Abstract

We are developing stylistic grammars to provide the basis for a French and English stylistic parser. Our stylistic grammar is a branching stratificational model, built upon a foundation dealing with lexical, syntactic, and semantic stylistic realizations. Its central level uses a vocabulary of constituent stylistic elements common to both English and French, while the top level correlates stylistic goals, such as clarity and concreteness, with patterns of these elements. Overall, we are implementing a computational schema of stylistics in French-to-English translation. We believe that the incorporation of stylistic analysis into machine translation systems will significantly reduce the current reliance on human post-editing and improve the quality of the systems' output.

2 A Definition of Style for Machine Translation

Our approach to a definition of style is non-literary, group-based, and, most important, goal-directed. A "group-based" approach emphasizes the stylistic standards shared by a body of writers, rather than the characteristics of an individual author. Examples of group styles can be found in newspaper reporting or scientific writing. By "goal-directed", we mean that we seek to explicitly correlate specific goals of style such as clarity, informality, or abstraction with particular syntactic structures and lexical choices.

Why focus on group-based style? For MT, where we expect to deal with large amounts of similar types of text, the analysis of group style is of more interest than the idiosyncratic style of any one writer. However, group style can be subdivided into two major types, literary and functional (or pragmatic). Analysing literary group style, as in a work of fiction, for example, is too ambitious a task for MT. Functional group styles, on the other hand, are correlated with particular types of situations and are a more realistic objective.

To fulfill both of the complementary aims, we require a view of style that is goal-directed. A recent example with this view of stylistics is the PAULINE system described in Hovy [1987]. PAULINE generates text that conforms to given pragmatic and stylistic constraints; the system is goal-directed, able to correlate such stylistic goals as formality, simplicity, and respect with the characteristics of the text produced.

In French-to-English translation, as an example, we must account for the fact that French is generally more abstract, static, and precise while English tends to be concrete, dynamic, and more lenient towards vagueness (Vinay and Darbelnet [1988]). For the same text, therefore, the French source language may express an abstract style while the English target language must be more concrete in its lexical and syntactic choices. If we are to properly translate from French style to English style, then we must understand how syntactic structures and lexical choices correlate with particular stylistic goals or aspects, such as abstraction, concreteness, staticness, and dynamism.
3 A Computational Schema of Stylistics in Translation

We have restricted our interest in style for MT purposes to concerns that are pragmatic, group-based, and goal-directed. The next question is: how is this restricted definition to be reflected in a computational schema of stylistics in machine translation? The schema that we propose is outlined in Figure 1. The process of stylistic translation is language-independent but French-to-English translation is the particular instantiation that we have studied in detail and from which we will give examples.

In the first stage of the process, French internal stylistics is considered. For each stylistic translation unit, i.e., the current section of text within which the style remains constant, in the French source text, we determine its stylistic context, i.e., a correlation of lexical choices and syntactic structures with a particular stylistic goal and a particular stylistic intensity.

In the second stage, we use knowledge of French-English comparative stylistics. For each stylistic translation unit in the French source text, given its stylistic context, we determine the appropriate English stylistic goal corresponding to the possibly different French style. Default rules of comparative stylistics are used but these can be broken if the situation warrants.

The third and final stage requires an understanding of English internal stylistics. For each stylistic translation unit in the unedited English translation, we edit lexical choices and syntactic structures to achieve an appropriate style that corresponds to the French source style, but is English style as well.

The main feature of the schema is that the translation of style, according to our definition, requires three distinct bodies of stylistic knowledge: French internal stylistics, French-English comparative stylistics, and English internal stylistics. The internal stylistics of a language tells us how to choose syntactic structures and lexical items to express a particular aspect of style, such as clarity or economy. From comparative stylistics, we gain an understanding of when the target text should express a different style from the source text and when the styles should remain the same.

The need for three kinds of stylistic knowledge may not initially appear obvious. Why not just edit the output of a MT system to remove awkward syntax and achieve a more natural flow of words? The answer is that to preserve the author's stylistic intent, while meeting the stylistic demands of the target language, we must determine the author's specific stylistic goals and consider how these different aspects can be realized syntactically and lexically in the target language. Thus, we must analyze the original source text to make certain we understand the source language style before trying to translate.

To implement the computational schema of stylistics in translation shown in Figure 1, knowledge about stylistics must be accumulated, created, and organized into a formal representation.

Our first task was the creation of a vocabulary of English stylistics. The definition of concepts and the attempt to organize them into a recognizable structure were necessary to understanding the problem of translating style. The development of the vocabulary proceeded in tandem with the collection and creation of stylistic rules.

A further complication in the codification of French, English, and comparative stylistics is the need to build different sets of rules for lexical choice, syntactic structure, and semantic structure. This knowledge goes to make up stylistic grammars — formal representations of the stylistic rules. These grammars provide a systematic description of the lexical, syntactic, and semantic patterns that differentiate the various stylistic goals.

With these formal bodies of rules to analyse the original source and unedited target texts, the strategy to be used by a French and English stylistic parser can be planned. Ultimately, all components will be integrated in a post-editor that will accomplish the actual translation of style.

4 Stylistic Grammars

As a first step towards constructing a grammar of style, we defined the basic vocabulary shown in Table 1.

We have therefore taken a more stylistically relevant approach by introducing stylistic elements, which provide a means of structuring the vocabulary and a link between these vocabulary terms and stylistic goals. Furthermore, a single vocabulary of stylistic elements will be used to describe lexical, syntactic, and semantic realizations.

Style, we believe, is achieved through the effects inherent in individual components (absolute shape), the effects created by the absolute position of components (whether, for example, the component is initial, medial, or final), and by relationships (relative shape) between the components of a stylistic constituent. These relationships fall into at least two basic types: balance and harmony, including the concordant and discordant elements in a constituent, and dominance, the number of central shapes in a constituent.

Each type of relationship is expressed by stylistic elements. Table 2 shows samples. All the stylistic elements that we have defined, about twenty, have been recognized in our sample English data. Given these elements, we have a means of organizing the terms in our stylistic vocabulary: each term is associated with one or more stylistic elements of the various types.
FIRST STAGE
Original
French source text
French internal stylistics
SECOND STAGE
French – English comparative stylistics
THIRD STAGE
English internal stylistics
Unedited
Poor-quality
English target translation

Figure 1: A Computational Schema of Stylistics in Machine Translation

| Stylistic constituent: | A passage of text, associated with a particular stylistic goal, within which the style remains constant, *e.g.*, a sentence or paragraph. |
|------------------------|---------------------------------------------------------------------------------------------------------------------------------|
| Stylistic component:   | A part of a stylistic constituent, *e.g.*, a phrase or clause.                                                                 |
| Stylistic shape:       | A stylistic component usually regarded as stylistically expressive, *i.e.*, having a particular stylistic effect.               |
| Stylistic texture:     | A stylistic constituent with an overall stylistic effect, formed by particular types of relationships between its individual shapes. |
| Stylistic equivalence class: | A set of stylistic constructions having the same type of stylistic shape.                                                       |
| Interval:              | The stylistic distance or difference between any two shapes in a stylistic constituent.                                          |
| Modulation:            | A shift within a stylistic constituent from one type of stylistic effect, *e.g.*, concord, to another, *e.g.*, discord.        |

Table 1: Basic Vocabulary of Style

| Monopoise:             | A syntactically complete stylistic texture with no disturbance in canonical order.                                           |
| Counterpoise:          | A stylistic texture containing an offset, a shape which perturbs the stylistic balance by disturbing the canonical order.   |
| Homopoise:             | A counterpoise in which the offset supports the overall stylistic balance.                                                 |
| Polypoise:             | A counterpoise in which the offset opposes the overall stylistic balance.                                                   |
| Concord:               | A stylistic shape expressing a unity of style, agreement, accord, stability, and not requiring resolution.                 |
| Discord:               | A stylistic shape expressing a disunity of style, disagreement, contention, conflict, incongruity, and requiring resolution. |
| Cycle:                 | A stylistic texture in which the initial and terminal shapes belong to the same stylistic equivalence class.               |
| Imitation:             | A stylistic texture in which two or more successive shapes belong to the same equivalence class.                             |
| Resolution:            | A terminal modulation which moves from a stylistic discord to a relative concord.                                           |
| Dissolution:           | A terminal modulation which moves from a stylistic concord to a relative discord.                                           |
| Aschematic:            | A stylistic constituent having a stylistic shape but syntactically incomplete.                                             |
| Monoschematic:         | A stylistic constituent with a single dominant shape and no accompanying subordinate shapes.                                |
| Diaschematic:          | A stylistic constituent in which the components are organized around a single dominant shape.                               |
| Polyschematic:         | A stylistic constituent organized around two or more dominant shapes.                                                      |

Table 2: Examples of Stylistic Elements
Now, for our goal-directed stylistic grammar, we propose a branching stratification model. As the foundation, we have three branches: lexical, syntactic, and semantic, with the有自己的 vocabulary of stylistic shapes and rules relating its type of shapes.

At the central level, we use a single vocabulary of constituent elements of style, as in Table 2, to build rules relating these elements to patterns of lexical, syntactic, and semantic stylistic shapes. This level is the unifying core of the overall stylistic grammar for we believe that people also use the same principles, the same constituent elements, to express style both in English and French.

Finally, at the top level, we construct rules to correlate individual stylistic goals with patterns of stylistic elements. Together, these levels form a language-independent goal-directed stylistic grammar for language translation.

4.1 A Syntactic Stylistic Grammar

For each of the branches in our goal-directed stylistic grammar, we must define a vocabulary of stylistic shapes and rules for putting these shapes together. However, a "stylistic shape" will be defined differently for the lexical, syntactic, and semantic branches.

For the syntactic stylistic grammar, we have built a catalogue of sentence components (syntactic stylistic shapes) organized into equivalence classes by stylistic effect, a quality inherent in a component which makes it a stylistic shape. But what, in syntactic terms, gives each component a particular stylistic effect? How should we define "stylistic shape" at the syntactic level?

We have adapted Quirk and Greenbaum’s [1979] use of syntactic integration as the basis for our definition of syntactic stylistic shape. They catalogue most adverbials, some prepositional phrases, and some clauses as either adjunct, conjunct, or disjunct depending on the degree of inherent integration. We have chosen to base the syntactic stylistic effect of all sentence components on their characteristic integration. We have therefore expanded the catalogue to define a syntactic stylistic vocabulary classifying components as either adjacent (strongly integrating), conjunct (less integrating, but still having a connective effect), disjunct (neutral), or "anti-junct" (having a disconnection effect).

We then adopted Crystal and Davy’s [1969] “stylistic” grammar as the foundation for our syntactic stylistic grammar. Their grammar is built upon a vocabulary of stylistically significant syntactic components. It also recognizes the role of pre- and post-modification in stylistic effects, a feature we have expanded in our grammar, a portion of which is shown in Table 3.

Now, having a base syntactic stylistic grammar which builds sentences out of syntactic stylistic shapes defined by their characteristic integrating effect, we can construct, at the next level, a grammar which relates constituent elements of style to patterns of syntactic stylistic shapes. Syntactic examples for this level are given in Table 4.
Sentence → Complete | Incomplete*
Simple Concordant Sentence IS-A Sentence with only Complete = Simple Concordant Complete
Complete → (Dependent clause)* | Minor
Simple Concordant Complete → Simple Concordant Major, (Concordant dependent clause)*
Major → (Conjunction)*, (A)*, (C), (VOC), S, P, (C), (A)*
Simple Concordant Major →
   (Conjunction)*, Simple Concordant S, P, (Simple Concordant C), (Concordant A)
Discordant Major →
   (Conjunction)*, (Discordant A)*, (C), (VOC), S, P, (C), (A)*
   (Conjunction)*, (A)*, (C), (VOC), (AuxVerb)*(C), S, P, (C), (A)*
   (Conjunction)*, (A)*, (C), (VOC), P, S, (C), (A)*
Discordant A → Disjunct or Antijunct adverb | Disjunct or Antijunct adverbial construction |
   Disjunct or Antijunct dependent [adverbial] clause
S → Nominal group | Pronoun | Dependent [noun] clause
Simple Concordant S IS-A S with only ((Nominal group = Simple Concordant Nominal group) | Pronoun)
Counterpoisal S IS-A S with only Nominal group = Counterpoisal Nominal group
Discordant Counterpoisal S IS-A Counterpoisal S with
   Nominal group = Discordant Counterpoisal Nominal group
Nominal group → (Premodification) Noun (Postmodification)* | Pronoun
Simple Concordant Nominal group IS-A Nominal group with
   (Premodification = Simple Concordant Premodification) and
   (Postmodification = (Simple Concordant Postmodification)
Counterpoisal Nominal group IS-A Nominal group with
   Postmodification = (Counterpoisal Postmodification)*
Discordant Counterpoisal Nominal group IS-A Counterpoisal Nominal group with
   Postmodification = (Discordant Counterpoisal Postmodification)*
Simple Concordant Premodification IS-A Premodification with only Adjunct or Conjunct Premodification
Premodification → Adjunct Premodification | Dependent [relative] clause | Nominal relative clause |
   Adjunct dependent [adverbial] clause
Conjunct Premodification → Non-finite construction | Conjunct dependent [adverbial] clause |
   Nominal "that" clause | Preposition + Nominal group
Disjunct Premodification → Verbless clause | Disjunct dependent [adverbial] clause
Antijunct Premodification → Adjective
Simple Concordant Postmodification →
   Adjunct or Conjunct Postmodification (Imitative Postmodification) | Imitative Postmodification
Imitative Postmodification →
   x, (Conjunction), y, (Conjunction), (z) where x, y, z are instances of Postmodification and
   are not instances of Antijunct Postmodification and are in the same stylistic equivalence class.
Counterpoisal Postmodification IS-A Postmodification with all except
   Conjunct Postmodification ≠ Preposition + Nominal group
Discordant Counterpoisal Postmodification IS-A Counterpoisal Postmodification with only
Antijunct Postmodification
Concordant dependent clause → Adjunct or Conjunct dependent clause | Imitative dependent clauses
Discordant dependent clause → Disjunct dependent clause | Antijunct dependent clause
Adjunct dependent [adverbial] clause →
   time-adverbial-clause → time-adverb, Major
   place-adverbial-clause → place-adverb, Major
   purpose-adverbial-clause → purpose-adverb, Major
time-adverb → since, usually, before, after, until, when, ...
place-adverb → where, wherever, ...
purpose-adverb → in order to, so as to, ...

Table 3: Sample Syntactic Stylistic Grammar with Modified Crystal and Davy Notation
With an explicit correlation between syntactic grammar rules, stylistic grammar rules, and stylistic goals, we can now propose a strategy for the English and French stylistic parser that we will construct. The syntactic rules do the driving but partial stylistic decisions can affect the direction of the syntactic parse. That is, once a stylistic rule has been partially recognized, we can narrow the search space for constructs to complete the rule. As the syntactic rules analyse sequences of syntactic structures, these structures can then be recognized as sequences of stylistic elements, the components of stylistic rules. In tandem, the stylistic rules will pick up a sequence of stylistic elements and associate it with a stylistic goal. The idea of tandem rules in language analysis is developed by Hirst [1987]. This approach to integrating syntax and semantics has been adapted to syntax and stylistics.

5 Summary

Our recent work has been the construction of a goal-directed stylistic grammar for English internal stylistics. Our current focus is an extension of the grammar for French internal stylistics. As well, we are using a single vocabulary of constituent stylistic elements as the guiding principle in the development of a semantic stylistic grammar, the second branch of our stratificational model.

Our use of stylistic grammars has allowed us to propose a strategy for the French and English stylistic parser which will eventually form part of an editor to accomplish the actual translation of style.

Acknowledgements

For discussion, helpful ideas, and comments on earlier drafts, we are grateful to Brian Fitch, Diane Horton, and Eduard Iov. This work was supported in part by a grant to the second author by the Natural Sciences and Engineering Research Council of Canada.

References

[1] Crystal, David and Davy, Derek. Investigating English Style. Longmans, Green and Co. Limited, 1969.
[2] Hirst, Graeme. Semantic interpretation and the resolution of ambiguity (Studies in natural language processing). Cambridge University Press, 1987.
[3] Iov, Eduard Hendrik. Generating Natural Language Under Pragmatic Constraints. PhD thesis, Department of Computer Science, Yale University, March 1987.
[4] Quirk, Randolph, and Greenbaum, Sidney. A University Grammar of English, Longman Group Limited, 1979.
[5] Vinay, J.-P. and Darbelnet, J. Stylistique comparée du français et de l'anglais. Didier, Paris, 1998.