactic choices. The three goals that have been fully developed are presented below with a brief description of each.

Clarity Writing clearly is emphasized by all the style and how-to-write books. To be clear is a goal usually sought by writers. Being clear is associated with simplicity, harmony, and a lack of ambiguity. Being clear is writing plainly, precisely, and predictably. Thus, the type of sentences that DiMarco and Hirst (1993a) consider clear are simple ones with only one independent clause, centred sentences with a central independent clause surrounded by dependent clauses, and parallel sentences that reduce ambiguity by stressing the same grammatical form.

Concreteness The effect of immediacy in concrete sentences is achieved by arranging sentence components to reflect the order of ideas or events. Syntax is used to mirror events. Certain syntactic components can be emphasized through the use of discordancy and parenthesis.

Staticness A feeling of staticness is associated with uniformity, predictability, and continuity. As these properties leave little room for stylistic variation, sentences with standard simple or strictly balanced structures are considered to be static. Balance within a sentence will be discussed below.

3.3.2 The Abstract Elements

The stylistic goals are much too abstract to be directly correlated with sentence syntax. The *abstract elements* describe the syntactic properties of sentences from a stylistic viewpoint.

They provide a bridge between low-level syntax and high-level goals.

The abstract elements were developed by first classifying commonly used stylistic terms according to three different syntactic properties that influence style. These properties were introduced in Section 3.1.3 and are described below.

Balance: The stylistic effects arising from the juxtaposing of similar of dissimilar sentence structures gives rise to the balance group.

Dominance: The stylistic effects arising from the structural hierarchy of a sentence are of importance in the dominance group.

Position: The stylistic effects produced by the linear position of syntactic elements within a sentence define the position group. Here the fundamental concepts of concord and discord are particularly important and are discussed in more detail when the position abstract elements are presented.

The result of this classification was a list of stylistically significant sentence types, which were then grouped into stylistically similar sentences. The abstract elements are the result of this grouping and are defined below.[37]

The Balance Abstract Elements

These abstract elements are used to describe the stylistic effects achieved by either perturbing or reinforcing the stylistic 'balance' of a sentence. There are two balance abstract elements.

Homopoise: This abstract element is used to describe sentences with interclausal coordination of syntactically similar components. Each component in a homopoisal sentence contributes equally to any underlying parallelism. Two simple clauses are conjoined in the example below.

(3-17) Great Britain opposes it and Holland endorses it.[38]

Heteropoise: The presence of one or more parenthetical components that are syntactically 'detached' and dissimilar from other components at the same level in the syntactic parse tree indicates a heteropoisal sentence. The parenthetical component may be initially, medially, or terminally placed.

(3-18) Stephen and Jennifer are not going to buy the house, according to a spokesman.[39]

[37] These definitions are taken directly from (DiMarco and Hirst, 1993a).
[38] Adapted from (DiMarco, 1990, p. 101).
[39] (DiMarco and Hirst, 1993a, p. 19).

The Dominance Abstract Elements

The dominance abstract elements categorize sentences by the presence and/or absence of independent and dependent clauses. There are three abstract elements in this category.

Monoschematic: A sentence that is monoschematic contains one main independent clause with no dependent clauses. There may be simple phrasal subordination. These sentences are very simple ones, as in the example below.

(3-19) Affinities play their role in this encounter.[40]

Centroschematic: Centroschematic sentences contain a dominant clause serving as the focus or centre of all other components. Such sentences are built up by the subordination and coordination of syntactic structures, such as complex phrasal ones or dependent clausal ones, which do not dominate or weaken the main clause. In sentence (3-20), the relative clause structure is subordinate to the main clause, but does not detract from it.

(3-20) Neither these devices nor the cramped viewing rooms which are too nar- row and whose ceilings are manage
to spoil the works.[41]

Polyschematic: A polyschematic sentence is last in the progression in complexity from monoschematic through centroschematic through polyschematic. These sentences have more than one central dominant clause and at least one dependent clause. As sentence (3-21) illustrates, polyschematics are often difficult to understand.

(3-21) If we consider the progress already achieved, the opposition that had to be overcome, for example, in order to open schools for girls, and the fact that Saudi Arabia is less than 60 years old, we could think that time will permit resolving the contradictions between the most liberal aspirations of one part of society and the ulemas' determination to keep the country as it is, and we should not forget to mention the Islamic fundamentalist movements which are threatening Saudi Arabia.[42]

The Position Abstract Elements

The position abstract elements are concerned with the stylistic effects arising from the linear positions of syntactic constituents. Of the three types of abstract elements, the concepts of concord and discord are most important to this group category. Whether the syntactic components in the initial, medial, and terminal parts of a sentence are concordant, following

[40](DiMarco, 1990, p. 211).
[41]*Manchester Guardian Weekly*, 14 February 1988, p. 15.
[42]Adapted from the *Manchester Guardian Weekly*, 14 February 1988, p. 14.

normal usage, or discordant, deviating from normal usage, determines the categorization of sentences with position abstract elements. They are *initial concord, initial discord, medial concord, medial discord, final concord*, and *final discord.* As an example, consider the initial placement of the adverb *entirely*, as in Sentence (3-22). It is not usually found there and the initial part of such a sentence is thus considered to be incongruous and discordant. It is, therefore, categorized by the abstract element *initial discord.* Additionally, if the rest of this sentence was concordant, then it would also be a *medial concord* and *final concord.*

There are two more abstract elements in this grouping. They are the result of the overall relationships between the concordant and discordant elements within a sentence. The above abstract elements describe local effects of concord and discord while these two describe global ones, in which there is a shift in stylistic effect at the end of the sentence.

Resolution: A sentence that begins discordantly and then shifts to a concord at the end is a resolution. In the example below, the initial discord is due to the initial placement of the adverb *entirely*, while the rest of the sentence is concordant as the main clause contains no such incongruities.

(3-22) Entirely in the spirit of protective support, could we suggest you pass on an appropriate comment to the personnel concerned.[43]

Dissolution: The complementary effect to resolution is that of dissolution. The final shift in stylistic effect is a move from concord to discord. In the following example, the terminal discord is produced by the uncommon syntactic inversion.[44]

(3-23) And the rain descended, and the floods came, and the winds blew, and beat upon that house; and it fell: and great was the fall of it.[45]

3.3.3 The Primitive Elements

The *primitive elements*, as described briefly in Section 3.1, form the basis of our stylistic theory. As in DiMarco's (1990) original theory, they correlate syntactic constituents of sentences with stylistic concepts. In addition, as in DiMarco's original theory, two different views of sentence structure are integral at this primitive-layer level. The first viewpoint is the *connective*, defined by linear bonds between syntactic components. The other viewpoint, the *hierarchical*, is also defined by syntactic bonds within the sentence. In this case, it is the subordinate and superordinate bonds between different levels of the sentence's syntactic hierarchy that are important.

[43](Quirk *et al.*, 1985, p. 652).

[44]The disruption in normal linear order by a terminal syntactic inversion has not yet been incorporated into our formalization of style.

[45]Matthew 7:27, Revised Standard Version (1952).

Each viewpoint has two kinds of primitive elements that categorize the stylistic effects of the relevant syntactic bonds. The connective primitive elements are the *conjuncts*, which identify syntactic components contributing to cohesive bonds within the sentence, and the *antijuncts*, which categorize syntactic components disruptive to the linear organization of the sentence. The hierarchical primitive elements are the *subjuncts*, which classify syntactic components dependent on other sentence components, and the *superjuncts*, which identify syntactic components detached in some way from the rest of the sentence. A particular syntactic component is labelled with a connective primitive element, either a conjunct or an antijunct, and a hierarchical primitive one, either a subjunct or a superjunct.

The primitive elements do more than identify the type of bonds that are formed by syntactic components. They also rank the relative strengths of these bonds. The degree of the bond is indicated with a superscript as in $conjunct^{1}$ where the superscript *1* indicates a mildly connective link with the rest of the sentence. Distinguishing between differing degrees of cohesiveness and subordinateness is necessary in order to obtain the precision needed for a flexible and robust stylistic theory.

I now present each viewpoint in turn, describing how syntactic components are classified with a particular primitive element. This classification is no longer based on intuitional heuristics, as in (DiMarco, 1990). Instead, it has been put on a sound linguistic foundation. The linguistic justifications described below are taken from (Green, 1992), except where noted.

The Connective View

As the connective view is concerned with the amount of cohesiveness that sentence components contribute, we look to Halliday and Hasan's (1976) work on cohesion for linguistic justification. They identified several cohesive relations, which are described and for which examples are given in Section 3.2.2. These relations are used as a starting point in assigning the degree of cohesiveness or conjunctness levels.

Halliday and Hasan rank the relations in terms of their relative degree of cohesiveness. Ellipsis and substitution are considered the most cohesive, followed by reference, and then by conjunction. Green (1992), however, feels that ellipsis is a more concise relationship than substitution and thus more cohesive. Using this information, the following *general* classification scheme for conjunctness is obtained: $conjunct^{4}$ for ellipsis, $conjunct^{3}$ for substitution, $conjunct^{2}$ for reference, and $conjunct^{1}$ for conjunction. The degree of connectness ranges from mildly connective at $conjunct^{1}$ to strongly connective at $conjunct^{4}$.

This general scheme loses some of the subtle distinctions that exist between some of the subrelations. As these distinctions are important at higher levels of the theory, the general scheme has been revised by Green (1992). I present a summary of each of the revised cohesive relations in turn.

Ellipsis Green analyzes the relative conciseness of the three types of ellipsis. He finds that *clausal ellipsis* is extremely cohesive, disruptively so, and classifies it as conjunct^5. His analysis of *verbal ellipsis* is restricted to cases where the entire verbal group has been elided or where only a modal verb remains. These types of verbal ellipsis are strongly connective and are given a conjunct^4 classification. As *nominal ellipsis* often consists of an elided noun with some other part of the nominal group taking on the function of Head, Green considers it to be the least cohesive ellipsis subrelation. He gives it a conjunct^3 classification.

Substitution In classifying the substitution subrelations, Green uses the same conciseness considerations as for ellipsis. And, as with the ellipsis subrelations, *clausal substitution* is considered to be the most concise, with one word, *so* or *not*, replacing a whole clause. Accordingly, clausal substitution is considered to be strongly cohesive at conjunct^4. *Verbal substitution* and *nominal substitution* are considered to be counterparts of one another as both replace one or more words in their respective groups. Both are considered to be moderately connective at conjunct^3.

Reference Green suggests that the relative cohesiveness of the reference subrelations should be based on the ease of referent recoverability. The easier the recovery, the more cohesive the subrelation. Using this criteria, *personal reference* is the most cohesive of the subrelations of reference and is considered to be moderately cohesive with a conjunct^3 classification. The subrelation *demonstrative reference* consists of two types: the deictics and the time/place adverbs. The deictics are more easily recovered, though not always from the present sentence, than the time/place adverbs, whose referents are usually extralinguistic. Thus the deictics are given a conjunct^2 classification and the time/place adverbs a conjunct^1 one.

Conjunction As conjunction is a semantic relationship, isolating the subrelations by syntax only is not possible. As Halliday and Hasan consider this relation to be the least cohesive, Green labels all conjunctive relations as conjunct^1.

Interpolation The effects of disruptive elements in a sentence are also considered in the connective view. They are labelled *antijunct* and are based on the concept of interpolation as developed by DiMarco and Hirst (1993a). All instances of interpolation are considered to be antijunct^2.

The Hierarchical View

The hierarchical view is concerned with the dependence that a particular syntactic element has in relation to other sentence components in a nested ordering. The main concept used in making the subjunct and superjunct classifications is that of *field of reference*, as in DiMarco's (1990, p. 57) original theory:

> The degree to which a syntactic component expands (makes more implicit) or limits (makes more explicit) the field of reference is correlated with its degree of superordination or subordination.

Green (1992) incorporates Halliday's (1985) work on Functional Grammar with the concept of field of reference. He discusses the effects of hypotaxis, rank-shifting or embedding, and the logical structure of the nominal group. We look at each of these in turn.

Hypotaxis Green (1992) interprets the three types of hypotaxis, which are described and for which examples are given in Section 3.2.1, in terms of field of reference. In doing so, he assigns each to a hierarchical primitive element:

- **Elaboration:** In elaboration, dependent clauses further specify, by description, the meaning of the dominant clause, narrowing its field of reference. Green thus considers elaboration to be moderately subordinating and gives it a subjunct^2 classification.

- **Extending:** In extension, a dependent clause extends the meaning of another by adding new information. As something has been added to the primary clause, its field of reference is expanded. Green thus gives this type of clause the mildly superordinating classification of superjunct^1.

- **Enhancement:** In this type of hypotaxis, the meaning of the primary clause is enhanced by specifying it with respect to time, place, manner, cause, or condition. The field of reference of the primary clause is thus reduced. Green, therefore, considers enhancement to be mildly subordination and classifies it as subjunct^1.

Embedding A sentence constituent is said to be *embedded*, that is a constituent functions from within another, when is it rank-shifted—a usually higher-ranked constituent realizes a function at a lower rank. In English, this happens almost always within the nominal group. There are two kinds of embedding: a clause acting as a nominal group and either a clause or prepositional phrase acting as a Qualifier in the postmodification. The embedded Qualifiers are totally dependent on the Head of the nominal group and they further specify it. As a result, Green (1992) considers embedding to be strongly subordinating and has classified it as subjunct^3. He considers the embedded nominal clause, however, to be related to hypotaxis as it is dependent on the rest of the sentence. And Halliday (1985) considers it reasonable to regard nominal clauses as instances of elaboration. To capture the difference between the two types of embedding, Green classifies the nominal clause as moderately subordinating and assigns a subjunct^2 classification to it.

The Logical Structure of the Nominal Group The logical structure of the nominal group is in its premodification. Each additional element to the left in the progression of modifiers further specifies the Head. Green captures this progressive narrowing of the Head's field of reference by having the subjunct-level classification of the adjectival component be correlated with the number of adjectives present, starting at a subjunct[1] classification for one adjective.

3.3.4 Summary

This section has described the theory of syntactic style underlying my stylistic analyzer. We have seen three contributors to the theory: DiMarco (1990), who provided the foundation; DiMarco and Hirst (1993a), who gave us a theory easier to conceptualize and who rooted the theory in formal linguistics; and Green (1992), who refined the primitive elements.

This syntactic theory of style is hierarchical. At the lowest level, the primitive elements are correlated directly with syntax. Here, two viewpoints of sentence structure are used. We saw that (Halliday and Hasan, 1976) was used to justify the classification of the connective sentence components and that (Halliday, 1985) was used to justify that of the hierarchical ones. At the top layer in the theory, we saw the stylistic goals that correspond directly to our view of goal-directed style. And the middle layer's abstract elements, which correlated the very concrete primitive elements with the very abstract stylistic goals, were also presented.

This theory of style has been the basis for the grammar of style, developed for my sentence-style analyzer. This grammar, along with examples, is presented in its entirety in the next chapter.

Chapter 4

A Grammar for Syntactic Style

As a pre-requisite to building a sentence-style analyzer, I developed the grammar for syntactic style that is presented in this chapter. For this task, I used three sources—(DiMarco, 1990), (Green, 1992), and (DiMarco and Hirst, 1993a)—and melded them together into one workable grammar. I used DiMarco as the baseline, keeping to the spirit presented there and using the syntactic coverage as a guide; DiMarco and Hirst for the abstract elements, refined and simplified from that of DiMarco; and Green for the primitive elements. Before outlining the grammar itself, I will give a few explanatory notes—a grammar guide.

4.1 A Grammar Guide

My syntactic-style grammar is based on the theory of style presented in the previous chapter. The hierarchical nature of this theory lends itself naturally to a context-free grammar, the dominant form that I have used for presenting the grammar. Envisioning a *style-tree*, analogous to the syntax tree, is useful because a context-free grammar is used and also because our theory of style closely mirrors sentence syntax. Just as a syntax tree is built up from individual words at the leaf level to a whole sentence at the root level, a style tree can be thought of as being built up from primitive elements at the leaf level to stylistic goals at the root level. The analogy stops here, however, as the stylistic terms do not correlate uniquely to the underlying syntactic constructions. As a sentence can fulfil more than one goal, the style tree may not have a single unique root. The syntax-tree analogy may thus be more useful if we think of the syntax-tree nodes as being annotated with stylistic terms, starting with the leaves and working up through the intermediate nodes to the root of the syntax tree.

An important aspect of the following grammar is my decision to split the primitive layer in two, forming an additional layer in the grammar—one based on practicality rather than theory. This was the direct result of combining three different grammars and trying to simplify (without loss of information) the resultant grammar while ensuring its consistency. This

layer—the *transitional layer*—refers to that part of the style tree that parallels the intermediate syntactic constituents. Here, these syntactic constituents are assigned abstract-element terms, the assignation being based on the primitive elements and/or the transitional elements associated with their composite syntactic constructions. For instance, the transitional elements assigned to the nominal group are based on those assigned to the premodification, the noun, and the postmodification composing the nominal group. The labelling of the premodification, is in turn, based on the primitive elements associated with its composite parts. This labelling with *transitional elements* provides us with information about the types of sentences of which these intermediate constituents can be a part. It is here in the transitional layer that the correlation of the abstract elements with the primitive elements is done. The advantage of the transitional layer, with its transitional elements, is a grammar that is more consistent, more modular, and easier to conceptualize.

The rest of this chapter discusses the full grammar for syntactic style. The next section includes both the primitive and transitional elements while the abstract elements and stylistic goals are presented in following sections. The reader is referred to Appendix A for a discussion on the terminology used in the grammar.

4.2 The Primitive and Transition Layers

In this section, the primitive and transitional elements are presented. I discuss justifications for the primitive elements, present explanatory material as needed, and include examples where appropriate.

The primitive elements and the transitional elements are discussed simultaneously because both are closely tied to the underlying syntax of the sentence. The grammar is therefore divided by the different syntactic categories found within a sentence. The primitive elements—from both the connective and the hierarchical views—and the transitional elements are given for each category. As the primitive elements apply only to basic syntactic units, the more complex syntactic constructions will have only transitional elements. The only category without any transitional elements is the adjectivals.

4.2.1 Adjectivals

adjectival $\longrightarrow$

(definite article OR indefinite article OR demonstrative determiner OR premodifying genitive) (adjective)*

The Primitive Elements—Connective View

There are three rules for the adjectivals in the connective view. The first one deals with the mildly connective adjectivals. As adjectives and non-demonstrative determiners presuppose

the presence of a noun, they are instances of conjunction. They are given the same conjunct[1] classification as given to the conjunctive cohesive relationship.

conjunct[1] adjectival $\longrightarrow$

adjective

medieval subject

definite article

the medieval subject

indefinite article

a medieval subject

As the demonstrative determiners are instances of deictic constructions, they are given the conjunct[2] classification.

conjunct[2] adjectival $\longrightarrow$

demonstrative determiner

this substantial selection

The final rule for the connective adjective is an example of personal reference. Thus, the premodifying genitive is given the conjunct[3] label.

conjunct[3] adjectival $\longrightarrow$

premodifying genitive

his religious works

The Primitive Elements—Hierarchical View

There are three rules for the adjectivals in the hierarchical view. The first deals with the subjunct[1] adjectivals. The premodifying genitive, demonstrative determiner, and the indefinite determiner all narrow the field of reference to a specific subset of things and thus are considered mildly subordinating.

subjunct[1] adjectival $\longrightarrow$

premodifying genitive

his religious works

demonstrative determiner

this substantial selection

indefinite article

a substantial selection

The definite article narrows the field of reference to a particular instance and so is given the moderately subordinating subjunct^2 level.

subjunct^2 adjectival $\longrightarrow$

definite article

The final rule deals with a series of adjectives used together as nominal modifiers. The addition of each adjective further narrows the field of reference. To capture the notion of increasing specification, the level of subjunct classification is increased by one with each additional adjective. A limit of four has been placed on the serial adjectives.

subjunct^i adjectival $\longrightarrow$

(adjective)^i *where* $1 \leq i \leq 4$

4.2.2 Premodification

premodification $\longrightarrow$

noun

adverbial

participle

adjectival

reduced sentence

The Primitive Elements—Connective View

The first two alternatives in the following rule for premodification cover the cases where none of the cohesive relations are exhibited. As these cases work neither *for* nor *against* cohesion, they are given the neutral conjunct^0 classification. The third alternative has been included to ensure that a sentence without any premodification in its nominal groups will be considered as a candidate for appropriate abstract elements and stylistic goals.

conjunct^0 premodification $\longrightarrow$

adverbial

increasingly arresting self-portraits

participle

arresting self-portraits

NO premodification

The premodifying noun contributes to the cohesion of a nominal group. The source of this cohesion, however, is more lexical than syntactic. Therefore, in this grammar the premodifying noun is considered as an adjective, resulting in a conjunct[1] classification.

conjunct[1] premodification $\longrightarrow$

noun

fake cloth backdrops

The conjunctness of the premodification in the second rule is directly determined by that of the adjectivals.

conjunct[i] premodification $\longrightarrow$

conjunct[i] adjectival *where* $1 \leq i \leq 3$

The third rule for premodification gives us an example that works against cohesion. A reduced sentence is an instance of interpolation and is given a antijunct classification.

antijunct[2] premodification $\longrightarrow$

reduced sentence

his pop-down-for-the-weekend cottage

The Primitive Elements—Hierarchical View

The first rule for premodification in the hierarchical view covers sentences without premodification in their nominal groups. As this is neither a subordinating nor a superordinating structure, a neutral subjunct[0] classification is given.

subjunct[0] premodification $\longrightarrow$

NO premodification

The second rule for subjunct adjectivals is self-explanatory.

subjunct[i] premodification $\longrightarrow$

subjunct[i] adjectival *where* $1 \leq i \leq 4$

The third rule covers the embedded noun, the adverbial, the participle, and the reduced-sentence structures. As the grammatical function of these syntactic constructs has been shifted *down* to that of a premodifier, they are embedded and given a subjunct^3 classification.

subjunct^3 premodification $\longrightarrow$

noun

adverbial

participle

reduced sentence

The Transitional Elements

The following rules begin to correlate primitive elements to abstract elements. The different types of premodification, as classified by the primitive elements, are in turn classified by transitional elements, identifying the abstract elements that can contain these syntactic constructions.

When premodification is conjunct and/or subjunct, without being excessively so, it is considered to be normal usage. As a centroschematic sentence is a *normal* sentence, these types of premodification become centroschematic premodification.

centroschematic premodification $\longrightarrow$

conjunct^i premodification *where* $0 \leq i \leq 3$

subjunct^i premodification *where* $0 \leq i \leq 3$

Monoschematic premodification includes all subjunct primitive elements that are also concordant, as these types of premodification only contain simple subordinate structures.

monoschematic premodification $\longrightarrow$

subjunct^i premodification *where* $0 \leq i \leq 3$

In the connective view, the use of non-excessive cohesion (conjunct^0 to conjunct^4) is normal usage and thus concordant. Interpolation and excessive cohesion (conjunct^5 and conjunct^6), on the other hand, are considered discordant. In the hierarchical view, discord is produced by excessive subordination (subjunct^4), or superordination (superjunct). This gives us the following rules for concordant and discordant premodification.

concordant premodification $\longrightarrow$

conjunct^i premodification *where* $0 \leq i \leq 3$

subjunct^i premodification *where* $0 \leq i \leq 3$

discordant premodification $\longrightarrow$

antijunct^i premodification *where* $1 \leq i \leq 3$

subjunct^4 premodification

4.2.3 Nouns

The lexical nouns do not contribute to syntactic style in our grammar. To incorporate the stylistic effects of nominal ellipsis and substitution, however, I have made use of the *noun* category, resulting in the following rule. As nominal ellipsis and substitution are phenomena in the connective view, there are no rules for the noun in the hierarchical view.

noun $\longrightarrow$

nominal ellipsis

nominal substitution

lexical_noun

pronoun

The Primitive Elements—Connective View

As the lexical noun does not contribute to sentence cohesion on its own, it has been assigned a conjunct^0 classification.

conjunct^0 noun $\longrightarrow$

lexical_noun

his long black <u>cloak</u>

The pronoun (a personal reference item), nominal substitution, and nominal substitution are moderately connective and are given the conjunct^3 classification.

conjunct^3 noun $\longrightarrow$

nominal substitution

his long black <u>one</u>

nominal ellipsis

The first expedition was quickly followed by another ∅.

pronoun

I told him that <u>he</u> was wrong

The Transitional Elements

The following rules show that nominal ellipsis, nominal substitution, and a pronoun can be part of centroschematic and concordant sentences. As the stylistic effect of these constructions supercede that of premodification and postmodification, which tends to be minimal in these cases, I have not included the lexical noun in the following rules. This will allow the conjunct^3 noun to be dominant in the grammar at the nominal-group level.

centroschematic noun $\longrightarrow$

conjunct^3 noun

concordant noun $\longrightarrow$

conjunct^3 noun

4.2.4 Postmodification

postmodification $\longrightarrow$

prepositional phrase

relative clause

nominal group

non-finite clause

verbless clause

adjectival

The Primitive Elements—The Connective View

There are five rules for postmodification in the connective view. The first handles the common case where postmodification is non-existent. As in premodification, it has been included to ensure the inclusion of sentences without postmodification in the appropriate abstract elements.

conjunct^0 postmodification $\longrightarrow$

NO postmodification

The second rule deals with a prepositional phrase acting as a postmodifier. Presupposing the existence of a modified noun, it is considered to be an instance of conjunction and is therefore classified as conjunct^1.

conjunct^1 postmodification $\longrightarrow$

prepositional phrase

his long black cloak with purple beading

As a relative clause is a reference item, referring back to the noun that it modifies, it is given the conjunct[2] classification.

conjunct[2] postmodification $\longrightarrow$

relative clause

the man who founded the religious association

There is one alternative for the conjunct[3] rule. It describes the case where the postmodification consists of an apposition which, due to the fact that it could stand for the appositive noun, is considered to be an instance of nominal substitution.

conjunct[3] postmodification $\longrightarrow$

nominal group

Paul Jones, the distinguished art critic, died in his sleep last night.

There are three alternatives for strongly connective postmodification. All are considered instances of verbal ellipsis.

conjunct[4] postmodification $\longrightarrow$

non-finite clause

You will look for vain for concrete measures emerging from this summit.

verbless clause

Norman Jones, then a student, wrote several best-sellers.

adjectival AND parenthesis

The people, very annoyed, are walking into the room.

There are two alternatives for the antijunct rule. The first, the adjectival, is an instance of both anastrophe and interpolation. Non-cohesive parenthesis is marked off by punctuation stronger than commas.[1] This type of parenthesis is very disruptive and supercedes any cohesive relationship, such as ellipsis, that might also be present.

antijunct[2] postmodification $\longrightarrow$

adjectival

a man always timid is unfit for this task.

non-cohesive parenthesis

Norman Jones—then a student—wrote several best-sellers.

[1]Examples of such punctuation would be the use of dashes or parentheses.

The Primitive Elements—Hierarchical View

There are four rules for postmodification in the hierarchical view. The first one deals with the case of no postmodification.

subjunct^0 postmodification $\longrightarrow$

NO postmodification

There is one alternative for the subjunct^1 rule. As the adjectival limits the field of reference by further specifying the noun being modified, it is considered subjunct^1.

subjunct^1 postmodification $\longrightarrow$

adjectival

The non-restrictive relative clause is considered to be an instance of hypotactic elaboration. Green (1992) describes these as clauses with *which*, *who*, or *whose*, whose domain is a nominal group.

subjunct^2 postmodification $\longrightarrow$

non-restrictive relative clause

Then he met Mary, who invited him to a party.

There are four alternatives for strongly subordinating postmodification. The restrictive relative clause and the non-finite clauses are clauses acting as nominal modifiers and thus are embedded. Likewise, the prepositional phrase, acting as a nominal modifier, is embedded. Green (1992) also considers apposition to be acting as a modifier within another nominal group. As such, it is an instance of embedding.

subjunct^3 postmodification $\longrightarrow$

restrictive relative clause

That's the girl that he met at the party.

non-finite clause

prepositional phrase

nominal group

The Transitional Elements

As in the premodification rules, subjunct and conjunct primitive elements that are not excessively cohesive or subordinate are part of centroschematic sentences.

centroschematic postmodification $\longrightarrow$

conjunct^{i} postmodification *where* $0 \leq i \leq 4$

subjunct^{i} postmodification *where* $0 \leq i \leq 3$

In the following alternatives for heteropoisal postmodification, I have specified that the postmodification must be parenthetical. This is to ensure that a sentence identified as an heteropoise contains the syntactically 'detached' parenthetical component that defines heteropoises.

heteropoisal postmodification $\longrightarrow$

conjunct^{i} postmodification WITH parenthesis *where* $1 \leq i \leq 6$

antijunct^{i} postmodification WITH parenthesis *where* $1 \leq i \leq 3$

Monoschematic sentences have no dependent clauses. Therefore, a lack of modification, as specified in the first alternative, and simple modification, as specified in the second alternative, are included in the following rule for monoschematic postmodification.

monoschematic postmodification $\longrightarrow$

subjunct^{0} postmodification

subjunct^{3} postmodification AND (nominal group OR prepositional phrase)

The following rules assign the appropriate primitive elements to concordant or discordant postmodification as in the rules for premodification.

concordant postmodification $\longrightarrow$

conjunct^{i} postmodification *where* $0 \leq i \leq 4$

subjunct^{i} postmodification *where* $0 \leq i \leq 3$

discordant postmodification $\longrightarrow$

conjunct^{i} postmodification *where* $5 \leq i \leq 6$

antijunct^{i} postmodification *where* $1 \leq i \leq 3$

subjunct^{4} postmodification

4.2.5 Nominal Groups

nominal group $\longrightarrow$

(premodification)* noun (postmodification)*

The Transitional Elements

In the rules for the nominal group, we see for the first time how the assignation of abstract-element labels depends on that of component syntactic constructs. The general rule is that the labelling of a syntactic constituent must agree with that of *all* its composite constituents. Discordant components, however, need only one internal discordant component. Other exceptions to this general rule are noted when encountered.

The first alternative in the following rule for the centroschematic nominal group allows nominal substitution and ellipsis to dominate the classification of the nominal group despite the presence of any premodification or postmodification.

centroschematic nominal group $\longrightarrow$

nominal group WITH centroschematic noun

nominal group WITH (centroschematic premodification AND centroschematic postmodification)

The following rule for the heteropoisal nominal group shows that the only contribution in this case comes from a heteropoisal postmodification.

heteropoisal nominal group $\longrightarrow$

nominal group WITH heteropoisal postmodification

The following rule for the monoschematic nominal group is self-explanatory.

monoschematic nominal group $\longrightarrow$

nominal group WITH (monoschematic premodification AND monoschematic postmodification)

As in the rule for the centroschematic nominal group, the first alternative in the rule for the concordant nominal group allows nominal substitution and ellipsis to supercede any premodification and postmodification that may be present.

concordant nominal group $\longrightarrow$

nominal group WITH concordant noun

nominal group WITH (concordant premodification AND concordant postmodification)

The following rule for the discordant nominal group is self-explanatory.

discordant nominal group $\longrightarrow$

nominal group WITH (discordant premodification OR discordant postmodification)

4.2.6 Nominal Clauses

For the purposes of my grammar, I consider any dependent clause that *takes the place of* a nominal group to be a nominal one. The rule below shows the types of clauses that can do this.

nominal clause $\longrightarrow$

restrictive relative clause

wh-interrogative clause

yes-no interrogative clause

non-finite clause

participle clause

The Primitive Elements—Connective View

The grammar looks at the internal structure of a nominal clause for only two reasons: to determine whether a subject is present or not and to determine whether there are any discordant internal components. These are the only two ways in which a nominal clause can contribute to the syntactic style of a sentence. The lack of a subject is an instance of nominal ellipsis; hence the conjunct^3 classification.

conjunct^0 nominal clause $\longrightarrow$

nominal clause WITH subject

conjunct^3 nominal clause $\longrightarrow$

nominal clause WITHOUT subject

The Primitive Elements—Hierarchical View

All nominal clauses are given a subjunct^2 classification due to their classification as an instance of hypotactic elaboration.

subjunct^2 nominal clause $\longrightarrow$

restrictive relative clause

That she is still alive is a consolation

wh-interrogative clause

How the book will sell depends on its author.

yes-no interrogative clause

Whether she likes the present is not clear to me.

non-finite clause

For you to tell everybody is the best thing.

participle clause

Telling lies is wrong.

The Transitional Elements

The following rule is to ensure the inclusion of sentences with nominal clauses acting as subjects as a legitimate centroschematic sentence.

centroschematic nominal clause $\longrightarrow$

conjunct^0 nominal clause

conjunct^3 nominal clause

subjunct^2 nominal clause

The following two rules for concordant and discordant nominal clauses are self-explanatory.

concordant nominal clause $\longrightarrow$

nominal clause WITH ALL components concordant

discordant nominal clause $\longrightarrow$

nominal clause WITH ANY component discordant

As a sentence with a subject is normal usage, a sentence is considered discordant if it lacks one. The following rule will allow us to identify, at higher levels of the grammar, the lack of the subject that occurs when a subjectless nominal clause functions as the sentence subject. It should be noted that a subjectless clause, itself, is *not* inherently discordant.

initial_discordant nominal clause $\longrightarrow$

conjunct^3 nominal clause

4.2.7 Noun Phrases

For the purposes of this grammar I have included both the nominal group, the traditional noun phrase, and the nominal clause under the syntactic category of noun phrase. All rules for the noun phrase follow from rules presented earlier.

noun phrase $\longrightarrow$

nominal group

nominal clause

The Transitional Elements

centroschematic noun phrase $\longrightarrow$

centroschematic nominal group

centroschematic nominal clause

heteropoisal noun phrase $\longrightarrow$

heteropoisal nominal group

monoschematic noun phrase $\longrightarrow$

monoschematic nominal group

concordant noun phrase $\longrightarrow$

concordant nominal group

concordant nominal clause

discordant noun phrase $\longrightarrow$

discordant nominal group

discordant nominal clause

initial_discordant noun phrase $\longrightarrow$

initial_discordant nominal clause

4.2.8 Adverbials

adverbial $\longrightarrow$

adverb

adverbial phrase

The Primitive Elements—Connective View

There is only one rule for adverbials in the connective view. As adverbials presupposes the existence of a component that is being modified, they are considered to be instances of a conjunctive cohesive relationship. Accordingly, they are given a conjunct[1] classification.

conjunct[1] adverbial $\longrightarrow$

ANY adverbial

The Primitive Elements—Hierarchical View

The rules for the adverbials in the hierarchical view are based on the classification of adverbs by Quirk *et al.* (1985). They use the terms *adjunct*, *subjunct*, *disjunct*, and *conjunct*. The terms *subjunct* and *conjunct* in the stylistic-grammar rules should not be confused with their use by Quirk *et al.*

As *adjuncts* are "similar in the weight and balance of their sentence role to other sentence elements such as subject and object"[2], they do not contribute to any subordination in the sentence. The following rule states this.

subjunct[0] adverbial $\longrightarrow$

adverbial WITH adjunct adverb

As *subjunct adverbs* are "subordinate to one or other of the sentence elements"[3], they are moderately subordinating. They are given a subjunct[2] classification.

subjunct[2] adverbial $\longrightarrow$

adverbial WITH subjunct adverb

Disjuncts are "syntactically more detached and in some respects 'superordinate'"[4]. As they are, by definition, detached from the rest of the sentence, they are classified as superjunct[2] in the rule below. In addition, *conjunct adverbs* also are "relatively detached and 'superordinate' "[5] and form the second alternative for the superjunct[2] adverbial.

superjunct[2] adverbial $\longrightarrow$

adverbial WITH disjunct adverb

adverbial WITH conjunct adverb

[2](Quirk *et al.*, 1985, p. 613).
[3](Quirk *et al.*, 1985, p. 613).
[4](Quirk *et al.*, 1985, p. 613).
[5](Quirk *et al.*, 1985, p. 631).

The Transitional Elements

As in other syntactic categories, the centroschematic adverbial is a typical one. Thus, the following rule applies.

centroschematic adverbial $\longrightarrow$

conjunct^1 adverbial

subjunct^0 adverbial

subjunct^2 adverbial

The following rule for the heteropoisal adverb stems from the definition of a heteropoisal sentence as one with syntactically 'detached' components.

heteropoisal adverbial $\longrightarrow$

superjunct^2 adverbial

4.2.9 Prepositional Phrases

prepositional phrase $\longrightarrow$

preposition nominal group

The Transitional Elements

The rules for the prepositional phrase are self-explanatory as they are completely dependent on the classification of the nominal group within the prepositional phrase.

centroschematic prepositional phrase $\longrightarrow$

prepositional phrase WITH centroschematic nominal group

heteropoisal prepositional phrase $\longrightarrow$

prepositional phrase WITH heteropoisal nominal group

monoschematic prepositional phrase $\longrightarrow$

prepositional phrase WITH monoschematic nominal group

concordant prepositional phrase $\longrightarrow$

prepositional phrase WITH concordant nominal group

discordant prepositional phrase $\longrightarrow$

prepositional phrase WITH discordant nominal group

4.2.10 Complement

complement $\longrightarrow$

adjectival

prepositional phrase

noun phrase

The Transitional Elements

As with the prepositional phrase, the grammar rules for complements are self-explanatory.

centroschematic complement $\longrightarrow$

centroschematic noun phrase

centroschematic prepositional phrase

adjectival

heteropoisal complement $\longrightarrow$

heteropoisal noun phrase

heteropoisal prepositional phrase

monoschematic complement $\longrightarrow$

NO complements

monoschematic noun phrase

monoschematic prepositional phrase

adjectival

concordant complement $\longrightarrow$

NO complements

concordant noun phrase

concordant prepositional phrase

adjectival

discordant complement $\longrightarrow$

discordant noun phrase

discordant prepositional phrase

4.2.11 Verbs

Though lexical verbs do not make any stylistic contribution to the grammar, I have included the following rule for verbs to allow for the inclusion of verbal substitution and verbal ellipsis.

verb $\longrightarrow$

lexical_verb

verbal substitution

verbal ellipsis

The Primitive Elements—Connective View

The first rule assigns a conjunct^0 classification to the lexical verb, as it makes no contribution to sentence cohesion.

conjunct^0 verb $\longrightarrow$

lexical_verb

The next two alternatives assign degrees of conjunctness to verbal substitution and ellipsis as described by Green (1992).

conjunct^3 verb $\longrightarrow$

verbal substitution

conjunct^4 verb $\longrightarrow$

verbal ellipsis

The Transitional Elements

As in the part of the grammar dealing with the nouns, verbal substitution and ellipsis can be part of centroschematic sentences. And, also as in the noun section, the lexical verb is not included to allow verbal substitution and ellipsis to dominate during the classification of verbal phrases later in the grammar.

centroschematic verb $\longrightarrow$

conjunct^i verb *where* $3 \leq i \leq 4$

concordant verb $\longrightarrow$

conjunct^i verb *where* $3 \leq i \leq 4$

4.2.12 Verb Phrases

verb phrase $\longrightarrow$

(adverbial)* verb (adverbial)* (complement)*

The Transitional Elements

The first alternative in the rule for centroschematic verb phrases incorporates verbal substitution and ellipsis. The effect of these constructions supercedes the effect of any complements that may be present. The second alternative shows that a verb phrase must have at least one centroschematic complement in order to be part of a centroschematic sentence.

centroschematic verb phrase $\longrightarrow$

verb phrase WITH centroschematic verb

verb phrase WITH AT LEAST ONE centroschematic complement

The following rule shows that the only construction that contributes to the heteropoisal verb phrase is an adverbial.

heteropoisal verb phrase $\longrightarrow$

verb phrase WITH heteropoisal adverbial

The following rule for monoschematic verb phrases is self-explanatory.

monoschematic verb phrase $\longrightarrow$

verb phrase WITH ALL complements monoschematic

The first alternative in the rule for a concordant verb phrase is to allow the concordant effects of verbal substitution and ellipsis to supercede the effects of a discordant complement, if present.

concordant verb phrase $\longrightarrow$

verb phrase WITH concordant verb

verb phrase WITH ALL complements concordant

The following rule for a discordant verb phrase is self-explanatory.

discordant verb phrase $\longrightarrow$

verb phrase WITH ANY discordant complement

verb phrase WITH heteropoisal adverbial

4.2.13 Adverbial Clauses

The Primitive Elements—Connective View

In the connective view, the adverbial clauses are conjunct[1] for the same reason as the adverbs—they are instances of the cohesive relationship of conjunction.

conjunct[1] adverbial clause $\longrightarrow$

ANY adverbial clause

The Primitive Elements—Hierarchical View

The classification of adverbial clauses in the hierarchical view is based on Green's (1992) view of how the different types of hypotaxis contribute to subordination and superordination. I have had to make some assumptions here due to the difficulty of distinguishing the different types of hypotaxis by syntax alone. These assumptions are discussed with the appropriate rule.

The first rule is for the subjunct[1] adverbial clause. Here, adverbs which introduce the adverbial clause are all considered hypotactic conjunctions. As these adverbs are all enhancing by manner, time, location, and cause, these adverbial clauses are hypotactic enhancements, which are subjunct[1] constructions.

subjunct[1] adverbial clause $\longrightarrow$

finite adverbial clause INTRODUCED BY hypotactic conjunction

Whenever the horse stopped, he fell off in front.

non-finite adverbial clause INTRODUCED BY hypotactic conjunction

Finally stopping the horse, he fell off in front.

In the following rule, I have assumed that a finite adverbial clause introduced by *when* or *where* is an elaboration. These adverbs can also introduce enhancements but the two types of hypotaxis cannot be distinguished without the aid of semantics. Thus, I give the following rule for subjunct[2] adverbial clause.

subjunct[2] adverbial clause $\longrightarrow$

finite adverbial clause INTRODUCED BY *when* OR *where*

You'll find the sugar where the coffee is.

I make the further assumption that *while*, though it can introduce an enhancement, introduces a finite extension. *Whereas* also introduces finite extensions. The result is the one alternative below for superjunct[1] hypotactic extensions.

superjunct[1] adverbial clause $\longrightarrow$

finite adverbial clause INTRODUCED By *whereas* OR *while*

William has poor eyesight <u>*whereas Sharon has poor hearing*</u>*.*

The Transitional Elements

The following two rules are self-explanatory. The reader will notice that the primitive-element analysis of the adverbial clauses is not used here at the transitional level. Our grammar is not yet refined enough to take advantage of all the information obtainable at the primitive-element level. This is true for the other dependent clauses dealt with by this grammar.

concordant adverbial clause $\longrightarrow$

adverbial clause WITH ALL components concordant

discordant adverbial clause $\longrightarrow$

adverbial clause WITH ANY component discordant

4.2.14 Non-finite Clauses

The Primitive Elements—Connective View

A non-finite clause without a subject is an instance of nominal ellipsis.

conjunct^0 non-finite clause $\longrightarrow$

non-finite clause WITH subject

conjunct^3 non-finite clause $\longrightarrow$

non-finite clause WITHOUT subject

The Primitive Elements—Hierarchical View

A non-finite enhancement is introduced by a preposition, such as *on, with*, or *by*, that is functioning conjunctively. This gives the subjunct^1 rule for non-finite clauses.

subjunct^1 non-finite clause $\longrightarrow$

non-finite clause INTRODUCED BY preposition

<u>*With his being away*</u>*, everyone works harder.*

As with the adverbial clauses, I have had to make assumptions in order to distinguish the different types of non-finite hypotaxis. Here, any non-finite clause that does not have an introducing preposition will be considered to be an elaboration, though it could also be an enhancement.

subjunct^2 non-finite clause $\longrightarrow$

non-finite clause WITHOUT introducing preposition

I worked for a local firm at that time, selling office equipment.

The Transitional Elements

The following two rules for concordant and discordant non-finite clauses are self-explanatory. As with the adverbial clauses, no use is made of primitive-element information.

concordant non-finite clause $\longrightarrow$

non-finite clause WITH ALL components concordant

discordant non-finite clause $\longrightarrow$

non-finite clause WITH ANY component discordant

The final rule for non-finite clauses allows the grammar to detect a subjectless clause in the initial position of a sentence. Though such a non-finite clause is not inherently discordant itself, its initial position is.

initial discordant non-finite clause $\longrightarrow$

conjunct^3 non-finite clause

4.2.15 Relative Clauses

relative clause $\longrightarrow$

wh-relative clause

He walks for an hour each morning, which would bore me.

The Primitive Elements—Connective View

The dependent relative clause is considered an instance of reference due to the presence of the relative pronoun. It has been given the conjunct^2 classification.

conjunct^2 relative clause $\longrightarrow$

wh-relative clause

The Primitive Elements—Hierarchical View

Because the dependent relative clause is a finite hypotactic elaboration, it is classified as subjunct^2.

subjunct^2 relative clause $\longrightarrow$

wh-relative clause

The Transitional Elements

The following rules for the relative clauses are self-explanatory. Again, note that primitive-element information is not used.

concordant relative clause $\longrightarrow$

relative clause WITH ALL components concordant

discordant relative clause $\longrightarrow$

relative clause WITH ANY component discordant

4.2.16 Dependent Clauses

I have included the following rule for dependent clauses for two reasons. First, we are not using the distinctive stylistic effects of each of the dependent clauses. We are only interested in whether a dependent clause is concordant or not. And second, the inclusion of clausal substitution and ellipsis in the grammar can be done in a painless and elegant manner.

clause $\longrightarrow$

clausal substitution

clausal ellipsis

adverbial clause

non-finite clause

relative clause

The Primitive Elements—Connective View

The following two rules for substituted and elided clauses are from (Green, 1992).

conjunct[4] clause $\longrightarrow$

clausal substitution

Brett's work is not yet consistent in style and quality, but will no doubt become so.

conjunct[5] clause $\longrightarrow$

clausal ellipsis

You can borrow my pen if you want $\oslash$.

The Transitional Elements

The only thing to note about the following two rules for concordant and discordant clauses is that the extreme cohesiveness of clausal ellipsis makes a sentence discordant.

concordant clause $\longrightarrow$

conjunct[4] clause

concordant adverbial clause

concordant non-finite clause

concordant relative clause

discordant clause $\longrightarrow$

conjunct[5] clause

discordant adverbial clause

discordant non-finite clause

discordant relative clause

The following rule allows the grammar to retain the information from the primitive-element level that decides whether a sentence with an initial dependent clause is subjectless or not.

initial_discordant clause $\longrightarrow$

initial_discordant non-finite clause

4.2.17 Majors

The following rule covers simple sentences, the *majors*, without any dependent clauses.

major $\longrightarrow$

(conjunction) (adjective)* (adverbial)* (prepositional phrase)* (nominal group)* noun phrase verb phrase

The Transitional Elements

As in the other more complicated syntactic constituents, the abstract-element labelling of a major depends on that of its constituent parts. The noun phrase and verb phrase make the most important contribution to the syntactic style of a major.

When both the noun phrase and the verb phrase can be part of a centroschematic sentence, then the major is one, as demonstrated by the following rule.

centroschematic major $\longrightarrow$

major WITH centroschematic noun phrase AND centroschematic verb phrase

As a heteropoisal contains a detached parenthetical constituent, it is sufficient to have either a noun phrase, verb phrase, or adverbial containing such a component. All three cases are covered by the rule below for a heteropoisal major.

heteropoisal major $\longrightarrow$

major WITH heteropoisal noun phrase

major WITH heteropoisal verb phrase

major WITH heteropoisal adverbial

The following three rules give information about the positional placement of the parenthetical component in the heteropoisal major.

initial heteropoisal major $\longrightarrow$

major WITH heteropoisal noun phrase

major WITH heteropoisal adverbial

medial heteropoisal major $\longrightarrow$

major WITH heteropoisal noun phrase AND NO heteropoisal adverbial

final heteropoisal major $\longrightarrow$

major WITH heteropoisal verb phrase

The following rule for the monoschematic major is self-explanatory.

monoschematic major $\longrightarrow$

major WITH monoschematic noun phrase AND monoschematic verb phrase

The rule for concordant major and the first alternative for the discordant major are self-explanatory. The second alternative is to cover the situation when there is no subject at all. This occurs when a nominal clause with a nominal ellipsis acts as subject in the sentence.

concordant major $\longrightarrow$

major WITH ALL components concordant

discordant major $\longrightarrow$

major WITH ANY component discordant

major WITH initial_discordant noun phrase

4.2.18 Completes

The following rules are for the more complicated sentences containing dependent clauses. It should be noted that the major is trivially complete.

complete $\longrightarrow$

(clause)* major (clause)*

The Transitional Elements

The rule for a monoschematic complete is self-explanatory.

monoschematic complete $\longrightarrow$

monoschematic major

In the following rule, note the use of a concordant, rather than a centroschematic, clause when defining a centroschematic complete. This is because the only information we currently have about the internal structure of a clause is its concordancy or discordancy.

centroschematic complete $\longrightarrow$

(concordant clause)* centroschematic major (concordant clause)*

The following rule for polyschematic complete ensures that there is at least one dependent clause present.

polyschematic complete $\longrightarrow$

(concordant clause)$^+$ centroschematic major (concordant clause)*

(concordant clause)* centroschematic major (concordant clause)$^+$

The following rules for the heteropoisal complete are self-explanatory.

heteropoisal complete $\longrightarrow$

(clause)* heteropoisal major (clause)*

initial heteropoisal complete $\longrightarrow$

initial heteropoisal major (clause)*

medial heteropoisal complete $\longrightarrow$

(clause)* medial heteropoisal major (clause)*

final heteropoisal complete $\longrightarrow$

(clause)* final heteropoisal major

The following rule for a concordant complete is self-explanatory.

concordant complete $\longrightarrow$

complete WITH ALL components concordant

The following rules are concerned with the concordancy of a particular linear position in the complete. The astute reader will notice that in the case of more than one clause in the initial and final positions for the initial concordant complete and final concordant complete, we require all of these clauses to be concordant. This somewhat simplifies matters, but is the result of the coarseness of our grammar.

initial concordant complete $\longrightarrow$

concordant major (clause)*

(concordant clause)* major (clause)*

medial concordant complete $\longrightarrow$

(clause)* concordant major (clause)*

finial concordant complete $\longrightarrow$

(clause)* concordant major

(clause)* major (concordant clause)*

The following rules for the discordant completes are analogous to those for concordant completes. The same note about the coarseness of the grammar applies to the positional discordant completes.

discordant complete $\longrightarrow$

complete WITH ANY component discordant

initial discordant complete $\longrightarrow$

discordant major (clause)*

(discordant clause)* major (clause)*

(initial discordant clause)* major (clause)*

medial discordant complete ⟶

(clause)* discordant major (clause)*

final discordant clause ⟶

(clause)* discordant major

(clause)* major (discordant clause)*

4.3 The Abstract Elements

In this section I give the grammar rules for the abstract elements. For the most part, the rules are self-explanatory extensions of the rules for the completes in the last section. I add comments where needed.

4.3.1 The Balance Abstract Elements

A homopoise is an "interclausal coordination of syntactically similar components".[6] In the following rule, we coordinate only very basic sentences, the monoschematic completes.

homopoise ⟶

monoschematic complete (monoschematic complete)$^+$

heteropoise ⟶

initial heteropoisal complete

medial heteropoisal complete

final heteropoisal complete

4.3.2 The Dominance Abstract Elements

monoschematic ⟶

monoschematic complete

centroschematic ⟶

centroschematic complete

[6](DiMarco and Hirst, 1993a, p. 17).

A polyschematic contains more than one central dominant clause, of which at least one contains a subordinate clause. Thus we have the following rule with its two alternatives.

polyschematic $\longrightarrow$

(centroschematic complete)$^+$ polyschematic complete (centroschematic complete)*

(centroschematic complete)* polyschematic complete (centroschematic complete)$^+$

4.3.3 The Position Abstract Elements

initial concord $\longrightarrow$

initial concordant complete

medial concord $\longrightarrow$

medial concordant complete

final concord $\longrightarrow$

final concord complete

initial discord $\longrightarrow$

initial discordant complete

medial discord $\longrightarrow$

medial discordant complete

final discord $\longrightarrow$

final discordant complete

The following two abstract elements contain a final shift in the concordancy or discordancy of the sentence. Note that these two abstract elements, unlike the others which build directly on the completes, are dependent on other positional abstract elements. For example, a resolution is a sentence containing both an initial discord and a final concord.

resolution $\longrightarrow$

initial discord AND final concord

dissolution $\longrightarrow$

initial concord AND final discord

4.4 The Stylistic Goals

This section gives the grammar rules for three basic stylistic goals. It is here that we identify what goal a sentence has fulfilled based on its syntactic attributes. As these syntactic attributes have been identified and classified by the abstract elements, the goals are in turn based on the abstract elements.

The first rule describes the stylistic goal of clarity. From the discussion on stylistic goals in the previous chapter, we know that clear sentences are simple ones with only one independent clause, described by the monoschematics; centred sentences with a central independent clause surrounded by dependent clauses, the centroschematics; parallel sentences, the homopoises; and sentences which reduce ambiguity, the resolutions. This gives us the following rule.

clarity $\longrightarrow$

monoschematic

centroschematic

homopoise

Concrete sentences emphasize syntactic components through the use of discordancy, giving the first four alternatives in the rule below, and through the use of parenthesis, giving us the final alternative.

concreteness $\longrightarrow$

initial discord

medial discord

final discord

dissolution

heteropoise

Static sentences have standard simple structures, the monoschematics, or strictly balanced structures, the homopoises.

staticness $\longrightarrow$

monoschematic

homopoise

Chapter 5

An Analyzer of Sentence Style

An actual system was built in order to illustrate our computational theory of style. ASSET, for Analyzing the Style of SEnTences, is written in Prolog and is based on the grammar of style described in Chapter 4. The systematic structure of this grammar, enforced by the consistency of its rules, made the task of implementation very straightforward. This chapter discusses the implementation.

5.1 The General Design

There were several constraints on the design of ASSET. I list them below and then discuss their implications.

Parser Independence: A stylistic analysis, based on syntax, of a sentence obviously includes a syntactic parse of that sentence. Thus, a parser is necessary for any system that does a stylistic analysis. The decision to make ASSET totally independent of the parser was in part theoretical—ASSET would not have to compromise theory because of limitations and/or methodology of the parser—and pragmatic—developing a parser from scratch was beyond the scope of my work.

Abstract Element Refinements: Future work on our theory of syntactic style includes the refinement of the abstract elements.[1]

Transition Element Refinements: The transition elements will need revising in the future. Reasons for this include the complete integration of the dependent clauses into the theory and the refinements that will be made to the abstract elements.

Efficiency: A goal of ASSET is that it be reasonably efficient.

[1] See (Green, 1992) for details.

The Uses of ASSET: The envisioned uses of the stylistical analysis of sentences includes Intelligent Computer Assisted Language Instruction (ICALI) and Machine Translation (MT). In addition, ASSET will be an essential tool for testing our theory of style.

The combination of the above constraints led to the following considerations in designing the stylistic analyzer. First and foremost, the near certainty of future revisions made modularity, good software engineering practice in any case, a must. Another consideration was the parser-independence specification. This implied that the syntactic parse of the sentence would be done before the stylistic analysis began. Thus, the parser acts as a front-end for ASSET, with the advantage that input to ASSET is independent of the parser. This allows the substitution of parsers within the system with the only requirement that a module be created to transform the output of a particular parser into the specified format for ASSET.

A major consideration in designing ASSET was its potential uses. At the present state of development of ICALI and MT, it is impossible to know exactly which information and what representation would be most useful. This implies that, in addition to letting the user know which stylistic goal(s), if any, have been met, all stylistic information used during the analysis must be part of the output of ASSET. This requirement was further reinforced by our intent to use ASSET to test and evaluate our grammar.

The need to have all stylistic information available further implied that the analysis of one part of the sentence, *e.g.,* the noun phrase, cannot constrain that of another, *e.g.,* the verb phrase. To obtain a degree of efficiency, in spite of the lack of constraints on the analysis, a bottom-up, or leaf-to-root, approach is used. A syntax tree that parallels the syntactic organization of our grammar is the basic structure of ASSET. To facilitate the bottom-up approach, the tree is upside-down and in reverse order. The input to ASSET for the simple sentence *The man in the park runs* is shown in the next two figures. Figure 5.1 shows the list structure that ASSET processes, while Figure 5.2 lays this structure out as an upside-down, reversed tree.

> [[[[[none], complement], [[[runs], lexical_verb], verb], verb_phrase], [[[[[[none], postmodifier], [[[park], lexical_noun], noun], [[[[the], definite_article], adjectival], premodifier], nominal_group], [[in], preposition], prepositional_phrase], postmodifier], [[[man], lexical_noun], noun], [[[[the], definite_article], adjectival], premodifier], nominal_group], noun_phrase], major], complete]

Figure 5.1: ASSET's input in its list-structure form.

The parser used in the development of ASSET is Pundit[2] (Prolog UNderstands Integrated Text), chosen because of its fairly large syntactic coverage and its comprehensive treatment of conjunctions. These are necessary features for the analysis of stylistically interesting sentences. Pundit uses a *restrictive grammar*, written as a set of BNF (Backus-Naur Form) rules.[3] Pundit's output consists of a syntactic tree, the printing of which may be turned

[2]Pundit is a system developed by the Unisys Corporation.

[3]See (Sager, 1981) for details of this grammar.

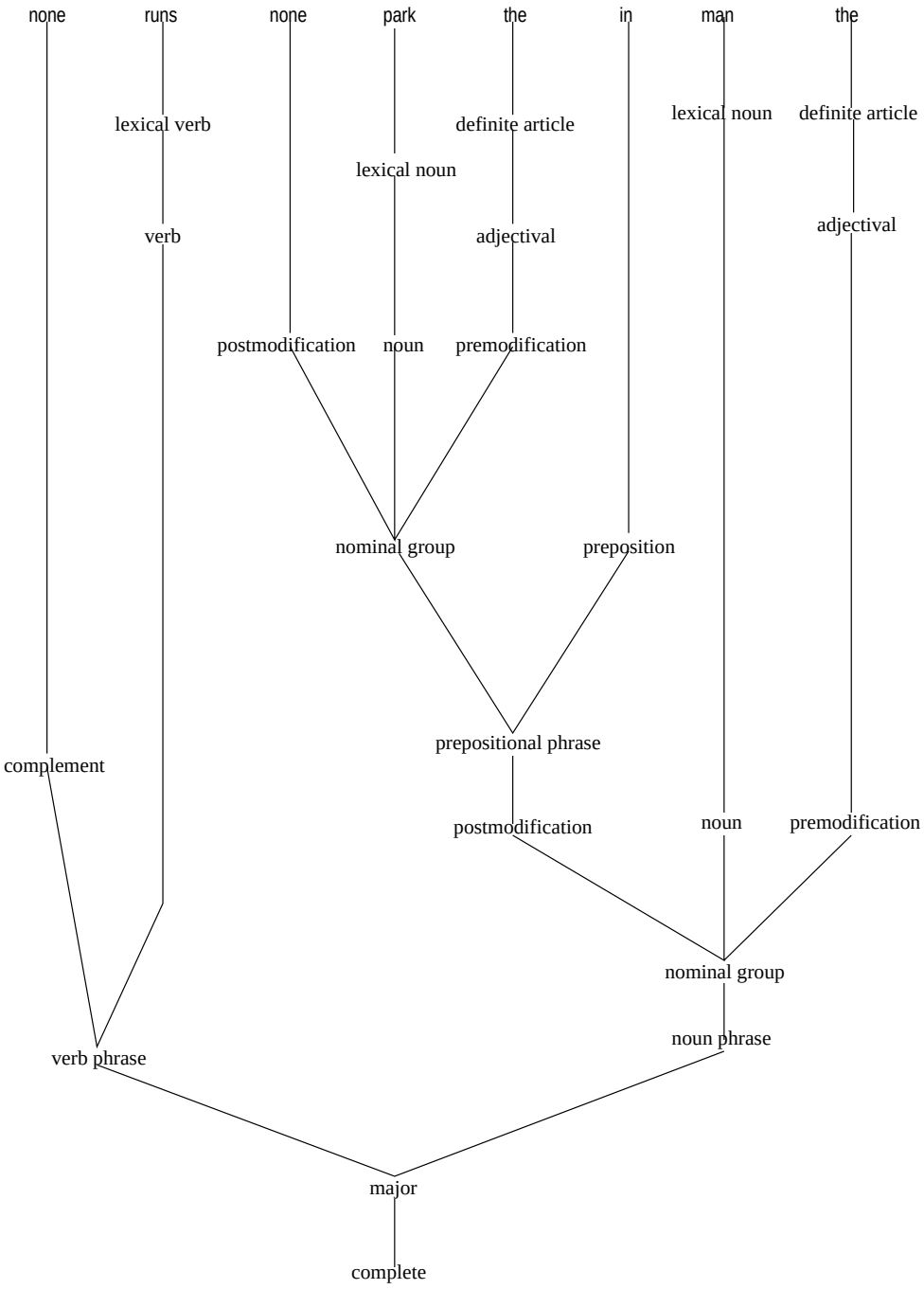

Figure 5.2: ASSET's input shown in tree form.

on or off with an interactive switch, plus an Intermediate Syntactic Representation (ISR), which specifies the type of sentence, the subject, the object, and any sentence adjuncts. It is this output that will transformed into a reversed, upside-down parse tree for input into ASSET.

1: Transform the parser output into format specified for ASSET.

2: Annotate the input tree with stylistic information (ANNOTATE MODULE).

3: Assign abstract elements to the input sentence (ABSTRACT ELEMENT MODULE).

4: Assign stylistic goal(s) to the input sentence (STYLISTIC GOAL MODULE).

5: Output the annotated tree structure.

Figure 5.3: A general algorithm for ASSET

A general algorithm for the system is shown in Figure 5.3 along with an accompanying illustration of its architecture in Figure 5.4. The next section describes the different modules of ASSET.

5.2 The Modules of ASSET

5.2.1 The Transformation Module

The transformation module is responsible for changing Pundit's output into the form, as shown in Figure 5.1, specified for ASSET. Part of the methodology for the design of this module involved examining parses of sentences chosen to isolate stylistically significant syntactic components. Building the module then consisted of correlating the syntax tree of Pundit with the syntax underlying the grammar used by ASSET.

5.2.2 The Annotate Module

The Annotate Module (AM) is responsible for the task of analyzing the style of the input sentence at the primitive-element and transition-element levels. The algorithm, shown in

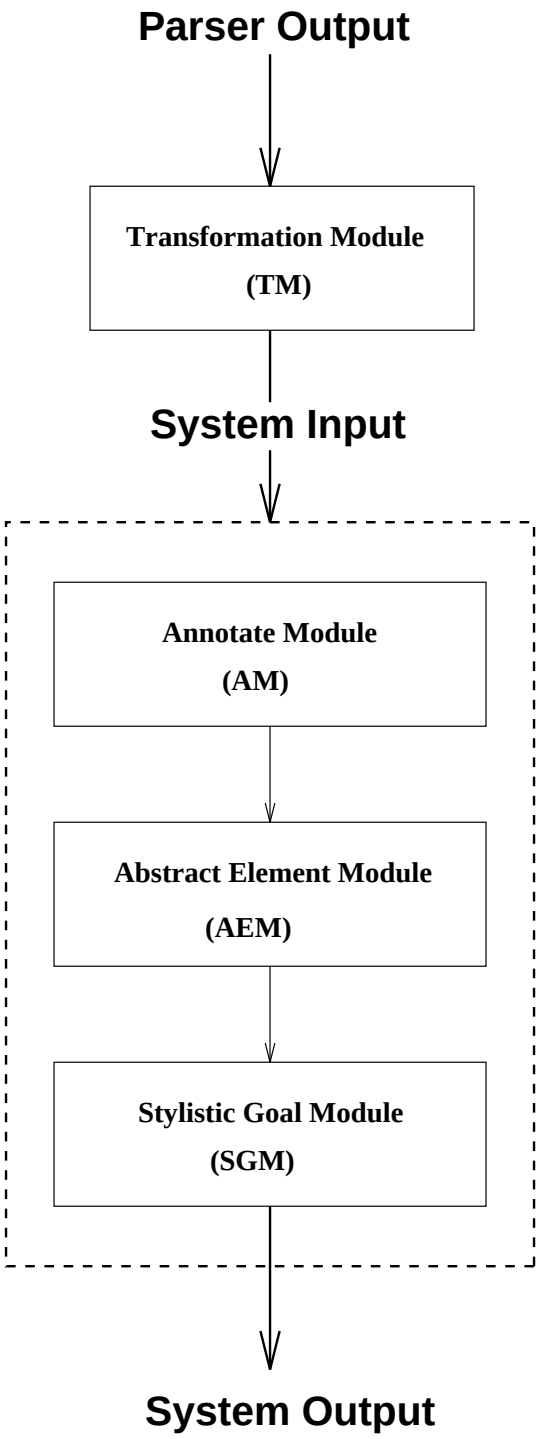

Figure 5.4: The overall architecture of ASSET

1: Annotate 'near' leaf nodes with primitive-element classifications (PRIMITIVE ELEMENT MODULE).

  1a. Analyze from the connective viewpoint (CONNECTIVE MODULE).

  1b. Analyze from the hierarchical viewpoint (HIERARCHICAL MODULE).

2: Annotate the rest of the nodes with transition-element classifications (TRANSITION ELEMENT MODULE).

Figure 5.5: An algorithm for the Annotate Module.

Figure 5.5, is very straightforward. There are two modules that annotate the nodes of the input tree with stylistic information.

The Primitive Element Module (PEM): This module is responsible for analyzing the appropriate nodes[4] by using the primitive element layer of our computational theory. Each node is analyzed from the connective viewpoint, by the Connective Module, and the hierarchical viewpoint, by the Hierarchical Module. The result of the analysis is a node annotated with the stylistic descriptions: either a conjunct or a antijunct primitive element and either a subjunct or a superjunct primitive element. Figure 5.6 shows the subparse tree for *the man in the park* with all the appropriate nodes annotated with primitive element terms.

The Transition Element Module (TEM): This module takes the parse tree, previously annotated by the PEM, and annotates the rest of the nodes with abstract element terms. The TEM uses information provided by the primitive element classification of nodes lower in the parse tree. Figure 5.7 shows the parse tree in Figure 5.6 now annotated with abstract element terms.[5]

It should be noted that the PEM and the TEM do not work sequentially, although the algorithm has described it thus for conceptual reasons. Because of the bottom-up processing,

[4]For the most part, the nodes that are analyzed by the primitive element module are near the leaf nodes of the syntax tree. The most notable exception is the dependent clause, also analyzed by the PEM.

[5]In Figures 5.7 and 5.9 the abstract element terms *monoschematic, centroschematic*, and *concordant* have been abbreviated to *mono, centro*, and *concord* respectively due to space restrictions.

```
nominal_group


    premodification([conjunct1],[subjunct2])
        adjectival(conjunct1],[subjunct2])
            definite_article(the)


    noun(conjunct0)
        lexical_noun(man)


    postmodification([conjunct1],[subjunct3])
        prepositional_phrase
            preposition(in)
            nominal_group

                premodification([conjunct1],[subjunct2])
                    adjectival([conjunct1],[subjunct2])
                        definite_article(the)

                noun(conjunct0)
                    lexical_noun(park)

                postmodification([conjunct0],[subjunct0])
                    postmodification(none)
```

Figure 5.6: A subparse tree annotated with primitive elements

```
nominal_group([centro,mono,concord])


    premodification([conjunct1],[subjunct2],[centro,mono,concord])
        adjectival([conjunct1],[subjunct2])
            definite_article(the)


    noun([conjunct0])
        lexical_noun(man)


    postmodification([conjunct1],[subjunct3],[centro,mono,concord])
        prepositional_phrase([centro,mono,concord])
            preposition(in)
        nominal_group([centro,mono,concord])

            premodification([conjunct1],[subjunct2],[centro,mono,concord])
                adjectival([conjunct1],[subjunct2])
                    definite_article(the)

            noun([conjunct0])
                lexical_noun(park)

            postmodification([conjunct0],[subjunct0],[centro,mono,concord])
                postmodification(none)
```

Figure 5.7: A completely annotated subparse tree.

calls to the TEM module come whenever the PEM module has annotated a sufficient number of nodes lower in the parse tree. Thus, calls to the PEM and the TEM are interspersed with each other. Figure 5.8 shows the architecture of the Annotate Module.

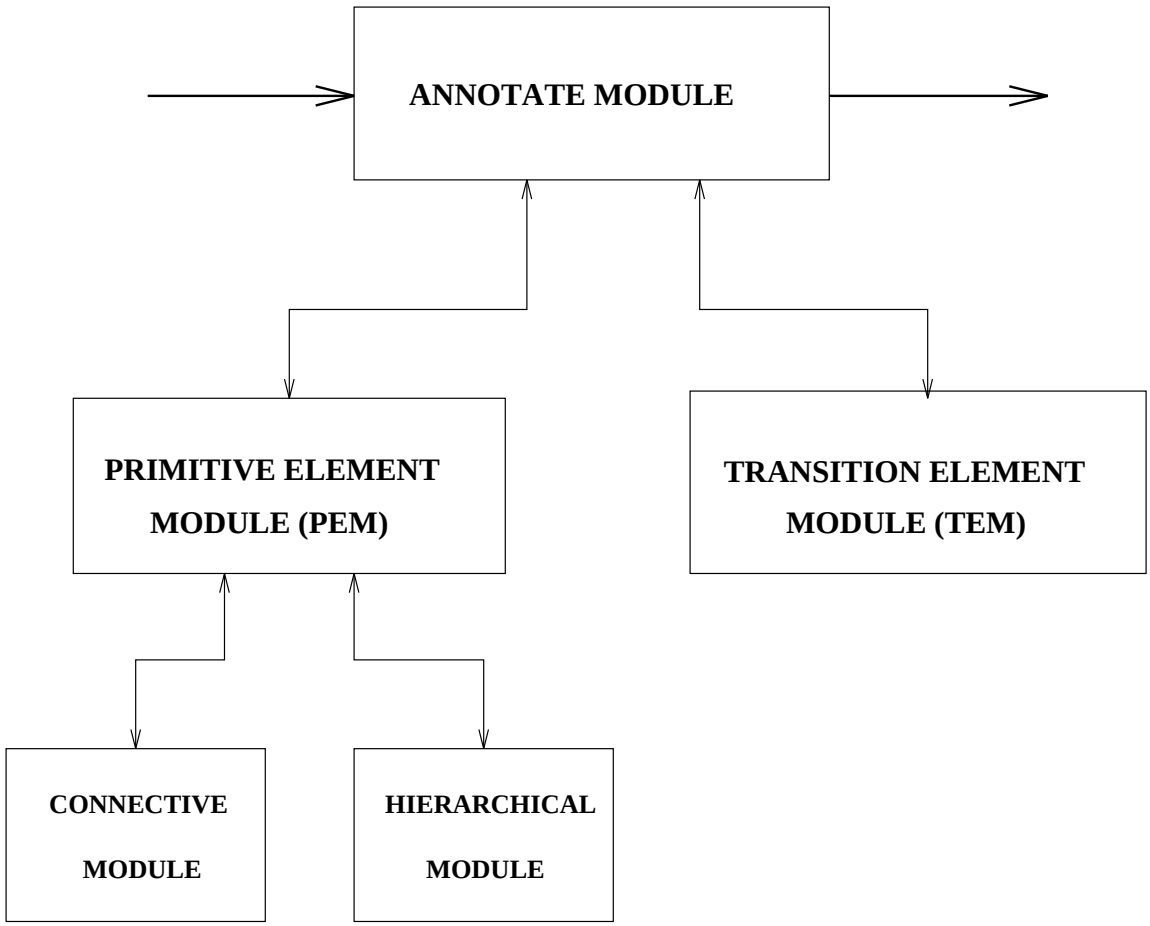

Figure 5.8: The architecture of the Annotate Module

5.2.3 The Abstract Element and the Stylistic Goal Modules

All that is left after the primitive element and transition element analyses is the classification of the input sentence in terms of the abstract elements and then the stylistic goals. This is a very straightforward task and is done by the Abstract Element and the Stylistic Goal Modules. Referring back to Figure 5.4, we see that a fully annotated parse tree is input to the Abstract Element Module, which then adds abstract element information to the structure and passes it on to the Stylistic Goal Module. Once the stylistic goal information has been

```
stylistic_goals(clarity,staticness)

abstract_elements(mono,concord,initial_concord,medial_concord,final_concord)

complete([mono,concord,initial_concord,medial_concord,final_concord])
  major([mono,concord])


      noun_phrase(([centro,mono,concord])
        nominal_group(([centro,mono,concord])

            premodification([conjunct1],[subjunct2],[centro,mono,concord])
              adjectival([conjunct1],[subjunct2])
                definite_article(the)

            noun([conjunct0])
              lexical_noun(man)

            postmodification([conjunct1],[subjunct3],[centro,mono,concord])
              prepositional_phrase([centro,mono,concord])

                preposition(in)
                nominal_group([centro,mono,concord])

                    premodification([conjunct1],[subjunct2],[centro,mono,concord])
                      adjectival([conjunct1],[subjunct2])
                        definite_article(the)

                    noun([conjunct0])
                      lexical_noun(park)

                    postmodification([conjunct0],[subjunct0],[centro,mono,concord])
                      postmodification(none)


      verb_phrase([mono,concord])

          verb(runs)

          complement([mono,concord])
            complement(none)
```

Figure 5.9: An example of ASSET's output.

determined, the structure is then passed out of ASSET. Figure 5.9 shows all the stylistic information contained in the output structure for the sentence *The man in the park runs.*

Chapter 6

Conclusion

6.1 Contributions of the Thesis

The main objective of this thesis was to combine the work of several researchers in computational style and, in so doing, to produce a computational theory of syntactic style solidly based in formal linguistic theory. Previous work (DiMarco, 1990) on computational style had been based on intuitive heuristics only. Revisions, in (DiMarco and Hirst, 1993a) and (Green, 1992), while providing the required linguistic basis, left the theory in an 'unfinished' state—the revisions at the primitive level had not been expanded over the full syntactic range of DiMarco's original work. As such, it could not be used to build a complete computational syntactic style analyzer. Another objective of this thesis was to demonstrate, by implementing the revised theory of syntactic style, the feasibility of incorporating stylistic analysis into NLP systems.

Both of these objectives have been met. A revised theory and a definitive grammar of style have been developed. This theory is firmly grounded in linguistic theory, allowing for the possibility of future extensions to the theory and for its reproducibility, due to the exact specification of the primitive elements.

The definitive grammar of style has extended the revised primitive elements to cover a range of syntax that enables the analysis of stylistically complex sentences. The addition of the transitional layer, thus splitting the primitive layer in two, has produced a conceptually simpler grammar as it is in this layer that the primitive elements are correlated with the abstract ones. In addition, the transitional layer has helped to ensure the grammar's consistency and also to enhance modularity.

ASSET was successful in showing that the definitive grammar of style is computationally tractable. One of the premises behind the whole notion of style used in this thesis is that style is a descriptive concept so that there is no 'bad' style. Because of this and because of the unknown amount of information that the envisioned uses of a style analyzer will need, all stylistic information gathered from the four levels of the theory is output by ASSET. This in turn allows ASSET to be used as a convenient means for testing the underlying theory of

style—either the present one or any future revisions.

An overall contribution made by this thesis has been the validation of our way of viewing style. Style is a matter of choice and a writer's sentences are intended to fulfill certain stylistic goals. These goals are realized by the linguistic choices made by the writers. The theory of style presented in this thesis views style as *compositional*. That is, the abstract-element labelling of sentences is dependent on their transitional-element labelling which is, in turn, dependent on the transitional element labelling of the underlying composite parts. ASSET has shown that a compositional analysis of style is practical and feasible.

6.2 Limitations and Future Work

The limitations in this thesis arise from two different sources: the theory itself and the implementation, ASSET. I first describe those limitations inherent in the theory itself.

6.2.1 Limitations of the Theory

The revised theory of style presented in this thesis is completely determined by the underlying syntax of the sentence being analyzed. Other influencing factors, namely lexical and semantic, are ignored. Thus, this theory cannot capture all the nuances of the style of a particular sentence. DiMarco *et al.* (1993) and DiMarco and Hirst (1993b) have begun investigations into lexical style. In addition, research is needed in semantic style and also in the interaction between the three types of style.

An interesting issue related to the above observation arises from the lack of distinction of the boundaries among the three types of style. During the development of the definitive grammar I had to deal with this demarcation issue—with both lexical style and semantic style. Lexical style becomes entangled with syntactic style in at least two contexts—the adverbials and the premodifying nouns. The hierarchical primitive-element classification of adverbials is based on their type as determined by (Quirk *et al.*, 1985). Unfortunately, the criteria for classifying the adverbs depend on the interaction of such things as which particular adverb is being analyzed and its linear position in the sentence. Thus, syntax alone cannot provide enough information to classify a particular adverb. A lexical theory of style is also needed.

The premodifying noun appears to draw its cohesiveness from its lexical properties and therefore it was difficult to determine its connective primitive-element classification. The decision was made to treat it as an adjective and to await its full treatment in a lexical theory of style.

The interaction of semantic and syntactic style became apparent with the adverbial and non-finite clauses in the hierarchical view. I had to make some assumptions when classifying the type of hypotaxis being exhibited by these dependent clauses. The reader is referred to 4.2.13 and 4.2.14 for full details of the assumptions used. Briefly, it was not always

possible to distinguish between the different types by syntax alone. The assumptions did allow Green's (1992) work to be incorporated into the definitive grammar of styler, but the full expressiveness of style due to hypotaxis will have to await the development of a semantic theory of style.

There are two other major limitations that arise from the theory itself. First, the primitive elements seem unable to capture the essence of either heteropoisal or monoschematic postmodification. In both cases, the transitional-element classification of postmodification is dependent on underlying syntax rather than on the primitive-element classification. A check for parenthesis is made for the heteropoisal postmodification and the presence of clausal postmodification is excluded in monoschematic postmodification.

The abstract elements have proven to be an obstacle to any attempted extensions of DiMarco's original work.[1] They are too 'coarse-grained' to capture all the subtleties of syntactic style. One effect of this coarseness is that two of the abstract elements, monoschematic and homopoise, are used to define two *different* stylistic goals. Green (1992) and DiMarco have done preliminary work in increasing the expressivity of the abstract elements. Future work includes relating these new elements to the primitive elements and also to the stylistic goals—tying them into the rest of the stylistic theory.

Future work on the grammar itself must include incorporating the analysis of dependent clauses into the definition of the complete sentences. In some cases, such as in postmodification, only the presence of the clause is important. In other cases, as with the adverbial clauses, the initial lexical item is considered or, as with the nominal clauses, the presence or absence of an internal subject is checked. It should be noted, however, that the concordancy or discordancy of a syntactic component is dependent on that of any included dependent clause.[2] The lack of an internal clausal analysis results in a grammar that cannot capture all the stylistic effects that arise from any dependent clause.

6.2.2 Limitations of the Implementation

In this section, I discuss the limitations of ASSET in relation to the definitive grammar of style. Implementational shortcomings arise from two sources, Pundit and ASSET itself. I will discuss the limitations of Pundit first.

A problem with Pundit stems from both its multiple parsing feature and its interactive parsing feature. In the former, the user is asked if additional parses are wanted and, in the latter, is asked to verify particular syntactic patterns of the input text. It should be noted that Pundit, to make the task of repeating a parse for a sentence less tedious, will save the patterns created by the interactive parsing feature. Moreover, this feature may be turned off altogether. In this case, however, a correct parse of the sentence may never be found.

A naive user, such as a second-language learner, would find such a system difficult as the onus for finding the correct parse is on the user. Moreover, as the system now stands, ASSET

[1] See (Mah, 1991), (Makuta-Giluk, 1991), and (BenHassine, 1992).

[2] The exception to this is the clausal postmodifier.

can only access one of Pundit's parses, the last one. Thus, a sophisticated user is needed to run ASSET.

The other shortcoming in Pundit lies with its inability to parse some of the stylistically interesting sentences covered by the definitive grammar. For instance, it cannot deal with clausal postmodifiers or postposed adjectival phrases that are parenthesized. Neither can it parse any type of non-cohesive parenthesis, such as the dash, semi-colon, or bracket. As a result, the only source of heteropoise in a nominal group available to ASSET is from a postposed nominal group. Other notable Pundit inabilities include the non-recognition of hyphens and postmodifying verbless clauses. Future work must include modifications to Pundit to increase the range of its grammar or the adoption of a more-robust and broader-coverage parser.

The limitations of ASSET arise from the incomplete implementation of the grammar of style. This problem can be divided into three main areas.

The first is that the adverbials weren't implemented, partly due to time constraints. However, further research is also needed in order to determine how to distinguish the different types of adverbials, using the classifications of (Quirk *et al.*, 1985). As the only source for a heteropoisal verb phrase arises from a heteropoisal adverbial, ASSET cannot now analyze any sentences that have heteropoisal verb phrase.

The two other unimplemented parts of the definitive style of grammar, left out due to time constraints, include the dependent clauses[3] and conjunctions. Thus, ASSET cannot handle polyschematic completes or the following abstract elements: polyschematics, homopoises, resolution, and dissolution. As this is somewhat limiting, it is imperative that ASSET be extended to handle these constructs.

6.3 Summary

In this thesis I have presented a revised computational theory of style, a definitive grammar built upon the theory, and an implementation, ASSET, built upon the grammar. The theory of style now has a solid linguistic foundation. The stylistic grammar is a consistent, understandable, and expressive codification of the theory. ASSET is an efficient style analyzer, capable of handling a variety of complex sentences. With the revised theory, the definitive grammar, and ASSET, I have shown that the incorporation of stylistic analysis into natural language processing systems is both feasible and practical. The application of this work in computational stylistics to the development of more-sophisticated NLP systems would promote their widespread use, especially by the naive user, by providing more-natural and effective interfaces.

[3]Those clauses which act as a postmodifier and those which act as a complement *can* be handled by ASSET.

Appendix A

Grammar Notation

This appendix defines and describes the notation used in the stylistic grammar presented both in Chapter 4 and Appendix B.

The left-hand side of the rule identifies what is being defined and the right-hand side lists one or more alternatives (one per line) of the rule. There may be examples with each alternative. The rule below defines the *conjunct[1] adjectival* with three different alternatives, each with an example. The first alternative specifies an adjective as *medieval* in *medieval subject.* The second alternative is the definite article, as *the* in *the medieval subject.* And third, the final alternative for the *conjunct[1] adjectival* is the *indefinite article*, such as the *a* in *a medieval subject.*

conjunct[1] adjectival $\longrightarrow$

adjective

medieval subject

definite article

the medieval subject

indefinite article

a medieval subject

I have used shorthand notations in the grammar rules in order to simplify their presentation.[1] These notations, each with examples, follow:

1. prepositional phrase $\longrightarrow$

   preposition nominal group

[1]Most of these shorthand notations have been adapted from (DiMarco, 1990).

A juxtaposition of terms on the right-hand side of the rule indicates a concatenation of instances of these terms. The rule above indicates that an instance of a prepositional phrase is defined by a preposition followed by a nominal group.

2. nominal group $\longrightarrow$

   (premodification) noun

A term in parenthesis indicates that it is optional. For instance, in the rule above, a nominal group could either consist of premodification followed by a noun or consist of a noun by itself.

3. homopoise $\longrightarrow$

   monoschematic complete (monoschematic complete)$^+$

The Kleene cross indicates one or more occurrences of the term within parenthesis.

4. nominal group $\longrightarrow$

   (premodification)* noun

The Kleene star indicates that there may be zero or more occurrences of the term within parenthesis.

5. heteropoisal nominal group $\longrightarrow$

   nominal group WITH heteropoisal postmodification

The *WITH* is used to further specify a right-hand alternative without having to write out all the components of the left-hand side. The example above specifies that a *heteropoisal nominal group* must have *heteropoisal postmodification.* If a nominal group is defined by

nominal group $\longrightarrow$

(premodification)* noun (postmodification)*

then the above rule is equivalent to

heteropoisal nominal group $\longrightarrow$

(premodification)* noun (heteropoisal postmodification)$^+$

6. conjunct[4] postmodification $\longrightarrow$

   adjectival AND parenthesis

AND indicates that all terms joined by the *AND* must be simultaneously satisfied by a single constituent. In the rule above, the postmodification must both be adjectival and parenthetical.

7. adjectival $\longrightarrow$

   (definite article OR indefinite article OR demonstrative determiner OR premodifying genitive) (adjective)*

OR indicates that one of the terms joined by *OR* must be true. The use of *OR* abbreviates a long sequence of alternatives. The above rule is equivalent to the one below.

adjectival $\longrightarrow$

definite article (adjective)*

indefinite article (adjective)*

demonstrative determiner (adjective)*

premodifying genitive (adjective)*

8. subjunct^{i} adjectival $\longrightarrow$

   (adjective)^{i} *where* $1 \leq i \leq 4$

The use of the variable i condenses a series of almost identical rules into one rule. For instance, the above rule is equivalent to the four below.

subjunct^{1} adjectival $\longrightarrow$

adjective

subjunct^{2} adjectival $\longrightarrow$

adjective adjective

subjunct^{3} adjectival $\longrightarrow$

adjective adjective adjective

subjunct^{4} adjectival $\longrightarrow$

adjective adjective adjective adjective

9. concordant verb phrase $\longrightarrow$

verb phrase WITH ALL complements concordant

If a verb phrase is defined as

verb phrase $\longrightarrow$

(adverbial)* verb (adverbial)* (complement)*

then the rule containing *WITH ALL* is an abbreviation for

concordant verb phrase $\longrightarrow$

(adverbial)* verb (adverbial)* (concordant complement)*

10. discordant verb phrase $\longrightarrow$

verb phrase WITH ANY discordant complement

The above rule is shorthand for

discordant verb phrase $\longrightarrow$

(adverbial)* verb (adverbial)* (complement)* discordant complement (complement)*

Appendix B

The Definitive Grammar

B.1 The Primitive and Transition Layers

B.1.1 Adjectivals

adjectival $\longrightarrow$

(definite article OR indefinite article OR demonstrative determiner OR premodifying genitive) (adjective)*

The Primitive Elements—Connective View

conjunct[1] adjectival $\longrightarrow$

adjective

definite article

indefinite article

conjunct[2] adjectival $\longrightarrow$

demonstrative determiner

conjunct[3] adjectival $\longrightarrow$

premodifying genitive

The Primitive Elements—Hierarchical View

subjunct^1 adjectival $\longrightarrow$

premodifying genitive

demonstrative determiner

indefinite article

subjunct^2 adjectival $\longrightarrow$

definite article

subjunct^i adjectival $\longrightarrow$

(adjective)^i *where* $1 \leq i \leq 4$

B.1.2 Premodification

premodification $\longrightarrow$

noun

adverbial

participle

adjectival

reduced sentence

The Primitive Elements—Connective View

conjunct^0 premodification $\longrightarrow$

adverbial

participle

NO premodification

conjunct^1 premodification $\longrightarrow$

noun

conjunct^i premodification $\longrightarrow$

conjunct^i adjectival *where* $1 \leq i \leq 3$

antijunct^2 premodification $\longrightarrow$

reduced sentence

The Primitive Elements—Hierarchical View

subjunct^0 premodification $\longrightarrow$

NO premodification

subjunct^i premodification $\longrightarrow$

subjunct^i adjectival *where* $1 \leq i \leq 4$

subjunct^3 premodification $\longrightarrow$

noun

adverbial

participle

reduced sentence

subjunct^4 premodification $\longrightarrow$

subjunct^4 adjectival

The Transitional Elements

centroschematic premodification $\longrightarrow$

conjunct^i premodification *where* $0 \leq i \leq 3$

subjunct^i premodification *where* $0 \leq i \leq 3$

monoschematic premodification $\longrightarrow$

subjunct^i premodification *where* $0 \leq i \leq 3$

concordant premodification $\longrightarrow$

conjunct^i premodification *where* $0 \leq i \leq 3$

subjunct^i premodification *where* $0 \leq i \leq 3$

discordant premodification $\longrightarrow$

antijunct^i premodification *where* $1 \leq i \leq 3$

subjunct^4 premodification

B.1.3 Nouns

noun $\longrightarrow$

nominal ellipsis

nominal substitution

lexical_noun

pronoun

The Primitive Elements—Connective View

conjunct^0 noun $\longrightarrow$

lexical_noun

conjunct^3 noun $\longrightarrow$

nominal substitution

nominal ellipsis

pronoun

The Transitional Elements

centroschematic noun $\longrightarrow$

conjunct^3 noun

concordant noun $\longrightarrow$

conjunct^3 noun

B.1.4 Postmodification

postmodification $\longrightarrow$

prepositional phrase

relative clause

nominal group

non-finite clause

verbless clause

adjectival

The Primitive Elements—The Connective View

conjunct^0 postmodification $\longrightarrow$

NO postmodification

conjunct^1 postmodification $\longrightarrow$

prepositional phrase

conjunct^2 postmodification $\longrightarrow$

relative clause

conjunct^3 postmodification $\longrightarrow$

nominal group

conjunct^4 postmodification $\longrightarrow$

non-finite clause

verbless clause

adjectival AND parenthesis

antijunct^2 postmodification $\longrightarrow$

adjectival

non-cohesive parenthesis

The Primitive Elements—Hierarchical View

subjunct^0 postmodification $\longrightarrow$

NO postmodification

subjunct^1 postmodification $\longrightarrow$

adjectival

subjunct^2 postmodification $\longrightarrow$

non-restrictive relative clause

subjunct^3 postmodification $\longrightarrow$

restrictive relative clause

non-finite clause

prepositional phrase

nominal group

The Transitional Elements

centroschematic postmodification $\longrightarrow$

conjunct^{i} postmodification *where* $0 \leq i \leq 4$

subjunct^{i} postmodification *where* $0 \leq i \leq 3$

heteropoisal postmodification $\longrightarrow$

conjunct^{i} postmodification WITH parenthesis *where* $1 \leq i \leq 6$

antijunct^{i} postmodification WITH parenthesis *where* $1 \leq i \leq 3$

monoschematic postmodification $\longrightarrow$

subjunct^{0} postmodification

subjunct^{3} postmodification AND (nominal group OR prepositional phrase)

concordant postmodification $\longrightarrow$

conjunct^{i} postmodification *where* $0 \leq i \leq 4$

subjunct^{i} postmodification *where* $0 \leq i \leq 3$

discordant postmodification $\longrightarrow$

conjunct^{i} postmodification *where* $5 \leq i \leq 6$

antijunct^{i} postmodification *where* $1 \leq i \leq 3$

subjunct^{4} postmodification

B.1.5 Nominal Groups

nominal group $\longrightarrow$

(premodification)* noun (postmodification)*

The Transitional Elements

centroschematic nominal group $\longrightarrow$

nominal group WITH centroschematic noun

nominal group WITH (centroschematic premodification AND centroschematic postmodification)

heteropoisal nominal group $\longrightarrow$

nominal group WITH heteropoisal postmodification

monoschematic nominal group $\longrightarrow$

nominal group WITH (monoschematic premodification AND monoschematic postmodification)

concordant nominal group $\longrightarrow$

nominal group WITH concordant noun

nominal group WITH (concordant premodification AND concordant postmodification)

discordant nominal group $\longrightarrow$

nominal group WITH (discordant premodification OR discordant postmodification)

B.1.6 Nominal Clauses

nominal clause $\longrightarrow$

restrictive relative clause

wh-interrogative clause

yes-no interrogative clause

non-finite clause

participle clause

The Primitive Elements—Connective View

conjunct^0 nominal clause $\longrightarrow$

nominal clause WITH subject

conjunct^3 nominal clause $\longrightarrow$

nominal clause WITHOUT subject

The Primitive Elements—Hierarchical View

subjunct^2 nominal clause $\longrightarrow$

restrictive relative clause

wh-interrogative clause

yes-no interrogative clause

non-finite clause

participle clause

The Transitional Elements

centroschematic nominal clause $\longrightarrow$

conjunct^0 nominal clause

conjunct^3 nominal clause

subjunct^2 nominal clause

concordant nominal clause $\longrightarrow$

nominal clause WITH ALL components concordant

discordant nominal clause $\longrightarrow$

nominal clause WITH ANY component discordant

initial_discordant nominal clause $\longrightarrow$

conjunct^3 nominal clause

B.1.7 Noun Phrase

noun phrase $\longrightarrow$

nominal group

nominal clause

The Transitional Elements

centroschematic noun phrase $\longrightarrow$

centroschematic nominal group

centroschematic nominal clause

heteropoisal noun phrase $\longrightarrow$

heteropoisal nominal group

monoschematic noun phrase $\longrightarrow$

monoschematic nominal group

concordant noun phrase $\longrightarrow$

concordant nominal group

concordant nominal clause

discordant noun phrase $\longrightarrow$

discordant nominal group

discordant nominal clause

initial_discordant noun phrase $\longrightarrow$

initial_discordant nominal clause

B.1.8 Adverbials

adverbial $\longrightarrow$

adverb

adverb phrase

The Primitive Elements—Connective View

conjunct[1] adverbial $\longrightarrow$

ANY adverbial

The Primitive Elements—Hierarchical View

subjunct^0 adverbial $\longrightarrow$

adverbial WITH adjunct adverb

subjunct^2 adverbial $\longrightarrow$

adverbial WITH subjunct adverb

superjunct^2 adverbial $\longrightarrow$

adverbial WITH disjunct adverb

adverbial WITH conjunct adverb

The Transitional Elements

centroschematic adverbial $\longrightarrow$

conjunct^1 adverbial

subjunct^0 adverbial

subjunct^2 adverbial

heteropoisal adverbial $\longrightarrow$

superjunct^2 adverbial

B.1.9 Prepositional Phrases

prepositional phrase $\longrightarrow$

preposition nominal group

The Transitional Elements

centroschematic prepositional phrase $\longrightarrow$

prepositional phrase WITH centroschematic nominal group

heteropoisal prepositional phrase $\longrightarrow$

prepositional phrase WITH heteropoisal nominal group

monoschematic prepositional phrase $\longrightarrow$

prepositional phrase WITH monoschematic nominal group

concordant prepositional phrase $\longrightarrow$

prepositional phrase WITH concordant nominal group

discordant prepositional phrase $\longrightarrow$

prepositional phrase WITH discordant nominal group

B.1.10 Complement

complement $\longrightarrow$

adjectival

prepositional phrase

noun phrase

The Transitional Elements

centroschematic complement $\longrightarrow$

complement WITH centroschematic noun phrase

complement WITH centroschematic prepositional phrase

adjectival

heteropoisal complement $\longrightarrow$

complement WITH heteropoisal noun phrase

complement WITH heteropoisal prepositional phrase

monoschematic complement $\longrightarrow$

NO complements

complement WITH monoschematic noun phrase

complement WITH monoschematic prepositional phrase

adjectival

concordant complement $\longrightarrow$

NO complements

complement WITH concordant noun phrase

complement WITH concordant prepositional phrase

adjectival

discordant complement $\longrightarrow$

complement WITH discordant noun phrase

complement WITH discordant prepositional phrase

B.1.11 Verbs

verb $\longrightarrow$

lexical_verb

verbal substitution

verbal ellipsis

The Primitive Elements—Connective View

conjunct^0 verb $\longrightarrow$

lexical_verb

conjunct^3 verb $\longrightarrow$

verbal substitution

conjunct^4 verb $\longrightarrow$

verbal ellipsis

The Transitional Elements

centroschematic verb $\longrightarrow$

conjunct^i verb *where* $3 \leq i \leq 4$

concordant verb $\longrightarrow$

conjunct^i verb *where* $3 \leq i \leq 4$

B.1.12 Verb Phrases

verb phrase $\longrightarrow$

(adverbial)* verb (adverbial)* (complement)*

The Transitional Elements

centroschematic verb phrase ⟶

verb phrase WITH centroschematic verb

verb phrase WITH AT LEAST ONE centroschematic complement

heteropoisal verb phrase ⟶

verb phrase WITH heteropoisal adverbial

monoschematic verb phrase ⟶

verb phrase WITH ALL complements monoschematic

concordant verb phrase ⟶

verb phrase WITH concordant verb

verb phrase WITH ALL complements concordant

discordant verb phrase ⟶

verb phrase WITH ANY discordant complement

verb phrase WITH heteropoisal adverbial

B.1.13 Adverbial Clauses

The Primitive Elements—Connective View

conjunct[1] adverbial clause ⟶

ANY adverbial clause

The Primitive Elements—Hierarchical View

subjunct[1] adverbial clause ⟶

finite adverbial clause INTRODUCED BY hypotactic conjunction

non-finite adverbial clause INTRODUCED BY hypotactic conjunction

subjunct[2] adverbial clause ⟶

finite adverbial clause INTRODUCED BY *when* OR *where*

superjunct[1] adverbial clause ⟶

finite adverbial clause INTRODUCED By *whereas* OR *while*

The Transitional Elements

concordant adverbial clause $\longrightarrow$

adverbial clause WITH ALL components concordant

discordant adverbial clause $\longrightarrow$

adverbial clause WITH ANY component discordant

B.1.14 Non-finite Clauses

The Primitive Elements—Connective View

conjunct^0 non-finite clause $\longrightarrow$

non-finite clause WITH subject

conjunct^3 non-finite clause $\longrightarrow$

non-finite clause WITHOUT subject

The Primitive Elements—Hierarchical View

subjunct^1 non-finite clause $\longrightarrow$

non-finite clause INTRODUCED BY preposition

subjunct^2 non-finite clause $\longrightarrow$

non-finite clause WITHOUT introducing preposition

The Transitional Elements

concordant non-finite clause $\longrightarrow$

non-finite clause WITH ALL components concordant

discordant non-finite clause $\longrightarrow$

non-finite clause WITH ANY component discordant

initial discordant non-finite clause $\longrightarrow$

conjunct^3 non-finite clause

B.1.15 Relative Clauses

relative clause $\longrightarrow$

wh-relative clause

The Primitive Elements—Connective View

conjunct[2] relative clause $\longrightarrow$

wh-relative clause

The Primitive Elements—Hierarchical View

subjunct[2] relative clause $\longrightarrow$

wh-relative clause

The Transitional Elements

concordant relative clause $\longrightarrow$

relative clause WITH ALL components concordant

discordant relative clause $\longrightarrow$

relative clause WITH ANY component discordant

B.1.16 Dependent Clauses

clause $\longrightarrow$

clausal substitution

clausal ellipsis

adverbial clause

non-finite clause

relative clause

The Primitive Elements—Connective View

conjunct[4] clause $\longrightarrow$

clausal substitution

conjunct[5] clause $\longrightarrow$

clausal ellipsis

The Transitional Elements

concordant clause $\longrightarrow$

conjunct[4] clause

concordant adverbial clause

concordant non-finite clause

concordant relative clause

discordant clause $\longrightarrow$

conjunct[5] clause

discordant adverbial clause

discordant non-finite clause

discordant relative clause

initial discordant clause $\longrightarrow$

initial discordant non-finite clause

B.1.17 Majors

major $\longrightarrow$

(conjunction) (adjective)* (adverbial)* (prepositional phrase)* (nominal group)*
noun phrase verb phrase

The Transitional Elements

centroschematic major $\longrightarrow$

major WITH centroschematic noun phrase AND centroschematic verb phrase

heteropoisal major $\longrightarrow$

major WITH heteropoisal noun phrase

major WITH heteropoisal verb phrase

major WITH heteropoisal adverbial

initial heteropoisal major $\longrightarrow$

major WITH heteropoisal noun phrase

major WITH heteropoisal adverbial

medial heteropoisal major $\longrightarrow$

major WITH heteropoisal noun phrase AND NO heteropoisal adverbial

final heteropoisal major $\longrightarrow$

major WITH heteropoisal verb phrase

monoschematic major $\longrightarrow$

major WITH monoschematic noun phrase AND monoschematic verb phrase

concordant major $\longrightarrow$

major WITH ALL components concordant

discordant major $\longrightarrow$

major WITH ANY component discordant

major WITH initial_discordant noun phrase

B.1.18 Completes

complete $\longrightarrow$

(clause)* major (clause)*

The Transitional Elements

monoschematic complete $\longrightarrow$

monoschematic major

centroschematic complete $\longrightarrow$

(concordant clause)* centroschematic major (concordant clause)*

polyschematic complete $\longrightarrow$

(concordant clause)$^+$ centroschematic major (concordant clause)*

(concordant clause)* centroschematic major (concordant clause)$^+$

heteropoisal complete $\longrightarrow$

(clause)* heteropoisal major (clause)*

initial heteropoisal complete $\longrightarrow$

initial heteropoisal major (clause)*

medial heteropoisal complete $\longrightarrow$

(clause)* medial heteropoisal major (clause)*

final heteropoisal complete $\longrightarrow$

(clause)* final heteropoisal major

concordant complete $\longrightarrow$

complete WITH ALL components concordant

initial concordant complete $\longrightarrow$

concordant major (clause)*

(concordant clause)* major (clause)*

medial concordant complete $\longrightarrow$

(clause)* concordant major (clause)*

finial concordant complete $\longrightarrow$

(clause)* concordant major

(clause)* major (concordant clause)*

discordant complete $\longrightarrow$

complete WITH ANY component discordant

initial discordant complete $\longrightarrow$

discordant major (clause)*

(discordant clause)* major (clause)*

(initial discordant clause)* major (clause)*

medial discordant complete $\longrightarrow$

(clause)* discordant major (clause)*

final discordant clause $\longrightarrow$

(clause)* discordant major

(clause)* major (discordant clause)*

B.2 The Abstract Elements

B.2.1 The Balance Abstract Elements

homopoise $\longrightarrow$

monoschematic complete (monoschematic complete)$^+$

heteropoise $\longrightarrow$

initial heteropoisal complete

medial heteropoisal complete

final heteropoisal complete

B.2.2 The Dominance Abstract Elements

monoschematic $\longrightarrow$

monoschematic complete

centroschematic $\longrightarrow$

centroschematic complete

polyschematic $\longrightarrow$

(centroschematic complete)$^{+}$ polyschematic complete (centroschematic complete)*

(centroschematic complete)* polyschematic complete (centroschematic complete)$^{+}$

B.2.3 The Position Abstract Elements

initial concord $\longrightarrow$

initial concordant complete

medial concord $\longrightarrow$

medial concordant complete

final concord $\longrightarrow$

final concord complete

initial discord $\longrightarrow$

initial discordant complete

medial discord $\longrightarrow$

medial discordant complete

final discord $\longrightarrow$

final discordant complete

resolution $\longrightarrow$

initial discord AND final concord

dissolution $\longrightarrow$

initial concord AND final discord

B.3 The Stylistic Goals

clarity $\longrightarrow$

- monoschematic
- centroschematic
- homopoise

concreteness $\longrightarrow$

- initial discord
- medial discord
- final discord
- dissolution
- heteropoise

staticness $\longrightarrow$

- monoschematic
- homopoise

Appendix C

Types of Sentences Analyzed

This appendix lists the types of sentences analyzed by ASSET with an example of each.

1. Nominal group without any premodification:
   Affinities play their role in this encounter.

2. Nominal premodification with four adjectives:
   California seems like an unsurpassed large beautiful multicultural proving ground.

3. A deictic and a noun in the premodification:
   That computer research on mind body connections accelerated.

4. Nominal substitution:
   The patriarch was muffled in his long black one with purple beading.

5. Participle in the premodification:
   They are inflicted by a prosecuting government.

6. Prepositional phrase in the postmodification:
   The first clear referendum on protectionism's appeal occurred in South Carolina over the weekend.

7. Relative clause in the postmodification:
   The artist provides a dreamy background which is done in bistre brush- strokes.

8. Non-finite clause in the postmodification:
   The artist provides a dreamy background done in bistre brushstrokes.

9. Adjectival in the postmodification:
   The people very happy are walking into the room.

10. Nominal group in the postmodification:
    California, an unsurpassed multicultural proving ground, has a large population.

11. Complex nominal groups in the postmodification:
Silvia, a commanding woman in her fifties, a shrew mellowed by religion, organizes prayer sessions on the lines of Tupperware meetings.

12. Embedding of a postmodifying nominal group in a nominal clause complement in a postmodifying relative clause in a postmodifying prepositional phrase:
They participated in a conference of 100 people who wanted to form a modest alternative, the Democratic Union.

Appendix D

Examples of Output from ASSET

This appendix displays three sample runs from ASSET.

```
sentence: Affinities play their role in this encounter.
>>


Style Goals:  clarity
              staticness


Abstract Elements:  centro
                    final_concord
                    medial_concord
                    initial_concord
                    mono
```

```
The Stylistic Parse Tree:
-------------------------
complete([mono,concord,initial_concord,medial_concord,final_concord,centro])
  major([centro,concord,mono])
      noun_phrase([centro,concord,mono])
        nominal_group([mono,concord,centro])
            premodification([conjunct0],[subjunct0],[centro,concord,mono])
              premod(none)
            noun([conjunct0],[subjunct0],[no_style])
              lexical_noun(affinities)
            postmodification([conjunct0],[subjunct0],[centro,concord,mono])
              postmod(none)
      verb_phrase([mono,concord,centro])
          verb([conjunct0],[subjunct0],[no_style])
            lexical_verb(play)
          complements([centro,concord,mono])
              complement_1([mono,concord,centro])
                noun_phrase([centro,concord,mono])
                  nominal_group([mono,concord,centro])
                      premodification([conjunct3],[subjunct1],[centro,concord,mono])
                        adjectival([conjunct3],[subjunct1])
                          premodifying_genitive(their)
                      noun([conjunct0],[subjunct0],[no_style])
                        lexical_noun(role)
                      postmodification([conjunct0],[subjunct0],[centro,concord,mono])
                        postmod(none)
              complement_2([mono,concord,centro])
                prepositional_phrase([centro,concord,mono])
                    preposition([no_conn_style],[no_hier_style],[no_style])
                      lex_preposition(in)
                    nominal_group([mono,concord,centro])
                        premodification([conjunct2],[subjunct1],[centro,concord,mono])
                          adjectival([conjunct2],[subjunct1])
                            demonstrative_determiner(this)
                        noun([conjunct0],[subjunct0],[no_style])
                          lexical_noun(encounter)
                        postmodification([conjunct0],[subjunct0],[centro,concord,mono])
                          postmod(none)




sentence: The artist provides a dreamy background which is done in bistre brushstrokes.
>>


Style Goals:  clarity


Abstract Elements:  centro
                    final_concord
                    medial_concord
                    initial_concord
```

```
The Stylistic Parse Tree:
-------------------------
complete([concord,initial_concord,medial_concord,final_concord,centro])
  major([centro,concord])
      noun_phrase([centro,concord,mono])
        nominal_group([mono,concord,centro])
            premodification([conjunct1],[subjunct2],[centro,concord,mono])
              adjectival([conjunct1],[subjunct2])
                definite_article(the)
            noun([conjunct0],[subjunct0],[no_style])
              lexical_noun(artist)
            postmodification([conjunct0],[subjunct0],[centro,concord,mono])
              postmod(none)
      verb_phrase([concord,centro])
          verb([conjunct0],[subjunct0],[no_style])
            lexical_verb(provides)
          complements([centro,concord])
            complement_1([concord,centro])
              noun_phrase([centro,concord])
                nominal_group([concord,centro])
                    premodification([conjunct1],[subjunct0,subjunct1],[centro,concord,mono])
                        adjectival([conjunct1],[subjunct1])
                          indefinite_article(a)
                        adjectival([conjunct1],[subjunct0])
                          adjective(dreamy)
                        adjectival([conjunct1],[subjunct1])
                          number_adjectives_1(dreamy)
                    noun([conjunct0],[subjunct0],[no_style])
                      lexical_noun(background)
                    postmodification([conjunct1],[subjunct3],[centro,concord])
                      relative_clause([initial_discord,concord])
                          wh_intro([wh_conjunct0],[wh_subjunct0],[no_style])
                            wh_word(which)
                          noun_phrase([no_subject],[no_subject],[no_subject_concord])
                            np(none)
                          verb_phrase([mono,concord,centro])
                              verb([conjunct0],[subjunct0],[no_style])
                                  lexical_verb(is)
                                  lexical_verb(done)
                              complements([centro,concord,mono])
                                complement_1([mono,concord,centro])
                                  prepositional_phrase([centro,concord,mono])
                                      preposition([no_conn_style],[no_hier_style],[no_style])
                                        lex_preposition(in)
                                      nominal_group([mono,concord,centro])
                                          premodification([conjunct1],[subjunct0,subjunct1],[centro,concord,mono])
                                              adjectival([conjunct1],[subjunct0])
                                                adjective(bistre)
                                              adjectival([conjunct1],[subjunct1])
                                                number_adjectives_1(bistre)
                                          noun([conjunct0],[subjunct0],[no_style])
                                            lexical_noun(brushstrokes)
                                          postmodification([conjunct0],[subjunct0],[centro,concord,mono])
                                            postmod(none)
```

```
sentence: The people very happy are walking into the room.
>>


Style Goals:  clarity
              concreteness


Abstract Elements:  final_discord
                    medial_discord
                    initial_discord
                    centro


The Stylistic Parse Tree:
-------------------------
complete([centro,discord,initial_discord,medial_discord,final_discord])
  major([discord,centro])
      noun_phrase([centro,discord])
        nominal_group([discord,centro])
            premodification([conjunct1],[subjunct2],[centro,concord,mono])
              adjectival([conjunct1],[subjunct2])
                definite_article(the)
            noun([conjunct0],[subjunct0],[no_style])
              lexical_noun(people)
            postmodification([antijunct2],[subjunct1],[centro,discord])
                adverb(very)
                adjective(happy)
      verb_phrase([mono,concord,centro])
          verb([conjunct0],[subjunct0],[no_style])
              lexical_verb(are)
              lexical_verb(walking)
          complements([centro,concord,mono])
            complement_1([mono,concord,centro])
              prepositional_phrase([centro,concord,mono])
                  preposition([no_conn_style],[no_hier_style],[no_style])
                    lex_preposition(into)
                  nominal_group([mono,concord,centro])
                      premodification([conjunct1],[subjunct2],[centro,concord,mono])
                        adjectival([conjunct1],[subjunct2])
                          definite_article(the)
                      noun([conjunct0],[subjunct0],[no_style])
                        lexical_noun(room)
                      postmodification([conjunct0],[subjunct0],[centro,concord,mono])
                        postmod(none)
```

Bibliography

(BenHassine, 1992) Nadia BenHassine. *A Formal Approach to Controlling Style in Generation.* Master's thesis, Department of Computer Science, University of Waterloo, 1992.

(Bolt, 1993) Philip Bolt. Grammar Checking Programs for Learners of English as a Foreign Language. In Masoud Yazdani, editor, *Multilingual Multimedia*, chapter 8, pages 140–197, Intellect Ltd., 1993.

(Cluett, 1976) Robert Cluett. *Prose Style and Critical Reading.* Teachers College Press, 1976.

(Corbett, 1971) Edward P. J. Corbett. *Classical Rhetoric for the Modern Student.* Oxford University Press, 1971.

(Crystal and Davy, 1969) David Crystal and Derek Davy. *Investigating English Style.* Indiana University Press, 1969.

(Crystal, 1987) David Crystal. *The Cambridge Encyclopedia of Language.* Cambridge University Press, 1987.

(Dahl, 1992) Deborah A. Dahl. PUNDIT – Natural Language Interfaces. In G. Comyn, N.E. Fuchs, and M.J. Ratcliffe, editors, *Logic Programming in action*, Heidelberg, Germany, September 1992. Springer-Verlag.

(DiMarco and Hirst, 1993a) Chrysanne DiMarco and Graeme Hirst. A Computational Theory of Goal-directed Style in Syntax. To appear in *Computational Linguistics*, September 1993.

(DiMarco and Hirst, 1993b) Chrysanne DiMarco and Graeme Hirst. Usage Notes as the Basis for a Representation of Near-synonymy for Lexical Choice. To appear in *Proceedings of 9^{th} Annual Conference of the University of Waterloo Centre for the New OED and Text Research*, Oxford, England, September 1993.

(DiMarco *et al.*, 1993) Chrysanne DiMarco, Graeme Hirst, and Manfred Stede. The Semantic and Stylistic Differentiation of Synonyms and Near-Synonyms. Presented at *AAAI*

Spring Symposium on Building Lexicons for Machine Translation, Stanford, California, March 1993.

(DiMarco, 1990) Chrysanne DiMarco. *Computational Stylistics for Natural Language Translation.* PhD thesis, Department of Computer Science, University of Toronto, 1990. Published as Technical Report CSRI-239.

(Green and DiMarco, 1993) Stephen J. Green and Chrysanne DiMarco. Stylistic Decision-Making in NLG. In *Proceedings of the Fourth European Workshop on Natural Language Generation*, pages 155–158, 1993.

(Green, 1992) Stephen J. Green. *A Functional Theory of Style for Natural Language Generation.* Master's thesis, Department of Computer Science, University of Waterloo, 1992. Published as University of Waterloo Faculty of Mathematics Technical Report cs-92-48.

(Halliday and Hasan, 1976) M.A.K. Halliday and Ruqaiya Hasan. *Cohesion in English.* Longman Group Ltd., 1976.

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

(Hirschman and Dowding, 1990) Lynette Hirschman and John Dowding. Restriction Grammar: A Logic Grammar. In P. Saint-Dizier and S. Szpakowicz, editors, *Logic and Logic Grammars for Language Processing*, pages 141–167, Ellis Horwood, 1990.

(Hovy, 1990) Eduard H. Hovy. Pragmatics and Natural Language Generation. *Artificial Intelligence*, **43**(2):153–197, 1990.

(Kane, 1983) Thomas S. Kane. *The Oxford Guide to Writing.* Oxford University Press, 1983.

(Lanham, 1974) Richard Lanham. *Style: An Anti-Textbook.* Yale University Press, 1974.

(Loffler-Laurian, 1987) Anne-Marie Loffler-Laurian. Enquète Stylistique sur la Traduction Automatique de Documents Techniques. *International Review of Applied Linguistics in Language Teaching*, **25**(2):154–168, 1987.

(Mah, 1991) Keith Mah. *Comparative Stylistics in an Integrated Machine Translation System.* Master's thesis, Department of Computer Science, University of Waterloo, 1991. Published as University of Waterloo Faculty of Mathematics Technical Report cs-91-67.

(Makuta-Giluk, 1991) Marzena Makuta-Giluk. *A Computational Rhetoric for Syntactic Aspects of Text.* Master's thesis, Department of Computer Science, University of Waterloo, 1991. Published as University of Waterloo Faculty of Mathematics Technical Report cs-91-54.

(Milic, 1982) Louis T. Milic. The Annals of Computing: Stylistics. *Computers and the Humanities*, **16**:19–24, 1982.

(Payette and Hirst, 1992) Julie Payette and Graeme Hirst. An Intelligent Computer-Assistant for Stylistic Instruction. *Computers and the Humanities*, **26**, 1992.

(Payette, 1990) Julie Payette. *Intelligent Computer-Assisted Instruction for Second-Language Teaching.* Master's thesis, Department of Computer Science, University of Toronto, 1990. Published as Technical Report CSRI-247.

(Quirk *et al.*, 1985) Randolph Quirk, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. *A Comprehensive Grammar of the English Language.* Longman Group Ltd., 1985.

(Sager, 1981) Naomi Sager. *Natural Language Information Processing: A Computer Grammar of English and its Applications.* Addison-Wesley Publishing Co, Inc., 1981.

(Tsutsumi, 1990) Taijiro Tsutsumi. Wide-range Restructuring of Intermediate Representations in Machine Translation. *Computational Linguistics*, **16**(2):71–78, June 1990.

(Vinay and Darbelnet, 1958) J.P. Vinay and J. Darbelnet. *Stylistique comparée du français et de l'anglais.* Didier, Paris, 1958.