A BASIS FOR A FORMALIZATION OF LINGUISTIC STYLE

Stephen J. Green
Department of Computer Science
University of Waterloo
Waterloo, Ontario, Canada N2L 3G1

INTRODUCTION

Style in language is more than just surface appearance, on the contrary, it is an essential part of the meaning conveyed by the writer. A *computational theory of style* could be of great use in many computational linguistics applications. A system that is 'stylistically aware' could analyze the writer's stylistic intent and understand the complex interaction of choices that produce a certain effect. In applications such as machine translation, a computational theory of stylistics would allow the preservation or modification of stylistic effects across languages. The theory would also be useful in computer-aided language instruction where, along with vocabulary and grammar, the individual writing style of the student could be analyzed and amended. The work described in this paper will be incorporated into into the Nigel grammar of the Penman system to provide a fine degree of stylistic control in language generation.

Drawing on both classical and contemporary rhetorical theory, we view style as *goal directed*: that is, texts are written for a purpose and this purpose dictates the stylistic choices that must be made. We find a computational counterpart to this view in the work of Hovy (1988), who used style as one of the pragmatic factors controlling generation in his PAULINE system. More recently, DiMarco (1990), the basis for this research, attempted to codify many of the elements of style that had previously been defined only descriptively and informally.

DiMarco presented a vocabulary of stylistic terms that was used to construct a *syntactic stylistic grammar* at three levels: *primitive elements*, *abstract elements*, and *stylistic goals*. At the base level of the grammar, the primitive elements describe the stylistic effects of individual sentence components. These primitive elements are then combined at a level of more abstract descriptions. These abstract elements comprise a stylistic 'metalanguage' that allows each element to be used to characterize a group of stylistically similar sentences. At the top level are the stylistic goals, such as clarity or concreteness, that are realized by patterns of the abstract elements.

The primitive-element level of DiMarco's stylistic grammar is divided into two views, *connective* and *hierarchic*. Here I will focus on the connective view, for which the stylistic effect of a sentence component is determined by considering its degree of *cohesiveness* within the sentence. The degrees of cohesiveness, or *connectivity*, vary on a scale from $conjunct^0$ (neither connective nor disconnective) through $conjunct^4$ (excessively connective).[1]

In more recent work, DiMarco and Hirst (1992) have provided a more formal basis for their theory of linguistic style, a basis that has its roots in the established linguistic theory of Halliday and Hasan (1976) and Halliday (1985). I am extending and refining their preliminary classifications of the primitive elements to provide a sounder basis for the entire computational theory of style. I will show how the connective primitive elements can be firmly tied to linguistic theory and how their properties are transmitted through the levels of the stylistic grammar.

A BASIS FOR LINGUISTIC STYLE

Drawing on the work of Halliday and Hasan (1976), a seminal work in textual cohesion, I will show how intrasentence cohesion, and its related stylistic effects, can be derived from the textual cohesive relations that Halliday and Hasan describe. Although there are undoubtedly significant stylistic effects at the text level, I feel that the codification of style at the sentence level has not yet been fully explored. For the most part, these cohesive relations function as well at the sentence level as they do at the text level. This is illustrated in Quirk *et al.* (1985), where all of the relations that Halliday and Hasan describe for texts are also demonstrated within single sentences.

Halliday and Hasan enumerate four major types of cohesive relations for English: *ellipsis, substitution, reference,* and *conjunction*. They classify

---

[1] There is also a scale of disconnectivity, or 'antijunctness', but I will not be using it in this discussion.

these relations in terms of their cohesive strengths relative to one another: ellipsis and substitution are the most cohesive relations, followed by reference, with conjunction being the least cohesive. One of the main objectives of my research is determining how all of these cohesive relations can be incorporated into the scale of 'conjunctness' described earlier. In this paper, I will deal only with ellipsis.[2]

Halliday and Hasan consider substitution to be equally as cohesive as ellipsis. I argue that ellipsis is more cohesive, after Quirk *et al.* (1985, p. 859) who state that for substitution and ellipsis "there are generally strong preferences for the most economical variant, *viz* the one which exhibits the greatest degree of reduction." Thus, the elliptical relations are more cohesive, due to the fact that they are generally more reduced. In DiMarco and Hirst, all forms of ellipsis are given a classification of *conjunct*[3] (strongly connective), but here I will look at the three types of ellipsis separately, assigning each its own degree of cohesiveness.[3] This assignment is made using by considering the most *reducing* relations to be the most cohesive, in the spirit of the above quote from Quirk *et al.* Since Halliday and Hasan provide a ranking for the four types of cohesive relation, and since ellipsis is considered to be the most cohesive relation, all of the degrees assigned for the different types of ellipsis will be ranked in the top half of the scale of cohesiveness.

The first type of ellipsis which Halliday and Hasan deal with is *nominal ellipsis*. This occurs most often when a common noun is elided from a nominal group and some other element of the nominal group takes the place of this elided noun. An example of this occurs in (1), where the noun *expedition* is elided, and the numerative *two* takes its place.

(1) The first expedition was quickly followed by another two ⊘.[4]

This is the least concise form of ellipsis, since only a single noun is elided. As such, it is given the lowest classification in this category: *conjunct*[3] (moderately-strong connective).

Next, we have *verbal ellipsis*. In instances of verbal ellipsis, any of the operators in the verbal group may be elided, as opposed to nominal ellipsis

where only the noun is elided. As Halliday and Hasan point out, many forms of verbal ellipsis are very difficult to detect, due to the complexity of the English verbal group. Because of this, I will deal only with two simple cases of verbal ellipsis: those in which the verbal group is removed entirely, as in (2), and those in which the verbal group consists of only modal operators, as in (3).

(2) You will speak to whoever I tell you to ⊘.[5]

(3) It may come or it may not ⊘.[6]

Both of these sentences are quite concise, as all, or nearly all, of the verbal group is elided. Verbal ellipsis is generally more concise than nominal ellipsis, and thus it has a higher level of cohesiveness: *conjunct*[4].

Finally, we look at *clausal ellipsis*, in which an entire clause is elided from a sentence. We see an example of this in (4).

(4) You can borrow my pen if you want ⊘.[7]

Since this form is more concise than either of the previous two verbal forms, we accord it a still higher level of cohesiveness: *conjunct*[5]. This classification gives clausal ellipsis a degree of cohesiveness verging on the extreme. The excessive amount of missing information tends to become conspicuous by its absence. Here we are beginning to deviate from what would be considered normal usage, creating an effect that DiMarco (1990) would call *stylistic discord.*

I will now present a short example to demonstrate how the effects of a foundation based on functional theory are built up through the three levels of the stylistic grammar.

A SIMPLE EXAMPLE

I will use the functional basis of style described above to illustrate how small variations in sentence structure can lead to larger variations in the stylistic goals of a sentence. This will be demonstrated by tracing the analysis of an example sentence through the levels of description of the stylistic grammar.

The first step in the analysis determines which connective primitive elements are present in the sentence and where they occur in our scale of cohesiveness. Next, the primitive elements are used to determine which abstract elements are present. Finally the abstract elements are examined to determine the stylistic goals of the sentence.

We start with sentence (4) as above. This sentence contains several connective primitive elements, the most prominent being the *conjunct*[5]

---

[2] When identifying the kinds of ellipsis, I use the terms defined by Halliday and Hasan and Quirk *et al.* All examples are taken from the appropriate sections of these references.

[3] I will be using a wider scale of cohesiveness than the one used by DiMarco and Hirst. Here *conjunct*[6], rather than *conjunct*[4], becomes the classification for the excessively connective. This change is made to allow for the description of more-subtle stylistic effects than is currently possible.

[4] Adapted from Quirk *et al.* example 12.54, p. 900.

[5] Quirk *et al.* example 12.64, p. 908.

[6] Adapted from Halliday and Hasan example 4:57, p. 170.

[7] Quirk *et al.* example 12.65, p. 909.

313

clausal ellipsis noted earlier, as well as instances of a *conjunct*[3] personal reference (*you*), a *conjunct*[2] deictic (*my*), and a *conjunct*[1] adversative (*if you want*). (Although I have completed the analysis for the other cohesive relations, here I am using the preliminary classifications given by DiMarco and Hirst (1992) for the other connective elements.)

Apart from the terminal ellipsis, all of these connective elements are *concordant*, that is, they represent constructions that conform to normal usage. The terminal ellipsis, due to its excessively high level of cohesiveness, is weakly discordant, a slight deviation from normal usage. Thus, this sentence contains *initial* and *medial concords*, followed by a *terminal discord*. In the terms of the stylistic grammar, this shift from concord to discord is formalized in the abstract element of *dissolution*. The presence of dissolution characterizes the stylistic goal of *concreteness*, which is associated with sentences that suggest an effect of specificity by an emphasis on certain components of a sentence. In this sentence, the emphasis is created by the terminal discord. The clausal ellipsis requires that a great deal of information be recovered by the reader and because of this it leaves her feeling that the sentence is unfinished.

The next example, sentence (5), is a modification of (4) and is an example of verbal ellipsis, as in (2).

(5) You can borrow my pen if you want to ⊘.

In this sentence, all of the previous connective elements remain except for the terminal clausal ellipsis. This ellipsis has been replaced by a verbal ellipsis that is *conjunct*[4], strongly but not excessively cohesive. This replacement consequently eliminates the terminal discord present in the previous sentence, changing it to a strong concord. Thus, (5) has *initial*, *medial*, and *terminal concords*, making it a fully concordant sentence. At the level of abstract elements, such a sentence is said to be *centroschematic*, that is, a sentence with a central, dominant clause with optional dependent clauses and complex subordination. Centroschematic sentences characterize the stylistic goal of *clarity*, which is associated with sentences that suggest plainness, preciseness, and predictability. In this sentence, the effect of predictability is created by removing the terminal discord, thus resolving the unfulfilled expectations of (4).

Thus, using the cohesive relations of Halliday and Hasan, it is possible, as I have shown, to provide a formal basis for the connective primitive elements of the syntactic stylistic grammar. These primitive elements can now be used as the components of more-precise abstract elements, with subtle variations in the primitive elements allowing more-expressive variations in the abstract elements

that constitute a sentence. These variations at the abstract-element level of the grammar are mirrored at the level of stylistic goals by large variations in the overall goals attributed to a sentence.

CONCLUSION

The research presented above is a part of a larger group project on the theory and applications of computational stylistics. I have completed the integration of all the connective primitive elements with Halliday and Hasan's theory of cohesion. My next step is to perform the same kind of analysis for the hierarchic primitive elements, giving them a solid basis in functional theory. In addition, I have completed refinements to the abstract elements, making them more expressive, and I will be able to formulate their definitions in terms of the new primitive elements.

The full theory of style will be implemented in a functionally-based stylistic analyzer by Pat Hoyt. This control of stylistic analysis combined with my work on the Penman generation system will allow us to begin exploring the myriad of applications that require an understanding of the subtle but significant nuances of language.

ACKNOWLEDGMENTS

REFERENCES
DiMarco, Chrysanne (1990). *Computational stylistics for natural language translation*. PhD thesis, University of Toronto.

DiMarco, Chrysanne and Hirst, Graeme (1992). "A computational approach to style in language." Manuscript submitted for publication.

Halliday, Michael (1985). *An introduction to functional grammar*. Edward Arnold.

Halliday, Michael and Hasan, Ruqaiya (1976) *Cohesion in English*. Longman.

Hovy, Eduard H. (1988). *Generating natural language under pragmatic constraints*. Lawrence Erlbaum Associates.

Quirk, Randolph, Greenbaum, Sidney, Leech, Geoffrey, and Svartvik, Jan (1985). *A comprehensive grammar of the English language*. Longman.