

A model of comparative stylistics  
for machine translation

Chrysanne DiMarco *and* Keith Mah  
Department of Computer Science  
University of Waterloo  
Waterloo, Ontario, Canada N2L 3G1  
[cdimarco@logos.uwaterloo.ca](mailto:cdimarco@logos.uwaterloo.ca)

April 12, 1994

### **Abstract**

The study of comparative stylistics attempts to catalogue and explain the differences in style between languages. Rules of comparative stylistics are commonly presented in textbooks of translation as simple ‘rules of thumb’, but if we hope to incorporate a knowledge of comparative stylistics into machine translation systems, we must take a more systematic approach. We develop a formal model of comparative syntactic stylistics to be used as a component of a general computational theory of style. We adapt textbook rules of human translation and study a small corpus of French–English translations to determine how these informal rules can be represented in our model as formal rules of translation. Our model of comparative stylistics could be implemented in a machine translation system, enabling the system to make a more informed decision about possible translation choices and their potential stylistic effects.

# 1 Introduction

## 1.1 The importance of style

Whenever a writer composes a piece of text, she uses a knowledge of style to help choose the most appropriate way to effectively communicate her meaning and intent. Style is achieved by linguistic choices on all levels: lexical, syntactic, and semantic. A writer chooses the appropriate words, syntactic structures, and semantic organization to convey her intended message to a particular audience in a particular situation. When writing a report for the CEO of a company, she may choose a formal, embellished style. But, when writing a personal letter, she will be simple and straightforward. These kinds of stylistic choices are often made unconsciously, but they are made nonetheless. And a translation, if it is to be faithful, will preserve these choices in the target language.

Style in this sense is not just a decorative veneer without any function. Rather, style is an essential part of the meaning that a writer or speaker is attempting to convey. By *style*, we mean more than the *prescriptive* sense, the familiar ‘correct’ mode of writing that is taught in schools and in textbooks of composition. Authoritarian prescriptive rules are not very useful in machine translation, for we have to deal with the text as it is, however complex, however idiosyncratic. We can’t impose prescriptive rules and rewrite the text.<sup>1</sup>

In contrast to the prescriptive view of style, we take the following points to be fundamental to computational stylistic analysis:

- Style is part of communication.
- Style is *goal-directed*, that is, a writer has specific intentions in mind when he composes a piece of text for a particular audience, and will therefore make appropriate linguistic choices to realize these intentions.
- Style is *codifiable*, so that a computational system can analyze and understand the stylistic import of text.

For a translator, whether human or machine, to ignore the importance of style is to risk creating a possibly ineffective or misleading translation. One sometimes has the experience of purchasing foreign-made products and finding that the owner’s manuals are awkward or unnatural, the result of a literal translation that did not allow for differences of style between the source and target languages. Although competent translators will not make these stylistic errors, a machine translation system runs the risk of doing so. A human post-editor must then take on the tedious task of correcting the style of the translated text. With the increasing demand for, and availability of, commercial MT products, automatic stylistic editing would be highly desirable.<sup>2</sup>

Students of translation theory have long recognized that stylistic differences between languages can be identified and catalogued. Textbooks of translation, like those of Guillemin-Flescher [1981] and Vinay and Darbelnet [1958], document various principles governing how the stylistic characteristics of one language can be maintained or modified in another. Vinay and Darbelnet, for

---

<sup>1</sup>Current grammar and style checkers, which adhere to the prescriptive view, are not relevant to our work. They simply enforce rigid rules and have no real understanding of the significance of stylistic variation. Similarly, systems that take a statistical approach to stylistic analysis of texts (*e.g.*, [Milic, 1982], [Cluett, 1976]), while useful for cataloguing many of the syntactic features that produce particular stylistic effects, have no understanding of the meaning of the statistics and must rely on human interpretation of the results computed.

<sup>2</sup>We foresee the day, in the not-too-distant future, when *embedded* MT systems for unilingual users will function within a word processing or e-mail environment.

example, discuss how the principles of *modulation* (changing point of view, as from the abstract to the concrete), *articulation* (making explicit syntactic links in text), *etc.*, apply in French to English translation.

Stylistic differences in translation occur at the lexical, syntactic, and semantic levels, as the examples below will illustrate. All are taken from Vinay and Darbelnet's [1958] classic textbook on French–English translation. At the *lexical level* we have situations such as the following, in which two languages differ in the choice of noun. In these sentences, the English version is quite concrete in the use of a specific amount, *a pint*, while the translation is more abstract, more natural for French, in the use of an indeterminate amount, *un peu*: *a little*.

- (1) Give a pint of your blood.
- (2) Donnez un peu de votre sang. (*lit.*, Give a little of your blood.)

Differences in *syntactic structure* between the source and target languages must also be considered. The structure can reflect the characteristic preferences of a language, as in the following pair of sentences. In French, adverbial phrases or clauses are placed by preference at the head of a phrase, especially if they have a causal sense, *i.e.*, the cause precedes the effect [Vinay and Darbelnet, 1958, 203]. English, on the other hand, is more concrete, presenting information in the order of importance in the text.

- (3) He waited unconcernedly for the opening of the case, as he felt sure to win.
- (4) Sûr d'obtenir gain de cause, il attendit sans inquiétude l'ouverture du procès. (*lit.*, Sure that he would win the case, he waited without anxiety for the opening of the trial.)

In addition to lexical and syntactic modifications in translation, there may be a change in *semantic structure*. In the examples below, English is more concrete and personal in its use of a deictic, *this*, while French is correspondingly more abstract and impersonal in its preferred use of nouns, *un pareil état d'esprit*.<sup>3</sup>

- (5) There is no future in the country if this is allowed to prevail.
- (6) Avec un pareil état d'esprit, le pays est voué à la stagnation. (*lit.*, With such a state of mind, the country is destined for stagnation.)

As these examples have shown, style affects the structure of a translation on several levels. An awareness of style allows the translator to produce text that is natural and that properly conveys the author's intent—and a *machine* translation system must have the same ability.

## 1.2 Computational stylistics for machine translation

In our earlier work [DiMarco, 1990; DiMarco and Hirst, 1990], we have developed French and English syntactic stylistic grammars using a preliminary version of our theory of style. We also developed a prototype stylistic parser, called *STYLISTIQUE*, that produced analyses of the stylistic

---

<sup>3</sup>Other kinds of changes in viewpoint may occur in translation, as in the following example:

- (i) Remember to use the Postal Code.
- (ii) N'oubliez pas d'indiquer le code postal. (*lit.*, Don't forget to indicate the postal code.)

[Note printed in a booklet of Canadian postage stamps.] (We thank Graeme Hirst for this example.)

characteristics of French and English input sentences.<sup>4</sup> While this work did demonstrate that our theory can be used to formalize both French and English syntactic style, we did not provide a mapping between the two stylistic grammars. Our aim now is to develop such a mapping in the form of a computational model of comparative stylistics that gives a reasoned account of the stylistic effects associated with different choices of syntactic structure in translation.<sup>5</sup>

Previous work in trying to account for style in machine translation has generally been limited to dealing with only very simple stylistic differences between languages. Although some practical issues of style in machine translation have been addressed, no formal model of style in translation had been developed.

Loffler-Laurian’s work [Laurian, 1986; Loffler-Laurian, 1983; Loffler-Laurian, 1987] attempted to characterize different types of *group style*. A group style is the set of characteristics that are common to the stylistic conventions of a group of writers, *e.g.*, authors of papers in technical journals. Loffler-Laurian attempted to determine experimentally the characteristics that readers have come to expect of particular genres of technical texts. She focused on how a knowledge of group style is used in the post-editing stage of translation. Loffler-Laurian’s research reinforced an important point for machine translation: although corresponding group styles may exist across languages, we still need to understand how languages differ in producing the same stylistic effect.

Tsutsumi [1990] took a different approach, looking at corresponding styles in source and target languages from different language groups: English and Japanese. He presented a methodology for recognizing, and accounting for, stylistic variations in syntactic structure at the sentence level. Four types of stylistic variations were considered: changes in viewpoint (grammatical constructions that cannot be directly translated between the two languages), variations in idioms and metaphors, variations in specific constructions using function words, and “others”. Each of these types of stylistic variation is dealt with through “wide-range restructuring of intermediate representations”; in effect, Tsutsumi rewrites the sentence before it is translated. For example, an idiom like *John kicked the bucket* might be rewritten as *John died*. The rewrites are performed through the use of an augmented context-free grammar. Each rule in the grammar specifies a tree structure to match and the corresponding replacement tree.

Although Tsutsumi’s work can be viewed as simple pattern-matching between source and target languages, this modest approach demonstrates that even simple, but strategic, techniques can be a significant aid in accounting for style in MT.

### 1.3 The role of comparative stylistics in machine translation

We know that for a translation to be faithful and effective, the author’s stylistic intent must be transferred between languages in translation. But this transfer involves complementary, and sometimes conflicting, goals:

- One wants to preserve the original author’s stylistic intent, the information being conveyed through the manner of presentation. But, as researchers in translation theory such as Guillemin-Flescher and Vinay and Darbelnet have catalogued, the same stylistic effect might be realized in different ways in different languages.
- As well, one wants to produce a text whose style is appropriate and natural to the particular target language. But languages differ as to the most ‘natural’ way to express

---

<sup>4</sup>Hoyt [1993] has since constructed a more comprehensive stylistic analyzer for English using the revised and extended theory of style described in [DiMarco and Hirst, 1993a] and [Green, 1992].

<sup>5</sup>An earlier version of this work appeared in [Mah, 1991].

an idea. For example, we have seen above that French tends to prefer more abstraction, English more concreteness. The best translation, therefore, might modify the original author's stylistic intent and express a different effect.

Sometimes, there is no way to resolve this dilemma, and one is left with an unsatisfactory translation. But, with a knowledge of the *comparative stylistics* of a language pair, and of the stylistic resources of each language and the possible range of effects they can create, one can substantially improve the quality of a translation.

## 2 A theory of syntactic style

In our earlier work [DiMarco, 1990; DiMarco and Hirst, 1993a; Green, 1992; Hoyt, 1993], we presented a computational theory of goal-directed syntactic style that provides a means of linking low-level linguistic choices to high-level stylistic goals. This section summarizes the details of this work as presented in [DiMarco and Hirst, 1993a]. For more details on the linguistic foundations of the grammar, the reader is referred to [Green, 1992]; Green also discusses an application of the theory to natural language generation. Hoyt [1993] presents the full syntactic stylistic grammar and also describes its implementation in a stylistic analyzer. In more recent work [DiMarco *et al.*, 1993; DiMarco and Hirst, 1993b], we have begun to address the problem of lexical style.

### 2.1 A vocabulary of style

In designing a grammar of style for computational use, we constructed a vocabulary of stylistic concepts by looking at the descriptive terms that stylists have developed over many centuries of use. We then organized these terms into three groups according to the general rhetorical properties of parallelism, structure nesting, and linear ordering.

#### 2.1.1 Abstract elements of style

We used the three groups as the divisions of our stylistic vocabulary and developed formal terms, *abstract elements of style*, in order to identify and make explicit general properties of style that were only implicit in the previous informal terminology. We will present the abstract elements in the three groups, according to their properties of *balance*, *dominance*, and *position*, which we define formally as follows:

**Balance:** A balance term characterizes a stylistic effect created by the juxtaposition of similar or dissimilar sentence components.

**Dominance:** A dominance term describes a stylistic effect created by the particular hierarchical structure of a sentence.

**Position:** A position term describes a stylistic effect created by the particular placement of a syntactic component within a sentence.

The abstract elements of style that we propose are based on effects of stylistic *concord* and *discord*:

**Concord:** A stylistic construction that conforms to the norm for a given genre.

**Discord:** A stylistic construction that deviates from the norm.<sup>6</sup>

The first group of abstract elements is related to effects of balance within a sentence. These elements describe syntactic relationships that tend to either perturb or reinforce the balance of a sentence. The first and simplest type of balance element is a *homopoise* (“same weight”):

**Homopoise:** A sentence with interclausal coordination of syntactically similar components.

---

<sup>6</sup>Discord is not necessarily ‘bad’. Indeed, it is the strategic use of discord, deviation from the norm, that can give expressiveness to writing.

In other words, there are one or more stylistic ‘shapes’ in a homopoisal sentence, each contributing the same type of effect to the concordant parallelism of the sentence. In the example below, two very simple clauses, with identically-shaped parse trees, are conjoined:

- (7) The style was formed and the principles were acquired.<sup>7</sup>

Sentences that are more complex can have their balance interrupted or perturbed by a *heteropoisal* (“different weight”) component:

**Heteropoise:** A sentence in which one or more parenthetical components are syntactically ‘detached’ and dissimilar from the other components at the same level in the parse tree.

The class of heteropoise sentences is divided into three types, depending upon the position of the parenthetical component: a heteropoise may be *medial*, *initial*, or *final*. Moreover, two subtypes can be distinguished at each position: *concordant* and *discordant*:

**Medial heteropoise:** A heteropoise in which the parenthetical component is in medial position.

**Concordant medial heteropoise:** A medial heteropoise in which the parenthetical component is cohesively linked to the rest of the sentence. (This notion will be made precise when we have introduced the *primitive elements* in Section 2.1.2.)

**Discordant medial heteropoise:** A medial heteropoise in which the parenthetical component is *not* cohesively linked to the rest of the sentence.

In the first example below, the relative clause *who had their drinking water tested* is a parenthesis, but is still cohesive, so that the sentence is a concordant medial heteropoise. But, in the second example, the interrupting phrase *and, in some instances, abolish* is set off by disruptive punctuation, so that the sentence is a discordant medial heteropoise.

- (8) Those individuals or municipalities, who had their drinking water tested, have largely been kept in the dark about the health risks.<sup>8</sup>

- (9) Since the Conservative government came to power, it has moved to reduce—and, in some instances, abolish—federal participation in these programs.<sup>9</sup>

**Initial heteropoise:** A heteropoise, concordant or discordant, in which the parenthetical component is in initial position.

An example of an initial heteropoise is given below:

- (10) Until a future arrangement is understood, there will be instability.<sup>10</sup>

**Final heteropoise:** A heteropoise, concordant or discordant, in which the parenthetical component is in final position.

An example of a final heteropoise is given below:

---

<sup>7</sup>Adapted from the *Manchester Guardian Weekly*, 14 February 1988, p. 15.

<sup>8</sup>[Friends of the Earth, 1986].

<sup>9</sup>[Canadian Council on Social Development, 1990b].

<sup>10</sup>[Unknown, 1991b].



- (11) They also heard that too few teachers can be described as outstanding or excellent, primarily because of attitudes towards teaching—not enough is done in the academic community to emphasize the value of teaching excellence.<sup>11</sup>

The second type of abstract element deals with stylistic dominance, which is concerned with the hierarchical structure of a sentence. A common type of dominance element is the *monoschematic*, a very simple sentence:

**Monoschematic:** A sentence with a single main clause with simple phrasal subordination and no accompanying subordinate or coordinate clauses.

The following sentence is a simple monoschematic sentence, with one postmodifying prepositional phrase but no clauses:

- (12) Part of the bill has run into a legal snag.<sup>12</sup>

The most common dominance element is the *centroschematic*:

**Centroschematic:** A sentence with a central, dominant clause with one or more of the following optional features: complex phrasal subordination, initial dependent clauses, terminal dependent clauses.

As the following example illustrates, a centroschematic sentence is built up by subordination, coordination, and dependent clauses:

- (13) Most women over 40 are partial to flowing muumuus, and the aloha-shirted crowd I found at Ken's Pancake House in Hilo lent the place a certain resemblance to the city's many orchid nurseries.<sup>13</sup>

Given that there is a progression in complexity from monoschematic to centroschematic sentences, a natural extension is to the *polyschematic*:

**Polyschematic:** A sentence with more than one central, dominant clause and at least one dependent clause.

Such sentences occur much less frequently than the monoschematic or centroschematic varieties, at least in the corpus used in this research. However, a sentence with obviously disparate components occasionally occurs as in the following example. There are two dominant clauses (*we could think ... and we should not forget ...*) and a dependent clause (*if we consider the progress ...*):

- (14) 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.<sup>14</sup>

---

<sup>11</sup>[Pierre, 1991b].

<sup>12</sup>[Canadian Council on Social Development, 1990a].

<sup>13</sup>[MacDonald, 1990a].

<sup>14</sup>Adapted from the *Manchester Guardian Weekly*, 14 February 1988, p. 14.

The third group of abstract elements are the position elements. The most common types of position element describe *concordant* or *discordant* stylistic effects in particular positions. The basic elements are *initial concord*, *medial concord*, and *final concord*, with a similar range of *discord* elements. The definitions of these elements are closely tied to the primitive-level descriptions of our vocabulary, so their precise characterizations will be delayed until after the primitive-level descriptions are introduced. Here, it is sufficient for the reader to know that the concords describe constructions, at a particular position in the sentence, that conform to normal usage, while the discords describe constructions that deviate from the norm.

### 2.1.2 Primitive elements of style

Now that a set of abstract stylistic elements has been defined, we can consider how to use them in practical stylistic analysis. In isolation, these elements are too high-level to be easily applied to the stylistic interpretation of an arbitrary sentence. Therefore, we relate these abstract *stylistic* elements to more-basic *syntactic* elements, whose stylistic characteristics are more specific and concrete. We observe that two analyses of stylistic structure are possible: *connective* and *hierarchic* orderings:

**Connective ordering:** The result of cohesive bonds drawing together components in a linear ordering.

**Hierarchic ordering:** The result of bonds of subordination and superordination drawing together components in a nested ordering.

These two complementary kinds of analysis are implicit in the work of most stylists and rhetoricians, such as Cluett [1976] and Bureau [1976] (Neil Randall, personal communication). Here, we will focus on how the connective ordering was used to guide the definition of more-primitive stylistic elements that provide a precise syntactic basis to the vocabulary, yet also allow a mapping to the abstract elements.

Primitive stylistic elements are individual constituents that have a particular ‘stylistic effect’. For the connective ordering, a component acquires its stylistic effect from its degree of *cohesiveness*, its *bonding* with other components in the sentence.

We introduce the terms *conjunct* and *antijunct*, and use superscripts on these terms to indicate the degree of connectivity or disconnectivity.<sup>15</sup> Syntactic components are classified as either *conjunct*<sup>5</sup> or *conjunct*<sup>6</sup> (excessively connective), *conjunct*<sup>3</sup> or *conjunct*<sup>4</sup> (strongly connective), *conjunct*<sup>2</sup> (moderately connective), *conjunct*<sup>1</sup> (mildly connective), and *conjunct*<sup>0</sup> (neutral). Similarly, the terms *antijunct*<sup>0</sup> through *antijunct*<sup>4</sup> are used to indicate increasingly disconnective effects; *conjunct*<sup>0</sup> and *antijunct*<sup>0</sup> are the same.

In assigning connective effects to syntactic components, we adapt the work of Halliday and Hasan [1976] on cohesion. Halliday and Hasan enumerate five types of cohesive relations: *substitution*, *ellipsis*, *reference*, *conjunction*, and *lexical cohesion*.<sup>16</sup> We use all of these except lexical cohesion and have added a new factor, *interpolation*, a disconnective relation that works *against*

---

<sup>15</sup>Green [1992] extended the range of connectivity to include excessively cohesive *conjunct*<sup>5</sup> and *conjunct*<sup>6</sup> components.

<sup>16</sup>Although Halliday and Hasan were mainly concerned with textual cohesion, their relations are equally applicable to intrasentence cohesion. In the development of our theory, we are especially concerned with the relationship between style and the cohesion of the nominal group.

cohesion. Interpolated elements are certain instances of parenthetical constructions, those that display none of the forms of cohesion listed earlier.<sup>17</sup>

The cohesive relations described above can be ranked from the most to the least cohesive. Halliday and Hasan [1976, 226–227] consider substitution, including ellipsis, to be the most strictly cohesive relation, followed by reference, and then conjunction. We adopted this ranking, and so we classify intrasentential substitution and ellipsis as strongly connective (*conjunct*<sup>3</sup>), reference as moderately connective (*conjunct*<sup>2</sup>), conjunction as mildly connective (*conjunct*<sup>1</sup>), and interpolation as disconnective (*antijunct*<sup>2</sup>).

In assigning a hierarchic classification to a syntactic component, we adapted Halliday's [1985] work on *subordination*, specifically, embedding and hypotaxis, and the definition of the term *superordination* by Quirk *et al.* [1985]. We classify embeddings as strongly subordinate, *subjunct*<sup>3</sup>, and hypotactic structures as only mildly subordinate, *subjunct*<sup>1</sup>.<sup>18</sup>

In Section 2.1.1, we gave an informal description of the abstract elements of *concord* and *discord*. We can now describe the precise nature of concords and discords in the connective view: the concords are associated with constructions that are connective, for it is assumed that cohesion is the default usage, while the discords are disconnective structures, as a lack of cohesion is less usual. Thus, a concord, which may be *initial*, *medial*, or *final*, is produced by a *conjunct* construction; a discord, which may also appear in any of these three positions, is associated with an *antijunct* construction.

## 2.2 Stylistic goals

As we have noted, the abstract elements are defined in terms of the lower-level primitive elements. In turn, the abstract elements are used as the basis for the definition of higher-level *stylistic goals*. The precise nature of stylistic goals, such as clarity or concreteness, has been difficult to articulate, so that, up to now, stylists have only been able to define such goals by means of examples and informal rules. But we have been able to generalize from the plethora of historical stylistic terms to recognize similarities between groups of sentences with common stylistic effects—thus, we can give formal definitions of stylistic goals in terms of the abstract elements.<sup>19</sup>

Stylistic goals can be organized along orthogonal dimensions. For example, a writer might try to be clear, or obscure, or make no effort either way. *Clarity* and *obscurity* are thus opposite ends of a stylistic dimension. Likewise, the goals of *concreteness* and *abstraction* form a dimension, and so do *staticness* and *dynamism*. Below, we look at one end of each of these dimensions.

**Clarity** Stylists have commonly taken clarity to be adherence to the stylistic norm and so we incorporate accepted definitions of norm, that is, cohesive and concordant structures, in our rule for clarity. Thus, we can give precise definitions to the kinds of sentences that stylists like Kane [1983], for example, associate with clarity:

---

<sup>17</sup>Another type of structure that seems to work against cohesion is a disruption in normal linear ordering, as the syntactic inversion at the end of the following text: *And the rains descended, and the floods came, and the winds blew, and beat upon that house; and it fell: and great was the fall of it* (Matthew 7:27, Revised Standard Version 1952). However, we have not yet incorporated this feature into our formalization of style.

<sup>18</sup>In this work, we do not use the full range of connectivity defined above; however, extensions of the theory by [Green, 1992] and [Hoyt, 1993] do. Also, we have not yet considered the role of the hierarchical elements in our model of comparative stylistics.

<sup>19</sup>We formalized the set of the most commonly used stylistic goals that appeared in Vinay and Darbelnet's [1958] textbook of French–English comparative stylistics.

- *Simple* sentences, which consist of one independent clause: these are our *monoschematic* sentences.
- *Centered* sentences, which consist of dependent constructions, followed by a main clause, followed by additional dependent clauses. We expand the notion of ‘centered sentence’ to include any sentence in which there is a dominant, concordant core: these are our *centroschematic* sentences.
- *Parallel* sentences, which reduce ambiguity by stressing the same grammatical form: these are our *homopoisal* sentences.

**Concreteness** Kane [1983] associates *concreteness* with sentences that suggest an effect of immediacy, in which the writer has arranged elements to reflect the natural order of events or ideas, so that syntax mirrors events. In our terms, concreteness is associated with sentences that emphasize a particular component, which may be highlighted either because it is *discordant* or because it is parenthetical, as in a *heteropoise*.

**Staticness** We associate *staticness* with sentences in which there is little opportunity for stylistic variation, that is, sentences that verge on being ‘fixed forms’. These are the *monoschematic* sentences, which are standard and simple structures, or the more elaborate but strictly balanced structures, the *homopoisal* sentences.

### 2.3 Structure of the computational model of style

In Figure 1, the overall structure of our computational model of style is displayed.

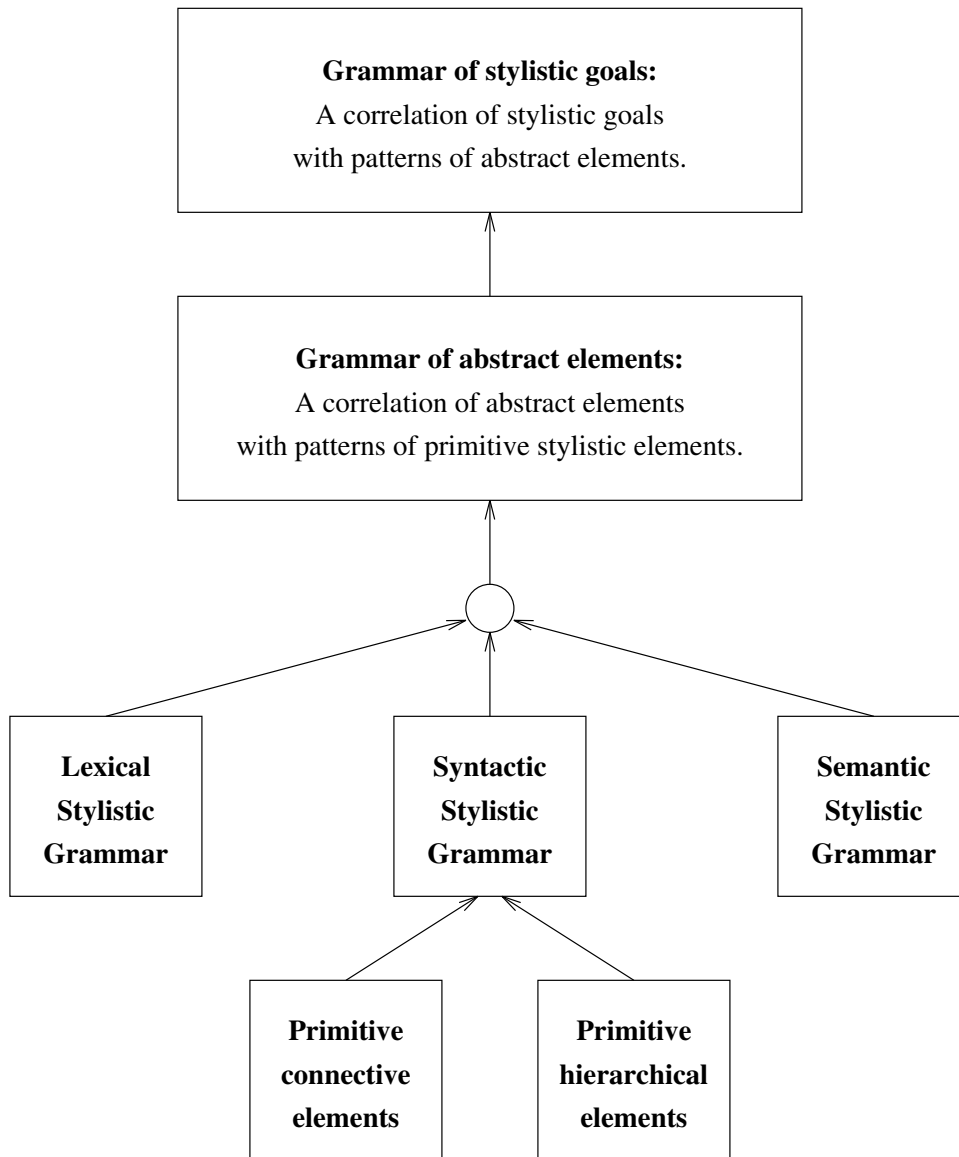


Figure 1: The computational model of style

### 3 A set of informal translation rules

#### 3.1 A catalogue of Guillemin-Flescher's rules

We now have a computational theory of syntactic style. We want to use this theory to develop formal rules of comparative stylistics that could be incorporated into an MT system. Such a formalization would provide information on how to make the best choice from among a variety of options in a particular translation situation. As we will see, deviations from a literal translation, different choices of syntactic structure, can affect both the naturalness and rhetorical effect of the resulting target-language text.

As the basic material for our formal rules, we selected a subset of traditional translation rules from those described by Guillemin-Flescher [1981].<sup>20</sup> She catalogued rules of French–English comparative stylistics that human translators implicitly use in transforming a French sentence into an appropriate English one. These transformations are actually restructurings that would be made in order to achieve a style that is natural to English, as a literal translation might not be. While her rules are quite clear and thorough, they are descriptive, not sufficiently precise to be used in an MT system. Therefore, our goal was to take Guillemin-Flescher's rules and encode them in a computational form, using the notation developed for our theory of style.

Guillemin-Flescher's rules of French–English comparative stylistics emphasize the clauses of a sentence and the relationships that exist between them. The format of the rules is as follows:

French syntactic structure  $\longrightarrow$  English syntactic structure

This is interpreted as a rewriting of a French structure (the left-hand side) into a different kind of English structure (the right-hand side).

The format of the rules we will present is as shown in Guillemin-Flescher, with minor changes to the names given to syntactic structures. The rules that we selected from her catalogue have a common underlying theme: a group of rules with a common left-hand side can be ordered so that the right-hand sides are constructs that express varying degrees of connectivity. As we stated in section 2.1.2, the concept of connectivity describes how closely two syntactic constructs are linked together.

##### 3.1.1 Anteposition

The first group of rules that we will present are the *antepositions*. An anteposition is a word or group of words that precedes the construct that it modifies. This is similar to our definition of premodification, but is not limited to the modification of nominal groups. A rule with a strong degree of connectivity would merge the anteposition with the construct it modifies, signifying the close relationship between the two concepts that are expressed. This is demonstrated in the two following rules and examples.<sup>21</sup>

Anteposition  $\longrightarrow$  Nominal group constituent

*Dans nos agences, nous nous faisons un plaisir de donner à nos clients tous conseils pour la meilleure utilisation de leurs pneus. (lit., In our offices, we take pleasure in giving to our customers all advice concerning the best use of their tires.)*<sup>22</sup>

---

<sup>20</sup>The rules that we omitted generally dealt with matters of lexical or semantic style that are not accounted for by our current theory.

<sup>21</sup>Each pair of parallel sentences in Guillemin-Flescher is labelled as a single example.

<sup>22</sup>[Guillemin-Flescher, 1981] (henceforth GF), 338, example 50.

*The staff of our Branches [sic] will be pleased to give advice on the best use of tires.*

**Anteposition** → **Principle clause constituent**

*En France, il y a 18 000 bureaux de Poste; il y en a toujours un, là où vous êtes. (lit., In France, there are 18,000 post offices; there is always one wherever you are.)*<sup>23</sup>

*There are 18,000 post offices in France and one wherever you are.*

In the two preceding rules, the French antepositions are merged into another construction in the English translation.

Stronger degrees of connectivity are possible. For example, the following rule extends past sentence boundaries to merge two sentences into one using a co-ordinating conjunction and punctuation (*and* and *;* in the example).

**Anteposition** → **Principle co-ordinate clause**

*La réservation s'étendra à toutes les catégories de places, y compris celles de voitures-lits, de restaurant et d'automobiles (trains auto couchettes et services auto express). Imprimé automatiquement, le titre délivré donnera la référence des places et comportera éventuellement le titre de transport, ce qui lui conférera alors valeur en tant que billet. (lit., The reservation will cover all categories of bookings, including sleepers, dining cars, and car-sleepers. Automatically printed, the ticket issued will give the reservation number and will also serve as the actual ticket, which will thus give him the value of a ticket.)*<sup>24</sup>

*Reservations will cover all categories of booking including sleepers, restaurant-cars and Motorail and will be automatically printed; the voucher handed over will give the reservation number and will include the actual ticket.*

### 3.1.2 Relative Clause

The next group of Guillemin-Flescher's rules, for the *relative clauses*, can also be ordered according to the degree of connectivity associated with each rule. The following four rules describe a strong degree of connectivity, since the relative clause, a distinct component of the French sentence, is merged with the main clause of the English sentence. Each of the four rules demonstrates a different way in which to merge the French relative clause into the English main clause, *e.g.*, as an adjectival, participle, or prepositional phrase.

**Relative clause** → **Principal clause constituent (adjectival)**

*Ce rêveur, qui doute de lui-même, n'a pas été aidé par les critiques qui lui reprochent de ne pas vivre avec son temps. (lit., This dreamer, who doubts himself, had not been helped by critics who reproach him for not living with the times.)*<sup>25</sup>

*He is a self-doubting dreamer and this has not been helped by critics who felt that he was not moving enough with the times.*

**Relative clause** → **Principal clause constituent (participle)**

*Les lignes non électrifiées, qui représentent 74% de la longueur du réseau, n'étaient que 22% du trafic total de la S.N.C.F. (lit., The non-electrified lines, which represent 74% of the length of the network, carry only 22% of the total traffic of the S.N.C.F.)*<sup>26</sup>

---

<sup>23</sup> GF, 334, example 33.

<sup>24</sup> GF, 338, example 48.

<sup>25</sup> GF, 341, example 59.

<sup>26</sup> GF, 342, example 62.

*Non-electrified lines representing 74% of the total network carry only 22% of the overall S.N.C.F. traffic.*

Relative clause → Truncated principle clause constituent (participle)

*Ce n'est rien, capitaine!.. Une branche morte qui s'est brisée.. (lit., It is nothing, captain!.. a dead branch that broke..)*<sup>27</sup>

*I'm all right, Captain.. only a rotten branch breaking..*

Relative clause → Principal clause constituent (prepositional phrase)

*Je vais vous lancer une corde à laquelle est attachée une ceinture de sauvetage. (lit., I am going to throw you a line to which is attached a life-jacket.)*<sup>28</sup>

*I'll throw you a line with a life-jacket attached.*

The next rule describes a weaker degree of connectivity than the preceding four, since the relative clause is translated as a temporal adverbial clause rather than being merged with the main clause.

Relative clause → Temporal clause

*Pendant quarante-cinq jours, il débarrassa les arbustes de leur première écorce et recueillit l'écorce intérieure en la découpant en lanières. Puis il fit longuement bouillir dans chaudron cette masse fibreuse et blanchâtre qui se décomposa peu à peu en un liquide épais et visqueux. Il le remit ensuite au feu et le répandit brûlant sur la coque du bateau. (lit., For 45 days, he rid the bushes of their outside bark and gathered the marrow by cutting it in strips. Then he boiled the fibrous whitish mass which decomposed little by little into a thick, viscous liquid. He returned it to the fire and spread it on the hull of the boat.)*<sup>29</sup>

*He boiled the pith in a small cauldron he had brought from the Virginia until the fibrous mass was rendered down to a thick, viscous liquid, which he reheated and smeared over the vessel's hull.*

The next rule describes an even weaker degree of connectivity. In this case, the relative clause is translated as a separate sentence, which is less connective than being translated as a clause of the original sentence.

Relative clause → Principle clause

*Oui, j'en ai laissé quelques-uns au laboratoire de police scientifique, où on les examine en ce moment. (lit., Yes, I left some at the laboratory at police headquarters, where they are examining them at this moment.)*<sup>30</sup>

*Yes, I've left some of them at the laboratory at police headquarters. They're working on them now.*

---

<sup>27</sup> GF, 342, example 63.

<sup>28</sup> GF, 342, example 64.

<sup>29</sup> GF, 340, example 57.

<sup>30</sup> GF, 340, example 56.



### 3.1.3 Juxtaposition

The next group of rules are the *juxtapositions*, which describe the placement of two syntactic units next to each other. In particular, a juxtaposition can be used for clarification or to add information. For example, *the Prime Minister, Jean Chrétien*, is a juxtaposition that adds information about which prime minister is being talked about, e.g., Jean Chrétien as opposed to Kim Campbell.

The rules for the various kinds of juxtapositions are as follows:

Juxtaposition → Principal clause (or constituent)

*Ces voitures, plus longues que celles d'aujourd'hui, offriront 11 compartiments de 6 places, soit 66 places en 2<sup>e</sup> classe. (lit. These cars, longer than those of today, will have 11 compartments of 6 places each, making 66 places in 2nd class.)*<sup>31</sup>

*These coaches will be longer than those in use now, and will have 11 × 6 seat compartments, viz 66 seats in 2nd class.*

Juxtaposition → Explicitly marked relation

*La jeune fille, plus grande que lui de quelques pouces, marchait un peu en avant. (lit., The girl, taller than him by several inches, was walking a little in front.)*<sup>32</sup>

*The girl, who was several inches taller than he, walked slightly in advance of him.*

The first rule involves a moderate degree of connectivity as it makes the juxtaposition a part of the main clause. The second rule can vary in its degree of connectivity; the instance shown is only an example. The degree of connectivity depends on whether the relationship between the two juxtaposed entities is explicitly marked, or is elliptic. As we described in section 2.1.2, ellipsis is a strongly cohesive relation. If a juxtaposition relationship is expressed as an ellipsis, as in *la jeune fille, plus grande que lui...*, then it is more connective than if the relationship is expressed by an explicit marker, as in *the girl, who was several inches taller...*

### 3.1.4 Embedding

The two remaining groups of rules require some explanation. They are both best described syntactically, but are influenced by lexical or semantic style. The first of these groups is the *embedding*. An embedding is an interruption of a construct. In translation, the interruption can either be sustained, or moved before (anteposition) or after (postposition) the construct that it interrupts. By moving the construct that was embedded, a sentence more connective than the original is produced because the disconnective effect of the interruption has been removed. The choice of whether to move the embedding before or after the construct that it interrupts appears to be partially governed by *thematization*, an aspect of semantic style.

Embedding → Anteposition

*L'Assurance garantit, sauf déclaration spéciale, la totalité des bagages de l'Assuré, y compris ceux conservés à la main. (lit., The Insurance guarantees, unless otherwise stated, all the Insured's luggage, including hand-luggage.)*<sup>33</sup>

*Unless otherwise stated, your insurance covers all your luggage including hand-luggage.*

---

<sup>31</sup>GF, 348, example 87.

<sup>32</sup>GF, 346, example 78.

<sup>33</sup>GF, 329, example 12.

Embedding → Postposition

*La première, réalisée depuis la fin de 1964, a été une phase préparatoire. (lit., The first, completed since the end of 1964, had been a preparatory phase.)*<sup>34</sup>

*The first was the preparatory stage, completed at the end of 1964.*

### 3.1.5 Inversion

The final group of rules, the *inversions*, covers a broad category and captures a large number of the differences between the source and target languages. These include differences in syntactic ordering (as in the example below), lexical influences, and semantic influences. An example of a semantic influence is an implied opposition, as in the following sentence:

- (15) In Athalie he condemned the ideas, in Phèdre the style, in Andromaque the characterisation.<sup>35</sup>

The rule below expresses the concept of reordering, *i.e.*, returning to the canonical or most accepted ordering, of components that seem awkward or out of order (inverted). In the example below, the reordered English translation is less awkward (and thus more connective) than a more literal translation.

Inversion → Canonical ordering

*Mettre en casserole profonde 75 g de beurre, à fondre sans brunir. (lit., Place in a deep saucepan 75 g of butter, to melt without browning.)*<sup>36</sup>

*Place 4 tablespoons of butter in a deep saucepan to melt but not brown.*

## 3.2 A small corpus study

We studied a small corpus of French–English translations to verify that the rules selected from Guillemin-Flescher’s catalogue were indeed capable of accounting for differences in the syntactic style of a representative set of parallel texts. We chose segments of parallel French and English magazine articles that could be classified as *political advocacy* texts, which include editorials, reviews, and promotional literature. The purpose of this type of text is to express the author’s opinion on a topic and to influence the reader. As a result, the writing tends to be high-quality, syntactically complex, and rhetorically expressive—all desirable characteristics for the application of our theory of style.

A corpus of sentences from political advocacy texts was constructed from three randomly selected paragraphs from each of nineteen articles and their translations. The articles used in the corpus include parallel sentences from *Le Monde* and the *Manchester Guardian Weekly*, consumer magazines, a textbook of translation, government publications, university publications, airline magazines, and miscellaneous magazines. The corpus contains a total of 162 pairs of sentences from 57 paragraphs.

After the corpus was constructed, it was examined for evidence of the translation rules presented in the preceding section. Figure 2 lists the 14 rules (presented in the preceding section) that were applied to pairs of sentences in the corpus and the number of occurrences of each rule. Of

---

<sup>34</sup> GF, 330, example 17.

<sup>35</sup> GF, 116.

<sup>36</sup> GF, 327, example 2.

1.	Anteposition	→ Nominal group constituent	3
2.	Anteposition	→ Principal clause constituent	2
3.	Anteposition	→ Principal co-ordinate clause	1
4.	Relative clause	→ Principal clause constituent (adjectival)	2
5.	Relative clause	→ Principal clause constituent (participle)	3
6.	Relative clause	→ Truncated principal clause constituent (participle)	1
7.	Relative clause	→ Principal clause constituent (prepositional phrase)	4
8.	Relative clause	→ Temporal clause	0
9.	Relative clause	→ Principal clause	1
10.	Juxtaposition	→ Principal clause constituent	2
11.	Juxtaposition	→ Explicitly marked relation	6
12.	Embedding	→ Anteposition	7
13.	Embedding	→ Postposition	5
14.	Inversion	→ Canonical ordering	51

Figure 2: The rules used in the corpus (with frequencies).

stylistic variations that are not covered by the rules, there were seven cases attributed to lexical style and 134 cases attributed to semantic style. Of the cases attributed to semantic style, 83 were the result of information being added or omitted in the translation, and 51 were a result of changes in the point of view. We also examined the corpus for evidence of counterexamples, that is, cases where an awkward but literal translation had been chosen in preference to a restructuring that would produce a natural-sounding sentence. As we cannot in most cases be sure as to which was the original source language for each pair of sentences, we can say only that all the English sentences had structures appropriate to the language, even if a good deal of reorganization from the French was required.

In looking at the results in Figure 2, we can see that there appear to be significant frequencies of rules that involve an increase of connectivity in French to English translation. For example, there is an indication that some of the most frequently occurring rules are those that involved the movement of a separate constituent into a more cohesive unit, like a nominal group (rule 1). For rules 12 and 13, an interrupting, disconnective construct was moved to a less dislocating position. For rule 14, a wide variety of cases was involved so that it is difficult to make a conclusive statement about the significance of the high frequency, but we can hypothesize that this rule indicates that literal translations were avoided even if it meant restructuring the sentence to achieve a more canonical ordering.

As we cannot simply infer from our results that English is an inherently more cohesive language than French, we can assume only that a translator, working in whichever direction, will tend to choose a more connective, smoother-sounding structure. But this suggests that any model of comparative stylistics that we develop must be one-way, for a structure that is connective in one language might not be so in another—and the rule of comparative stylistics that describes this situation as a decrease in connectivity cannot be assumed to be reversible. What we can hope to construct, however, is a general model of comparative stylistics, with a format and a vocabulary of concepts that can be adapted to different languages and different directions in translation. In the next section, we present an instance of such a model for French–English translation, with the expectation that the approach is more generally applicable.

## 4 A grammar of French–English comparative stylistics

### 4.1 Format of the grammar

In this section, we will present a grammar of computational rules derived from the informal textbook rules identified in Figure 2 that will form a model of French–English comparative stylistics. The first step in the process of deriving the computational rules was to use our theory of style to produce stylistic parses for each of the sentences in the corpus described in Section 3.2. Each of the sentences was parsed by hand. (Although *STYLISTIQUE*, the first implementation of the stylistic grammar, was available, it proved to be too inefficient to parse many of the sentences.<sup>37</sup>) After the stylistic parsing had been completed, each of the sentence pairs in the corpus, *i.e.*, a French sentence and its corresponding English translation, was labelled according to the applicability of the textbook rules in Figure 2.

For each instance of Guillemin-Flescher’s rules, the French and English primitive-element descriptions of the sentence components corresponding to the left-hand and right-hand sides of the rule were recorded. The correspondences between French and English primitive-element descriptions will be written in the following form:

French primitive-element description  $\mapsto$  English primitive-element description

For example, the correspondence between the underlined components in sentences (16) and (17) below would be written in our grammar as follows:

conjunct<sup>2</sup>-postmodification  $\mapsto$  conjunct<sup>4</sup>-postmodification

- (16) Mais maintenant, les chercheurs présumant que des radicaux libres d’oxygène, métabolites naturels, mais nocifs, pourraient être les principaux déclencheurs de la réaction en chaîne qui aboutit à la cataracte. (lit., But now, the researchers presume that oxygen free-radicals, natural, but harmful, metabolic by-products, may be the main trigger of the chain reaction that leads to cataracts.)
- (17) But now researchers feel free radicals of oxygen, natural but harmful bi-products of our metabolism, are the prime suspects in starting the domino reaction leading to cataracts.

In this example, a relative clause, which is classified as conjunct<sup>2</sup> as it is an instance of reference, is translated as a participle clause, classified as conjunct<sup>4</sup>, more strongly cohesive, as it is an instance of ellipsis.

These correspondences are grouped according to the kind of stylistic transformation from French to English syntactic structures that they describe. A rule of comparative stylistics is formed from the pairing of a type of transformation, *i.e.*, an *abstract element of comparative stylistics*, with all the correspondences that can produce it. For example:

Abstract element of comparative stylistics  $\longrightarrow$

French primitive-element description  $\mapsto$  English primitive-element description

---

<sup>37</sup>Since this part of the study was done, an efficient and more comprehensive stylistic analyzer, named ASSET, has been completed [Hoyt, 1993].

Therefore, each formal rule of comparative stylistics that we will present will have a left-hand side that gives the *name* of the rule, an abstract element of comparative stylistics, and a right-hand side that lists a *correspondence* between the primitive-element descriptions of a French sentence and its English translation. For conciseness, rules with a common left-hand side will be presented together—a single name will be given, followed by a list of alternative correspondences.

## 4.2 A formalization of Guillemin-Flescher’s rules

### 4.2.1 Concord

We will first define abstract elements of comparative stylistics that are associated with effects of *concord*. We will concentrate on cohesive effects related to the nominal group, as the stylistic characteristics of the nominal group are fundamental to our theory of style. We will consider situations where the degree of connectivity<sup>38</sup> is *augmented* in translation, or simply *sustained*, or even *diminished*.

The first group of rules to be described is the *augmented concords*. The term *augmented* refers to the increase in connectivity in translating from French to English, *i.e.*, the degree of conjunctness increases from French to English.

The rule for *augmented concord* has two basic kinds of alternatives. In the first case, a separate constituent is integrated into the nominal group, *i.e.*, it becomes part of a more cohesive component in translation. For example, an independent French noun might become part of the premodification of the English nominal group. In the second case, the cohesion *within* the nominal group is increased in the translation. For example, a mildly connective French postmodifying adjective might be replaced by a more strongly connective English relative clause.

In the rule below, we present two representative alternatives of each kind of *augmented concord*:  
augmented-concord →

adjective  $\mapsto$  conjunct<sup>*i*</sup>-premodification<sup>39</sup>

*Seuls, quelques hommes discutent à voix basse, faisant les cent pas sur la route en lacets qui traverse ce village situé à 8 kilomètres à l’est de la route Beyrouth-Damascus. (lit., Alone, several men talk in low voices, pacing on the winding road that goes through the village located eight kilometres to the east of the Beirut-Damascus road.)*<sup>40</sup>

*Only a few men stood around talking in low voices on the road that runs through the village some eight kilometres to the east on the Beirut-Damascus road.*<sup>41</sup>

noun  $\mapsto$  conjunct<sup>*i*</sup>-premodification<sup>42</sup>

*L’efficacité et accessibilité de ce régime en font l’un des meilleurs au monde. (lit., The effectiveness and accessibility of this system makes it one of the best in the world.)*<sup>43</sup>

*It is among the most economical, effective, and accessible health care systems in the world.*<sup>44</sup>

<sup>38</sup>As explained in Section 2.1.2, connectivity is the closeness with which two syntactic constructs are bound together.

<sup>39</sup>*i* can take on any value from 1 to 6.

<sup>40</sup>[Chipaux, 1990b].

<sup>41</sup>[Chipaux, 1990a].

<sup>42</sup>*i* can take on any value from 1 to 6.

<sup>43</sup>[Conseil canadien de développement social, 1990].

<sup>44</sup>[Canadian Council on Social Development, 1990a].

conjunct<sup>1</sup>-postmodification  $\mapsto$  conjunct<sup>2</sup>-postmodification

*Les personnes pauvres du Canada devraient être les dernières à souffrir des compressions budgétaires. (lit., The poor people of Canada should be the last to suffer the budget cutbacks.)*<sup>45</sup>

*Canadians who are poor should least expect to suffer further from the government's belt-tightening.*<sup>46</sup>

conjunct<sup>2</sup>-postmodification  $\mapsto$  conjunct<sup>4</sup>-postmodification

*Mais maintenant, les chercheurs présument que des radicaux libres d'oxygène, métabolites naturels, mais nocifs, pourraient être les principaux déclencheurs de la réaction en chaîne qui aboutit à la cataracte. (lit., But now, the researchers presume that oxygen free-radicals, natural, but harmful, metabolic by-products, may be the main trigger of the chain reaction that leads to cataracts.)*<sup>47</sup>

*But now researchers feel free radicals of oxygen, natural but harmful bi-products of our metabolism, are the prime suspects in starting the domino reaction leading to cataracts.*<sup>48</sup>

Each of the alternatives for augmented concord describes an increase of connectivity in translation. For the first two alternatives, independent constituents become integrated into the more cohesive nominal group in translation. In the first one, the independent French adjective *seuls* is translated as part of the premodification, *only*, in the English nominal group. And, in the second one, the French nouns *l'efficacité* and *accessibilité* become the English premodifying adjectives *effective* and *accessible*.

For the third and fourth alternatives, connectivity is increased within the nominal group in translation. The conjunct<sup>1</sup> postmodification, *pauvre*, becomes the more cohesive conjunct<sup>2</sup> postmodifying relative clause, *who are poor*. And, finally, the conjunct<sup>2</sup> postmodification, *qui aboutit à la cataracte*, is translated as the more elliptical, and therefore more cohesive, conjunct<sup>4</sup> postmodifying non-finite clause, *leading to cataracts*.

The following rule defines the various types of *diminished concord*. Here the term *diminished* describes the decrease in connectivity in the translation from French to English. The structure of the rule for diminished concord is complementary to that for augmented concord, also having two basic kinds of alternatives. For the first kind, an element of the nominal group is 'broken out' in translation, *i.e.*, it is removed from the cohesive group structure. For example, a postmodifying relative clause might be moved out of the nominal group into a separate clause. For the second kind, the cohesion within the nominal group is reduced in the translation. For example, a strongly connective French postmodifying non-finite clause might be replaced by a weakly connective English premodifier.

The rule for *diminished concord* is as follows:

diminished-concord  $\longrightarrow$

conjunct<sup>*i*</sup>-nominal-group  $\mapsto$  clause<sup>49</sup>

<sup>45</sup>[Conseil canadien de développement social, 1990].

<sup>46</sup>[Canadian Council on Social Development, 1990a].

<sup>47</sup>[Dappen, 1991a].

<sup>48</sup>[Dappen, 1991b].

<sup>49</sup>*i* can take on any value from 1 to 6.

*A l'opposé, la description de modèles sociologiques, systématiques et abstraits, inspirés de conceptions a priori, aboutit à dresser l'écran d'une construction artificielle devant les événements et les individus. (lit., On the other hand, the description of sociological, systematic, and abstract models, inspired by a priori conceptions, leads to the raising of a screen of artificial construction before the events and the individuals.)*<sup>50</sup>

*On the other hand, simply to describe a series of systematized sociological models, constructed on an a priori basis, is to interpose a dark screen between the reader and the events and the people he is reading about.*<sup>51</sup>

**conjunct<sup>2</sup>-postmodification**  $\mapsto$  **clause**

*Il semble que Cook était plutôt fâché contre les indigènes qui lui avaient volé l'un de ses navires et qu'il leur avait vaillamment déclaré la guerre. (lit., It seems that Cook was rather angry at the natives who had stolen one of his ships and that he had valiantly declared war on them.)*<sup>52</sup>

*It seems that Cook was rather peeved at the natives for stealing one of his ships, so he made war upon them.*<sup>53</sup>

**conjunct<sup>4</sup>-postmodification**  $\mapsto$  **conjunct<sup>2</sup>-postmodification**

*Un « bilingue fonctionnel » formé par l'école québécoise est, comme le veut l'objectif global du ministère de l'Éducation, un élève non anglophone qui est capable d'utiliser la langue anglaise pour communiquer dans des situations correspondant à ses intérêts. (lit., A functional bilingual trained at a Quebec school is, as the general objectives of the Ministry of Education require, a non-English-speaking student who is capable of using the English language to communicate in situations corresponding to his/her interests.)*<sup>54</sup>

*"Functionally bilingual" students trained by a Quebec school are, according to the general objectives of the Department of Education, non-Anglophone students who are able to use English to communicate in situations that correspond to their needs and interests.*<sup>55</sup>

**conjunct<sup>4</sup>-postmodification**  $\mapsto$  **conjunct<sup>1</sup>-premodification**

*Au début 1990, un correspondant anglophone de la région d'Ottawa-Hull nous informait qu'il avait posé sa candidature à un poste bilingue offert pour une période indéterminée par un ministère dans cette région. (lit., At the beginning of 1990, an anglophone correspondent from the Ottawa-Hull region informed us that he had presented his candidacy for a bilingual position offered for an unspecified period of time by a ministry in that region.)*<sup>56</sup>

*Early in 1990 an Anglophone correspondent from the Ottawa-Hull area told us that he had applied for an indeterminate bilingual position with a department in the region.*<sup>57</sup>

Each of the alternatives for diminished concord describes a decrease of connectivity in translation. For the first two alternatives, a cohesive nominal group is broken out into separate constituents in translation. In the first one, the French nominal group, *la description de modèles*

<sup>50</sup>[Dupeux, 1988b].

<sup>51</sup>[Dupeux, 1988a].

<sup>52</sup>[MacDonald, 1990b].

<sup>53</sup>[MacDonald, 1990a].

<sup>54</sup>[Beauparlant, 1990a].

<sup>55</sup>[Beauparlant, 1990b].

<sup>56</sup>[Duhaim, 1990b].

<sup>57</sup>[Duhaim, 1990a].

*sociologiques, systématiques et abstraits*, is translated as the English infinitival clause, *to describe a series of systematized sociological models*. In the second, one part of the nominal group, the postmodifying relative clause, *qui lui avaient volé l'un de ses navires*, becomes the English clause, *for stealing one of his ships*.

For the third and fourth alternatives, connectivity is decreased within the nominal group in translation. The conjunct<sup>4</sup> postmodification, *correspondant à ses besoins et à ses intérêts*, becomes the non-elliptical, and therefore less cohesive, conjunct<sup>2</sup> postmodifying relative clause, *that correspond to their needs and interests*. And, finally, the conjunct<sup>4</sup> postmodification, *offert pour une période indéterminée* is translated as the weakly cohesive conjunct<sup>1</sup> premodification, *indeterminate*.

The next rule defines the abstract element, *sustained concord*, which describes the preservation of the same degree of connectivity across the translation from French to English. The rule for sustained concord describes instances of four general cases: a premodifier is translated as a premodifier, a postmodifier as a postmodifier, a postmodifier as a premodifier, and a premodifier as a postmodifier.<sup>58</sup>

sustained-concord →

conjunct<sup>3</sup>-premodification ↔ conjunct<sup>3</sup>-premodification

*Sa candidature fut rejetée parce que son français n'était pas apparu satisfaisant à l'intervieweur. (lit., His candidacy was rejected because his French didn't appear satisfactory to the interviewer.)<sup>59</sup>*

*His candidacy was rejected because, in the interviewer's opinion, his French was not adequate.<sup>60</sup>*

conjunct<sup>2</sup>-postmodification ↔ conjunct<sup>2</sup>-postmodification

*Ces questions, sauf celles portant sur le harcèlement, ne se fondent peut-être pas directement sur la violence, mais elles révèlent la nécessité de remédier à la discrimination systématique qui contribue à désavantager les femmes... (lit., These questions, except those pertaining to harassment, are perhaps not based directly on violence, but they reveal the need to remedy the systematic discrimination that contributes to disadvantaging women...)<sup>61</sup>*

*Except for harassment, these issues may not relate directly to violence, but they are indicative of the need to correct the systemic discrimination that functions to disadvantage women...<sup>62</sup>*

conjunct<sup>1</sup>-postmodification ↔ conjunct<sup>1</sup>-premodification

*Nous avons présenté des modifications à l'entente avec le gouvernement fédéral et le Cabinet tranchera la question. (lit., We have introduced modifications to the agreement with the federal government and the Cabinet will resolve the question.)<sup>63</sup>*

*We are asking for changes in our agreement and the federal government has taken these changes before Cabinet for a decision.<sup>64</sup>*

<sup>58</sup> We include the last case as it is theoretically feasible, even though there were no examples in our corpus.

<sup>59</sup> [Duhaine, 1990b] edited for length.

<sup>60</sup> [Duhaine, 1990a] edited for length.

<sup>61</sup> [Unknown, 1990a] edited for length.

<sup>62</sup> [Unknown, 1990b] edited for length.

<sup>63</sup> [Unknown, 1991a].

<sup>64</sup> [Unknown, 1991b].



conjunct<sup>i</sup>-premodification  $\mapsto$  conjunct<sup>i</sup>-postmodification

Each of the alternatives of sustained concord describes the preservation of the same degree of connectivity in translation. In the first alternative, the French adjective *son* is a simple possessive with a default classification of conjunct<sup>3</sup>. The corresponding English adjective *his* is also a simple possessive. This translation maintains the same connectivity between these French and English adjectives.

In the second alternative, the French postmodifying relative clause, *qui contribue à dévantagez les femmes*, has a classification of conjunct<sup>2</sup>. The English clause, *that functions to disadvantage women*, also receives a classification of conjunct<sup>2</sup>.

In the third alternative, the French postmodifying prepositional phrase *à l'entente* is translated as the English postmodifier *in our agreement*; and they both have a classification of conjunct<sup>1</sup>.

#### 4.2.2 Discord

Stylistic *discord* can be produced in two ways: by a lack of connectivity or by an excess of it.<sup>65</sup> Thus, we can propose rules that describe *augmented discord*, *diminished discord*, or *sustained discord*, according to these two different situations. From the form of the rules and correspondences that we observed in the corpus, we were able to hypothesize certain additional rules and correspondences for a general model of comparative stylistics. The rules that were observed are presented with examples from the corpus, while those that were hypothesized do not have an associated example.

The rule describing augmented discord is as follows:

augmented-discord  $\longrightarrow$

conjunct<sup>i</sup>-nominal-group  $\mapsto$  antijunct<sup>j</sup>-nominal-group<sup>66</sup>

conjunct<sup>i</sup>-nominal-group  $\mapsto$  conjunct<sup>k</sup>-nominal-group<sup>67</sup>

The general rule expressing augmented discord is intended to describe situations where concordant, cohesive components become discordant in translation either because they are non-cohesive or excessively cohesive. For example, a French postmodifying adjective, which is entirely normal, might be translated literally as an English postmodifying adjective, which is an instance of interpolation and therefore discordant:

(18) Un homme toujours timide n'est pas en état de faire ce travail. (*lit.*, A man always timid is not in shape to do this job.)

(19) A man always timid is unfit for this task.

---

<sup>65</sup>As we are focusing on variations in cohesion of the nominal group, we cannot yet account for excessive cohesive effects created or removed in translation at the sentence level. For example, in French, we might have the sentence, *Vous m'emprunter mon stylo si vous le voulez.* (*lit.*, *You can borrow my pen if you wish it.*), which is perfectly cohesive and concordant. But, in English, the sentence might become, *You can borrow my pen if you want*  $\emptyset$ , which contains an instance of clausal ellipsis, which is excessively cohesive according to our theory.

<sup>66</sup> $i$  can take on any value from 0 to 4.  $j$  can take on any value from 0 to 4.

<sup>67</sup> $i$  can take on any value from 0 to 4.  $k$  can take on any value from 5 to 6.

We may also expect to see cases where a discordant interpolation is introduced into a postmodifier in translation:

(20) Norman Jones, un étudiant à cette époque-là, a écrit plusieurs best-sellers. (*lit.*, Norman Jones, a student at that time, wrote several best-sellers.)

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

The following rule describes the dual of augmented discord, *diminished discord*. The rule is intended to describe the situations where discordant components become concordant in translation. As we did earlier, we describe a general rule, but now one based on the replacement of non-cohesive components within the nominal group.

diminished-discord  $\longrightarrow$

antijunct<sup>i</sup>-nominal-group  $\mapsto$  conjunct<sup>j</sup>-nominal-group<sup>68</sup>

conjunct<sup>k</sup>-nominal-group  $\mapsto$  conjunct<sup>j</sup>-nominal-group<sup>69</sup>

The final rule in this group defines *sustained discord*, which predicts situations where discords are maintained across translation. As the rule indicates, the exact degree of antijunctness may vary in translation, but we still consider the discord to have been maintained.

sustained-discord  $\longrightarrow$

antijunct<sup>i</sup>-nominal-group  $\mapsto$  antijunct<sup>j</sup>-nominal-group

### 4.2.3 Heteropoise

The next series of rules describes the various kinds of *heteropoises* used in comparative stylistics. Recall from Section 2.1.1 that a heteropoise is a sentence with an interruption, or parenthesis. The first two groups of heteropoises, the *augmented heteropoises* and the *diminished heteropoises*, describe the introduction and removal, respectively, of parentheses in translation.

augmented-heteropoise  $\longrightarrow$

centroschematic  $\mapsto$  initial-heteropoise

*L'instabilité régnera jusqu'à la mise en place de nouveaux arrangements.* (*lit.*, *The instability will remain until the putting in place of new arrangements.*)<sup>70</sup>

*Until a future arrangement is understood, there will be instability.*<sup>71</sup>

centroschematic  $\mapsto$  medial-heteropoise

<sup>68</sup> *i* can take on any value between 0 and 4. *j* can take on any value between 0 and 4.

<sup>69</sup> *k* can take on any value between 5 and 6. *j* can take on any value between 0 and 4.

<sup>70</sup> [Unknown, 1991a].

<sup>71</sup> [Unknown, 1991b].

centroschematic  $\mapsto$  final-heteropoise

*Le Macvin est une autre des spécialités du Château d'Arley qui s'obtient par le mélange de 30 p. cent de marc à maturité avec du moût fraîchement pressé qu'on laisse ensuite vieillir entre 6 et 12 mois dans des tonneaux. (lit., Macvin is another of the specialties of the Chateau d'Arley which is made by the mixing of 30% of mature marc with freshly pressed grape must that is left next to age between 6 and 12 months in barrels.)<sup>72</sup>*

*Macvin is another of Château d'Arley's specialties, made by blending 30% of mature marc with freshly pressed grape must, followed by 6 to 12 months of barrel aging.<sup>73</sup>*

There are three alternatives for augmented heteropoise. The first is the situation where a centroschematic sentence, a canonical structure, is replaced by one with an initial parenthesis. In an analogous manner, the other two alternatives are transformations of centroschematic sentences into sentences with medial and final interruptions.

The dual of augmented heteropoise, *diminished heteropoise*, is associated with situations in which sentences with parentheses are smoothed out into stable centroschematic sentences.

diminished-heteropoise  $\longrightarrow$

initial-heteropoise  $\mapsto$  centroschematic

*Assurément, nous abordons cette décennie mieux prévenus que jamais; nous recherchons des produits « écologiques » et nous participons aux programmes de recyclage. (lit., Assuredly, we approach this decade better prepared than ever; we research ecological products and we participate in recycling programs.)<sup>74</sup>*

*We Canadians certainly enter the new decade more willing than ever to consider the earth when we shop. We're seeking out products we perceive to be "environmentally friendly" and dutifully participating in recycling programs.<sup>75</sup>*

medial-heteropoise  $\mapsto$  centroschematic

*Quand elle a essayé, en fin d'après-midi, de le conduire à hôpital, c'était trop tard. (lit., When she tried, near the end of the afternoon, to drive him to the hospital, it was too late.)<sup>76</sup>*

*When she did get him to the hospital towards the end of the afternoon, it was too late.<sup>77</sup>*

final-heteropoise  $\mapsto$  centroschematic

*Le compromis proposé aurait permis aux professeurs et aux étudiants de jouer des pièces en classe, sans auditoire. (lit., The compromise proposed would have allowed professors and students to play the performances in class, without an audience.)<sup>78</sup>*

*The compromise would have allowed for the performance of plays in class by students and teachers without any audience.<sup>79</sup>*

---

<sup>72</sup>[Chatto, 1990a].

<sup>73</sup>[Chatto, 1990b].

<sup>74</sup>[Reynolds, 1990a].

<sup>75</sup>[Reynolds, 1990b].

<sup>76</sup>[Chipaux, 1990b].

<sup>77</sup>[Chipaux, 1990a].

<sup>78</sup>[Kerr, 1990b].

<sup>79</sup>[Kerr, 1990a].

There are three alternatives for diminished heteropose. In the first one, the initial interrupting, parenthetical construction *assurément* is translated by *certainly*, which is moved into an interior, more syntactically integrated position. In the second one, the medial phrase *en fin d'après-midi*, which seems to be a digression, is transformed into the seemingly more essential *towards the end of the afternoon*. Lastly, in the third alternative, the final detached prepositional phrase *sans auditoire* is integrated into the body of the sentence as *without any audience*.

The next four rules describe four types of *sustained heteropose*. A *non-shifted heteropose* describes a situation in which the position of the parenthesis remains the same in translation. For example, in the second alternative, the medial interruption, *par exemple*, is sustained in the English translation by *for example*. A *front-shifted heteropose* is a situation in which the parenthesis moves to the beginning of the sentence. For example, in the first alternative, the medial interruption, *quant à moi* becomes the initial phrase *but for me* in the English translation. *Medial-shifted* and *end-shifted heteroposes* are cases where the parenthesis moves, respectively, either to the middle or end of the sentence.

non-shifted-heteropose →

initial-heteropose ↦ initial-heteropose

medial-heteropose ↦ medial-heteropose

*Au Languedoc, par exemple, la chaleur du soleil du sud transparait dans le savoureux marc de muscat, presque liquoreux tant la saveur du raisin s'y fait présente. (lit., In Languedoc, for example, the heat of the southern sun penetrates into the flavoursome marc de muscat, almost syrupy so much because the flavour of the grape makes itself known.)*<sup>80</sup>

*In Languedoc, for example, the warmth of the southern sun is embodied in a luscious marc de muscat so full of the flavour of the grape that it is almost sweet.*<sup>81</sup>

final-heteropose ↦ final-heteropose

front-shifted-heteropose →

medial-heteropose ↦ initial-heteropose

*Je préfère, quant à moi, les produits moins connus des maisons moins grandes. (lit., I prefer, in my opinion, lesser known products from smaller houses.)*<sup>82</sup>

*But for me, the fascination lies in the less familiar products of smaller houses.*<sup>83</sup>

final-heteropose ↦ initial-heteropose

medial-shifted-heteropose →

initial-heteropose ↦ medial-heteropose

*Dans sa petite robe noire, les yeux cernés par les larmes et la fatigue, Germaine raconte: « Mon frère perdait son sang, et je n'ai rien pu faire. » (lit., In her little black dress, her eyes shadowed by tears and fatigue, Germaine said, "My brother was losing blood, and I couldn't do anything.")*<sup>84</sup>

<sup>80</sup>[Chatto, 1990a, 30].

<sup>81</sup>[Chatto, 1990b, 30].

<sup>82</sup>[Chatto, 1990a, 30].

<sup>83</sup>[Chatto, 1990b, 30].

<sup>84</sup>[Chipaux, 1990b, 3].

*“My brother was losing blood, and I couldn’t do anything for him,” said Germaine, a woman in a little black dress, tearful and hollow-eyed with fatigue.*<sup>85</sup>

final-heteropose  $\mapsto$  medial-heteropose

end-shifted-heteropose  $\longrightarrow$

initial-heteropose  $\mapsto$  final-heteropose

medial-heteropose  $\mapsto$  final-heteropose

As we saw in Section 2.1.1, a heteropose can be either concordant or discordant, depending on whether the parenthetical component is cohesively linked to the rest of the sentence. Thus, when heteroposes are created or shifted in translation, we can also define *concordant* and *discordant* augmented or sustained heteroposes. For example, in the translation below, an interpolation—constructed using disruptive punctuation—is created at the end of the English sentence, producing an instance of a *discordant augmented heteropose*:<sup>86</sup>

discordant-augmented-heteropose  $\longrightarrow$

centroschematic  $\mapsto$  discordant-final-heteropose

*On leur a dit aussi que trop peu de professeurs méritent le qualificatif de remarquables ou d’excellents, surtout à cause des attitudes vis-à-vis de l’enseignement car la communauté universitaire ne souligne pas assez la valeur de l’excellence à ce chapitre. (lit., We also told them that too few professors warranted the status of remarkable or excellent, particularly because of attitudes with regard to education for the university community doesn’t stress enough the importance of this issue.)<sup>87</sup>*

*They also heard that too few teachers can be described as outstanding or excellent, primarily because of attitudes towards teaching—not enough is done in the academic community to emphasize the value of teaching excellence.*<sup>88</sup>

### 4.3 Summary of the abstract elements of comparative stylistics

A summary of the abstract elements of French–English comparative stylistics is presented in Table 1.

---

<sup>85</sup>[Chipaux, 1990a] edited.

<sup>86</sup>This is only one way in which a *discordant augmented heteropose* might be created; other alternatives are certainly possible.

<sup>87</sup>[Pierre, 1991a].

<sup>88</sup>[Pierre, 1991a].

<b>Concord:</b>	Elements associated with effects of cohesion
Augmented-concord	Degree of connectivity is increased in translation
Diminished-concord	Degree of connectivity is decreased in translation
Sustained-concord	Same degree of connectivity is maintained in translation
<b>Discord:</b>	Elements associated with too little or too much cohesion
Augmented-discord	Degree of connectivity is decreased or increased too much in translation
Diminished-discord	Degree of connectivity becomes concordant in translation
Sustained-discord	Too little or too much connectivity is maintained in translation
<b>Heteropoise:</b>	Elements associated with an interruption or parenthesis
Augmented-heteropoise	Introduction of a parenthesis in translation
Diminished-heteropoise	Removal of a parenthesis in translation
<b>Sustained-heteropoise:</b>	Elements associated with maintenance of parenthesis in translation
Non-shifted heteropoise	Parenthesis remains in same position
Front-shifted heteropoise	Parenthesis moves to start of sentence
Medial-shifted heteropoise	Parenthesis moves to middle of sentence
End-shifted heteropoise	Parenthesis moves to end of sentence

Table 1: Summary of abstract elements of comparative stylistics

## 5 Applying the grammar in machine translation

Now that we have developed a set of formal rules, which collectively form a grammar of French to English comparative stylistics, the question remains: how can these rules be used in an MT system? We can begin to answer this question by considering how our abstract model of comparative stylistics could be integrated into a general model of machine translation. This augmented model of machine translation is shown in Figure 3.<sup>89</sup> Thus, the original model is now modified as follows:

- The analysis stage includes stylistic analysis. The stylistic analysis may be done in tandem with ordinary parsing, as in the *STYLISTIQUE* system [DiMarco, 1990], or in the final stage of analysis, as in Hoyt's [1993] *ASSET* system.
- The analysis stage requires a theory of the source-language stylistics.
- The intermediate stage indicates a set of optional stylistic structures for the target-language text.
- The intermediate stage requires a theory of the source language-target language comparative stylistics.
- The generation stage requires a theory of the target-language stylistics.
- The generation stage includes a stylistic control mechanism that will choose the appropriate stylistic rules to apply. This control mechanism takes into account the stylistic options produced during the intermediate stage.

Figure 4 gives an example of the stylistic analysis stage in the augmented model of translation.<sup>90</sup> The first section of the example shows the sentence that is analyzed.<sup>91</sup> The second section shows the stylistic goals that are the final result of the stylistic analysis. They are organized along three dimensions: clarity/obscurity, abstraction/concreteness, staticness/dynamism.<sup>92</sup> The next two sections list the abstract elements and primitive elements that describe the characteristics of the syntactic style of the sentence. The last section of the example shows the resulting set of candidate rules of comparative stylistics.

The first three candidate rules suggest ways in which the style of the nominal groups in the sentence could be handled in translation. The conjunct<sup>1</sup> definite articles will very likely be translated directly. But, the conjunct<sup>1</sup> postmodifying prepositional phrase could either be translated directly or could be transformed into a conjunct<sup>1</sup> premodifier. That is, *les resultats par le courrier* could be translated either as *the results by mail* or *the mailed results*. Deciding which option to choose will depend on the exact degree of intrasentence cohesion that should be incorporated in the target-language sentence. The choice will also depend on the exact stylistic goal that is to be achieved: should the original author's intent be replicated or should the intent be modified to be more representative of the inherent character of the target language? It might be that a premodifier, rather than a postmodifier, might be more natural and more cohesive in one language, while the reverse might be true for the other language. However, accounting for such subtle distinctions will require more refinement of the primitive-element classifications.

---

<sup>89</sup>The diagram in Figure 3 is based on the diagram of a generic MT system given in [Hutchins, 1986].

<sup>90</sup>The format of the primitive-element description is a hand-simulation using the same style as *ASSET* [Hoyt, 1993], which currently handles only English input.

<sup>91</sup>[Les ami(e)s de la terre, 1986], minor editing of punctuation.

<sup>92</sup>A default setting of *neutral* is assigned if the sentence cannot be definitely classified as one of the end-points for a dimension.

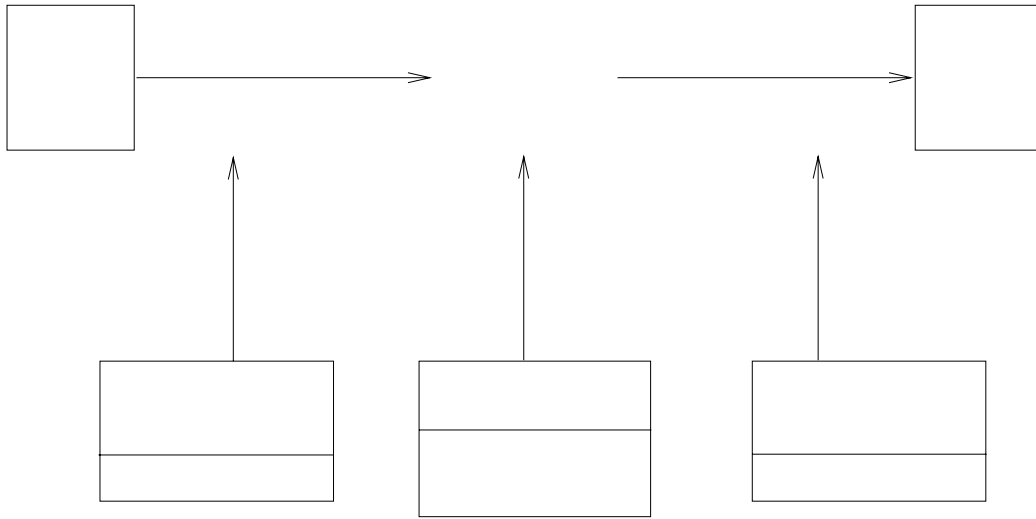


Figure 3: An augmented model of machine translation that incorporates comparative stylistics

The next three candidate rules suggest three different ways in which the dependent clause, *même lorsque les niveaux de contamination étaient inacceptables*, could be positioned in the English translation: either moved to the front of the sentence, inserted in the middle, or kept in the same end position, but set off by punctuation. In all cases, the sentence will be transformed from a centroschematic to a heteropoise so that, according to our rules for stylistic goals given in Section 2.2, the style of the sentence will shift from merely clear to definitely concrete. Therefore, the last rule might be chosen to maintain the centroschematic style of the sentence, so that both French and English sentences will be classified as clear.

As we described in Section 1.1, a key feature of our approach is that our theories of style are goal-directed, *i.e.*, linguistic choices are considered to be made to achieve specific stylistic goals. In the example above, the English sentence could be made heteropoisal and concrete, as in the version below:

- (22) They were informed of the results by mail but, even though the levels of contamination were unacceptable, they were not informed about the health hazards.

Such a concrete style would be in keeping with the stylistic preferences of English [Vinay and Darbelnet, 1958]. However, there is also the option of making the English sentence a *discordant* heteropoise using disruptive punctuation as follows:

- (23) They were informed of the results by mail, but not informed about the health hazards—even though the levels of contamination were unacceptable.

In this case, the decision as to whether the final clause should be made parenthetical and discordant is rhetorical: should a strictly faithful translation be avoided in favour of a strongly emphatic sentence that may be more persuasive to the intended audience?



**Sentence to be analyzed:**

Ils ont reçu les résultats par le courrier mais sans aucune information sur les dangers pour la santé même lorsque les niveaux de contamination étaient inacceptables.

**Stylistic goals of this sentence:**

Clarity, neutral, neutral.

**Abstract-element description:**

Initial concord, medial concord, final concord, centroschematic.

**Primitive-element description:** centroschematic complete, concordant complete, centroschematic major, concordant major, monoschematic noun phrase, concordant noun phrase, conjunct0 noun, centroschematic verb phrase, concordant verb phrase, centroschematic noun phrase, concordant noun phrase, conjunct1 premodification, conjunct1 determiner, conjunct0 noun, concordant postmodification, conjunct1 postmodification, conjunct1 prepositional phrase, subordinating conjunction, concordant prepositional phrase, concordant prepositional phrase, concordant prepositional phrase, concordant clause, conjunct1 adverb, centroschematic noun phrase, concordant noun phrase, conjunct1 premodification, conjunct1 determiner, conjunct0 noun, conjunct1 postmodification, conjunct1 prepositional phrase, monoschematic verb phrase, concordant verb phrase.

**Applicable rules of comparative stylistics:**

conjunct1 premodification  $\mapsto$  conjunct1 premodification (sustained concord)  
conjunct1 postmodification  $\mapsto$  conjunct1 premodification (sustained concord)  
conjunct1 postmodification  $\mapsto$  conjunct1 postmodification (sustained concord)  
centroschematic  $\mapsto$  initial heteropoise (augmented heteropoise)  
centroschematic  $\mapsto$  medial heteropoise (augmented heteropoise)  
centroschematic  $\mapsto$  final heteropoise (augmented heteropoise)  
centroschematic  $\mapsto$  centroschematic

Figure 4: A sample stylistic analysis

In order to make such a decision, we need to understand how stylistic goals influence higher-level *rhetorical goals*—Makuta-Giluk [1991] has taken the first steps towards constructing such a formal model of rhetoric. Makuta-Giluk’s model of rhetoric is a natural extension of our theory of style: rhetorical goals are expressed as combinations of stylistic goals. In turn, as we have described, stylistic goals are expressed in terms of the abstract elements, and the abstract elements in terms of the primitive elements. Ultimately, the theory is founded on the linkage of primitive elements of style to specific syntactic choices. That is, the relationship between syntactic cohesiveness and stylistic effect determines the choice of syntactic structure.

Our application of cohesiveness to a formal model of comparative stylistics is supported by Guillemin-Flescher’s consideration of the strengths of intrasentence syntactic relationships in her description of the rules of comparative stylistics summarized in Section 3.1. These strengths of syntactic relationships correspond to the concept of connectivity that is used in our formal rules. Thus, we can expect that syntactic connectivity, which is the basis of the stylistic grammar, can be used to choose the most appropriate rules from the candidate set.

## 6 Conclusion

### 6.1 Contributions of the work

The aim of this research was to develop a set of formal rules of syntactic French–English comparative stylistics. These rules were developed as the syntactic component of a computational model of comparative stylistics for improving machine translation. We anticipate that this model could be incorporated into an MT system as part of a general theory of style.

The process of encoding the theoretical rules of comparative stylistics found in textbooks involved identifying instances of the theoretical rules in a corpus of French sentences and corresponding English translations. The portions of the French sentences and English translations identified by an instance of a theoretical rule were parsed and represented using DiMarco's [1990] stylistic grammar. The correspondences between the stylistic properties of the French and English sentences, as identified by DiMarco's stylistic grammar, were recorded. These correspondences were then organized into formal rules of comparative stylistics with the alternatives of each rule defining an abstract element of comparative stylistics.

### 6.2 Related work

DiMarco's original theory of style has been refined and expanded by DiMarco and Hirst [1993a] and by others. Green [1992] extended the theory and developed an approach for incorporating stylistic control into the Penman generation system.<sup>93</sup> Hoyt [1993] built a new stylistic analyzer that uses the revised theory of style to produce a description of the stylistic characteristics of the input sentences.

In addition, Ryan [1989] extended the theory to the text level to correlate the thematic structure of paragraphs with their stylistic effects. Payette [1990] and Payette and Hirst [1992] applied a portion of the theory to develop a prototype of a computer-assisted second-language instruction system that could analyze sentences for clarity. And, recently, Makuta-Giluk [1991] and Makuta-Giluk and DiMarco [1993] expanded the theory to account for the additional level of rhetorical goals.

### 6.3 The next step

As mentioned earlier, we expect that a complete model of comparative stylistics—for lexical, syntactic, and semantic aspects of style—will eventually be developed and integrated into an MT system. DiMarco *et al.* [1993] and DiMarco and Hirst [1993b] have taken the first steps towards a theory of lexical style that will eventually complement the syntactic theory. At present, we are developing a representation of lexical knowledge that will allow us to describe the subtleties and nuances that distinguish sets of near-synonyms. This representation could then be used in an MT system in deciding which target-language word most closely matches the sense of the given source-language word. And, of course, the target language might offer no single word corresponding to the exact specifications of the source-language text; or there might be several words to choose from, differing in style, emphasis, shade of meaning, or collocational requirements.

---

<sup>93</sup>Penman is a natural language generation system developed at the Information Sciences Institute, University of Southern California. For technical details of the Penman system, the reader is referred to: Penman documentation, The Penman Natural Language Generation Group, Information Sciences Institute, University of Southern California, 1988.

We expect that the integration of our computational model of comparative stylistics with an MT system will help to improve the quality of the translated output. Our model will provide additional information enabling the MT system to make a more informed decision about the potential modulations to the translated text and their resulting stylistic effects.

## **Acknowledgements**

We would like to express our gratitude to Graeme Hirst for careful and comprehensive reviews of earlier drafts of this paper. We also thank Lisa Theodore for supplying the English literal translations of the French example sentences.

We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada and the Information Technology Research Centre.

## References

- [Beuparlant, 1990a] Michel Beuparlant. L'école québécoise forme des bilingues fonctionnels. *Langue et société*, (33):28–29, hiver 1990.
- [Beuparlant, 1990b] Michel Beuparlant. Quebec schools produce functionally bilingual students. *Language and Society*, (33):28–29, Winter 1990. [translator unknown].
- [Bureau, 1976] Conrad Bureau. *Linguistique fonctionnelle et stylistique objective*. Presses Universitaires de France, 1976.
- [Canadian Council on Social Development, 1990a] Canadian Council on Social Development. Bill C-69 and how it will affect Canadians. *Canada's Social Programs are in Trouble*, pages 3–6, 1990.
- [Canadian Council on Social Development, 1990b] Canadian Council on Social Development. Canada's social programs are in trouble. *Canada's Social Programs are in Trouble*, pages 1–2, 7, 1990.
- [Chatto, 1990a] James Chatto. Eaux-de-vie haut de gamme. *enRoute*, 18(12):29–30, décembre 1990.
- [Chatto, 1990b] James Chatto. Singular spirits. *enRoute*, 18(12):29–30, December 1990.
- [Chipaux, 1990a] Françoise Chipaux. Bsouss, the martyred village. *Manchester Guardian Weekly*, 143(17):13, 28 October 1990.
- [Chipaux, 1990b] Françoise Chipaux. Bsouss, village libanais martyr. *Le Monde*, pages 1, 3, 21/22 octobre 1990.
- [Cluett, 1976] Robert Cluett. *Prose style and critical reading*. Teachers College Press, Columbia University, New York, 1976.
- [Conseil canadien de développement social, 1990] Conseil canadien de développement social. Comment le projet de loi C-69 nous nuira-t-il? *Les programmes sociaux du Canada en péril*, pages 3–6, 1990.
- [Dappen, 1991a] Andy Dappen. À l'attaque de la cataracte. *Canadien*, 4(9):15–16, janvier 1991.
- [Dappen, 1991b] Andy Dappen. Preventing cataracts. *Canadian*, 4(9):15–16, January 1991.
- [DiMarco and Hirst, 1990] Chrysanne DiMarco and Graeme Hirst. Accounting for style in machine translation. In *Proceedings of the Third International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, pages 148–153, Austin, June 1990.
- [DiMarco and Hirst, 1993a] Chrysanne DiMarco and Graeme Hirst. A computational theory of goal-directed style in syntax. *Computational Linguistics*, 19(3):448–497, 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. In *Proceedings of the Ninth Annual Conference of the University of Waterloo Centre for the New Oxford English Dictionary and Text Research*, pages 33–43, Oxford, September 1993.

- [DiMarco *et al.*, 1993] Chrysanne DiMarco, Graeme Hirst, and Manfred Stede. The semantic and stylistic differentiation of synonyms and near-synonyms. In *Proceedings of the AAAI Spring Symposium on Building Lexicons for Machine Translation*, pages 114–121, Stanford, March 1993.
- [DiMarco, 1990] Chrysanne DiMarco. *Computational stylistics for natural language translation*. PhD thesis, Department of Computer Science, University of Toronto, 1990. [Also available as Technical Report CSRI-239].
- [Duhaime, 1990a] Colette Duhaime. The linguistic ombudsman in action. *Language and Society*, (32):23–24, Fall 1990.
- [Duhaime, 1990b] Colette Duhaime. L’ombudsman linguistique à l’œuvre. *Langue et société*, (32):23–24, automne 1990.
- [Dupeux, 1988a] G. Dupeux. French society. In Newmark [1988], pages 272–273. [translation by Peter Wait].
- [Dupeux, 1988b] G. Dupeux. La société française. In Newmark [1988], pages 272–273.
- [Friends of the Earth, 1986] Friends of the Earth. Groundwater pollution in New Brunswick. *in-foetox*, 3(2):15–16, June 1986.
- [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 Department of Computer Science Technical Report CS-92-48].
- [Guillemin-Flescher, 1981] Jacqueline Guillemin-Flescher. *Syntaxe comparée du français et de l’anglais*. Ophrys, Paris, 1981.
- [Halliday and Hasan, 1976] M.A.K. Halliday and Ruqaiya Hasan. *Cohesion in English*. Longman Group Limited, London, 1976.
- [Halliday, 1985] M.A.K Halliday. *Introduction to functional grammar*. Edward Arnold, 1985.
- [Hoyt, 1993] Patricia A. Hoyt. A goal-directed functionally-based stylistic analyzer. Master’s thesis, Department of Computer Science, University of Waterloo, 1993. [Published as University of Waterloo Department of Computer Science Technical Report CS-93-48].
- [Hutchins, 1986] W.J. Hutchins. *Machine translation: Past, present, future*. Ellis Horwood Limited, Chichester, 1986.
- [Kane, 1983] Thomas S. Kane. *The Oxford guide to writing: A rhetoric and handbook for college students*. Oxford University Press, Oxford, 1983.
- [Kerr, 1990a] Robert Kerr. Copyright: Masse declares war. *CAUT Bulletin*, 37(10):21, December 1990.
- [Kerr, 1990b] Robert Kerr. Le droit d’auteur: Masse déclare la guerre. *CAUT Bulletin*, 37(10):22, décembre 1990.
- [Laurian, 1986] Anne-Marie Laurian. Stylistics and computing: Machine translation as a tool for a new approach to stylistics. *Computers and Translation*, 1:215–223, 1986.

- [Les ami(e)s de la terre, 1986] Les ami(e)s de la terre. Pollution des eaux souterraines au N.B. *infoetox*, 3(2):16, juin 1986.
- [Loffler-Laurian, 1983] Anne-Marie Loffler-Laurian. Typologie des discours scientifiques: Deux approches. *Études de linguistique appliquée*, 51:8–20, juillet–septembre 1983.
- [Loffler-Laurian, 1987] Anne-Marie Loffler-Laurian. Enquête stylistique sur la traduction automatique de documents techniques. *International Review of Applied Linguistics in Language Teaching*, 25(2):154–168, May 1987.
- [MacDonald, 1990a] Cynthia MacDonald. The big picture. *enRoute*, 18(12):38–50, December 1990.
- [MacDonald, 1990b] Cynthia MacDonald. L'île aux mille mystères. *enRoute*, 18(12):38–52, décembre 1990.
- [Mah, 1991] Keith Mah. A computational formalism for French–English comparative stylistics. Master's thesis, Department of Computer Science, University of Waterloo, 1991. [Also available as University of Waterloo Department of Computer Science Technical Report CS-91-67].
- [Makuta-Giluk and DiMarco, 1993] Marzena Makuta-Giluk and Chrysanne DiMarco. A computational formalism for syntactic aspects of rhetoric. In *Proceedings of the First Conference of the Pacific Association for Computational Linguistics*, pages 63–72, Vancouver, April 1993.
- [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. [Also available as University of Waterloo Department of Computer Science Technical Report CS-91-54].
- [Milic, 1982] Louis T. Milic. The annals of computing: Stylistics. *Computers and the Humanities*, 16:19–24, 1982.
- [Newmark, 1988] Peter Newmark. *A textbook of translation*. Prentice Hall International (UK) Ltd, Hertfordshire, 1988.
- [Payette and Hirst, 1992] Julie Payette and Graeme Hirst. An intelligent computer-assistant for stylistic instruction. *Computing and the Humanities*, 26(2):87–102, 1992.
- [Payette, 1990] Julie Payette. Intelligent computer-assisted instruction in syntactic style. Master's thesis, Department of Electrical Engineering, University of Toronto, 1990. [Also available as Technical Report CSRI-247].
- [Pierre, 1991a] Gloria Pierre. Encore à propos d'enseignement et de recherche. *Affaires universitaires*, 32(2):2, février 1991.
- [Pierre, 1991b] Gloria Pierre. What about teaching? and publishing? *University Affairs*, 32(1):2, January 1991.
- [Reynolds, 1990a] Paul Reynolds. Le prix à payer. *Le consommateur canadien*, (2):4, 1990.
- [Reynolds, 1990b] Paul Reynolds. Paying the price. *Canadian Consumer*, (2):4, 1990.

- [Ryan, 1989] Mark B.P. Ryan. The computational codification of the semantic aspects of style. Master's thesis, Department of Computer Science, University of Toronto, 1989. [Also available as Technical Report CSRI-231].
- [Tsutsumi, 1990] Taijiro Tsutsumi. Wide-range restructuring of intermediate representations in machine translation. *Computational Linguistics*, 16(2):71–78, June 1990.
- [Unknown, 1990a] Unknown. Derrière nous—devant nous. *CAUT Bulletin*, 37(10):2, 6, décembre 1990.
- [Unknown, 1990b] Unknown. Looking back—Looking ahead. *CAUT Bulletin*, 37(10):2, December 1990.
- [Unknown, 1991a] Unknown. L'avenir du pays. *Canadien*, 4(9):67–75, janvier 1991.
- [Unknown, 1991b] Unknown. Looking ahead. *Canadian*, 4(9):67–76, January 1991.
- [Vinay and Darbelnet, 1958] J.P. Vinay and J. Darbelnet. *Stylistique comparée du français et de l'anglais*. Beauchemin Itée, Saint-Laurent, Québec, 1958.