Comparative Stylistics in an Integrated Machine Translation System

by

Keith Mah

A thesis
presented to the University of Waterloo
in fulfilment of the
thesis requirement for the degree of
Master of Mathematics
in
Computer Science

Waterloo, Ontario, Canada, 1991

Abstract

Comparative stylistics is a subfield of stylistics that attempts to account for the differences in style between languages. Rules of comparative stylistics are commonly presented, in textbooks of translation, as "rules-of-thumb", but if we hope to incorporate a knowledge of comparative stylistics into natural language understanding systems, we must take a more formal approach. In particular, we will develop a computational model of comparative stylistics for machine translation that could be used to guide translation and thereby improve the quality of the translated output. An implementation of this model would provide additional information to the machine translation system about the potential modulations to the translated text and their effects, enabling it to make a more informed decision.

In this thesis, we develop a set of formal rules of syntactic French–English comparative stylistics to be used as a component of a model of comparative stylistics. As the foundation for the formal rules, we adapt theoretical rules of syntactic French–English comparative stylistics compiled by Guillemin-Flescher [1981] and the formal representation of syntactic style developed by DiMarco [1990]. A corpus of French sentences and their English translations is analyzed to convert the theoretical rules to a set of formal rules that builds on DiMarco's grammars of syntactic style. Thus, we present a formal grammar of French–English comparative stylistics. We also suggest a method for incorporating these rules into an existing machine translation system.

Acknowledgements

Writing these acknowledgements is much like accepting an Oscar. There are so many people who provided assistance and support who should be acknowledged and so little space in which to do it. To the many people I do not mention below: Thanks!

- First, I would like to thank my supervisor, Chrysanne DiMarco, for suggesting the topic for this thesis, and her guidance and patience over the last year and a half.

- Next, my two second readers: Pierre H. Dubé and Sergei Nirenburg for reading my thesis on short notice and their inspiring ideas and suggestions during the research stage.

- The past and present members of the Logic Programming and Artificial Intelligence Group (LPAIG) including Robin Cohen, Cameron Shelley, Jennifer Chu, Marzena Makuta-Giluk, Nadia BenHassine, Pat Hoyt, Scott Goodwin, Steve Green and Steve Woods for many stimulating conversations and interesting ideas.

- John Sellens for putting up with my queries and requests regarding the LPAIG equipment and software.

- Sue Thompson who cut through the red tape and with the help of Linda Norton finally convinced the University that I really did exist!

- Carl Kaufmann for introducing me to log-linear models and explaining them in moderate detail.

- Hamish MacDonald, UNIX "Xpert" and GNU Emacs guru, for taking the time to figure out and fix my mistakes, and pointing out better ways to do things.

- Dexter Bradshaw, my officemate, for many discussions on random topics.

- Wendy Rush for advice about the world at large, listening, periodically feeding me—basically being a "mom away from mom".

- Bart Domzy my roommate for two long years and my guide to the intricate workings of the Department of Computer Science, thesis writing and the complex maze known as Kitchener-Waterloo.

- Ali Hammoud, Binnie Langley, Bob Mercer, Dennis Fox, Floyd Harriott, Jacqueline van Dyk, Lawrence Monck, Margaret Ho, Mike Bauer, and Simon Coles for encouraging me to pursue graduate studies at Waterloo, and periodically "checking up on me" along the way.

- Anette Happel who also encouraged me to pursue graduate studies at Waterloo, periodically "checked up on me," and who I hold responsible for renewing my interest in sports.

- And finally, I would like to recognize the support and encouragement of my parents and my sister, Mae Jean, over the past six years.

Contents

List of Figures

Chapter 1

Introduction

Style! I have no style, I merely wait til the mud settles.
—Goldwin Smith, attributed

1.1   Why Style is Important

Traditionally, linguists and computational linguists have studied syntax and semantics in their attempts to formulate models for understanding natural language utterances. Some researchers have extended this trend of studying overt aspects of language to include pragmatic issues.

Recently, a handful of researchers in natural language understanding have recognized the importance of *style*, a subtle but significant aspect of language. These researchers have looked beyond the common view that style is merely a non-functional veneer. Instead, they see style as an essential part of the meaning that an author (or speaker) attempts to convey.

Style is influenced by linguistic choices on all levels: lexical, syntactic, and semantic. For example, consider the differences in meaning between the following two sentences.

(1-1)   She is overweight.

(1-2)   She is full-figured.

Both sentences have identical meanings on the surface, but in the first sentence the lexical choice *overweight* conveys the additional information that the author is direct and possibly being impersonal. In the second sentence, the lexical choice *full-figured*, as it is used here, is indirect and suggests that the author is attempting to be polite. Differences in style can also be more subtle, as shown in the following two sentence fragments.

(1-3)   the black cat

(1-4)   the cat, which is black

The first fragment is a common method of expressing the fact that the cat is black. The second fragment is less common. The postposing of the fact that the cat is black in a relative clause places more emphasis on the cat than its blackness.

When a text is written, it is written for a particular audience. The author carefully chooses the lexical items, syntactic constructions, and semantic content that is appropriate for an effective communication. When writing a report for the CEO of a company, the author will choose a formal style. But, when writing a scientific article for a general audience, the author will choose a simple, clear style. These choices are often made unconsciously, but they are made nonetheless.

The value of these choices of style is very apparent in the field of natural language generation. One of the questions currently being studied in the field of generation is determining *what to say.* But even after that problem has been solved an equally difficult problem remains: *how to say it.* That is, once it has been decided that something should be said (*e.g.,* the proposition that the cat is black), what form should the English representation of the proposition take (*e.g., the black cat*; *the cat, which is black*; or, *the cat is black*)? It is clear that the form of the utterance should not be randomly, but carefully, chosen after considering the stylistic effects that are desired. Hovy [1988] has examined some possible pragmatic effects (*e.g.,* emotional and interpersonal relationships) and how they can be produced by a natural language generation system. His work is described later in section 3.5.1.

We use style every day, whether consciously or unconsciously, to help us choose the most appropriate words for what we write or say and the most appropriate way to use these words to correctly and effectively convey the meanings and effects that we intend. We ignore the importance of style at our peril.

1.2   Why Style is Important for Machine Translation

In the preceding section we gave a brief account of the importance of style for understanding and generating language. In this section we will see the value of style for translation.

Many of us who have purchased foreign-made products may have noticed peculiar phraseology in the owner's manuals. These peculiarities are a result of differences in the way concepts are expressed in the original language and English. The peculiarities that make the text seem unnatural exist because the translator used a literal translation and did not allow for the differences of style (*i.e.,* choices of lexical items, syntactic constructions, and semantic content) between the original language and English.

Although experienced translators will not usually make these stylistic errors, a machine translation (MT) system will. One of the problems with MT is that the translated output, while usually understandable, is often plagued with unnatural-sounding phrases that make the translation difficult to read. A human post-editor must then take on the tedious task of correcting these errors. With the increasing demand for, and availability of, commercial MT products, it is desirable for MT system to automatically correct these stylistic differences between languages and thereby reduce the post-editor's workload.

There are many examples of stylistic differences between languages. For someone who stays in (Canadian) hotels while traveling, a common example would be the phrase *do not disturb.* A common literal translation of this phrase into French is *ne pas déranger,* but the correct translation is *prière de ne pas déranger* (*lit.,* I pray you not to disturb me). This difference in

translation does not impede the understanding of the sentence whether translated correctly or not. The blunt, direct nature of the literal translation of the English sentence would only seem unnatural relative to the more circuitous, indirect sentence that a Francophone would expect to see.

Occasionally, not accounting for stylistic differences can cause confusion. There are stylistic differences between American English (*i.e.,* that used in the United States) and Canadian English. For example the American term used for the indigenous people of the Arctic is *Eskimo*, whereas the Canadian term is *Inuit* (or *Innuit*). Confusion could result if an American tourist came to Canada and asked to see *the Eskimos*. In Canada, the phrase *the Eskimos* is often used to refer to the (North American) football team based in Edmonton, Alberta—which may not be the type of Eskimo the tourist had intended to see.

These examples show that an understanding of style is important for translation, and therefore machine translation. This awareness of style allows the translator to produce translations that are natural and that properly convey the author's intent—and a *machine* translation system should have the same ability.

1.3   A Computational Approach to Comparative Stylistics

Now that we have established that an awareness of style and how it changes between languages is important for MT, we can examine what is needed in order to provide a knowledge of style, particularly differences in style, in a MT system.

Comparative stylistics is a subfield of stylistics that attempts to describe the differences of style between languages. One of the rules that we might find in a textbook of French–English comparative style, such as [Vinay and Darbelnet 1958] or [Guillemin-Flescher 1981], is "French words tend to be more abstract than the corresponding English words." This rule cannot be directly used by existing MT systems since it is difficult to explicitly define terms like *tend* and *abstract*.

The first step towards the development of a computational approach of comparative stylistics is to develop a method for recognizing the style of a sentence and representing this stylistic information in a form that is amenable to computational applications such as MT. In this thesis, we will use DiMarco's [1990] formal stylistic grammars. Her grammars are described in section 3.5.2.

The next step is to encode the rules of comparative stylistics that are found in textbooks in a manner that is useful for MT. This is the aim of this thesis. The process of encoding the theoretical rules of comparative stylistics found in a textbook involved identifying instances of the theoretical rules of translation 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 stylistic grammars. The correspondences between the stylistic properties of the French and English sentences, as identified by DiMarco's stylistic grammars, were recorded and later organized into formal rules of comparative stylistics that we propose as a useful addition to MT systems.

Before proceeding to fulfill this primary aim, the development of a formal representation for comparative stylistics, we will take a detailed look at the fields of machine translation and stylistics in the next two chapters of this thesis. Then, in chapter 4, we will look at a description of the theoretical rules of French–English comparative stylistics and the corpus that was used to derive the corresponding formal rules. Next, in chapter 5, the process of deriving the formal rules will be described in detail, the formal rules will be presented and explained, and the chapter will end with a description of how the formal rules could be incorporated into a MT system. Finally, this thesis will conclude in chapter 6 with a summary of contributions and a description of future work.

Chapter 2

Overview of Machine Translation

*A translation is no translation, he said, unless it will give you
the music of a poem along with the words of it.*
—John Millington Synge, The Aran Islands

2.1   A Brief History

Automatic translation of language has been a dream throughout history. A new era of automatic translation attempts began in 1933 when patents for translating machines were issued [Hutchins 1986]. With the application of computers to the task of translation and generous funding, excitement over the prospect of fully automatic high quality translation (FAHQT) grew through the 1950's.

Even in a period of optimism, machine translation (MT) was not without its critics. In the late 1950's, the critics started to become more vocal. In 1966, the Automatic Language Processing Advisory Committee (ALPAC) released a report criticizing MT. The first of three major criticisms was that there was no real demand for MT, as the demand for translations could be met by humans. The second criticism was that all current systems required post-editing, a process where a human corrects the translated output. The third criticism involved the high error rate of existing MT programs. The report destroyed confidence in the potential of FAHQT and the rich sources of funding evaporated.

Some MT researchers turned to the development of *machine-aided translation* (MAT). Here a human translator *interacts* with the machine during the translation process. In the *human-aided machine translation* variant of MAT, the computer stops and requests assistance from the translator when it is unable to decide among several possibilities. The role of the human is similar to that of a *pre-editor* who assists during, rather than before, the actual translation process. For the complementary case, *machine-aided human translation*, the machine acts as a translator's workstation. The human does the translation with the machine assisting in the processing. The machine provides support for text manipulation or, in effect, word processing and for translation-dictionary lookup. The dictionary-lookup facility is particularly effective when technical terminology that is not familiar to the human is used.

Other researchers continued to work on fully automatic MT, but by developing theoretically sound methods for MT. They developed *indirect* and *artificial intelligence-based* systems that

have attempted to integrate knowledge from MT, linguistics, and computational linguistics.

All these systems have a three-stage translation process. The *analysis* stage breaks down the source text to determine and represent its morphological and syntactic structure. This analysis identifies the part of speech of each component word. This information is used to resolve some of the ambiguities inherent in natural language[1]. The part of speech analysis is also used to help determine relations between words[2]. The *transfer* stage manipulates the representation of syntactic structure for each sentence produced by the analysis stage to produce a representation of the corresponding syntactic structure for the translated text. The translation of words also occurs at this stage. The *generation* stage produces, from the representation produced by the transfer stage, the target-language text. The theoretical bases for MT and the structure that has been added to MT systems since the ALPAC report (*i.e.,* the development of modular components and other software engineering techniques) has reduced the error rate in the generated text.

Of the two remaining points of contention raised by the ALPAC report, the necessity of post-editing remains a problem. Since the report was issued, the demand for translation of documents has escalated beyond the capacity of human translators. This demand has resulted from both an explosion of knowledge and from the "global village" attitude regarding information sharing that has arisen during the evolution of today's society. This vast increase in the amount of information and knowledge has increased the amount of material that researchers must read to keep up in their respective fields. As well, the material may be written in one of any number of languages. Similarly, increased trade between nations (*e.g.,* the creation of the European Economic Community) has augmented the demand for translation of manuals and instructions.

After the ALPAC report, progress in MT research and development slowed, but did not stop. TAUM-METEO was developed during the mid-1970's. It is described in [Slocum 1988, 19] as "the world's only example of a fully-automatic MT system." With an accuracy rate of 90–95%, METEO translates Canadian weather reports from English to French at a rate of over 24 000 words per day [Slocum 1988] (see also [Hutchins 1986]). With this success and an increased demand for translation, interest in MT resurfaced once again in the 1980's.

In the next section, I will present brief descriptions of three major strategies (paradigms) that have been used in MT systems. I will then describe an actual MT system that has been developed for research purposes.

2.2   Strategies

This section will describe the three primary strategies employed for constructing MT systems. These strategies are: *direct*, *indirect*, and *artificial intelligence-based*.

---

[1]In the sentence *The boy picked up the bat*, the word *bat* could refer to either an instrument used to strike a ball in a baseball game or a small animal capable of flying.

[2]In the sentence *Time flies like an arrow*, the word *flies* could be either a noun or a verb. The semantic interpretation (meaning) of the sentence changes depending on the part of speech selected. If *flies* is a noun, then there is a type of fly that likes a particular arrow. Otherwise, if *flies* is a verb, then time passes by like an arrow (quickly?).
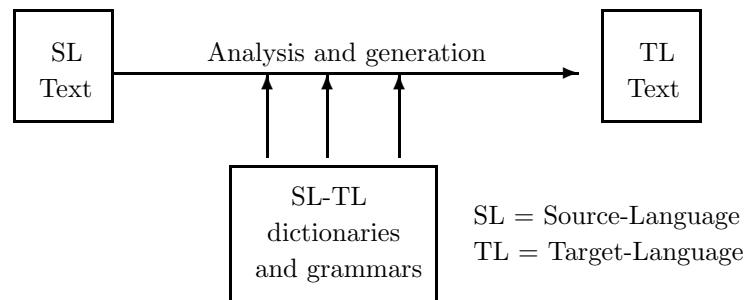
Figure 2.1: The direct approach.

2.2.1 Direct Translation

Many of the early MT efforts took a *direct* approach to the MT problem. Although the direct approach has survived, primarily in the SYSTRAN system, this strategy was most prominent during the 1950's and 1960's [Hutchins 1986]. In its crudest form, "direct translation" is little more than a dictionary-lookup scheme. As shown in figure 2.1, the translation process proceeds directly from analysis of the source-language (SL) sentence to generation of an equivalent, in some sense, sentence in the target-language (TL). During translation, a dictionary supplies the target-language equivalents of words, and a grammar may be employed to sequence the words appropriately.

These word-for-word translations were ridiculed. Some of the resulting mistranslations have become enshrined in MT folklore. Two of the most popular mistranslations resulted from an attempt to translate the phrases "Out of sight, out of mind" and "The spirit is willing, but the flesh is weak" from English to Russian and then back to English again [Hutchins 1986]. There are differing accounts, but all relate incorrect, but humorous, phrases. Two accounts report the translations as *Invisible insanity*, and *The whiskey is all right but the meat has gone bad*.

To understand why such absurd translations were produced, we can examine how the process of translating the English word "alcohol" might be translated into "fire water" in some native North American (NNA) language, then back to English again. The translation system would look in its English–NNA dictionary to find the correct translation of *alcohol* as *fire water*. When the reverse translation is made, the system finds two words to translate and translates each separately, not realizing the intended meaning of the compound noun. This unfortunate method of translation would occur when the phrase to be translated had not been anticipated by the system's maintainers and, so, the existing dictionaries and rules do not provide an adequate "guessing" ability in the translation process. As the direct approach matured, better translation dictionaries were developed, and grammars for the source and target languages were introduced to produce more acceptable translations.

A major failing of the direct approach is the literalness of the word-for-word translations. While a translation that is faithful to the original text is desirable, a completely literal translation may not be appropriate. In some cases, a more appropriate way of expressing the sentence, through word choice and style, in the target language may exist. Another flaw is the high cost of maintaining the dictionaries and rules, which is due, in part, to the lack of modularity within

the system. Adding an entry or rule may cause unwanted side effects that result from applying the rule at an inappropriate time or preventing the application of correct rules.

These problems are, however, a result of the empirical development of direct translation systems. The development process was one of trial and error where the system would be patched to correct problems in the output. Many iterations through the patching process destroyed any structure that may once have existed and obscured the effects of future modifications. A more principled approach to the development of MT systems is provided by indirect translation systems.

2.2.2   Indirect Translation

The second generation of MT systems is characterized by the use of *indirect translations*. Indirect translation methods introduce one or more intermediate representations into the translation process. Indirect translation first became popular in the late 1960's, after the ALPAC report, and continues to be commonplace today. Employing an intermediate representation adds a degree of modularity to the translation system. This modularity allows the analysis stage to deal only with the conversion of the source-language text to its representation and for the generation stage to convert from (a possibly modified version of) the same representation to its corresponding target-language text. Any translations that must be made between the source-language and target-language representations occur in an intermediate transfer stage.

The separation of the three stages contrasts with the direct translation approach where the three stages are combined. Modularity provides many benefits, such as the ability to modify or replace components of a system without affecting the other components. It can also significantly reduce the effort required to translate between multiple (more than three) languages. For instance, the European Economic Community (EEC) has commissioned the EUROTRA project to develop MT systems to produce translations between any pair of the seven official languages of the EEC's member nations. If direct translation methods were used, 42 ($n(n-1)$) complete systems would need to be developed. Using an indirect method, one analysis and one generation module would have to be developed for each language, and, if necessary, 42 transfer modules. Ignoring the development of the intermediate representation(s), this is the equivalent to developing 7 ($2n/2$) complete systems if transfer modules are not needed, or 19 ($n(n+1)/3$) complete systems if they are. In other words, an indirect method should require less than half the development effort of a direct method.

Use of intermediate representations allowed researchers to claim that their systems dealt with language semantics, in the sense of relations between the words of a sentence. This leads to the question: "How much of the source-language text should be represented?" There are two basic answers to this: a sentence and the text. (The text is effectively multiple sentences.) The former had the advantage of being relatively easy to translate, as most, if not all, work in linguistics and computational linguistics at the time dealt with words and sentences. However, sentence-by-sentence translation cannot resolve all the ambiguities that may occur in a sentence. For instance, in the sentence, *The boy picked up the bat*, the meaning of the phrase, *the bat*, is ambiguous, as it may be translated into French as *la chauve-souris* (*lit.,* bat—the flying animal), or as *la batte (de base-ball)* (*lit.,* (baseball) bat). Working with
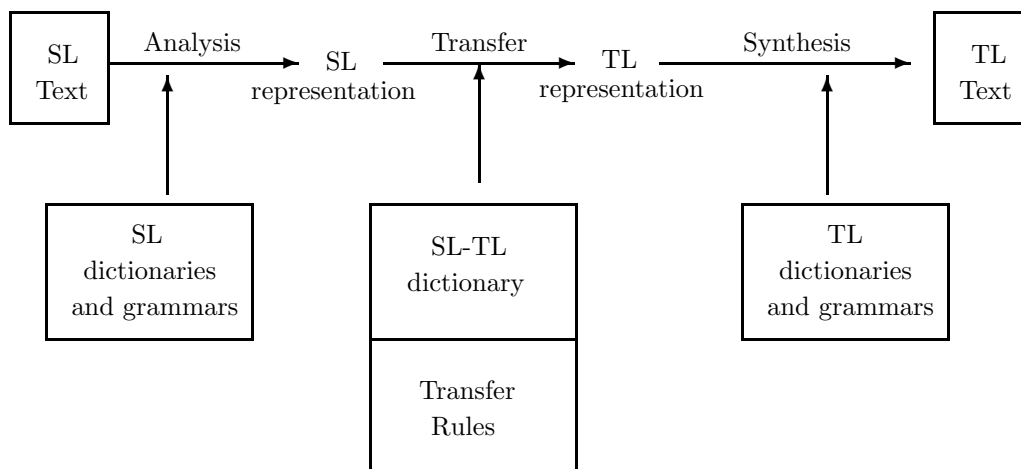
Figure 2.2: The transfer approach.

the text as a whole is more difficult, but this would resolve most of the ambiguities, thus resulting in a more accurate translation.

There are two types of indirect translation. The first is the *transfer* approach, and the second is the *interlingual* approach.

The Transfer Approach

The transfer approach to indirect translation (figure 2.2) uses two intermediate representations and a transfer stage. The translation process first analyzes the source text to produce its representation. This representation is used by the transfer step to produce an equivalent representation for the target-language. Finally, the target representation is used to generate the translated version of the source text.

For most existing systems, the source and target representations are in the form of *structure trees*. These structure trees are initially parse trees for the sentences they represent. Additional information is added to resolve any remaining ambiguities and anaphora. Only those ambiguities that can occur within the language being represented are resolved within the representation. A simple example of one of these types of ambiguity is the polysemy of *bat* shown in the previous example.

The relationship between parse trees and structure trees is similar to the relationship between a transformational grammar's surface structures and its deep structures. The deep-structure form of a sentence represents the basic meaning or the "generic" form of the sentence. The surface-structure form of the sentence represents the realization of the sentence. Similarly, the structure tree represents the intended meaning of what was said and the parse tree represents what was actually said.

The transfer stage is used to resolve ambiguities that exist between the languages being translated. This includes those concepts (words) that are distinguished in one language and not the other. For example, French has two distinct words, *connaître* and *savoir*, for the English word *know*. When translating a sentence containing the word *know* from English to French,

Figure 2.3: The interlingua approach.

the representation for the English sentence does not distinguish between the two French words. It is in the transfer stage that the distinction is made when creating the French representation.

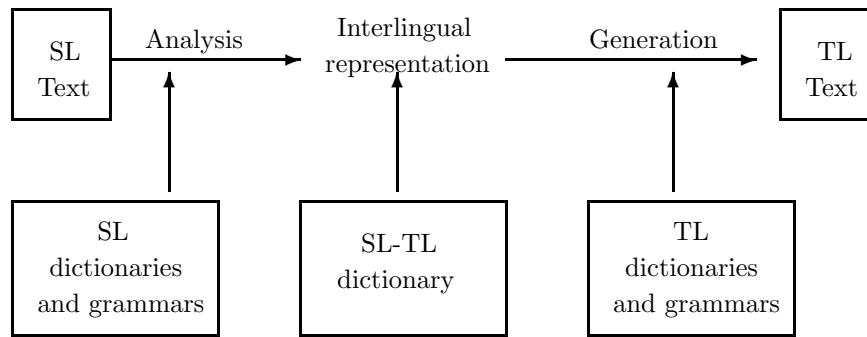To place the approaches in context, the transfer approach sits between the direct and the interlingual (described below). The degree of similarity between these approaches is governed by the complexity of the intermediate representation. For example, direct translation can be simulated by eliminating the representation, causing the transfer stage to become more complex. In this case, the transfer stage must do the work of the analysis and generation stages as well. Similarly, the interlingual approach can be simulated by eliminating the transfer stage and using a single representation for the source and target representations.

The Interlingual Approach

An alternative to the transfer approach is the interlingual approach (figure 2.3). The interlingual approach employs a single representation for all the languages to be translated. Using this strategy, the translation process analyses the source text to produce an intermediate representation of the text. The generation stage can then directly manipulate this representation to produce the equivalent target text. The lack of a transfer stage requires that work that would otherwise be completed in the transfer stage be shifted into the analysis and generation stages.

The use of a single representation is similar to the concept of *linguistic universals* often attributed to Noam Chomsky [Chomsky 1981]. Chomsky proposes that all humans are born with the same innate language capacity, and that each individual's environment "conditions" him for the language(s) that he speaks. If a universal language can be found, then it can be used as an intermediate representation. As this language would be the basis for all other languages, any language can be mapped to or derived from it. To have the capability to represent polysemy, this language would contain the union of all languages.

In practice, the representation that is used is simpler than a "universal language." There is no need to represent the aspects that are common to the languages being translated as they will not need to be distinguished at a later point. Thus, there is a stronger link between the source-and target-languages than with a transfer approach. Adding a new language to a transfer-based MT system does not affect the existing components of the system. A new analysis module, generation module, and set of transfer modules could be added without affecting the existing

system. However, the case can be different for an interlingua-based system as the interlingua may have to be modified to account for distinctions between the new language and any of the languages already being translated. Not only will the interlingua be affected, but also the analysis and generation modules, as they must now allow for the newly distinguished aspects.

The interlingual approach to MT has had a history of failures. After the ALPAC report, research strategies went from one extreme, the direct approach, to the other, the interlingual[3]. At the time, however, not enough research into semantics had been completed to properly support the interlingual approach. The use of *metalanguages* as interlingua has reappeared in recent years. Researchers have attempted to use two types of interlingua, *natural language* and *metalanguage.*

The Distributed Language Translation (DLT) system [Hutchins 1986; Neijt 1986] is a recent attempt to implement a MT system based on a natural language interlingua. The natural languages that are used as interlingua are normally "artificial." Artificial languages are constructed to avoid some of the problems that are encountered in the translation process. There are at least three of these "natural languages" currently in use for human communications: Esperanto, Loglan, and Lojban. A completed prototype of DLT, using Esperanto as its interlingua, was expected in 1990.

In contrast to natural language interlinguas, a metalanguage interlingua can be any representation of the source text. A metalanguage avoids the morphological and syntactic restrictions that are inherent in any natural language. It removes the need for maintaining the representations of sentences as a "well-formed" series of words. In this approach, a tree is a common structure for representing a sentence. Such a tree will have an appearance similar to that of a parse tree, but, unlike a parse tree, would not be required to represent a syntactically well-formed construction in any natural language.

The "ideal" metalanguage would be a representation of the (correct) semantic interpretation of the text. This representation would describe the content of the text without introducing the constraints inherent in the structure of a natural language. However, this removal of structure introduces an important problem in the generation stage, that of knowing what to say and how to say it. This problem is created because the structure and organization of the original text is no longer available at the generation stage. Given the semantic content of the text, the generation stage must find a starting point for producing the target language equivalent, and decide how to structure the presentation. The following translation (produced by SAM [Carbonell *et al.* 1981]) is an example of what might be produced:

> Original text:
> Friday evening a car swerved off Route 69. The vehicle struck a tree. The passenger, a New Jersey man, was killed. David Hall, 27, was pronounced dead at the scene by Dr. Dana Blauchard, medical examiner. Frank Miller, 32, of 593 Foxon Rd, the driver, was taken to Milford Hospital by Flanagan Ambulance. He was treated and released. No charges were made. Patrolman Robert Onofrio investigated the accident.

---

[3]They are at opposite ends of the spectrum since the direct approach translates without using an intermediate representation whereas the interlingual approach places its emphasis on the representation.

Translation (English to English):
An automobile hit a tree near highway 69 four days ago. David Hall, age 27, residence in New Jersey, the passenger, died. Frank Miller, age 32, residence at 593 Foxon Road in New Haven, Connecticut, the driver, was slightly injured. The police department did not file charges.

This "translation" is actually a paraphrase of the original text. For instance, in the translated text, some details, like the name of the investigating officer, have been omitted, and other inferred information, like "the driver was slightly injured", added. While this may be acceptable in some applications, it is inappropriate in translations of technical journals or in legal materials, where faithfulness to the structure and to the details of the original is desired.

2.2.3 Artificial Intelligence-Based Translation

Artificial intelligence (AI)-based approaches to MT are recent attempts to use metalanguage interlinguas as intermediate representations. These attempts are characterized by their use of both semantics and "world knowledge" as part of the translation process. The earliest of these attempts was that of Wilks [1973] in the early 1970's. The remainder did not appear until a decade later.

In the past decade, many AI-based MT systems have been developed. These include Lytinen's MOPTRANS ([Lytinen and Schank 1982; Lytinen 1984]), which used conceptual dependency and Memory Organization Packets, and Rosetta ([Leermakers and Rous 1986; Landsbergen 1987]), which used a modified Montague grammar.

With the exception of the newer systems, the indirect translation systems do not attempt to understand the sentences that are being translated. In these systems, the meaning of a sentence is considered secondary. In the newer systems, the semantics of a sentence is an integral part of the translation, as it can be used to reduce the number of errors produced during the translation of the text. For instance, when there are several possible translations for a word, semantics can be used to assist in selecting the most appropriate one. In AI-based MT systems, theories of semantics (like semantic decomposition, conceptual dependency, and Montague grammar) are borrowed from computational linguistics to represent semantics.

General world knowledge is sometimes necessary in translation, as in the resolution of pragmatics problems. For example, in the sentence *The soldiers fired at the women and I saw several fall*, it is ambiguous whether *several* refers to *soldiers* or *women*. If we assume that the soliders are male, this ambiguity becomes important when the target-language has gender markers. In order for the system to determine the correct form for the translation of the word *several*, it would be necessary for the system to determine the antecedent of *several*. Native speakers of English do not have difficulty resolving this ambiguity. They can infer that it is likely the women who fell, by relying on general knowledge that "If animate things have an object projected at them, they may well drop downwards."

2.3   A State of the Art System

A trend in MT research is to adopt theoretical bases from linguistics, computational linguistics, and AI. This trend is apparent from the introduction of representations of syntax, semantics, and world knowledge in the indirect and AI-based strategies for MT. One system that recognizes this is KBMT-89 (Knowledge Based Machine Translation) developed at Carnegie Mellon University in 1989.

KBMT-89 is an AI-based interlingual system that translates computer manuals from English (E) to Japanese (J) and from Japanese to English. The architecture of KBMT-89 is shown in figure 2.4. It has five major modules: the analyzer, the generator, the augmentor, the concept lexicon, and the development support tools.

The analyzer is similar to that of an interlingual system. It contains a parser that parses the source-language (SL) text with the help of the analysis grammars and lexicons. The parser uses a lexical functional grammar [Kaplan and Bresnan 1982], thus it produces F-structures (F-s). The mapping rule interpreter uses a semantic interpreter, and structural and feature mapping rules (MRs) are used to perform the transformations necessary to produce the interlingua texts (ILTs).

The generator is also similar to that of an interlingual system. Using the interlingual texts and structural mapping rules, it produces a F-structure representing the sentence. The translated sentence is produced through a process of syntactic and lexical selection.

The augmentor is unique to KBMT-89. Since natural language contains many inherent ambiguities, the analyzer is not always able to disambiguate a sentence based on semantic, syntactic, and lexical constraints. In these cases, KBMT-89's analyzer retains all of the potential interpretations for processing by the augmentor. The augmentor has two phases: automatic and interactive. The automatic augmentation phase attempts to complete the disambiguation process by applying additional partial linguistic theories. For example, a strategy for anaphora resolution (resolving pronoun antecedent references) based on linguistic theory has been implemented in KBMT-89. Since KBMT-89 was intended to be a research vehicle, the automatic augmentation phase was designed to facilitate experimentation with partial linguistic theories, including those for anaphora resolution and style. If the automatic augmentation phase is unable to completely disambiguate the sentence, the interactive augmentation phase calls on the user for assistance.

As mentioned in the preceding section, AI-based MT systems incorporate and use knowledge about the world. In KBMT-89, this world knowledge is encoded as a knowledge base in the concept lexicon. This knowledge is used by each of the modules described above (the analyzer, the generator, and the augmentor). Since the world is very complex, it is not possible to construct and maintain a knowledge base containing knowledge about the world in general, at least not with current technologies. Thus, KBMT-89 is only provided with knowledge about the small part of the world that is relevant to its domain of translation. The remaining module of KBMT-89 provides tools for developing, testing, and maintaining the knowledge base for a domain.

2.4   The Story So Far

In this chapter I have presented a brief overview of machine translation. This overview included the history of MT, the strategies that have been used, and a brief description of the architecture of a state-of-the-art MT system. A more complete discussion of MT (with detailed examples) can be found in [Hutchins 1986].

In chapter 1, we observed that errors in style can be a contributing factor to the errors in MT output. We also claimed that a formal representation is required to integrate a stylistic theory into a MT system. Evidence for this claim is provided by the history of improvements to MT output as successive improvements to the representation of syntax, semantics, and world knowledge were incorporated.

In the next chapter, we will examine three views of style. We will also describe some computational applications of these views of style, including MT. Chapter 3 will conclude with a description of DiMarco's formal representation of style. This representation will be used in later chapters to develop a set of rules for French–English comparative stylistics that are suitable for computational applications.
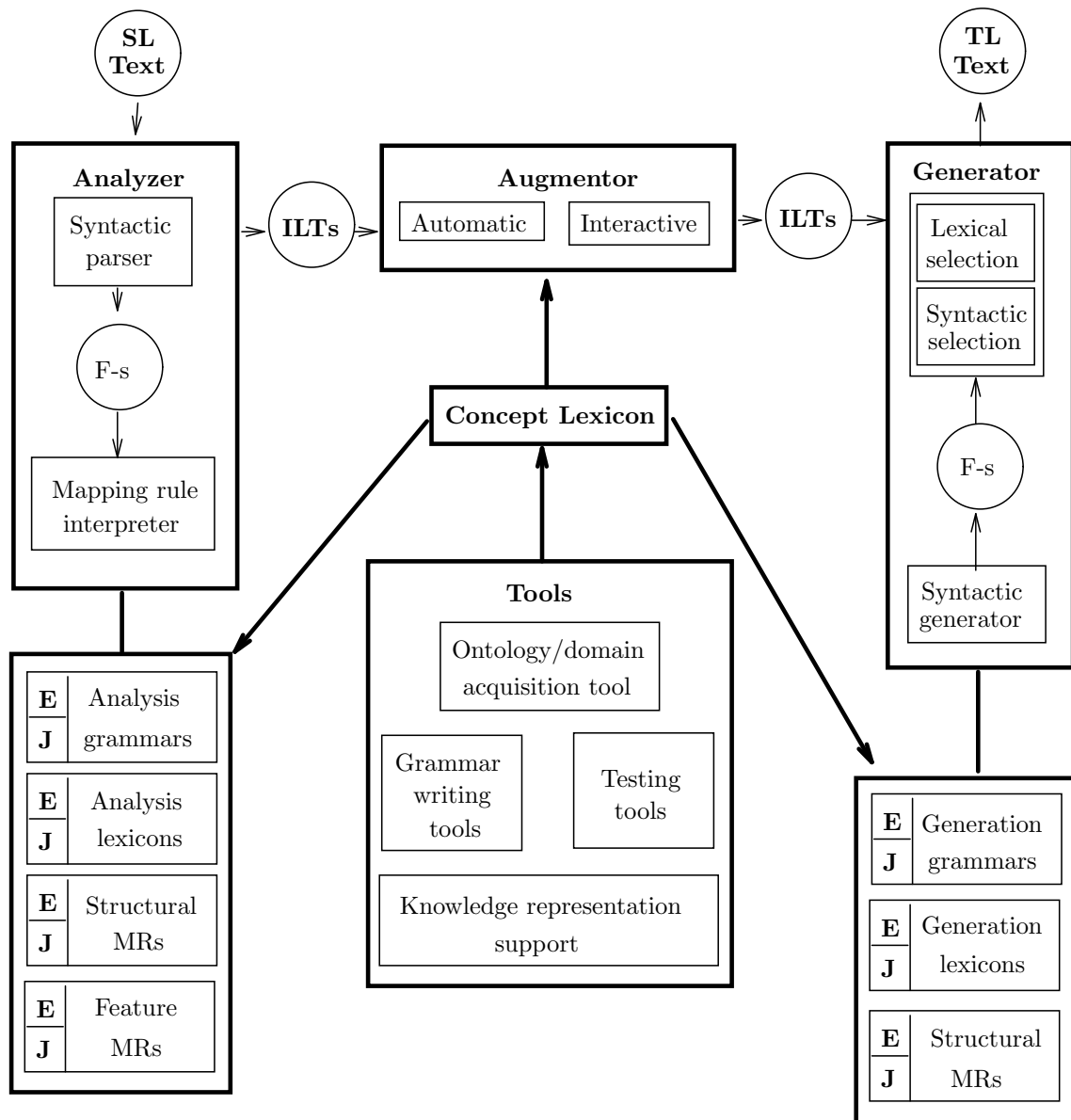
Figure 2.4: KBMT-89 system diagram.

Chapter 3

An Evolution of Computational Style

Ces choses sont hors de l'homme, le style est l'homme même. (These things [subject matter] are external to the man; style is the man.)
—Georges Louis De Buffon, Discours sur le style.

3.1 What is Style?

What is *style*? Using the fashion industry as an example, many people see style as the glamour of the high-fashion model or the appearance of the person on the street. Fashion trends are determined by the choices of the various designers, be it Saint Laurent or Armani. We can generalize our descriptions of fashion by talking about French style or American style. For the average person, dressing fashionably has additional rules that underlie our choices—"do not wear stripes with plaids."

For language, many people think of the apparent parallel situation. Linguistic style is considered by many to be the use of flowery language. Also, we can discuss the style of famous writers: the style of Shakespeare and the style of the Augustan poets. For the average person, style can be the grammar school rules dictating the usage of commas. But, linguistic style extends beyond this veneer.

Linguistic style is the sum of the many choices that we make in writing or speaking. For example, should the salutation of a letter be *Dear Kathleen* or *Dear Ms. Blair*? The answer, of course, depends on both the situation and the effect that we wish to convey—casual for a personal letter and formal for business or serious personal letter. The choices can be overt, as in the preceding example, or subtle, as in whether the reader *agrees* or *doesn't disagree* with my claims. Whether overt or subtle, the choices are there and they are an important part of language, influencing our choices of words, syntactic constructions, semantic content, and textual organization.

In this chapter, we will look briefly at three views of style: the prescriptive view where correct style is dictated to us, the descriptive view where accounts of our style are developed, and goal-directed style which attempts to correlate an author's lexical, syntactic, and semantic

choices with particular stylistic goals. The order in which these general views of style are presented represent one of the evolutionary paths that stylistic research has taken. Additionally, we will look at some of the computational applications of these views of style.

3.2   Prescriptive Style

Prescriptive style is the familiar "correct" style that is taught in school and in books of style (*e.g.,* Fowler [1968], Strunk and White [1979], and Kane [1983]). When we were taught English, we learned rules of usage like "Do not end a sentence with a preposition"[1] and "The adverb always follows the verb." These rules establish social conventions that facilitate the clear communication of ideas. By establishing common forms for expressing ideas we reduce the chance of miscommunication. These conventions are not fixed. They change as society changes. Use of double negatives is currently unpopular, but double negatives were used for emphasis in Shakespeare's time.

   The authoritarian nature of prescriptive style is not very useful for computational applications. It is the careful combination of conforming and deviating from the accepted stylistic norms that gives writing its expressiveness. A deviation may take the form of a misspelling or a subtle grammatical error to attract attention. For example, the unusual placement of the adjective "inane" *after* the noun it modifies emphasizes the writer's opinion of the government's actions.

   (3-1)   A change of heart by the government with respect to such common sense provisions
            seems to us <u>inane</u>.[2]

3.3   Descriptive Style

The English language has many sublanguages (or varieties) (*e.g.,* spoken versus written, dialects (the Queen's English, Canadian English, Australian English, etc.), accents, and those sublanguages determined by situation (those used between a parent and a baby, two scientists, or two lawyers)). Fortunately, the sublanguages of English are similar enough that not using the most appropriate sublanguage will usually result only in awkwardness and, in some situations, a lack of credibility. In actual practice, the appropriateness and correctness of an utterance is governed by the intuitions and conventions of native speakers of English rather than an arbitrary set of rules. Instead of prescribing an appropriate style for each variety of English (or an appropriate style for all varieties of English), *descriptive style* accounts for the style that is actually used by native speakers of English (or of the language being described).

   Crystal and Davy [1969] have developed a formal, descriptive account of English style. Their account was developed with three objectives:

---

[1]E. Gowers, in *Plain Words* (chapter 9, "Troubles with Prepositions") credits Sir Winston Churchill with responding "This is the sort of English up with which I will not put" when Churchill's grammar was corrected.
   [2][Kerr 1990a].

- To identify the linguistic features having stylistic significance and to develop a metalanguage for discussing these features;

- To organize the features so that comparisons can be made across languages; and

- To classify the features according to their extra-linguistic purpose.

Crystal and Davy's description of style was presented in the form of a grammar, a stylistic grammar of the English language. The grammar classifies stylistic features into five major categories: intersentence relationships, sentence types, clause types, groups, and words. The intersentence-relationships category includes semantic, pragmatic, and prosodic features, together with their relationships. The sentence-type category describes the form of a sentence (*e.g.,* complete or incomplete (sentence fragment)) and its complexity (*e.g.,* simple, complex, compound, etc.). The clause-type category describes the function and relationships between the clauses of a sentence. The groups category describes the functions and the relationships between the elements that form a clause (*e.g.,* nominal groups, premodifications, etc.). Lastly, the word category describes the morphological and phonological features of the words of a sentence. Using this grammar, Crystal and Davy claimed that they were able to account for the sublanguages of English by considering the variations in the relative frequencies of the stylistic features.

Later, in section 3.5.2, we will take a detailed look at the stylistic grammar developed by DiMarco [1990]. DiMarco follows the same philosophy as Crystal and Davy towards the construction of a stylistic grammar. However, DiMarco's grammar makes it possible for a *computer* to analyze and represent the syntactic style of a sentence (the sentence type, clause type, and group categories of Crystal and Davy's grammar).

3.4 Computational Stylistics

As the name suggests, *computational stylistics* is the application of computers to the processing of style. The manual analysis of style for a large amount of text can be time-consuming and tiresome. Computers can be used to automate the tedious task of analysis, leaving humans the more interesting task of interpreting the results. Currently, there are three main applications of computational style: the statistical analysis of style, the grammar and style checkers, and the application of style to machine translation. These applications will be described in the three following sections.

3.4.1 Statistical Analysis

Statistical analyses of style involves the accumulation of statistical data on the patterns of usage for text samples. Figure 3.1 is an example of the types of statistics that can be collected. It was produced by running the `style` program [Cherry and Vesterman 1981] from AT&T's Documenter's Workbench on chapter 2 of this thesis.

These statistics provide information about the text at the word and sentence levels. The readability grades are typically computed from the number of letters or syllables per word,

```
readability grades:
        (Kincaid) 12.9  (auto) 13.3  (Coleman−Liau) 14.1  (Flesch) 15.0 (36.4)
sentence info:
        no.  sent 224 no.  wds 4270
        av sent leng 19.1 av word leng 5.34
        no.  questions 2 no.  imperatives 0
        no.  nonfunc wds 2487  58.2%   av leng 7.16
        short sent (<14) 22% (49) long sent (>29)  10% (23)
        longest sent 48 wds at sent 25; shortest sent 4 wds at sent 163
sentence types:
        simple  58% (129) complex  30% (68)
        compound   4% (10) compound−complex   8% (17)
word usage:
        verb types as % of total verbs
        tobe  42% (182) aux  21% (92) inf  17% (75)
        passives as % of non−inf verbs  26% (91)
        types as % of total
        prep 12.0% (513) conj 2.7% (115) adv 2.9% (124)
        noun 28.9% (1235) adj 18.0% (767) pron 3.2% (137)
        nominalizations   4 % (190)
sentence beginnings:
        subject opener:  noun (30) pron (20) pos (0) adj (38) art (60) tot  66%
        prep  16% (35) adv   7% (16)
        verb   1% (3)  sub_conj   8% (17) conj   0% (0)
        expletives   2% (5)
```

Figure 3.1: An example of statistical stylistic analysis.

and the number of words per sentence. These grades are then normalized for some genre of texts. In the example, the Kincaid formula (used for technical documents) indicates that the text has an average level of reading difficulty, and the Flesch formula indicates that the text is written at a grade 15 reading level. The remaining statistics describe the sentences of the text. The basic premise is that long and complex sentences are more difficult to understand and should be avoided.

Statistical analyses are useful for examining the writing characteristics of a person or a group of people. They are not particularly useful for producing new texts[3]. Nevertheless, they can provide important information that would otherwise be overlooked. For example, in the French novel *Madame Bovary* ([Flaubert 1857]), the reader is left with a sense of boredom. This sense is not created through the use of words having meanings related to boredom, but more subtly through the intentional overuse of the imperfect (verb) tense[4]. The imperfect tense expresses a state of being (*e.g., It was raining as we read our newspapers*), as opposed to the perfect tense which expresses a completed action (*e.g., It rained just before we had lunch*). Thus, when the imperfect is overused we are left with the feeling that all the actions span long

---

[3]Some interesting exceptions exist. See [Milic 1982] for a description.
[4]Pierre H. Dubé, Department of French, University of Waterloo, personal communication.

periods of time—a very boring prospect!

Analyses like the preceding example make up a typical component of the grammar and style checking programs on the market today. These programs will be described in section 3.4.2. Literary stylists, like Milic and Cluett, have employed additional statistics that are more interesting and sophisticated.

Milic was a pioneer in the field of computational stylistics [Milic 1982]. He was the first to use computers for the analysis of style. His thesis claimed that "an objective description of Swift's style was possible if it was based on quantitative measures of specific linguistic units and compared with those of a set of related writers." [Milic 1982, 19]. The analysis was based upon the frequencies of words, phrases, sentences, parts of speech, and the first and last words of sentences. For the texts that were analyzed in his thesis, Milic discovered that the frequency distribution patterns were constant. In other words, the authors of the texts used the same patterns of words and syntactic structures. Milic discovered, unexpectedly, that it was the number of different patterns, the "D" statistic, that varied between these authors.

Milic analyzed literary prose for his thesis. His later work include the analysis of poetry and the generation of new poems based on the analyses of other poems—in effect, imitating the style of the author whose poems were analyzed.

Milic's method of statistical analysis was adopted by Cluett as the basis for the York Computer Inventory of Prose Style[5]. The main objectives of the York Inventory [Cluett 1976] were:

- To discover data about the syntactic habits of individual writers;

- To collect this data along with the corresponding texts;

- To make the collected data and the programs to manipulate the data accessible to researchers; and

- To provide theories and hypotheses about the nature and history of prose style.

Texts are added to the York Inventory according to the needs of researchers. For each text that is added, approximately 3500 words are selected to be analyzed. These words are randomly selected as part of 5 to 10 *cuts* (continuous sequences), with 350 to 700 words per cut. Each word and punctuation marker is encoded by a three-digit numeric code. Each of the 97 possible codes[6] represents a category encoding a part of speech and function of a word. For example, the code for "miscellaneous" nouns is 011, and the code for an attributive noun (*e.g., the railroad car*) is 012. So, the phrase *after leaving the ship* is encoded as "513 071 311 011" (preposition gerund determiner noun). This coding scheme was designed such that the first two digits form a prefix representing a class of categories (*e.g.,* all categories of nouns have a prefix of 01).

After the text samples have been encoded and the encoding key-punched, a computer program is used to determine the frequency of each of the 97 categories, 31 classes (*i.e.,*

---

[5]For brevity, I will refer to the York Computer Inventory of Prose Style as the York Inventory.
[6]See table 3.2 in [Cluett 1976] for a complete list.

two-digit prefixes), and 12 of the possible *prepositional phrase strings* (*e.g.,* preposition noun preposition, preposition gerund preposition, preposition pronoun preposition, etc.). Next, Milic's "D" statistic is computed to measure the syntactic variety of the text. The value of the "D" statistic is the number of different sequences of three classes[7] (*e.g.,* 51 31 01 (preposition determiner noun)). The program then produces frequency tables listing the 1000 most frequent sequences of three classes, the sequences of two and three classes used to start a sentence, the sequences of three classes used to terminate a sentence, and other patterns of stylistic interest. The program ends after computing the statistical reliability of the sample and statistics on sentence length.

A detailed discussion of the interpretation of these statistics and their role in the description of prose style is given in [Cluett 1976].

3.4.2   Grammar and Style Checkers

In contrast to the descriptive approach to style used in the statistical analysis of style, grammar and style checkers take a prescriptive approach to style. Grammar and style checkers are intended to help writers during the proofreading stage of the writing process. By comparing the writers' work to some norm, they are able to detect and report some types of spelling, grammatical, and stylistic errors.

They usually have two primary components: a dictionary and a set of rules. The rules can range in form from a simple table of phrases to match, along with the corresponding diagnosis, to a formal grammar of the English language. The `diction` and `explain` programs from AT&T's Documenter's Workbench use this simple approach. The `diction` program attempts to match phrases from its list of examples of poor diction against the text. Each matching phrase is displayed at the writer's terminal and the `explain` program is used to print the diagnosis. For example, if the writer were to use the word *implement*, `diction` would print out the word *implement*, and `explain` would suggest that the writer use the phrase *carry out* in place of *implement*. Most grammar and style checkers make some attempt to parse the sentences of the writer's text to produce a more accurate and precise analysis of the text.

The grammatical component of the checkers are, in part, extended spelling checkers. By using a grammar and part of speech information, a grammar checker can detect "misspellings" that a traditional spelling checker will miss. For example, *the read wagon* would be reported as an error since an article (*the*) may not be followed by a verb (*read*)[8]. The actual categories of grammatical error that are reported differ from program to program. Some commonly reported errors include:

- Subject-verb agreement;
  Julie, as well as her assistants, are *entitled to the commission.* (should be *is* (singular))

---

[7]Due to memory limitations, the maximum value of the "D" statistic is 1000 [Cluett 1976, 294]. In practice, this was not a severe limitation as the value of the statistic tends to lie between 820 and 960 [Cluett 1976, 24].

[8]Actually, this is a poor example since *read* can also be used as a noun. For example, *This thesis is a good read.*

- Pronoun case;
  If you have any further questions, please call either Chrysanne or <u>myself</u>. (should be <u>me</u>)

- Noun-modifier agreement; and
  Several of the missing <u>photograph</u> were found. (should be *<u>photographs</u>* (plural))

- Verb forms.
  The class was <u>gave</u> by Julie. (should be *<u>given</u>* (present participle))

The categories of stylistic errors that are reported also vary from program to program. Some common reported errors include:

- Diction (*e.g.,* legal-ese, business-ese);
  prioritize, dollarization

- Redundant phrases;
  each individual, merge together

- Double (or more) negatives;
  I <u>don't</u> know <u>nothing</u>.

- Excessive sentence length and complexity.

Detailed information on the development of particular grammar and style checking programs can be found in [Heidorn *et al.* 1982] and [Richardson and Braden-Harder 1988]. Reviews of commercially available grammar and style checking programs can be found in [Eglowstein 1991], [Neuman 1991], and [Madden 1991].

3.4.3   Style for Machine Translation

An area of current research is the application of computational stylistics to machine translation (MT). In the remainder of this section, I will describe two research directions in this area, that of Loffler-Laurian and of Tsutsumi. The next section (3.5) will describe a particular computational view of style that also has applications to MT.

Loffler-Laurian

Loffler-Laurian's[9] work [Laurian 1986; Loffler-Laurian 1983; Loffler-Laurian 1987] attempts 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 for technical journals). For example, the features that make that make a technical paper recognizable as such. In [Loffler-Laurian 1987], Loffler-Laurian proposes the following characteristics of technical style: the presence of figures and tables; particular usages of terminology, syntax, and morphology; and the absence of other usages of terminology, syntax, and morphology.

---

[9]Ms. Loffler-Laurian has published papers under the names Laurian and Loffler-Laurian. The latter is the more commonly used.

Loffler-Laurian uses post-editing as an instrument for MT research. Human post-editors were asked for their comments and opinions on the difficulties that they encountered during the post-editing stage. Their responses provided information about how a group of people characterize their own language, particularly "the ideal style that they always try to approach" [Laurian 1986]. Two types of post-editing were defined from this information: *rapid post-editing* and *conventional post-editing*. Rapid post-editing is used when a fast translation is required, and it represents minimal stylistic correction for intelligibility. On the other hand, conventional post-editing would be used if the translation were to be published, and it represents the "correct" style that is expected by the reader (consumer).

Both computational linguists and computational stylists have traditionally ignored the iconographic and typographic elements of text. Loffler-Laurian [Laurian 1986] claims that the proportion of diagrams and figures with respect to text is a characteristic difference that is expected by the reader. For example, a highly technical scientific journal such as *Comptes-Rendus de l'Academie des Sciences de Paris* has a higher proportion of figures to text than a popular scientific journal such as *Pour la Science*—the French edition of Scientific American.

Similarly, Loffler-Laurian claims that readers are sensitive to patterns of punctuation usage. Additionally, she indicates that punctuation usage in English is different from that of French. One situation where English conjunctions are realized in French by punctuation is discussed in [Guillemin-Flescher 1981] is presented below.

(3-2)   Another line of patients passed in the road <u>and</u> several of them pointed at the small car.[10]

(3-3)   Une autre file de malades traversa la route<u>;</u> plusieurs montrèrent la petite voiture du doigt.[11]

The conjunction *and* in the English sentence (3-2) is realized as a semicolon in the French sentence (3-3). Loffler-Laurian cautions researchers to distinguish between the compulsory differences in usage between English and French, and the differences that are a result of the group style.

Loffler-Laurian's research concentrates on characterizing types of group style. She attempts to experimentally determine the characteristics readers have come to expect of particular genres of technical texts. With respect to (machine) translation, one important point was emphasized: although corresponding group styles may exist across languages, the realization of the style can be different for each language.

Tsutsumi

Tsutsumi [1990] presents a methodology for bridging *stylistic gaps* (stylistic differences) between the source and target languages in MT. In Tsutsumi's work, style is limited to variations in syntax. Four types of stylistic gaps were identified: stylistic gaps in viewpoint (grammatical constructions that cannot be directly translated between the two languages), stylistic gaps

---

[10][O'Connor 1946, 440] after [Guillemin-Flescher 1981, 350] example 97.
[11][Gresset and Richard 1975, 153] after [Guillemin-Flescher 1981, 350] example 97.

in idioms and metaphors, stylistic gaps in specific constructions using function words, and "others".

Each of these types of stylistic gap is dealt with through *wide-range restructuring of intermediate representations.* In effect, Tsutsumi re-writes the sentence before it is translated. For example, an idiom like *John kicked the bucket* might be re-written as *John died.* The re-writings 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.

These rules have been incorporated into SHALT, a transfer-based MT system developed by Tsutsumi. The transfer stage of SHALT was divided into two steps: the first dealt with the implementation of the restructuring rules, and the second dealt with basic transfer (as described in section 2.2.2 starting on page 9).

An important contribution of Tsutsumi's work is the recognition that computational stylistics is useful for the translation of pairs of languages that are not in the same language group. All other work in applying computational stylistics to MT has involved languages in the same group (*e.g.,* Indo-European—English, French, German, etc.). It is, however, difficult to discover generalizations from Tsutsumi's rules, as they are integrated into the code, and are dependent on SHALT. We will see, in the following section, that one of the strengths of DiMarco's work is that her theory of computational stylistics can easily be separated from its implementation.

3.5 Goal-Directed Computational Style

Good writers know that they must understand their audience in order to write effectively. Knowing the audience allows writers to modify their style to effectively convey the content of their labour in a form acceptable to their audience. For example, a physicist explaining quantum mechanics to a general audience would probably choose to write clearly and simply, in order to be understandable to the audience. The same physicist writing a paper for a specialized conference would probably choose to write formally and somewhat obscurely, in order to appear authoritative. That is, the physicist chooses stylistic and rhetorical goals (*e.g.,* clarity, simplicity, formality, obscurity) according to the expectations of the audience.

A goal-directed view of style attempts to associate lexical choices, syntactic choices, and semantic choices with particular stylistic goals. In this section, we will look at Hovy's use of goal-directed style for (computerized) generation of text, and DiMarco's use of goal-directed style to construct a formal representation for style.

3.5.1 Hovy

The main contribution of Hovy's PAULINE text generation system was that it was the first to address the inclusion of pragmatic factors in natural language generation. These pragmatic factors allow PAULINE to vary the text being produced to suit the hearer and the situation. These factors also allow PAULINE to encode information beyond literal meaning into the texts that it produces. The following examples [Hovy 1988, 99] demonstrate how pragmatic factors might be used to affect the hearer's opinions.

(3-4)   Well, so Carter lost the primary to Kennedy by 1335 votes.

(3-5)   I am pleased to inform you that Carter has improved his chances of winning the nomination. At the present time, Carter has many more delegates than he had in the past; also, Carter has many more than Kennedy does.

Both examples were produced from the same input, but with different settings for the pragmatic factors. In particular, (3-4) was intended to be informal and not explicitly state the speaker's biases. On the other hand, (3-5) was intended to be formal and to explicitly exhibit the speaker's partiality. This effect is produced by choices of words, syntactic constructions, and overall text organization.

The effects of pragmatic factors are created through the use of rhetorical goals. As the name suggests, Hovy's rhetorical goals (*i.e.,* formality, simplicity, timidity, partiality, haste, force, colour, openmindedness, and respect) are used to produce a particular impression in the hearer. These impressions are produced through choices of words, phrases, and syntactic constructions. In addition to their role in constraining decision-making, rhetorical goals provide an intermediate level of representation that is used to account for different methods of creating pragmatic factors. For example, if a speaker wants to produce an effect that makes the hearer feel close, he may use the rhetorical goal "humorous." Hovy [1990, 164] gives an example of when this goal may not produce the required effect:

> Usually this will work well, but it will backfire if the hearer has just heard of his mother's death. In this case, the speaker's best strategy is to activate the rhetorical goals to be serious and slightly formal—which, under normal circumstances, would tend to alienate the hearer.

Combinations of rhetorical goals interact to produce pragmatic effects. For example, low formality, high force, and high partiality together produce the pragmatic effect of no-nonsense text.

The set of pragmatic factors and associated values proposed in PAULINE was not intended to make a "serious claim" [Hovy 1990]. The only claim that Hovy made was that "they are the kinds of features necessary for language generation" [Hovy 1990]. The lack of a formal basis for his factors is one of the shortcomings of PAULINE.

3.5.2   DiMarco

DiMarco [DiMarco 1990; DiMarco and Hirst 1990] presented a formalization for syntactic style. The formalization of syntactic style had two components: a vocabulary of style and stylistic grammars for English and French. This approach to the representation of style adopts the methodology of Crystal and Davy [1969], with DiMarco's vocabulary of style and stylistic grammar corresponding to Crystal and Davy's metalanguage and stylistic grammar, respectively.

The vocabulary of style describes style at three levels: stylistic goals, abstract elements, and primitive elements. The stylistic goals represent, as one would expect of a goal-directed approach to style, effects that are to be created in a sentence. The three pairs of goals that DiMarco examined were:

- Abstraction and concreteness;

- Staticness and dynamism; and

- Clarity and obscurity.

Each of the stylistic goals is defined by the presence of particular *abstract elements of style* which (attempt to) categorize high-level stylistic properties of sentences. For example, DiMarco claims that the stylistic goal of *clarity* is achieved for a sentence if the sentence is either *monoschematic*, *centroschematic*, *homopoisal*, or is a *resolution*. DiMarco observed that the abstract elements[12] of style can be classified by effects associated with position, balance, and dominance. These stylistic terms will be based on effects of *concord* and *discord*, which DiMarco defines as follows:

Concord: A stylistic arrangement that expresses a unity of style, agreement, accord, stability, and does not require resolution (to be defined below).

Discord: A stylistic arrangement that expresses a disunity of style, disagreement, contention, conflict, incongruity, and does require resolution.

The first type of abstract element deals with the effects created by the particular *position* of a component within a sentence. The most common type of position element describes an *imitative* effect:

Imitation: A stylistic arrangement in which two or more successive stylistic effects have similar syntactic forms.

For example, in the following sentence, successive initial participle clauses are used to create a concordant effect:

(3-6) Universally celebrated since 1830, regarded as the equal of Velasquez and for many superior to Murillo, sought by art collectors and museums, Francisco de Zurbarán is the despair of historians, one of those painters more admired than really known.[13]

In addition to an effect created by the strategic placement of similar shapes, we can observe other kinds of effects related to changes in the nature of shapes over the length of a sentence:

Resolution: A shift in stylistic effect that occurs at the end of a sentence and is a move from a relative discord to a stylistic concord.

We can see an example of a resolution in the following example. There is an initial discord, created by the misplacement of the adverb *entirely*, which is not usually found in the initial position. However, there is a final concord as the subsequent main clause contains no incongruities and restores the sense of harmony:

---

[12]The following description of the DiMarco's abstract elements has been abridged from section 3.2.4 of [DiMarco 1990] with the permission of the author.

[13]*Manchester Guardian Weekly*, 14 February 1988, 15 after [DiMarco 1990, 42].

(3-7) Entirely in the spirit of protective support, could I suggest you pass on an appropriate comment to the personnel concerned.[14]

And the complementary effect is dissolution:

Dissolution: A shift in stylistic effect that occurs at the end of a sentence and is a move from a relative concord to a stylistic discord.

We can see an example of a dissolution in the following example. A strong initial concord, created by imitative clauses, is set against a final discord, producing a syntactic inversion:

(3-8) And the rain descended, and the floods came, and the winds blew, and beat upon that house; and it fell: and great was the fall of it.[15]

The second group of abstract elements is related to effects of *balance* within a sentence. Here, we shall be looking at relationships that tend to either reinforce or perturb the harmony of a sentence. The distinction between effects related to position and those related to balance is that the latter are not restricted to one position in the sentence, but can occur more freely. Thus, resolution and dissolution are effects on the harmony of the sentence, but are dependent on their occurring in a sentence-final position.

The first and simplest type of balance element is a *homopoise* ("same weight"):

Homopoise: A stylistic arrangement with a single stylistic weight, allowing coordination of stylistically equivalent shapes, but no subordination.

In other words, there are one or more stylistic shapes in a homopoisal sentence, each contributing the same stylistic "weight" to the overall concordant effect of the sentence. A homopoisal sentence may consist of a single uninterrupted clause which contributes the sole weight to the stylistic effect of the sentence:

(3-9) True, posterity has not been kind to him.[16]

Or, a homopoisal sentence may contain a coordination, a conjunction of stylistically similar components, with each component contributing the same relative weight to the overall weight of the coordinate structure:

(3-10) The style was formed and the principles acquired.[17]

More complex sentences can have their balance interrupted or perturbed by a *heteropoisal* ("different weights") component:

Heteropoise: A stylistic arrangement with more than one stylistic weight that contributes either to an overall balance or imbalance.

---

[14][Quirk *et al.* 1985, 652] after [DiMarco 1990, 43].

[15]Matthew 7:27, Revised Standard Version [1952] after [DiMarco 1990, 43].

[16]*Manchester Guardian Weekly*, 14 February 1988, 15 after [DiMarco 1990, 43].

[17]*Manchester Guardian Weekly*, 14 February 1988, 15 after [DiMarco 1990, 44].

A heteropoise can appear in initial, medial, or final position. Thus, an initial, medial, or final heteropoise is a heteropoise in which an initial, medial, or final stylistic shape, respectively, provides a weight that creates an imbalance within the sentence. In the following example of an initial heteropoise, the sentence contains an initial offset, an appositive noun phrase, that contributes a stylistic weight separate from that of the main clause:

(3-11) <u>The heir to a fortune</u>, her friend did not care about passing examinations.[18]

The third type of abstract element deals with stylistic *dominance*, which is concerned with the relative weight contributed by each sentence component to the overall stylistic effect. A common type of dominance element is the *monoschematic*, a very simple shape.

Monoschematic: A sentence with a single dominant shape and no accompanying subordinate or coordinate shapes.

We have already seen an example of a monoschematic sentence in the discussion of homopoise:

(3-12) True, posterity has not been kind to him.[19]

The most common dominance element in the texts DiMarco studied is the *centroschematic*:

Centroschematic: A sentence in which the components are organized around a single dominant component.

Centroschematic sentences can be quite varied in structure, but their shared characteristic is a predominant component that serves as the organizational centre for all other components. Such structures are built with subordination and coordination. In the example below, the main clause is supported by a complex, but subordinate and coordinate, relative clause structure:

(3-13) Neither these devices nor the cramped viewing rooms <u>which are too narrow and whose ceilings are much too low for the big altar pieces</u> manage to spoil the works.[20]

Each of the abstract elements is defined in turn by *primitive elements of style.* At the level of primitive elements, the stylistic grammar describing the abstract elements diverges into two views: the *connective* and *hierarchic* views. The connective view of a sentence describes the degree of integration of the components of the sentence. DiMarco defines nine degrees of connectiveness ranging from conjunct^4 (strongly integrating) to conjunct^{-4} (weakly integrating). These degrees of connectiveness (conjunct^x) have a companion naming scheme (antijunct^{-x}) where the superscripts are negated (*i.e.,* conjunct^{-4} is equivalent to antijunct^4, and antijunct^{-4} is equivalent to conjunct^4). This alternate naming scheme is used to describe the degree of "anti-integration." For convenience, I will alternate between the two naming schemes so that the superscripts are always positive. The degrees of connectiveness were assigned by DiMarco based on frequency of usage and the presence of syntactic cues (*e.g.,* subject-verb agreement)

---

[18][Quirk *et al.* 1985, 1314] after [DiMarco 1990, 44].
[19]*Manchester Guardian Weekly*, 14 February 1988, 15 after [DiMarco 1990, 46].
[20]*Manchester Guardian Weekly*, 14 February 1988, 15 after [DiMarco 1990, 44].

for a particular syntactic construction. For example, in English it is more common for an adjective to precede the nominal group it modifies than to follow it. Thus, the adjective *inane* would be assigned a degree of conjunct[1], mildly connective, in the fragment *seems <u>inane</u> to us* and a degree of antijunct[1], mildly disconnective, in the fragment *seems to us <u>inane</u>.*

The hierarchic view of a sentence describes the degree of syntactic subordination of each sentence component. The range of the nine degrees of subordination are defined in a manner similar to the connective view, from subjunct[4] (equivalent to superjunct^{-4} representing a high degree of subordination to subjunct^{-4} representing a low degree of subordination.

The primitive elements of style exist at the bottommost level of the vocabulary and of the grammar. These elements describe the stylistically significant features of syntax and how they combine to produce stylistic effects in a sentence. The syntactic elements that DiMarco recognizes as having stylistic significance are adjectives, adverbs, premodifications (of nominal groups), postmodifications (of nominal groups), nominal groups, noun phrases, verb phrases, dependent clauses, major sentences, minor sentences, and complete sentences. These choices of stylistically significant elements are supported by [Crystal and Davy 1969], [Cluett 1976] and [Quirk *et al.* 1985].

As with any pioneering work, DiMarco's grammars are not without their limitations. In [DiMarco 1990], she indicates that the current grammars have limited descriptive capabilities and discusses how the English and French grammars were constructed with different underlying philosophies. Both of these limitations can lead to anomalous results when the grammars are used. The limited descriptive capabilities produce imperfect results when the stylistic effects produced by features are too subtle for the grammars to clearly detect, and the differences in the English and French grammars produce anomalies when comparing the stylistic parses produced using each of the grammars (I will discuss this in section 5.1). Another limitation is that the grammar is restricted to analyzing only aspects of style that can be encoded in the syntax of a sentence. In particular, this overlooks the effects due to the use of punctuation, and the effects of lexical and semantic style on syntactic style.

3.6   The Story So Far

In this chapter, we have looked at three views of style: prescriptive, descriptive, and goal-directed. We have also looked at how computers can be applied to the processing of style, particularly in analysis. We have also seen that computers have applications beyond just analyzing style, to using an understanding of style to produce variations in text and to improve translation.

In the next chapter, we will examine how style varies between languages, English and French in particular. Then we will see how a component embodying this information about *comparative stylistics* can be integrated into DiMarco's grammars and applied to machine translation.

Chapter 4

A Theory of Syntactic French–English Comparative Stylistics

> *English prose style is notoriously difficult to discuss without falling into either subjective impressionism or excessive technicality.*
> —*Louis T. Milic*

4.1   Introduction

Comparative stylistics is a subfield of stylistics that attempts to describe the differences in style between languages. Rules of comparative style are commonly presented, in textbooks of translation, as "rules-of-thumb." We may see rules like "French words tend to be more abstract than the corresponding English words" or "English tends to use the passive voice more often than French."

Translators use this type of rule to produce translations that appear more natural and appropriate to a particular context. In developing a computational model of comparative stylistics for machine translation (MT), we hope to improve the quality of the translated output. An implementation of this model will provide additional information to the MT system about the potential modulations to the translated text and their effects. This additional information will enable the MT system to make a more informed decision.

In this chapter, I will briefly describe three branches of style: lexical, syntactic, and semantic. This description is followed, in section 4.3, by the presentation of a set of rules for syntactic comparative stylistics for French–English translation.

4.2   Branches of Stylistic Description

When working with large fields of study, such as stylistics, it is often convenient to divide the field into several smaller branches. Both DiMarco [1990] and Vinay and Darbelnet [1958] have divided style into three branches: lexical, syntactic, and semantic. In the remainder

of this section, I will describe each branch and the fuzzy boundaries between the branches, particularly with respect to comparative stylistics.

4.2.1   Lexical Style

Lexical style can be characterized as diction, a choice between similar words. The choice could be made on the basis of appropriateness to a context. For example, *John kicked the bucket* would be more appropriate in a colloquial context and *John passed away* would be appropriate in a more formal context. More often, the choice will be made on the basis of aspects that are distinguished in one language but not the other. French, for example distinguishes between *guichet* (*lit.,* counter, wicket), *fenêtre* (*lit.,* window), and *devanture* (*lit.,* a store window) for the English word *window*. Conversely, English distinguishes between *cot* and *bed* for the French *lit* (*lit.,* bed).

Some of these aspects have been classified by Vinay and Darbelnet ([1958]). Their categories (and examples) include:

- Duration.

  The English word *never* is normally translated into French using the construction *ne . . . jamais*. The following examples are cases where *never* does not express a duration, and is translated accordingly.

English	We never asked.
French	Nous avons oublié de demander.
	(*lit.,* We forgot to ask.)
English	He never thanked me.
French	Il ne s'est pas donné la peine de me remercier.
	(*lit.,* He did not make an effort to thank me.)

- Literal versus figurative sense.

  As the following examples show, the meaning of a word can be influenced by both the intended interpretation (literal or figurative) and the position of the word.

	Literal	Figurative
English	thin, lean	meager
French	maigre	maigre
English	the old prime minister	the former prime minister
French	le premier ministre ancien	l'ancien premier ministre

- Technical versus usual language.

  This is similar to literal and figurative senses, where the usage (context) influences the choice of words. For example, where English uses the word *funeral*, French uses *obsèques* (*lit.,* funeral) in technical situations and *enterrement* (*lit.,* burial) in everyday situations.

	Technical	Usual
English	funeral	funeral
French	obsèques	enterrement
English	condolences	sympathy
French	condoléances	condoléances

- Connotative versus denotational meaning.

  Connotative and denotational meanings are more difficult to capture. A denotational meaning captures association between a word and its literal meaning (*i.e.,* its dictionary definition). So, a *denotational* meaning of "field" is "an area of cleared, enclosed land used for cultivation or pasture." A *connotative* meaning captures the hearer's or the reader's associations with the word. The connotative meaning of "field" may, depending on the hearer or reader, include: the sounds, sights, and smells of a baseball game; or the idyllic image of the grasslands near a lake.

These are only a small number of the aspects of lexical style. More detailed descriptions of these aspects and others can be found in [Vinay and Darbelnet 1958].

4.2.2   Syntactic Style

Syntactic style can be characterized as the choices made in the construction of a clause or a sentence. The following examples from [Corbett 1971, 434] demonstrate some of the possible variations.

(4-1)   The boy loves the girl.

(4-2)   The <u>tall, handsome</u> boy <u>sincerely</u> loves the <u>short, homely</u> girl.

(4-3)   The boy, <u>tall and handsome</u>, loves <u>sincerely</u> the short, homely girl.

(4-4)   <u>Tall and handsome</u>, the boy loves the short, homely girl <u>sincerely</u>.

The first example is a simple and ordinary sentence. The second example adds modifiers to the first, but still leaves us with an ordinary sentence. The third and fourth sentences vary the positions of the modifiers to produce some stylistic variation by placing the modifiers in unusual positions. The positioning of *tall and handsome* in the last two examples places the emphasis on *the boy*, and the positioning of *sincerely* places the emphasis on *love* in the third example, and *the girl* in the fourth.

The general approach of placing modifiers in unusual positions can be applied to other languages, like French, but the actual positions may be different. In a later section, a detailed description of some syntactic variations that exist between French and English will be given.

4.2.3   Semantic Style

Semantic style is difficult to characterize as little is known about it. One possible characterization is as information flow. Two components that are useful in describing the information flow are *focus* (or thematization) and *argument structure*.

Focus describes a reader's centre of attention and how it changes from sentence to sentence and from paragraph to paragraph. It is a tool that facilitates coherence and is used to order a series of related ideas coherently. For example, the following sequence of sentences from [McKeown 1985] appears incoherent.

Skiing, golf, and tennis are cases in point.
Many sports are a rich man's domain.

Yet, reversing the ordering makes the sentences coherent.

Some researchers, like Fries[1] and Hovy[2], believe that semantic style should not be limited to focus. Based on the corpus that was examined for this thesis (see section 4.4), argument structure, which subsumes focus, is a logical successor. It includes what to talk about, how much to say, what evidence to include, and how to organize it. Classical rhetoricians have developed a framework for describing the structure of argumentative essays. The framework divides the development of an essay into five functional parts: the exordium (introduction), the narratio (statment of fact), the confirmatio (confirmation or proof), the refutatio (refuting opposing arguments) and the peroratio (conclusion). They have also defined standard methods for developing each part to produce known effects.

The full expressiveness of this framework does not appear to be required for French–English translation. In most of the texts that I examined, the structure of the translated text closely follows the original. The primary exception that was encountered was the magazine *Protégez-vous* and its English counterpart *Protect Yourself*[3]. Many of the structural differences between the translated texts can be attributed to the amount of description and clarification, the arrangement of the propositions that compose a sentence, and the amount of information in a sentence.

Detailed descriptions of focus can be found in [Sidner 1979] and [McKeown 1985]. A description of the application of focus to semantic style that applies DiMarco's [1990] approach appears in [Ryan 1989]. Hovy [1988] discusses aspects of selecting the level of detail, and the amount of information to include in computer-generated texts. [Alvarado *et al.* 1990a] and [Alvarado *et al.* 1990b] describe a computational theory for the representation of editorial text.

4.2.4   Interactions: The shades of grey in-between

Each of the preceding branches of style does not exist in isolation. Each branch can influence one or both of the others. Focus is a component of semantic style, but a choice of unusual words (lexical style) or a choice of unusual constructions (syntactic style) can influence the focus by emphasizing the choice of words or syntax.

There does not appear to be a straightforward method of determining the hierarchy of the branches. A hierarchy that places semantic style over syntactic style, and syntactic style over lexical style is logical. It emphasizes the cultural influences, like acceptable organizations of

---

[1]Peter H. Fries, Department of English Language and Literature, Central Michigan University, personal communication.

[2]Eduard H. Hovy, Information Sciences Institute, personal communication.

[3]The English edition is no longer published.

the text, over syntactic and lexical concerns. The reverse ordering, lexical style over syntactic style over semantic style, is also a logical ordering. It emphasizes content (*i.e.,* finding the right words to get the ideas across). These two hierarchies would suggest that the genre and purpose of the text, in addition to the differences in the languages, will play a role determining which hierarchy to use.

4.3   An Initial Set of Rules

As a starting point for the formalization of syntactic comparative style, I selected a set of traditional rules based on those described by Guillemin-Flescher [1981]. Her characterization of French and English syntactic structures emphasizes the clauses of a sentence and the relationships that exist between them. Guillemin-Flescher gives the following example:

> **(4-5)** Mais, <u>quand il eut la plume entre les doigts</u>, <u>il ne sut rien trouver</u>, si bien que, s'appuyant sur ses deux coudes, <u>il se mis à réfléchir</u>.[4]

There is a temporal relationship between the clauses "quand il eut la plume entre les doigts" and "il ne sut rien trouver" (*i.e., when* he was unable to think of what to write). Also, there is a causative relationship between "il ne sut rien trouver" and "il se mit à réfléchir" (*i.e.,* he was unable to think of what to write, *so* he pondered). Some of the other relationships that Guillemin-Flescher uses are point of reference, spatial localization, composition relation, and consequence relation.

4.3.1   Anteposition

The rules that were selected have a common underlying theme. Each 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 connectiveness. The concept of connectiveness describes how closely two syntactic constructs are linked together. For example, *black* and *cat* in *the black cat* have a higher degree of connectiveness than *the cat which is black*. In the first case, *black* and *cat* are parts of one cohesive unit (a noun phrase). Whereas in the second case, they are components of separate constructs (a noun phrase and a relative clause).

   The first group of rules that I 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 DiMarco's definition of premodification, but is not limited to the modification of nominal groups. A strong degree of connectiveness 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.

---

[4][Flaubert 1857, II, XIII, p. 228] after [Guillemin-Flescher 1981, 108]. In [Flaubert 1957, II, XIII, p. 226], the sentence is translated as "But when he took up his pen he couldn't think of what to write, and he leaned on his elbows and pondered."

Anteposition $\longrightarrow$ Integrated into a nominal group
> *Dans nos agences, nous nous faisons un plaisir de donner à nos clients tous conseils pour la meilleure utilisation de leurs pneus.*[5]

> *The staff of our Branches [sic] will be pleased to give advice on the best use of tyres.*[6]

Anteposition $\longrightarrow$ Integration into a principle clause
> *En France, il y a 18 000 bureaux de Poste; il y en a toujours un, là où vous êtes.*[7]

> *There are 18 000 post offices in France and one wherever you are.*[8]

In the two preceding rules, the French antepositions are merged into another construction in the English translation. In the example for the first rule, the French anteposition *dans nos agences* is integrated into the English nominal group *the staff of our Branches*. In the example for the second rule, the French anteposition *En France* is integrated into the main clause of the English translation.

Stronger degrees of connectiveness are possible. The following rule merges two sentences into one using a co-ordinating conjunction (*and* in the example).

Anteposition $\longrightarrow$ 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.*[9]

> *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.*[10]

These stronger degrees of connectiveness often extend beyond sentence boundaries. DiMarco's stylistic grammars (which will be used in the next chapter to formalize the traditional rules given in this chapter) are limited to representing single sentences. For this reason, the strongest degrees of connectiveness will be omitted.

The last rule of the anteposition group demonstrates a weak degree of connectiveness. The rule describes French antepositions that are so weakly connected to the rest of the sentence that they can be separated to form another sentence.

Anteposition $\longrightarrow$ Principle clause
> *Que ce soit un mot, une lettre ou une phrase entière, Tipp-Ex fluid corrige aussi bien les originaux que les copies.*[11]

---

[5][Michelin 1972] after [Guillemin-Flescher 1981, 338] example 50.
[6][Michelin 1972] after [Guillemin-Flescher 1981, 338] example 50.
[7][Pub. PTT —] after [Guillemin-Flescher 1981, 334] example 33.
[8][Pub. PTT —] after [Guillemin-Flescher 1981, 334] example 33.
[9][S.N.C.F. 1971, 51] after [Guillemin-Flescher 1981, 338] example 48.
[10][S.N.C.F. 1973] after [Guillemin-Flescher 1981, 338] example 48.
[11][Pub. Tippex —] after [Guillemin-Flescher 1981, 337] example 47.

> *Corrects typing errors, drawings, ballpoint ink, (on originals or carbon copies, either in or out of the typewriter).*[12]

The example for this rule, although it was the best example I could find, does not show the intent of the rule very clearly. (See the example for the rule, Relative clause ⟶ Principle clause, for a better example of the intent.) In the above example, the meaning of the anteposition (*i.e.,* any type of error—whether it is a word, a letter, or an entire sentence[13]) becomes the main clause and the focus of the sentence (*i.e.,* any type of error—whether it was typed, drawn or hand-written) in the English translation.

4.3.2   Relative Clause

The next group of rules, the *relative clauses*, can also be ordered according to the degree of connectivity expressed by each rule. The following four rules express a strong degree of connectiveness, 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 demonstrate 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   ⟶   Integrated into a principle clause (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.*[14]

> *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.*[15]

Relative clause   ⟶   Integrated into a principle clause (participle)
> *Les lignes non électrifiées, qui représentent 74% de la longueur du réseau, n'écoulent que 22% du trafic total de la S.N.C.F.*[16]

> *Non-electrified lines representing 74% of the total network carry only 22% of the overall S.N.C.F. traffic.*[17]

Relative clause   ⟶   Integrated into a truncated principle clause (participle)
> *Ce n'est rien, capitaine! . . . Une branche morte qui s'est brisée . . .* [18]

> *I'm all right, Captain . . . only a rotten branch breaking . . .* [19]

Relative clause   ⟶   Integrated into a principle clause (prepositional phrase)
> *Je vais vous lancer une corde à laquelle est attachée une ceinture de savetage.*[20]

---

[12][Pub. Tippex ⸺] after [Guillemin-Flescher 1981, 337] example 47.

[13]My literal paraphrase.

[14][Mayoux 1979, 67] after [Guillemin-Flescher 1981, 341] example 59.

[15][Blake and Crewe 1978, 67] after [Guillemin-Flescher 1981, 341] example 59.

[16][S.N.C.F. 1971, 16] after [Guillemin-Flescher 1981, 342] example 62.

[17][S.N.C.F. 1973] after [Guillemin-Flescher 1981, 342] example 62.

[18][Hergé 1948, 41] after [Guillemin-Flescher 1981, 342] example 63.

[19][Lonsdale-Cooper and Turner 1962, 41] after [Guillemin-Flescher 1981, 342] example 63.

[20][Hergé 1947, 58] after [Guillemin-Flescher 1981, 342] example 64.

I'll throw you a line <u>with a life-jacket attached.</u>[21]

The next rule expresses a weaker degree of connectiveness than the preceding four since the relative clause is translated as an adverbial clause expressing time (*i.e.,* a temporal 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 <u>qui se décomposa peu à peu en un liquide épais et visqueux.</u> Il le remit ensuite au feu et le répandit brûlant sur la coque du bateau.[22]

He boiled the pith in a small cauldron he had brought from the Virginia <u>until the fibrous mass was rendered down to a thick, viscous liquid,</u> which he reheated and smeared over the vessel's hull.[23]

The next rule expresses an even weaker degree of connectiveness. 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, <u>où on les examine en ce moment.</u>[24]

Yes, I've left some of them at the laboratory at police headquarters. <u>They're working on them now.</u>[25]

And, finally, a rule expressing the weakest degree of connectiveness—where the relative clause is omitted altogether.

Relative clause   ⟶   Suppressed

Puis il fit un brusque demi-tour et s'enfuit ventre à terre dans les taillis <u>où il disparut.</u>[26]

He turned abruptly and bolted into the wood.[27]

4.3.3   Juxtaposition

The next group of rules are the *juxtapositions*. A juxtaposition is the placement of two syntactic entities next to each other. In particular, a juxtaposition is used for clarification or to add information. For example, *the Prime Minister, Brian Mulroney,* is a juxtaposition that adds information about which prime minister (*i.e.,* Brian Mulroney as opposed to John Major). In

---

[21][Lonsdale-Cooper and Turner 1961, 58] after [Guillemin-Flescher 1981, 342] example 64.

[22][Tournier 1967, 34–35] after [Guillemin-Flescher 1981, 340] example 57.

[23][Denny 1974, 32] after [Guillemin-Flescher 1981, 340] example 57.

[24][Hergé 1948, 18] after [Guillemin-Flescher 1981, 340] example 56.

[25][Lonsdale-Cooper and Turner 1962, 18] after [Guillemin-Flescher 1981, 340] example 56.

[26][Tournier 1967, 32] after [Guillemin-Flescher 1981, 338] example 51.

[27][Denny 1974, 30] after [Guillemin-Flescher 1981, 338] example 51.

many case, juxtapositions will be more easily explained by lexical style (*i.e.,* knowing when the added information is necessary) than syntactic style.

The juxtaposition group of rules are:

Juxtaposition $\longrightarrow$ Integrated into the principal clause

> *Toutes les chambres, <u>d'une superficie de 14m^2</u>, disposent d'une petite terrasse privée où vous pourrez vous reposer à l'abri des regards indiscrets.*[28]

> *All rooms <u>are a comfortable size</u> and have their own balcony where you can sit in complete privacy.*[29]

Juxtaposition $\longrightarrow$ Explicitly stated relation

> *Il était là, <u>les ciseaux à la main,</u> l'air mal assuré.*[30]

> *He stood <u>holding the scissors</u> uncertainly.*[31]

The first rule has an average degree of connectiveness as it makes the juxtaposition a part of the main clause. The second rule varies in its degree of connectivity, since it is a default case. The degree of connectivity depends on how the relationship between the two juxtaposed entities is expressed (*i.e.,* explicitly stated). If the relationship is expressed as part of a nominal group, then it is more connective than the first rule since it is tightly integrated into the sentence. If the relationship is expressed as a separate sentence, then it is less connective than the first rule since it is not integrated into the sentence at all.

4.3.4   Embedding

The two remaining groups 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 the translation, the interruption can either be sustained, or moved before (anteposition) or after (postposition) the construct that it interrupts. By moving the embedding, we are left with a sentence that is more connective than the original since 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 governed by coherence, an aspect of semantic style.

Embedding $\longrightarrow$ Anteposition

> *L'Assurance garantit, <u>sauf déclaration spéciale</u>, la totalité des bagages de l'Assuré, y compris ceux conservés à la main.*[32]

> *<u>Unless otherwise stated</u>, your insurance covers all your luggage including hand-luggage.*[33]

---

[28] [Pub. Najac ——] after [Guillemin-Flescher 1981, 348] example 88.
[29] [Dibben ——] after [Guillemin-Flescher 1981, 348] example 88.
[30] [Gresset and Richard 1975, 136] after [Guillemin-Flescher 1981, 345] example 76.
[31] [O'Connor 1946, 431] after [Guillemin-Flescher 1981, 345] example 76.
[32] [Pub. C.E.A. ——] after [Guillemin-Flescher 1981, 329] example 12.
[33] [Dibben ——] after [Guillemin-Flescher 1981, 329] example 12.

Embedding $\longrightarrow$ Postposition

> *La première, <u>réalisée depuis la fin de 1964</u>, a été une phase préparatoire.*[34]

> *The first was the preparatory stage, <u>completed at the end of 1964</u>.*[35]

4.3.5 Inversion

The final group, *inversion*, is a broad category. It attempts to capture a large number of the differences between the two languages including syntactic ordering (as in the example below), lexical influences, and semantic influences. An example of a lexical influence was given in the earlier section on lexical style (*i.e.,* where the positioning of the adjective *ancien* before or after the noun affected the interpretation). An example of a semantic influence is the choice of active or passive voice. In this case, the form of the argument may dictate whether the active voice (involving the writer in the argument) or the passive voice (distancing the writer from the argument) should be used.

The rule below expresses the concept of re-ordering components of a sentence (*i.e.,* returning to the canonical or most accepted ordering) 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—*Place <u>in a deep saucepan</u> 4 tablespoons of butter, to melt but not brown.*

Inversion $\longrightarrow$ Canonical ordering

> *Mettre <u>en casserole profonde</u> 75 g de beurre, à fondre sans brunir.*[36]

> *Place 4 tablespoons of butter <u>in a deep saucepan</u> to melt but not brown.*[37]

4.4 A Corpus Study

The rules in the preceding section are a subset of those given by Guillemin-Flescher [1981]. Since DiMarco's grammars will be used to construct a formal representation of this subset of rules, the major consideration used in selecting the subset was that the situations represented by the rules be representable using DiMarco's grammars (see section 3.5.2 for a description of the limitations). A corpus study was used to verify that the rules that were selected are capable of accounting for the syntactic style of a specific genre of texts.

Work in style is often limited to a particular genre (or domain). It is difficult to build a model that applies to all genres, since the genres vary widely. Within a particular genre, style continues to vary, but the number of possibilities is limited to a non-trivial but manageable number.

The genre that I have selected is *political advocacy*. The purpose of texts in this genre is to express the author's opinion on a topic and to influence the reader. Examples of political

---

[34][S.N.C.F. 1971, 73] after [Guillemin-Flescher 1981, 330] example 17.
[35][S.N.C.F. 1973] after [Guillemin-Flescher 1981, 330] example 17.
[36][La Reynière 1971, 45] after [Guillemin-Flescher 1981, 327] example 2.
[37][Coltman 1973, 51] after [Guillemin-Flescher 1981, 327] example 2.

Anteposition	→ Integrated into a nominal group	3
Anteposition	→ Integration into a principle clause	2
Anteposition	→ Principle co-ordinate clause	1
Relative clause	→ Integrated into a principle clause (adjectival)	2
Relative clause	→ Integrated into a principle clause (participle)	3
Relative clause	→ Integrated into a truncated principle clause (participle)	1
Relative clause	→ Integrated into a principle clause (prepositional phrase)	4
Relative clause	→ Principle clause	1
Relative clause	→ Suppressed	1
Juxtaposition	→ Integrated into the principal clause	2
Juxtaposition	→ Explicitly stated relation	6
Embedding	→ Anteposition	7
Embedding	→ Postposition	5
Inversion	→ Canonical ordering	51

Figure 4.1: The rules occurring in the corpus (with frequencies).

advocacy texts include editorials, reviews, and promotional literature. The choice of genre was influenced by recent world events (*e.g.,* glasnost, the Gulf War) which have highlighted the amount of information available in languages other than English. The choice was also influence by the increasing availability of commercial machine translation (MT) systems [Benton 1991; Valigra 1991].

To verify that the rules presented in the preceding section are applicable to the political advocacy genre, a corpus of political advocacy texts was constructed from three randomly selected paragraphs from each of nineteen articles[38]. The corpus contains a total of 162 sentences from 57 paragraphs.

After the corpus had been constructed, it was examined for occurrences of the rules presented in the preceding section. Figure 4.1 lists the fourteen rules of the sixteen (presented in the preceding section) that occurred in the corpus and the number of occurrences of each rule. Of stylistic variations that are not covered by the rules, there were 20 cases attributed to punctuation (which is not representable by DiMarco's grammars), 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 for reasons of clarity, and 51 were a result of changes in the point of view.

---

[38]The articles used in the corpus are: [Chipaux 1990a], [Chipaux 1990b], [Greilsamer 1990b], [Greilsamer 1990a], [Reynolds 1990b], [Reynolds 1990a], [Murray 1990a], [Murray 1990b], [Dupeux 1988a], [Dupeux 1988b], [Canadian Council on Social Development 1990b], [Conseil canadien de Développement social 1990a], [Canadian Council on Social Development 1990a], [Conseil canadien de Développement social 1990b], [Friends of the Earth 1987], [Les Ami(e)s de la Terre 1987], [Friends of the Earth 1986], [Les Ami(e)s de la Terre 1986], [Duhaime 1990a], [Duhaime 1990b], [Beauparlant 1990b], [Beauparlant 1990a], [Pierre 1991d], [Pierre 1991c], [Pierre 1991b], [Pierre 1991a], [Unknown 1990b], [Unknown 1990a], [Kerr 1990a], [Kerr 1990b], [Chatto 1990b], [Chatto 1990a], [MacDonald 1990a], [MacDonald 1990b], [Dappen 1991b], [Dappen 1991a], [Unknown 1991b], and [Unknown 1991a].

4.5   The Story So Far

In this chapter, we looked at descriptions of comparative stylistics and of three branches of style. These descriptions were followed by the presentation of a set of rules giving a theoretical description of syntactic French–English comparative stylistics. A corpus of sentences from political advocacy texts was used to verify that these theoretical rules applied to this particular genre.

In the next chapter, the theoretical rules that we represented in this chapter will be formalized. The chapter describes the procedure used to formalize the theoretical rules listed in figure 4.1 using both the corpus described in the preceding section and DiMarco's stylistic grammars (described in section 3.5.2). This is followed by the presentation of the formal rules and a description of how the rules might be used in a machine translation system.

Chapter 5

A Grammar of Syntactic French–English Comparative Stylistics

> *Give us the tools, and we will finish the job.*
> — *Winston Churchill, radio broadcast, February 9, 1941*

5.1   Introduction

In this chapter, I will present a set of formal rules for French–English comparative stylistics (derived from the theoretical rules given in figure 4.1) and a description of how these formal rules can be applied to machine translation.

The process of deriving the formal rules was straightforward, but time-consuming. The first step in the process was to use DiMarco's grammars (see [DiMarco 1990]) to produce stylistic parses for each of the sentences in the corpus described in section 4.4. Each of the sentences was parsed by hand. Although STYLISTIQUE, DiMarco's implementation of the stylistic grammars, was available, it proved to be too inefficient to parse many of the sentences. After the stylistic parsing had been completed, each of the sentence pairs (*i.e.,* a French sentence and its corresponding English translation) in the corpus was inspected for instances of the theoretical rules in figure 4.1.

For each instance of a theoretical rule, the stylistic features[1] of the sentence components corresponding to the left and right hand sides of the rule were recorded. For example, the left hand side of the following rule corresponds with the underlined portion of the French sentence in figure 5.1 and the right hand side corresponds with the underlined portion of the English translation.

      Inversion $\longrightarrow$ Canonical ordering

---

[1] The use of the term *feature* corresponds with the terms *primitive element* and *abstract element* in DiMarco's terminology.

Figures 5.2 and 5.3 show the lowest level of the stylistic parse corresponding to the two under-lined sentence fragments in figure 5.1. Figure 5.4 summarizes the stylistic features that would be recorded for these parses, and shows the correspondences between the French and English stylistic features.

In DiMarco's grammars, prepositions and nouns have relatively little stylistic significance. I chose to ignore such features since they appear (in most cases) to remain constant in translation or are attributable to lexical style (*e.g.,* a noun is replaced by a verb, or vice versa).

In the next section, the correspondences between the French and English stylistic features that were detected in the corpus will be presented. These correspondences are written in the form:

French stylistic feature $\mapsto$ English stylistic feature

For example, a correspondence between a conjunct^2-adjective and a conjunct^1-adjective would be written as:

conjunct^2-adjective $\mapsto$ conjunct^1-adjective

These correspondences are grouped according to kind of effect that they produce. A rule of comparative stylistics is formed from the pairing of an effect (*i.e.,* an abstract element of comparative stylistics) with a correspondence. For example:

Effect $\longrightarrow$

French stylistic feature $\mapsto$ English stylistic feature

To make the assignment of the correspondences to their groupings appear more intuitive, a minor adjustment was made to DiMarco's grammars. This adjustment serves to "normal-ize" the treatment of adjectives and premodifications in the French stylistic grammar with their counterparts in the English grammar. More specifically, the French grammar assigns a value of conjunct^2 to an ordinary adjective whereas the English grammar would assign a value of conjunct^1. A similar relationship exists between French and English premodifiers. Thus, to normalize the treatment of adjectives and premodifications in the French and English grammars, the degree of connectiveness of French adjectives and premodifications was reduced by one. In other words, French adjectives or premodifiers which would be assigned a value of conjunct^2 by DiMarco's grammar (*e.g.,* conjunct^2-adjective or conjunct^2-premodification) would appear in the correspondences with a value of conjunct^1 (*e.g.,* conjunct^1-adjective or conjunct^1-premodification). Thus, the last two correspondences in figure 5.4 would appear in the following section as:

conjunct^1-adjective $\mapsto$ conjunct^1-adjective
conjunct^2-premodification $\mapsto$ conjunct^2-premodification

These rules appear with the group for *sustained-concord* as they indicate that the degree of connectiveness should be the same (sustained) in French and English.

La réforme s'est étendue <u>au niveau secondaire en 1983 pour le premier cycle</u> (secondaire I, II) puis, en 1986, au deuxième cycle.[a]

The reform was extended <u>to the first cycle of the secondary level in 1983</u> (Secondary I, and II) and then, in 1986, to the second cycle.[b]

Figure 5.1: Sample sentence with translation.

au	niveau	secondaire	en	1983	pour	le	premier	cycle
prep	noun	conjunct^2-adjective	prep	noun	prep	conjunct^3-premod	conjunct^2-adjective	noun

Figure 5.2: Partial stylistic parse (French).

to	the	first	cycle	of	the	secondary	level	in	1983
prep	conjunct^2-premod	conjunct^1-adjective	noun	prep	conjunct^2-premod	conjunct^1-adjective	noun	prep	noun

Figure 5.3: Partial stylistic parse (English).

French Feature	English Feature	Example (French/English)
prep (preposition)	prep (preposition)	au/of en/in pour/to
noun	noun	niveau/level 1983/1983 cycle/cycle
conjunct^2-adjective	conjunct^1-adjective	secondaire/secondary premier/first
conjunct^3-premod (premodification)	conjunct^2-premod (premodification)	le/the

Figure 5.4: Correspondences of stylistic features (a summary of figures 5.2 and 5.3).

---

[a][Beauparlant 1990a, 28].
[b][Beauparlant 1990b, 28].

5.2   The Formal Rules of French–English Comparative Stylistics

In the previous section, we described the format of the formal rules of comparative stylistics. The left hand side of a rule is the *name* of the rule. The right hand side of a rule is a *correspondence*. For conciseness, rules with a common left hand side are presented together. They are written with a single left hand side and a list of alternative correspondences as the right hand side. Thus, the three rules defining augmented-concord are:

augmented-concord $\longrightarrow$

     conjunct^1-adjective $\mapsto$ conjunct^2-adjective

     conjunct^1-adjective $\mapsto$ conjunct^2-relative-clause

     conjunct^1-adverb $\mapsto$ conjunct^2-adverb

5.2.1   Concord

The first group of rules to be described are the *augmented-concords*. This name was derived from two aspects of the correspondences it describes. The "augmented" aspect of the name arises from the increase in connectiveness[2] as reflected in the correspondence (*i.e.,* the value of the superscript increases (on the conjunct scale) when translating from French to English). The "concord" aspect of the name arises from DiMarco's recognition that it is the interplay between stylistic concord and discord that is responsible for the variations in stylistic effect within sentences.

Before the rules are presented, the stylistic features that are used will be briefly described. The first feature is the *adjective.* The grading of adjectives in both English and French is based on both position and form. In English, adjectives normally precede the nominal group they modify. An adjective that follows the nominal group that it modifies is unusual and therefore considered discordant (having a disconnective effect). In French, it is usual for an adjective to follow the nominal group it modifies. However, the position of the adjective, in terms distance from the nominal group it modifies, is still important. For example, an adjective is normally adjacent to the nominal group that it modifies and is rarely separated from the nominal group by a verb (*i.e.,* in the complement). In terms of form, ordinary adjectives (in both languages) are assigned a connectiveness level of conjunct^1. The types of adjectives that are considered more connective than usual are possessive adjectives (genitives) (*e.g., <u>his</u> car*), adjectives that are themselves modified (*e.g., <u>very</u> substantial meal*) and nouns that are used as adjectives (*e.g., long <u>distance</u> rates*).

The next feature is the *adverb.* French adverbs are graded on the basis of their inherent strength and the presence of syntactic cues binding them to the object they modify. For example, an adverb like *ne ... pas*, which is closely bound with its verb, is considered more connective than an ordinary adverb like *hâtivement.* English adverbs are graded on their

---

[2]As explained in section 4.3.1, the concept of connectiveness describes how closely two syntactic constructs are bound together.

positions within the sentence (*i.e.*, beginning, middle, end) and the frequency with which they occur in that position.

In both the English and French grammars, the category of *clauses* subsumes several types of more specific classes of clauses (*e.g.*, adverbial, nonfinite, relative, and verbless). The degree of connectiveness for a general clause is the same as that for the specific type of clause. That is, a conjunct^3-adverbial-clause is also considered to be a conjunct^3-clause and a conjunct^2-relative-clause is also a conjunct^2-clause. Adverbial clauses are assigned the same connectiveness as the adverb contained in the clause (*i.e.*, a conjunct^2-adverbial-clause contains a conjunct^2-adverb).

Now we can look more closely at the rule for augmented-concord, which has three three alternatives.

augmented-concord $\longrightarrow$

conjunct^1-adjective $\mapsto$ conjunct^2-adjective

> *Le coût de l'abonnement au service téléphonique est resté stable et les tarifs des appels <u>interurbains</u> ont chuté, alors que les frais d'abonnement au câble ont augmenté.*[3]

> *While cable rates have climbed, local telephone rates have remained static and <u>long distance</u> rates have actually dropped.*[4]

conjunct^1-adjective $\mapsto$ conjunct^2-relative-clause

> *Les personnes <u>pauvres</u> du Canada devraient être les dernières à souffrir des compressions budgétaires.*[5]

> *Canadians <u>who are poor</u> should least expect to suffer further from the government's belt-tightening.*[6]

conjunct^1-adverb $\mapsto$ conjunct^3-adverb

> *Sur un mur contigu à sa maison, une inscription <u>hâtivement</u> peinte:* «Oui à notre chef Hafez El-Assad».[7]

> *One [sic] the wall adjoining her house, someone had scribbled <u>hurriedly</u>: "Yes to our leader Hafez Al-assad."*[8]

Each of the alternatives for augmented-concord describes an increase of connectiveness in translation. In the first alternative, the French adjective *interurbains* is a simple adjective with a default classification of conjunct^1. Since the corresponding English adjectival, *long distance*, is itself a nominal, it binds more closely to the noun it modifies than a simple adjective. This difference produces the increase in connectiveness between French and English adjectives that this alternative describes.

In the second alternative, the French adjective *pauvres* is also an ordinary adjective. The English relative clause *who are poor* is, like all relative clauses, classified by default as conjunct^2

---

[3][Murray 1990b].
[4][Murray 1990a].
[5][Conseil canadien de Développement social 1990b, 3].
[6][Canadian Council on Social Development 1990a, 3].
[7][Chipaux 1990b, 3].
[8][Chipaux 1990a].

since it contains a moderately connective relative pronoun. This difference produces the increase in connectiveness between French adjectives and English relative clauses.

In the third alternative, the French adverb *hâtivement* is a simple adverb with a default classification of conjunct[1]. The English adverb *hurriedly* receives a default classification of conjunct[3] due to its final position in the sentence, which has an inherently strong degree of connectivity. Thus, we see the increase in connectiveness between these French and English adverbs.

The following rule enumerates the types of *diminished-concord*. The name was derived using the same philosophy as for augmented-concord, but here the term "diminished" describes the decrease in connectiveness in the translation from French to English.

We have not previously described the feature of *postmodification*. The term, as used by DiMarco, applies to modifiers of nominal groups (*i.e.,* nouns, noun phrases, pronouns) that occur after the object they modify. English and French postmodifications include the categories: clause, nominal group, prepositional phrase, and adjective.

The rule for diminished-concord is:

diminished-concord $\longrightarrow$

     conjunct[2]-adjective $\mapsto$ conjunct[1]-adjective

        Dans sa <u>petite</u> 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.»[9]

        "My brother was losing blood, and I couldn't do anything for him," said Germaine, a woman in a <u>little</u> black dress, tearful and hollow-eyed with fatigue.[10]

     conjunct[2]-adjective $\mapsto$ nominal-group

        Il est mort à <u>son</u> arrivée à l'hôpital Jeitaoui, à Beyrouth.[11]

        By the time <u>he</u> reached the Jeitaoui Hospita [sic], he was dead.[12]

     conjunct[3]-adverb $\mapsto$ conjunct[2]-adverb

        Pour <u>y</u> arriver, nous devons opter pour des produits durables, à embalage écologique ou biodégradable, ou encore réutiliser les produits de consommation ...[13]

        Reductions can come <u>through</u> choosing products that are more durable, packaged more responsibly, or that are biodegradable. Reusing products in the home is another way to cut down on what we throw away.[14]

     conjunct[2]-postmodification $\mapsto$ conjunct[1]-postmodification

        Quand elle a essayé, en fin <u>d'après-midi</u>, de le conduire à hôpital, c'était trop tard.[15]

---

[9][Chipaux 1990b, 3].
[10][Chipaux 1990a] edited.
[11][Chipaux 1990b, 3].
[12][Chipaux 1990a].
[13][Reynolds 1990a].
[14][Reynolds 1990b].
[15][Chipaux 1990b, 3].

When she did get him to the hospital towards the end <u>of the afternoon</u>, it was too late.[16]

conjunct[2]-clause $\mapsto$ conjunct[1]-clause

Il se pourrait bien <u>que la France juge un jour Paul Touvier</u>, ancien milicien roué quoique ordinaire, tortionnaire présumé, à tout le moins gardien en chef de geôles de sinistre mémoire.[17]

France might well be able <u>to deliver its verdict on Paul Touvier</u>, a former militiaman, an ordinary though wily man, suspected torturer, and at any rate the chief custodian of some sinister prisons.[18]

conjunct[2]-clause $\mapsto$ conjunct[1]-postmodification

Résultat: les câblodistributeurs peuvent <u>ajouter de nouveaux canaux, par exemple</u>, et ce sont encore les abonnés au service de base que écouperont lorsqu'un service ne sera pas rentable.[19]

Discretionary, or non-basic services (<u>new channels, for example</u>), can be added by cable companies at will and if a service proves unprofitable, subscribers to basic service can wind up paying the bill.[20]

conjunct[2]-clause $\mapsto$ conjunct[1]-clause

«Il semble qu'elles aident à <u>éliminer les radicaux libres</u>», précise Dr John Trevithick, professeur de biochimie participant à l'étude de l'UWO.[21]

"They appear <u>to be instrumental in 'scavenging' free radicals</u>," says Dr. John Trevithick, a professor of biochemistry and one of three researchers involved in the UWO study.[22]

Each of the alternatives for diminished-concord describes a decrease of connectiveness in translation. In the first alternative, the French adjective *petite* is a characterizing adjective with a classification of conjunct[2]. Such an adjective describes an inherent property of the noun and thus binds more closely than an ordinary adjective. The English adjective *little* is a less connective, simple adjective with a default classification of conjunct[1].

In the second alternative, the French adjective *son* is a possessive adjective with a classification of conjunct[2]. By definition, a possessive adjective binds more closely to a noun than a simple adjective. The English pronoun *he* is considered a nominal group, with a neutral degree of connectivity.

In the third alternative, the French adverb *y* is almost completely assimilated by the verb; this strong integration with the verb results in a high degree of connectivity, conjunct[3]. The English adverb *through*, although moderately connective by virtue of its customary medial

---

[16][Chipaux 1990a].
[17][Greilsamer 1990a].
[18][Greilsamer 1990b].
[19][Murray 1990b].
[20][Murray 1990a].
[21][Dappen 1991a, 15–16].
[22][Dappen 1991b, 15].

position in the sentence, is not assimilated by the verb and so is less connective than the French adverb.

In the fourth alternative, the French prepositional phrase *d'après-midi* is a characterizing postmodification with a classification of conjunct[2]. Again, a characterizing construction describes an intrinsic quality of the object it modifies and thus binds more closely than a simple postmodification. In English, the prepositional phrase *of the afternoon* is considered to be only a simple postmodification, classified as conjunct[1], with no special characterizing properties.

In the fifth alternative, the French relative clause *que la France juge un jour Paul Touvier* has a default classification of conjunct[2]. The English nonfinite clause *to deliver its verdict on Paul Touvier* would be moderately connective due to the presence of the connective word, "to." However, without a subject, it lacks the solidarity and inherent connectivity of a subject-verb block. Thus, it is less connective than the French relative clause.

In the sixth alternative, the French nonfinite clause *ajouter de nouveaux canaux, par exemple* has a default classification of conjunct[2]. The English postmodification *new channels, for example* is a simple nominal group that receives a default classification of conjunct[1]. This difference in connectivity reflects the difference between the parenthetical character of the nominal group, which could be easily left out, and the essential nature of the French clause.

And finally, in the seventh alternative, the French nonfinite clause *éliminer les radicaux libres* has a default classification of conjunct[2]. The English nonfinite clause *to be instrumental in 'scavenging' free radicals* would be moderately connective due to presence of the connective word, "to." But again the lack of a subject makes it less connective than the French clause.

The next rule defines another abstract element, *sustained-concord*. The name was derived using the same philosophy as for augmented-concord. The "sustained" component of the name describes the preservation of connectiveness across the translation from French to English.

We have not previously described the feature of *premodification*. Like postmodifications, premodifications apply only to modifiers of nominal groups that occur before the object they modify. English and French premodifications include the categories: participle, adverb, adjective, and determiner.

The rule for sustained-concord is:

sustained-concord $\longrightarrow$

    conjunct[1]-adjective $\mapsto$ conjunct[1]-adjective

        Dans sa petite robe <u>noire</u>, les yeux cernés par les larmes et la fatigue, Germaine raconte: «Mon frère perdait son sang, et je n'ai rien pu faire.»[23]

        "My brother was losing blood, and I couldn't do anything for him," said Germaine, a woman in a little <u>black</u> dress, tearful and hollow-eyed with fatigue.[24]

    conjunct[2]-adjective $\mapsto$ conjunct[2]-adjective

        Sa candidature fut rejetée parce que <u>son</u> français n'était pas apparu satisfaissant à l'intervieweur.[25]

---

[23][Chipaux 1990b, 3].
[24][Chipaux 1990a] edited.
[25][Duhaime 1990b, 24] edited for length.

His candidacy was rejected because, in the interviewer's opinion, <u>his</u> French was not adequate.[26]

conjunct²-adverb ↦ conjunct²-adverb

Sa candidature fut rejetée parce que son français <u>n'était</u> <u>pas</u> apparu satisfaissant à l'intervieweur.[27]

His candidacy was rejected because, in the interviewer's opinion, his French was <u>not</u> adequate.[28]

conjunct³-adverb ↦ conjunct³-adverb

Dans le centre du Québec, M. Normand Dubé, directeur général de la Commission scolaire de Trois-Rivières, souligne que «la région n'est pas en contact avec l'anglais» et qu'on <u>y</u> forme que «difficilement» des bilingues.[29]

As for central Quebec, Normand Dubé, director general of the Trois-Rivières school board, notes that "the region is not in contact with English," and that it is difficult to produce bilingual students <u>there</u>.[30]

conjunct²-clause ↦ conjunct²-postmodification

Par exemple, comment justifier des véhicules polluants que ne servent la plupart du temps qu'à <u>transporter leur conducteur</u> sur un réseau urbain vieillissant et de plus en plus congestionné?[31]

Is it really sensible for most of us to drive air-polluting private vehicles, most <u>carrying just the driver</u>, around decaying and increasingly clogged urban roads?[32]

conjunct²-premodification ↦ conjunct²-premodification

Sur <u>un</u> mur contigu à sa maison, une inscription hâtivement peinte: «Oui à notre chef Hafez El-Assad».[33]

One [sic] <u>the</u> wall adjoining her house, someone had scribbled hurriedly: "Yes to our leader Hafez Al-assad."[34]

conjunct²-premodification ↦ conjunct²-adverb

Ainsi la justice démontre-t-elle, <u>une nouvelle fois</u>, sa dépendance, tandis que le chef de l'Etat perd l'occasion, lui si légitimement prompt à s'expliquer sur sa politique étrangère, d'expliquer aux Français les raisons qu'il a de leur épargner le procès de Vichy.[35]

---

[26][Duhaime 1990a, 24] edited for length.
[27][Duhaime 1990b, 24] edited for length.
[28][Duhaime 1990a, 24] edited for length.
[29][Beauparlant 1990a, 28].
[30][Beauparlant 1990b, 28].
[31][Reynolds 1990a].
[32][Reynolds 1990b].
[33][Chipaux 1990b, 3].
[34][Chipaux 1990a].
[35][Greilsamer 1990a].

> <u>Once again</u> *French justice has demonstrated its lack of independence, while the Head of State who is so prompt—and rightfully so—to explain his foreign policy, has lost the opportunity to tell the French public why he thinks they should be spared the trial of Vichy.*[36]

Each of the alternatives of sustained-concord describes the preservation of the same degree of connectiveness in translation. In the first alternative, the French adjective *noire* is a simple adjective will a default classification of conjunct[1]. The corresponding English adjective *black* is also a simple adjective receiving a default classification of conjunct[1]. This similarity maintains the same connectiveness between these French and English adjectives.

In the second alternative, the French adjective *son* is a possessive adjective which is more connective than a simple adjective and thus receives a classification of conjunct[2]. The English possessive adjective *his* receives a classification of conjunct[2] for the same reason.

In the third alternative, the French adverb *ne … pas* has a classification of conjunct[2] due to its close association with the verb. The English adverb *not* receives a default classification of conjunct[2] due to its customary medial position in the sentence. In both cases, the adverbs have the same moderate degree of connectivity.

In the fourth alternative, the French adverb *y* is assimilated by the verb. This strong integration with the verb results in a classification of conjunct[3]. The English adverb *there* occurs in the sentence-final position, which has an inherently strong connectivity of conjunct[3]. In both cases, the adverbs have the same strong degree of connectivity.

In the fifth alternative, the French nonfinite clause *transporter leur conducteur* receives a default classification of conjunct[2]. The English adverbial clause *carrying just the driver* is also classified as conjunct[2], in this case, because it occurs in a moderately connective, postmodifying, position.

In the sixth alternative, the French article *un* receives a default classification of conjunct[3] as does the English article *the*. In both cases the article binds strongly to the noun it modifies.

And finally, in the seventh alternative, the French premodification *une nouvelle* has a higher degree of connectiveness, conjunct[2], than a ordinary premodification due to its strong link to the noun *fois*: it forms part of a unit whose figurative meaning takes precedence. The English adverb *once* is classified as conjunct[2] due to its sentence-initial position. The initial position sets up the basis on which the rest of the sentence will be built; thus the initial position is moderately connective.

5.2.2   Discord

Now, we will describe the dual of concord, the various types of *discord*. The discords appear to be duals to the concords (*i.e.,* the preceding three rules) because an increase in connectiveness (concord) implies a decrease in disconnectiveness (discord), and vice versa. To avoid duplicate rules, only those rules that explicitly contain discords will be included in the *augmented-discords, diminished-discords* and *sustained-discords*.

---

[36][Greilsamer 1990b].

Before continuing the presentation of rules, two cosmetic anomalies must be explained. The first anomaly is the lack of examples for some rules. From the form of the rules and correspondences that we observed in the corpus, we were able to predict the probable existence of additional rules and correspondences. The rules that were observed are presented with examples from the corpus, while those that were predicted do not have an associated example.

The second anomaly is the use of the meta-stylistic feature *imf-discord*. It is used as an abbreviation for the situation where any one of initial-, medial- or final-discord can appear, but only one at a time. That is:

> imf-discord ↦ initial-discord
>
> imf-discord ↦ medial-discord
>
> imf-discord ↦ final-discord

Two stylistic features that figure prominently in the rules for augmented-discords are *centroschematic* and *monoschematic*. The centroschematic feature is more abstract than the features that we have used up to now. It describes entire sentences rather than individual components. As we defined in section 3.5.2, a centroschematic sentence consists of a central independent clause and one or more dependent clauses. One other important characteristic is that all the components of the sentence are concordant.

Similarly, the monoschematic feature is used to describe entire sentences. A monoschematic sentence can be thought of as the special case of centroschematic sentences where there is only a single, simple, independent clause and no dependent clauses.

The rules describing augmented-discord are:

augmented-discord ⟶

> centroschematic ↦ medial-discord
>
> > *Le compromis proposé aurait permis <u>aux professeurs et aux étudiants de jouer des pièces en classe</u>, sans auditorie.*[37]
> >
> > *The compromise would have allowed <u>for the performance of plays in class by students and teachers</u> without any audience.*[38]
>
> conjunct[1]-adjective ↦ antijunct[1]-adjective
>
> > *Le changement d'attitude adopté par le gouvernement à l'égard de cette disposition logique nous semble tout simplement <u>insensé</u>.*[39]
> >
> > *A change of heart by the government with respect to such common sense provisions seems to us <u>inane</u>.*[40]
>
> monoschematic ↦ imf-discord
>
> centroschematic ↦ imf-discord

---

[37][Kerr 1990b].
[38][Kerr 1990a].
[39][Kerr 1990b].
[40][Kerr 1990a].

The alternatives expressing augmented-discord are intended to describe the situations where concordant components become discordant in translation. In the first alternative, the overall concordance of the French centroschematic sentence is disrupted in the English sentence by a medial discord, caused by the excessive imitation of prepositional phrases. DiMarco bases her definition of discord on three sources: excessive imitation (*i.e.,* excessive structure), minimal structure, and the frequency of usage of a construction.

In the second alternative, the French postmodifying adjective *insensé* corresponds to the English adjective *inane*. In English, unlike French, a postposed adjective is unusual and therefore connective, so that it is classified as antijunct and discordant.

The final two alternatives predict cases where the structure of a concordant sentence, whether simple or complex, is rearranged in translation to place a component in an unusual, discordant, position.

The following rule describes the dual of augmented-discord, *diminished-discord*. The alternatives expressing diminished-discord are intended to describe the situations where discordant components become concordant in translation. As we did earlier, we predict two rules based on the rearrangement of sentence components. Now, discordant components will be removed or transformed into a more connective, concordant, organization.

diminished-discord $\longrightarrow$

    imf-discord $\mapsto$ monoschematic

    imf-discord $\mapsto$ centroschematic

The following rule defining *sustained-discord* predicts situations where discords are maintained across translation. Sustained-discords include the situations where the discords are in the same position (*e.g.,* initial-discord $\mapsto$ initial-discord) or in different positions (*e.g.,* initial-discord $\mapsto$ final-discord).

sustained-discord $\longrightarrow$

    imf-discord $\mapsto$ imf-discord

5.2.3   Heteropoise

The next series of rules describes the various kinds of *heteropoises* used in comparative stylistics. Intuitively, a heteropoise is an interruption in the flow of a sentence. For example, in *Politicians, being rather eloquent, capture our interest*[41] the heteropoise is the interrupting participle clause, *being rather eloquent*.

The first two groups of heteropoises, the *augmented-heteropoises* and the *diminished-heteropoises*, describe the introduction and removal, respectively, of heteropoises in translation.

augmented-heteropoise $\longrightarrow$

---

[41][DiMarco 1990, English parse 43].

centroschematic ↦ initial-heteropoise

L'instabilité régnera jusqu'à la mise en place de nouveaux arrangements.[42]

Until a future arrangement is understood, there will be instability.[43]

centroschematic ↦ medial-heteropoise

centroschematic ↦ final-heteropoise

Le Macvin est une autre des spécialités du Château d'Arlay 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.[44]

Macvin is another of Château d'Arlay's specialties, made by blending 30% of mature marc with freshly pressed grape must, followed by 6 to 12 months of barrel aging.[45]

There are three alternatives to augmented-heteropoise. The first describes the situation where a canonical, concordant sentence arrangement is replaced by one with an initial interruption. In an analogous manner, the other two alternatives describe the transformation of concordant sentences into forms with medial and final interruptions.

The dual of augmented-heteropoise, *diminished-heteropoise*, describes situations in which sentences with interruptions are smoothed out into stable centroschematic sentences.

diminished-heteropoise ⟶

initial-heteropoise ↦ centroschematic

Assurément, nous abordons cette décennie mieux prévenus que jamais; nous recherchons des produits «écologiques» et nous participons aux programmes de recyclage.[46]

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.[47]

medial-heteropoise ↦ centroschematic

Quand elle a essayé, en fin d'après-midi, de le conduire à hôpital, c'était trop tard.[48]

When she did get him to the hospital towards the end of the afternoon, it was too late.[49]

final-heteropoise ↦ centroschematic

Le compromis proposé aurait permis aux professeurs et aux étudiants de jouer des pièces en classe, sans auditorie.[50]

---

[42][Unknown 1991a, 71].
[43][Unknown 1991b, 70].
[44][Chatto 1990a, 30].
[45][Chatto 1990b, 30].
[46][Reynolds 1990a].
[47][Reynolds 1990b].
[48][Chipaux 1990b, 3].
[49][Chipaux 1990a].
[50][Kerr 1990b].

> *The compromise would have allowed for the performance of plays in class by students and teachers <u>without any audience</u>.*[51]

There are three alternatives to diminished-heteropoise. In the first alternative, the initial interrupting, parenthetical, construction *assurément* is translated by *certainement*, which is moved into an interior, more syntactically integrated position. In the second alternative, 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 interrupting 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-heteropoise*. The first group describes the case where the position of the heteropoise remains the same. For example, in the second alternative, the medial heteropoise, *par exemple*, is sustained in the English translation by the interrupting *for example.*

The second group describes the case where the heteropoise moves to the beginning of the sentence. For example, in the first alternative, the medial heteropoise, *quant à moi* becomes the initial phrase *but for me* in the English translation.

The last two groups describes the cases where the heteropoise moves, respectively, either to the middle or end of the sentence.

non-shifted-heteropoise   ⟶

     initial-heteropoise ↦ initial-heteropoise

     medial-heteropoise ↦ medial-heteropoise

> *Au Languedoc, <u>par exemple</u>, la chaleur du soleil du sud transparaît dans le savoureux marc de muscat, presque liquoreux tant la saveur du raisin s'y fait présente.*[52]

> *In Languedoc, <u>for example</u>, 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; while in nearby Lirac, in the Côtes du Rhône, Jean-Claude Assémat makes a superb marc de syrah.*[53]

     final-heteropoise ↦ final-heteropoise

front-shifted-heteropoise   ⟶

     medial-heteropoise ↦ initial-heteropoise

> *Je préfère, <u>quant à moi</u>, les produits moins connus des maisons moins grandes.*[54]

> *<u>But for me</u>, the fascination lies in the less familiar products of smaller houses.*[55]

---

[51][Kerr 1990a].
[52][Chatto 1990a, 30].
[53][Chatto 1990b, 30].
[54][Chatto 1990a, 30].
[55][Chatto 1990b, 30].

final-heteropoise ↦ initial-heteropoise

medial-shifted-heteropoise ⟶

initial-heteropoise ↦ medial-heteropoise

> *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.»[56]

> *"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.*[57]

final-heteropoise ↦ medial-heteropoise

end-shifted-heteropoise ⟶

initial-heteropoise ↦ final-heteropoise

medial-heteropoise ↦ final-heteropoise

5.2.4   Resolution

As we saw in section 3.5.2, a *resolution* is a shift from a stylistic discord to a relative concord at the end of a sentence. For example, in the following sentence, an initial discord is produced by the unusual use of the adverb *entirely* at the beginning of a sentence. The remainder of the sentence is usual and concordant, producing a relative concord.

(5-1)   Entirely in the spirit of protective support, could I suggest you pass on an appropriate comment to the personnel concerned[58].

The following three groups of rules express the creation, destruction, and preservation of resolutions. An *augmented-resolution* describes a situation where various types of sentences are all transformed into the strict sequence of a stylistic discord followed by a concord. For example, in the second alterative the French sentence contains a discordant medial interruption, *il y a quelques années*. But in the English translation the corresponding phrase, *some years ago*, consolidates the discord at the beginning of the sentence, so that the rest of the main clause is concordant and a resolution is formed.

augmented-resolution ⟶

initial-heteropoise ↦ resolution

medial-heteropoise ↦ resolution

> *Un historien contemporain écrivait, il y a quelques années, que l'histoire sociale de la France au XIXe siècle était encore 'terra incognita'.*[59]

---

[56][Chipaux 1990b, 3].
[57][Chipaux 1990a] edited.
[58][Quirk *et al.* 1985, 652] after [DiMarco 1990].
[59][Dupeux 1988b].

> *Some years ago* it was remarked by a contemporary historian that the social history of France in the nineteenth century was still uncharted territory.[60]

final-heteropoise ↦ resolution

dissolution ↦ resolution

centroschematic ↦ resolution

The next rule for *diminished-resolution* predicts cases in which the strict sequence of a resolution is "scrambled" in translation into various types of sentence structure.

diminished-resolution ⟶

resolution ↦ initial-heteropoise

resolution ↦ medial-heteropoise

resolution ↦ final-heteropoise

resolution ↦ dissolution

resolution ↦ centroschematic

And finally, a resolution may be preserved in translation.

sustained-resolution ⟶

resolution ↦ resolution

5.2.5 Dissolution

The counterpart of resolution, a *dissolution* is a shift from a stylistic concord to a relative discord at the end of a sentence. For example, in the following sentence, the discord at the end of the sentence is produced by the final inverted clause, a deviation from the parallel structure of the first two clauses of the sentence.

(5-2) And the rains descended and the house fell and great was the fall of it[61].

The final three groups of rules express the creation, destruction, and preservation of dissolutions.

augmented-dissolution ⟶

initial-heteropoise ↦ dissolution

medial-heteropoise ↦ dissolution

final-heteropoise ↦ dissolution

resolution ↦ dissolution

---

[60][Dupeux 1988a].
[61][DiMarco 1990, English parse 35] edited from Matthew 7:27 (Revised Standard Version [1952]).

centroschematic ⟼ dissolution

diminished-dissolution ⟶

dissolution ⟼ initial-heteropoise

dissolution ⟼ medial-heteropoise

dissolution ⟼ final-heteropoise

dissolution ⟼ resolution

dissolution ⟼ centroschematic

sustained-dissolution ⟶

dissolution ⟼ dissolution

5.3   Applying the Grammar to Machine Translation

Now that we have developed the rules which collectively form a grammar of French–English comparative stylistics, one major question remains: how would the rules be used in a machine translation (MT) system? The first step is for the MT system to select candidate rules from the complete set of rules. This is a straightforward application of forward-chaining. In other words, whenever the left hand side (a feature) of a correspondence appears in the stylistic parse, all rules containing that feature are candidate rules to be used in the translation.

The following is an example of a stylistic parse and the resulting set of candidate rules. The first section of the example shows the sentence that was parsed by STYLISTIQUE. The sentence that is shown contains four markers (*i.e.,* complement, cmais, cpvirgule, and clause) called *pseudo-cuts*, which are explicit markers of punctuation used as an aid in disambiguation. Pseudo-cuts are used purely for efficiency and have no theoretical value. They are discussed more fully in section 7.2.2 of [DiMarco 1990]. The next three sections of the example, show the stylistic goals, abstract elements, and primitive elements that describe the syntactic style of the sentence. The last section of the example shows the resulting set of candidate rules.

```
_____
ils ont recu les resultats complement par le courrier cmais sans aucune informa
tion sur les dangers pour la sante cpvirgule clause meme lorsque les niveaux de
contamination etaient inacceptables
_____
Stylistic goals of this sentence:
(clarity/obscurity, abstraction/concreteness, staticness/dynamism)
_____
[clarity,neutral,neutral]

Abstract stylistic elements (Connective view):
_____
[[initial_and_medial_concord,initial_concord],[centroschematic],[]]
```

Connective stylistic primitive elements:
——
[adj,adjectival_phrase,adv,adverbial_clause,c_adverbial_phrase,c_clause,c_comple
ment,c_complete,c_major,c_nominal_group,c_noun_phrase,c_postmodification,c_premo
dification,c_prep_phrase,c_sentence,c_verb_phrase,cl,clauses,comp,complete,conju
nct0_adverb,conjunct0_clause,conjunct1_noun,conjunct1_subject,conjunct2_adjectiv
e,conjunct2_adverb,conjunct2_prep,conjunct3_adjective,conjunct3_determiner,conju
nct3_direct_object,copula,det,n,ngroup,postmod,pour,predet,premod,prep,prepositi
on,pronoun,verb,verbmods,xcopula,xverb]

Applicable Rules of Comparative Stylistics:
——
centroschematic ——> imf_discord (augmented_discord)
centroschematic ——> medial_discord (augmented_discord)
centroschematic ——> initial_heteropoise (augmented_heteropoise)
centroschematic ——> medial_heteropoise (augmented_heteropoise)
centroschematic ——> final_heteropoise (augmented_heteropoise)
centroschematic ——> resolution (augmented_resolution)
centroschematic ——> dissolution (augmented_dissolution)
conjunct2_adjective ——> conjunct1_adjective (diminished_concord)
conjunct2_adjective ——> nominal_group (diminished_concord)
conjunct2_adjective ——> conjunct2_adjective (sustained_concord)
conjunct2_adverb ——> conjunct2_adverb (sustained_concord)
conjunct2_postmodification ——> conjunct1_postmodification (diminished_concord)
conjunct2_premodification ——> conjunct2_premodification (sustained_concord)
conjunct2_premodification ——> conjunct2_adverb (sustained_concord)

Due to the coarseness of DiMarco's grammar and therefore my rules of comparative stylistics, there may be a large number of candidate rules (as the preceding example demonstrates). When DiMarco's grammars are refined, it should be possible to reduce the size of the set of candidate rules by considering the connectiveness of the sentence as a whole (*e.g.,* high, average, or low connectiveness). This application of connectiveness was suggested by Guillemin-Flescher's ([1981]) use of the strengths of syntactic relationships between the syntactic components of a sentence in her description of the rules of theoretical comparative stylistics that were described in section 4.3 of this thesis. These strengths of syntactic relationships correspond to the concept of connectiveness that is used in the formal rules. Thus, we can use the concept of connectivity, which is the basis of the stylistic grammar, to choose the most appropriate rules from the candidate set. As a simple example, the French phrase *une nouvelle fois* can be translated into English as *once again*[62]. Although each of the words in the French phrase have an ordinary degree of connectiveness, the entire phrase has a higher degree of connectiveness relative to its English equivalent. This relative increase in connectiveness is due to the use of a non-literal meaning of *nouvelle* (*i.e.,* the influence of semantic aspects of style). In its literal

---

[62]The complete sentences can be found on page 51 as an example for one of the rules expressing sustained-concord.

sense, *nouvelle* means *new*, but, in this case, it is used to imply that the action has happened before. This additional information could be used to eliminate the candidate rules that do not normally apply to that level of connectiveness. This would involve developing an algorithm to determine the connectiveness of a sentence and re-classifying the rules to take advantage of the new information.

Once the set of candidate rules has been selected, the MT system must use a planner to decide which rule, if any, to apply. The planner would choose a rule from the candidate set based on information from various sources (*e.g.,* English and French contexts (both within the text and also the world at large), lexical style, syntactic style, and semantic style) and their interactions. A statistical approach that might be used by the planner is to construct an n-dimensional contingency table (for n sources of information) from a corpus. A log-linear model[63] can be used to to analyze the statistical significance of each source of information and their combinations for a particular candidate rule. This analysis can be pre-computed using a statistical analysis package such as SAS[64] (using "PROC CATMOD") and would provide information about the relative importance of the various sources of information and when they are most useful. In this manner, a machine translation system could effectively apply our grammar of comparative stylistics to select the most appropriate rule to use in a particular translation.

5.4   The Story So Far

With the development of the formal rules of French–English comparative stylistics, we have fulfilled the major aim of this thesis. The chapter began with a description of the process through which the formal rules were derived. This process made use of the theoretical rules and corpus described in chapter 4, and the formal stylistic grammars of DiMarco described in chapter 3. After the rules were presented, a description of how the formal rules might be incorporated into a machine translation system was described.

In the next chapter, I will present a summary of this thesis that will include the contributions of the research and some directions for future work.

---

[63]See [Dobson 1990] for an introduction to log-linear models. The seminal reference for log-linear models is [Bishop *et al.* 1975].

[64]SAS is a registered trademark of SAS Institute Inc.

Chapter 6

Conclusion: That's all folks

> *Please watch out for each other and love and forgive everybody. It's a good life, enjoy it.*
> — *Jim Henson*

6.1 Contributions

The aim of this thesis was to develop a set of formal rules of syntactic French–English comparative stylistics. These rules were developed as part of a computational model of comparative stylistics for improving machine-translated output. It is anticipated that this model will be incorporated into a machine translation (MT) system as part of a 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 grammars. The correspondences between the stylistic properties of the French and English sentences, as identified by DiMarco's stylistic grammars, 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.

While constructing the corpus, I examined a fair number of French political advocacy texts and their English translations. I was able to observe many variations in lexical, syntactic, and semantic style. I found that many of the structural differences between the translated texts can be attributed to the amount of description and clarification, and the arrangement of the propositions that compose a sentence. From these observations, I propose that Ryan's [1989] definition of semantic style as focus be extended to include an account of argument structure.

6.2 Limitations and Other Work

There are three major limitations in this thesis with respect to the development of a computational model of comparative stylistics.

The first limitation of this work is the coarseness of the formal rules of comparative stylistics. This weakness is due to the inherited asymmetry and coarseness of DiMarco's stylistic grammars. The asymmetry of the French and English grammars made it awkward to compare the concrete elements describing style; instead, more abstract elements had to be used. For example, DiMarco's English grammar allows for explicit description of the various clause types (*e.g.,* conjunct^n-nonfinite-clauses, and conjunct^n-relative-clauses) where the French grammar allows only an abstract description (*e.g.,* conjunct^n-clause).

DiMarco's theory of style is currently being refined and expanded at the University of Waterloo by DiMarco and her students. Pat Hoyt and Steve Green are currently refining the grammars and removing the asymmetries by evolving the English stylistic grammar using the same functional approach that influenced the French grammar.

Marzena Makuta-Giluk [1991] and Cameron Shelley are expanding the theory to account for the additional phenomena of rhetorical goals and speech acts, respectively. And finally, Nadia BenHassine [forthcoming] is developing a generation system capable of making use of stylistic information.

When DiMarco's grammars have been sufficiently refined to reduce the coarseness and remove the asymmetries, the process of deriving the formal rules of comparative style will have to be repeated in order to determine the new correspondences. The abstract elements (*e.g.,* augmented-, diminished-, and sustained-concord) will remain the same. When the rules are re-derived, a more extensive corpus should be used in order to confirm the existence of the predicted rules.

The second limitation is the restriction of this work to syntactic style. Rules of lexical and semantic comparative stylistics must be developed to complete the model of comparative style. After all three complementary sets of rules are developed, it should be possible to clearly define the interactions of the three branches of style.

The third limitation is the lack of an implemented planner for selecting rules from a set of candidate rules. Although the planner, as described previously in section 5.3, can be developed now, its development has been delayed until rules of lexical and semantic comparative style have been developed. At that time, the planner will have access to a more complete stylistic representation of a sentence and be truly capable of choosing the appropriate rules.

6.3   Conclusion

As mentioned earlier, it is expected that a complete model of comparative style will be developed and integrated into a MT system. The development of the model of comparative stylistics should be sufficiently complete as the limitation described in the preceding section are resolved.

The integration of our model of comparative stylistics in a MT system with the same architecture as KBMT-89 (see figure 6.1 and section 2.3) should be straightforward. The integration will required the addition of a stylistic analyzer and a set of stylistic mapping rules (*i.e.,* the formal rules of comparative stylistics) to the analyzer of the MT system. The automatic phase of the augmentor would be augmented by a stylistic planner like that described in section 5.3
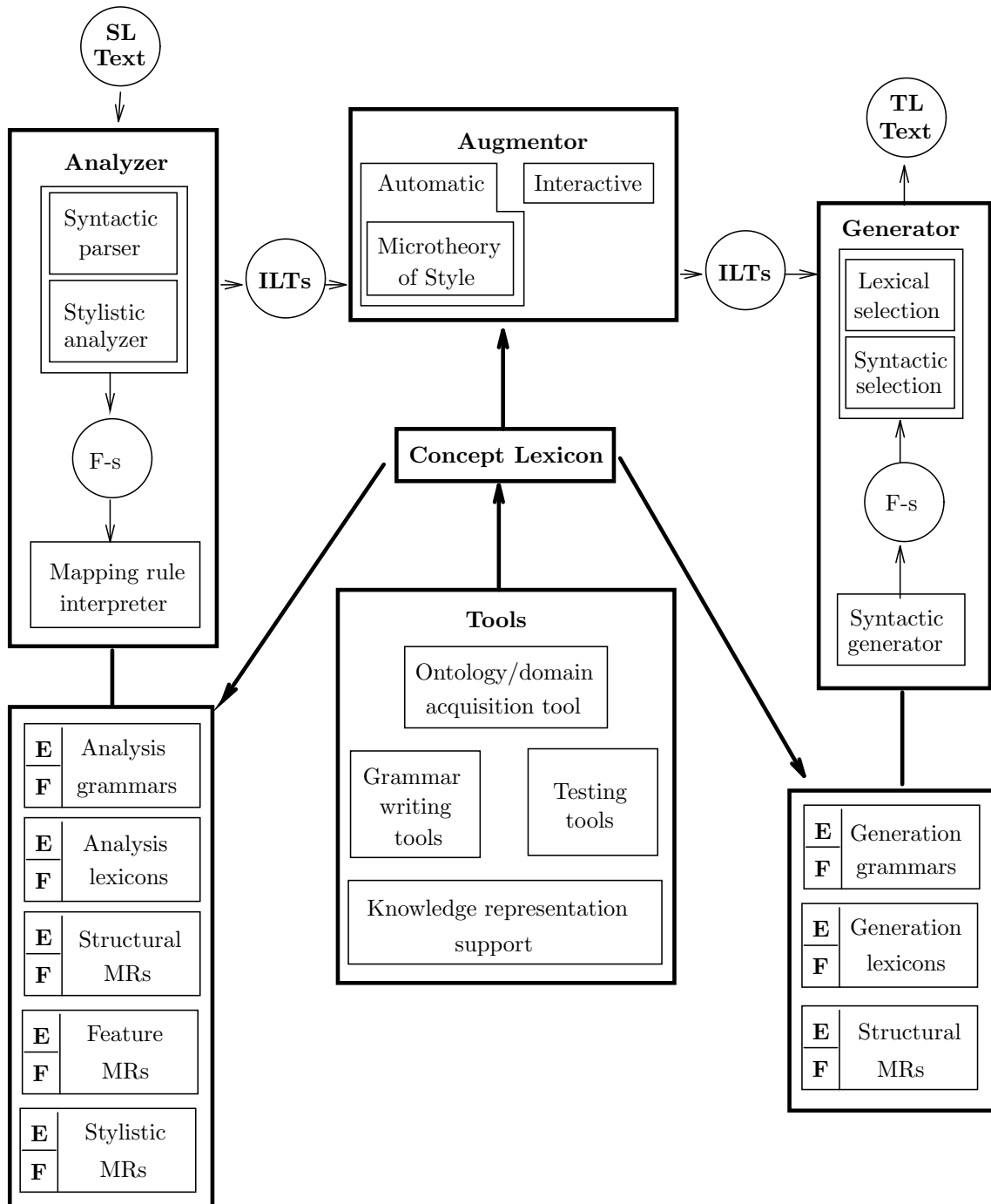
Figure 6.1: System diagram for a machine translation system that integrates style.

as part of a microtheory of style. Lastly, the generation stage would be added to as necessary ([BenHassine forthcoming] will discuss the issues involved in adding an understanding of style to a generation system).

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

> *Every now and then, go away, take a little relaxation, because when you come back to your work, your judgement will be surer. To remain constantly at work will cause you to lose power of judgement. Go some distance away, because then the work appears smaller and more of it can be taken in at a glance, and a lack of harmony of proportion is more readily seen.*
>
> *— Leonardo da Vinci*

Appendix A

The Corpus

[Chipaux 1990b; Chipaux 1990a]

Les rues sont quasi désertes, les visages sont fermés: les habitants du petit village de Bsouss, non loin de Baabda — dans la banlieue chrétienne de Beyrouth — ont peur, très peur.

The streets were almost empty and people's expressions inscrutable. The inhabitants of the small Christian village of Bsouss, not far from Ba'abda — in Beirut's Christian suburb — were scared, very scared.

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-Damas.

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.

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.*»

"My brother was losing blood, and I couldn't do anything for him," said Germaine, a woman in a black dress, tearful and hollow-eyed with fatigue.

Quand elle a essayé, en fin d'après-midi, de le conduire à l'hôpital, c'était trop tard.

When she did get him to the hospital towards the end of the afternoon, it was too late.

Celui du Sacré-Cœur, près de Baabda, touché par les bombardements du matin, débordait de blessés et n'a pu l'accueillir.

The Sacred Heart Hospital, near Ba'abda, had been hit by shells; it was overflowing with casualties and could not take him.

Il est mort à son arrivée à l'hôpital Jeitaoui, à Beyrouth.

By the time he reached the Jeitaoui Hospita [sic], he was dead.

Germaine a perdu dans ce massacre ses trois frères et trois de ses neveux de dix-neuf, dix-sept et treize ans.

Germaine lost her three brothers in the massacre along with three nephews aged 19, 17, and 13 years.

Sur un mur contigu à sa maison, une inscription hâtivement peinte: «*Oui à notre chef Hafez El-Assad*».

One [sic] the wall adjoining her house, someone had scribbled hurriedly: "Yes to our leader Hafez Al-assad."

La Croix-Rouge libanaise n'est arrivée ici que dimanche après-midi et a aidé à l'enterrement des victimes dans une fosse commue.

The Lebanese Red Cross arrived here on Sunday afternoon and helped to buried [sic] the victims in a communal grave.

[Greilsamer 1990a; Greilsamer 1990b]

Il se pourrait bien que la France juge un jour Paul Touvier, ancien milicien roué quoique ordinaire, tortionnaire présumé, à tout le moins gardien en chef de geôles de sinistre mémoire.

France might well be able to deliver its verdict on Paul Touvier, a former militiaman, an ordinary though wily man, suspected torturer, and at any rate the chief custodian of some sinister prisons.

Il apparaît en revanche de plus en plus improbable que des juges aient à instruire les crimes imputés à René Bousquet, ancien grand commis de l'Etat, haut fonctionnaire de la collaboration administrative.

On the other hand, it appears increasingly improbable that the crimes alleged to have been committed by René Bousquet, a former mandarin and a leading figure in the Vichy administration's collaboration with the occupying German power will ever be examined by judges.

Autant dire que le dossier Bousquet est promis à de très longues discussions devant la Cour des cassation.

This means the Bousquet case is bound to become the subject of very protracted discussions in the Court of Cassation.

Autant rappeler ce que révélait *le Monde* du 26 septembre: le president de la République, en conscience, considère qu'il serait dommageable pour la paix civile que René Bousquet soit jugé.

One may just as well point out what Le Monde has already revealed: President François Mitterand believes in all conscience that putting René Bousquet on trial would be damaging to civil peace.

Tout le reste n'est qu'habillage juridique conçu à la chancellerie et relayé par le procureur général.

All the rest is legal dressing decided at the Justice Ministry and executed by the Public Prosecutor.

Ainsi la justice démontre-t-elle, une nouvelle fois, sa dépendance, tandis que le chef de l'Etat perd l'occasion, lui si légitimement prompt à s'expliquer sur sa politique étrangère, d'expliquer aux Français les raisons qu'il a de leur épargner le procès de Vichy.

Once again French justice has demonstrated its lack of independence, while the Head of State who is so prompt — and rightfully so — to explain his foreign policy, has lost the opportunity to tell the French public why he thinks they should be spared the trial of Vichy.

[Reynolds 1990a; Reynolds 1990b]

L'environnement est censé être *la* question des années 1990.

The environment, goes the popular wisdom, will be the issue of the 1990s.

Assurément, nous abordons cette décennie mieux prévenus que jamais; nous recherchons des produits «écologiques» et nous participons aux programmes de recyclage.

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.

Et pourtant, nos choix sont-ils judicieux?

Yet how much do these choices really help?

Le recyclage ne suffit pas, estime Ruth Lotzkar, porte-parole de l'ACC sur l'environnement: «Il faut absolument réduire la quantité de déchets que nous produisons.»

"We cannot recycle our way out of the garbage crisis," says Ruth Lotzkar, a CAC spokesperson on the environment. "The real key lies in reducing the amount of waste we generate in the first place."

Pour y arriver, nous devons opter pour des produits durables, à embalage écologique ou biodégradable, ou encore réutiliser les produits de consommation . . .

Reductions can come through choosing products that are more durable, packaged more responsibly, or that are biodegradable. Reusing products in the home is another way to cut down on what we throw away.

Mais il reste que le moyen le plus efficace, c'est sans doute de consommer moins.

And, perhaps most effectively, we can generate less trash by consuming less, period.

Par exemple, comment justifier des véhicules polluants que ne servent la plupart du temps qu'à transporter leur conducteur sur un réseau urbain vieillissant et de plus en plus congestionné?

Like our cars. Is it really sensible for most of us to drive air-polluting private vehicles, most carrying just the driver, around decaying and increasingly clogged urban roads?

Combien de temps continuerons-nous d'«électrifier nos vies», en réclamant toujours plus d'électricité pour la construction de nouvelles usines polluantes?

And our power consumption. How long can we further electrify our lives and thereby place demands on power utilities to build new plants that pose potential environmental threats?

[Murray 1990b; Murray 1990a]

D'après l'ACC, il y a un rapport entre l'expansion du secteur de la câblodistribution et la «surexploitation» de ses clients traditionnels.

Is there a connection between recent corporate expansion in the cable industry and the fleecing of its traditional customers?

Elle fera bientôt valoir aux audiences publiques du Conseil de radiodiffusion et des télécommunications canadiennes (CRTC) que RCI et d'autres grandes entreprises de câblodistribution se sont engagées dans des activités coûteuses ne relevant pas de la surveillance du CRTC, et que ce sont les abonnés qui en feront les frais.

CAC believes there is. At the public hearings of the Canadian Radio-television and Telecommunications Commission (CRTC) taking place now, CAC will argue that RCI and other large cable companies have been dabbling in costly new activities that fall outside of CRTC's regulatory control, and that cable subscribers have been footing the bill.

Leur rendement (la mesure de rentabilité d'une entreprise) a été de 32,8 p. 100 en 1988, contre 15 p. 100 pour Bell Canada et British Columbia Telephone, qui sont aussi régies par le CRTC en raison de leur monopole.

Return on capital — the measure of a company's profitability — was 32.8 per cent for the cable industry in 1988. That compares with 15 per cent in the same period for Bell Canada and British Columbia Telephone, whose monopoly status also brings them under the jurisdiction of the CRTC.

Le coût de l'abonnement au service téléphonique est resté stable et les tarifs des appels interurbains ont chuté, alors que les frais d'abonnement au câble ont augmenté.

While cable rates have climbed, local telephone rates have remained static and long distance rates have actually dropped.

Le CRTC a aggravé le problème en n'essayant même pas de réglementer les services discrétionnaires.

Having lost control of basic service, the CRTC compounded the problem by making no attempt to regulate discretionary and competitive services.

Résultat: les câblodistributeurs peuvent ajouter de nouveaux canaux, par exemple, et ce sont encore les abonnés au service de base que écouperont lorsqu'un service ne sera pas rentable.

Discretionary, or non-basic services (new-channels, for example), can be added by cable companies at will and if a service proves unprofitable, subscribers to basic service can wind up paying the bill.

[Dupeux 1988b; Dupeux 1988a]

Un historien contemporain écrivait, il y a quelques années, que l'histoire sociale de la France au XIX$^{\text{e}}$ siècle était encore 'terra incognita'.

Some years ago it was remarked by a contemporary historian that the social history of France in the nineteenth century was still uncharted territory.

Un jugement aussi sévère ne pourrait plus être porté aujourd'hui.

Nobody could make such a drastic statement today.

La publication récente de thèses de doctorat tout orientées vers l'historie sociale et d'excellents ouvrages de synthèse ont grandement amélioré nos connaissances, au moins pour certaines régions françaises et pour une partie du XIX$^{\text{e}}$ siècle.

In recent years the publication of doctoral theses devoted to social history and some excellent works of synthesis have enormously increased our knowledge of several regions of France, and of much of the nineteenth century.

Il n'est pas moins vrai que des lacunes considérables subsistent, et qu'il reste beaucoup à faire.

But there are still big gaps and plenty of scope for further research.

Les raisons de ce retard sont diverses.

There are various reasons for this.

Il se pourrait que l'une des plus importantes tienne à la difficulté de définir avec précision le domaine de l'historie sociale.

One of the principal ones may be the very difficulty of defining the scope of social history.

Le désir, fort louable, de décrire dans tous ses détails la vie quotidienne des Français à diverses époques, aboutit trop souvent à une accumulation de détails pittoresques, mais dont on n'est nullement assuré qu'ils expriment une réalité vécue par tel ou tel groupe.

A praiseworthy ambition to describe in detail the daily life of Frenchmen can simply end up with an accumulation of picturesque details without any certainty that they reflect the realities of life for any particular social group.

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.

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.

Nous avons conçu cette historie de la société française à l'epoque contemporaine comme une histoire des groupes sociaux, définis par la place occupée dans le processus de production et la division sociale du travail, comme une histoire aussi de leurs rapports, et de l'evolution, dans le temps, de ces rapports.

The conceptual aim of the present work is to provide a history of the various groups that constitute French society, classified according to the position they occupy in the production process and in the social division of labour, of their relationship with each other, and with the way these relationships have changed in the course of time.

Nous avons cherché à montrer comment, dans cette évolution, certains groupes sociaux défavorisés et mécontents ont réussi à améliorer leur position et assuré, pour quelque temps au moins, leur domination, parcourant ainsi toute la courbe qui les a menés de la revendication au conservatisme, puis à la réaction et à la peur sociale; comment d'autres groupes, autrefois prépondérants, ont été rejetés définitivement dans l'ombre tandis que d'autres encore se constituaient et exigeaient leur part de bien-être, sinon de pouvoir.

We have tried to show how certain underprivileged and malcontent social groups have managed to better their social position and even for a time achieve a sort of domination, running the whole course from radical challenge to conservative acceptance, ending with reactionary fears for the future of society; how other groups that once dominated the scene have been totally eclipsed, while others again are emerging from the shadows and demanding their share of the good things of life and even power.

[Conseil canadien de Développement social 1990a; Canadian Council on Social Development 1990b]

Une partie du projet de loi fait l'objet d'une bataille juridique.

Part of the bill has run into a legal snag.

Le gouvernement de la Colombie-Britannique a poursuivi le gouvernement fedéral à propos du plafonnement du RAPC et a affirmé qu'Ottawa n'avait pas le pouvoir de limiter le cofinancement du RAPC.

The B.C. government took the federal government to court over the cap on CAP, arguing that the federal government has no authority to limit cost-sharing for CAP.

Cette province a aussi soutenu être «legitimement en droit» de s'attendre à ce que le gouvernement fédéral ne dépose pas un texte de loi comme le project C-69 sans le consentement de la province.

It also argued that B.C. has a "legitimate expectation" that a bill like Bill C-69 would not be introduced without the consent of that province.

La Cour d'appel de la Colmbie-Britannique a donné raison au gouvernement de cette province sur les deux questions et la cause a été portée en appel devant la Cour suprême du Canada.

The B.C. Court of Appeal agreed with the B.C. government on both questions, and the matter is being appealed to the Supreme Court of Canada.

Le Canada ne doit pas s'engager dans cette voie sans le consentement éclairé de le majorité de sa population.

This is not a direction Canada should follow, not without the informed consent of a majority of Canadians.

L'adoption du projet de loi C-69 doit être précédée d'une sensibilisation du public et d'un grand débat national.

Public education and widespread debate must precede the enactment of Bill C-69.

Est-il besoin de rappeler en cette période post-Meech que les premiers ministres ont unanimement promis que jamais plus d'importantes orientations nationales ne devraient êtres arrêtées à huis clos?

In Canada's post-Meech-Lake period we must remember that all first ministers agreed that never again should important national directions be set behind closed doors.

Le projet de loi C-69 fait fi de ce principe.

Bill C-69 ignores this.

Il résulte du budget fédéral de 1990, le premier à être élaboré sans consultation publique depuis 1984. Le ministre des Finances a refusé de consulter ou de rencontrer de nombreux groupes et organismes comme par le passé.

It is the product of the 1990 federal budget, the first since 1984 to be drafted behind closed doors — the finance minister refused to consult or meet with groups and organizations as in past years.

[Conseil canadien de Développement social 1990b; Canadian Council on Social Development 1990a]

Les personnes pauvres du Canada devraient être les dernières à souffrir des compressions budgétaires.

Canadians who are poor should least expect to suffer further from the government's belt-tightening.

Mais l'adoption du projet de loi C-69 aggravera probablement leur situation car les provinces sont peu disposées à supporter des coûts accrus reliés aux services sociaux et à l'aide sociale, ou en sont incapables.

Yet, if this bill is passed, they probably will suffer, as provinces are unwilling or unable to absorb any more social services and welfare costs.

On s'attend déjà à ce que les divers services sociaux soient mis à rude épreuve sur le plan budgétaire, puisque le Canada doit faire face au libre-échange, à la recrudescence du chômage, aux nouvelles restrictions qui frappent l'assurance-chômage, à la TPS, à l'arriéré des demandes de statut de réfugié et à la désagrégation de plus en plus marquée des liens familiaux.

The pressure on social-services budgets is already expected to grow as Canada adjusts to free trade, higher unemployment, new restrictions on unemployment insurance, the GST, the backlog of refugee claimants and increasing family breakdown.

Plus de 85 p. 100 des Canadiens pensent que notre régime d'assurance-maladie est le meilleur et le plus précieux de nos services.

More than 85 per cent of all Canadians believe that medicare is our best and most valuable service.

L'efficacité et l'accessibilité de ce régime en font l'un des meilleurs au monde.

It is among the most economical, effective, and accessible health care systems in the world.

Les aînés du Canada sont fiers des programmes sociaux qu'ils ont contribué à mettre sur pied afin de mettre les personnes pauvres, handicapées ou âgées à l'abri de la malnutrition, de la maladie, des mauvaises conditions de logement et de l'isolement social.

Older Canadians are proud of the social programs which they helped set up, and which help protect those who are poor, disabled or old against malnutrition, ill health, poor housing and social isolation.

Depuis son arrivée au pouvoir, le parti conservateur s'est employé à reduire — et dans certains cas à abolir tout à fait — la participation du gouvernement fédéral à ces programmes.

Since the Conservative government came to power, it has moved to reduce — and, in some instances, abolish — federal participation in these programs.

Sans tambour ni trompette, celui-ci a réduit sa contribution au régime d'assurance-maladie, à l'enseignement postsecondaire et à d'autres programmes.

With little or no publicity, the government has been withdrawing from participation in medicare, postsecondary education and other areas.

Le projet de loi C-69 pousse encore plus avant le désengagement du gouvernement fédéral.

Bill C-69 further erodes federal involvement.

La Fédération nationale des retraités et citoyens âgés s'est jointe aux nombreux organismes qui s'opposent résolument à ce que le gouvernement supprime petit à petit nos programmes sociaux.

The National Pensioners and Senior Citizens Federation stands four-square with the many other organizations which oppose this government's slow strangulation of social programs.

[Les Ami(e)s de la Terre 1987; Friends of the Earth 1987]

L'accident de Bhopal met en relief un aspect important des questions environnementales de l'heure.

An examination of the Bhopal accident and its aftermath reveal some important aspects of contemporary environmental issues.

Vous n'avez qu'à mentionner Bhopal pour visualiser l'image de la souffrance et de l'angoisse de ses victimes car aujourd'hui, la santé et le bien-être des gens sont au coeur du problème écologique que représentent les produits chimiques toxiques.

Mention Bhopal and foremost in everyone's mind is the suffering and anguish of the victims. As with other environmental issues involving chemicals, people, and their health and well being, are increasingly at the centre of concern.

Dans cette édition, nous verrons comment contrer certains inquiétudes en mettant sur pied des enquêtes médicales sur la santé.

In this issue, readers will find out how to deal with some of these concerns through health surveys.

C'est aussi ce lien environnement-économie qui nous amène à mettre en doute la nécessité des pesticides fabriqués à l'usine de Bhopal.

Taken from another angle, the environment-economy linkage leads to questioning the need for the pesticides manufactured at the Bhopal plant.

Les usines comme celle de Bhopal pourraient devenir désuètes si on cherchait vraiment à remplacer les pesticides à base de produits hautement toxiques.

Given a greater attention to alternatives to pesticides, plants such as the one at Bhopal could be made obsolete.

Au Canada, la réaction est beaucoup plus lente.

In Canada things move more slowly.

Notre analyse du décret proposé pour la protection de l'environnement semble indiquer plus de promesses que d'actes.

As our analysis of the proposed Environmental Protection Act shows, more is promised than is delivered.

[Les Ami(e)s de la Terre 1986; Friends of the Earth 1986]

Les individus ou les municipalités chez qui l'eau potable a été mise à l'épreuve pour la plus grande partie n'ont pas été avertis des risques à la santé.

Those individuals or municipalities, who had their drinking water tested, have largely been kept in the dark about the health risks.

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.

They were informed of the results by mail, but not informed about the health hazards, even when the levels of contamination were listed as 'unacceptable'.

Ceci a créé beaucoup d'anxiété et de malaise chez la population, et de consultation auprès des médecins, surtout chez les femmes en grossesses et celles ayant de jeunes enfants.

This has created considerable anxiety and uneasiness among people, and sent a number of pregnant women, and women with young children, to their doctors for advice.

Au moment d'écrire ceci, l'attention du public s'éloignait de la menace faite par la présence de produits chimiques dans l'eau potable pour s'arrêter sur la découverte d'un dissolvant de nettoyeur à sec dans l'eau de puits de plus de 70 domiciles de Fairvale, une communauté dortoir de Saint-Jean au Nouveau Brunswick.

At the time of writing, public attention had shifted away from the threat posed by agricultural chemicals to drinking water, and focused on the discovery of dry-cleaning solvent in the well water of more than 70 households in Fairvale, a bedroom community of Saint John.

Les fermiers endurent des menaces sans fin à leurs moyens d'existence, allant du mauvais temps et des fléaux jusqu'à des dettes et des marchés en baisse pour leurs produits.

Farmers endure endless threats to their livelihoods, from bad weather and blight, to crushing debt loads and poor markets for their produce.

Cependant, la menace chimique faite à leurs sources d'eau ne peut être résolue avec de la patience et de la tenacité.

However, the chemical threat to their water supplies will not be overcome with patience and a dogged determination.

S'il n'y a pas d'action bientôt, les coûts pourraient être très importants.

If action is not forthcoming, the costs could be very great.

[Duhaime 1990b; Duhaime 1990a]

Même si une partie de la documentation était disponible en français, aucun cours ne se donnait encore entièrement en français au sein des Commandements aérien et maritime, et la fiabilité des données relatives aux cours bilingues dispensés par la Force mobile et le Service de l'instruction laissait à désirer.

While some of the documentation was available in French, there were no courses given entirely in French in Air Command or Maritime Command, and the reliability of the data concerning bilingual courses given by Mobile Command and Training Command was questionable.

La plupart des stagiaires francophones interrogés par le Commissariat ont mis en doute la pertinence d'une documentation en français lorsque les cours eux-mêmes ne sont dispensés qu'en anglais.

Most of the Francophone trainees interviewed by the Office of the Commissioner questioned the relevance of providing documentation in French when the courses themselves are given in English only.

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 déterminée par un ministère dans cette région.

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.

Notre correspondant affirme que l'entrevue s'est déroulée en français seulement — ce qui ne manqua pas de nous étonner — et que sa candidature fut rejetée parce que son français n'était pas apparu satisfaissant à l'intervieweur.

He said that the interview took place exclusively in French — which greatly surprised us — and that his candidacy was rejected because, in the interviewer's opinion, his French was not adequate.

[Beauparlant 1990a; Beauparlant 1990b]

Pour les uns, ce degré de bilinguisme est satisfaisant puisqu'il constitue une bonne base qui permettra aux élèves d'acquérir les éléments formels de la langue.

This degree of bilingualism is sufficient for some because it constitutes a firm foundation on which to graft the formal elements of the language.

Pour d'autres, il est nettement insuffisant.

For others, it is clearly insufficient.

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 besoins et à ses intérêts.

"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 their needs and interests.

La réforme s'est étendue au niveau secondaire en 1983 pour le premier cycle (secondaire I, II) puis, en 1986, au deuxième cycle.

The reform was extended to the first cycle of the secondary level in 1983 (Secondary I, and II) and then, in 1986, to the second cycle.

Les étudiants devaient donc consacrer 100 heures à l'étude de l'anglais par année pendant leurs cinq années d'etudes.

Students thus were required to devote 100 hours a year to the study of English during their five years of courses.

Les résultats scolaires des élèves démontrent que la réforme Laurin forme maintenant des bilingues fonctionnels à la fin études secondaires.

Students' academic results showed that Laurin's reforms were now turning out functionally bilingual students at the end of their secondary curriculum.

Dans le centre du Québec, M. Normand Dubé, directeur général de la Commission scolaire de Trois-Rivières, souligne que «la région n'est pas en contact avec l'anglais» et qu'on y forme que «difficilement» des bilingues.

As for central Quebec, Normand Dubé, director general of the Trois-Rivières school board, notes that "the region is not in contact with English," and that it is difficult to produce bilingual students there.

Par ailleurs, à Victoriaville, on a constaté qu'une activité hors programme, le Théâtre d'anglais, constituait un excellent stimulant pour les étudiants les plus doués des secondaires III et IV.

In Victoriaville, however, it has been found that an activity outside the program, English theatre, was an excellent means of motivating the most gifted Secondary III and IV students.

[Pierre 1991c; Pierre 1991d]

Le problème ne date pas d'hier.

This concern is longstanding.

Il y a deux ans, des cadres supérieurs d'universités s'entendaient dire au cours d'un débat de l'Association des Universités et Collèges du Canada sur l'efficacité de l'enseignement que l'une des raisons pour lesquelles l'enseignement pâtit est l'exigence de publier à tout prix.

Two years ago senior university administrators at an Association of Universities and Colleges of Canada meeting on effective teaching were told that one reason teaching suffers is the requirement to publish or perish.

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.

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.

Qu'en pensez-vous?

What do you think?

Nous aimerons bien voir un débat sur ces questions dans nos pages.

We'd like to see these issues debated in these pages.

Donnez-nous votre avis en moins de 300 mots.

Please send use your thoughts in 300 words or less.

[Pierre 1991a; Pierre 1991b]

Le mois dernier nous avons abordé un aspect de l'enseignement et de la recherche auquel il semble opportun de revenir.

Last month we touched on one aspect of teaching and research and believe it's timely to look again at this topic.

En fait, nous souhaitons amorcer un débat qui amènera le sujet au premier rang des préoccupations de l'enseignement supérieur.

We'd like to start a discussion that would bring this topic to the top of the higher education agenda.

Par exemple, on dirait que les petites univeristés se sentent obligées d'en imiter de plus grandes pour être prises au sérieux.

One has the impression, for example, that smaller universities feel obligated to model themselves on larger institutions in order to have status.

Tout le monde semble vouloir d'importants programmes d'études supérieures, n'accepter pour professeurs permanents que des titulaires de doctorat ayant largement publié et compter sur un gros budget de recherches.

Everybody seems to want extensive graduate programs, only PhDs with good publication records as tenured faculty and lots of research money.

Et dans la hâte d'égaler les universités de grande taille, quel est le grand perdant?

And in all this push to emulate the largest institutions in the country, what gets lost?

L'enseignement are premier cycle.

The teaching of undergraduates.

Pourquoi n'y aurait-il pas d'excellents établissements de premier cycle où se feraient suffisamment de travaux de niveau supérieur pour créer un climat intellectuel stimulant?

Why can't we have outstanding undergraduate institutions with enough graduate work to provide a stimulating intellectual environment?

Où l'on ferait de la recherche parce qu'on ne devient pas un excellent professeur sans elle? Où l'on exigerait la plus haute qualité d'enseignement?

Where research is carried out because it is essential to being a top performer as a teacher and where the highest standards of teaching are demanded.

Et pourquoi ces établissements n'auraient-ils pas autant de considération que les établissements de recherche, chacun faisant ce qu'il réussit le mieux?

And why can't such institutions be considered as worthy as research institutions — with each doing what it does best?

[Unknown 1990a; Unknown 1990b]

Pour plusieurs d'entre nous, cette date ne sera plus jamais une date ordinaire.

This will never again be just a date on a calendar for many of us.

Le 6 décembre évoque une pensée globale, une série d'images rémemorées ou imaginées, un serrement furieux des poings, de l'esprit ou du coeur, un tremblement de peur.

"December 6" has become a complete thought, a series of images remembered or imagined, an angry clenching of the fists or the mind or the heart, a tremor of fear.

Depuis le 6 décembre, des initiatives ont été prises sur les campus, au sein des organismes nationaux de recherche et de financement et à certains paliers de governement pour commémorer les 14 étudiantes de l'École Polytechnique.

Since December 6, efforts have been made on campuses, in national research and funding organizations and in some levels of government to commemorate the 14 women students from the École Polytechnique.

Diverses universités de la province ont créé un fonds commémoratif.

Numerous memorial funds were established across the country by individual universities.

Ces fonds serviront à plusieurs fins, notamment a l'attribution de bourses d'études, à la rétribution de conférencières et de conférenciers, au soutien des études sur les femmes et à la tenue de séminaires.

Money from these funds will be used in various ways, among them, scholarships, speakers funds, support for women's studies, and seminars.

Les associations de professeurs et le personnel administratif de nombreux campus s'intéressent particulièrement aux questions d'action positive concernant les femmes et d'autres groupes défavorisés, d'équité salariale et de harcèlement sexuel.

On many campuses, particular attention is being directed by faculty associations and administrations to issues of positive (affirmative) action for women and other previously disadvantaged groups, pay equity, and sexual harassment.

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 systémique qui contribue à déavantager les femmes et à créer une ambiance «froide» dans les universités.

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 and contributes to the "chilly climate" that exists at universities.

[Kerr 1990b; Kerr 1990a]

Rendement en classe.

Performances in class.

Le compromis proposé aurait permis aux professeurs et aux étudiants de jouer des pièces en classe, sans auditorie.

The compromise would have allowed for the performance of plays in class by students and teachers without any audience.

Il est difficile de comprendre pourquoi on écarte cette exemption.

It is hard to see why this exemption is removed.

Maintenant, il est fort probable que les professeurs seront persuadés qu'il est plus facile de monter une pièce de Shakespeare en classe, puisque ce dernier ne peut percevoir de redevances et qu'il ne les fera pas passer par un système bureaucratique stupide.

It will probably have the effect of persuading teachers that it is simpler to stage Shakespeare in class since he cannot collect royalties and won't require silly bureaucracy prior to putting on one of his plays.

Ce n'est certainement pas là une façon de stimuler l'intérêt à l'égard des auteurs contemporains qui pourraient, à leur tour, bénéficier d'une amélioration générale de la protection du droit d'auteur en vertu de la nouvelle loi.

This is hardly the way to encourage an interest in the contemporary playwrights who might in turn benefit from the general improvement of the copyright protection under the new law.

Est-ce réellement là le résultat que vous visiez?

Was this really the result you had in mind?

Il sera également sans doute illégal pour les professeurs de copier quelque chose au tableau sans obtenir au préalable la permission au chapitre du droit d'auteur. Il s'agit ici d'une autre exemption proposée qui n'a proposée qui n'a presque pas soulevé de controverse.

Presumably this also means it will be against the law for teachers to copy something on the blackboard without copyright permission, another of the relatively uncontroversial exemptions proposed.

Le changement d'attitude adopté par le gouvernement à l'égard de cette disposition logique nous semble tout simplement insensé.

A change of heart by the government with respect to such common sense provisions seems to us inane.

Le compromis proposé prévoyait un certain nombre d'exemptions parfaitement raisonnables, par exemple, le droit de copier les bulletins de nouvelles pour les utiliser en classe, utiliser des citations dans les examens, etc., exemptions qui devraient toutes, selon nous, faire partie de la nouvelle loi.

There were a number of other perfectly reasonable exemptions in the compromise such as the right to copy news broadcasts for classroom use, put quotes in examinations, etc., all of which in our view should be part of your legislation.

[Chatto 1990a; Chatto 1990b]

Il n'est guère étonnait, vu ce processus long et particulier, que deux marcs ne soient jamais totalement identiques.

Given such a long and idiosyncratic process, it is hardly surprising that no two marcs are ever quite the same.

Le connaisseur distingue immédiatement les marcs de grands bourgognes, Les Hospices de Beaune ou le Domaine de la Romanée-Conti par exemple, aussi distincts l'un de l'autre que les cognacs les plus fins.

The great Burgundian examples, from Les Hospices de Beaune or Domaine de la Romanée-Conti, are instantly recognizable to the connoisseur, as distinct from one another as the finest cognacs.

Je préfère, quant à moi, les produits moins connus des maisons moins grandes.

But for me, the fascination lies in the less familiar products of smaller houses.

Ce sont des marcs distillés depuis des siècles pour le plaisir seul des vignerons et les partager avec eux est en quelque sorte un privilège, un coup d'oeil sur les goûts personnels de leurs créateurs.

These are the marcs that have been distilled for centuries for the private enjoyment of the wine makers, and to share them is still something of a privilege, an insight into the personal tastes of their creators.

En Grèce, où le marc porte le nom de *tsipouro*, ma curiosité insatiable m'a fait goûter à des boissons qu'il aurait mieux valu laisser là où elles se trouvaient.

In Greece, where marc is called *tsipouro*, my pestering curiosity has fetched liquids from the cellar that might have been better left undisturbed — viscous, colourless firewater with the odour of petroleum and all the subtlety of a sledgehammer.

De nombreuses *grappa* italiennes ou *aguardiente* espagnoles m'ont fait monter les larmes aux yeux et pousser des soupirs d'approbation hypocrite. J'ai toujours eu plus de chance en France.

Many a farmhouse Italian *grappa* or Spanish *aguardiente* has brought tears to my eyes, and gasps of hypocritical approval from my numbed and shriveled tongue — but I have always been luckier in France.

Au Languedoc, par exemple, la chaleur du soleil du sud transparaît dans le savoureux marc de muscat, presque liquoreux tant la saveur du raisin s'y fait présente. À proximité à Lirac, dans les Côtes du Rhône, Jean-Claude Assémat fait un marc de la Syrah superbe.

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; while in nearby Lirac, in the Côtes du Rhône, Jean-Claude Assémat makes a superb marc de syrah.

Le marc ne se prête guère aux melanges. Il en existe cependant un dans lequel il se distingue.

Marc does not lend itself to mixes, but there is one concoction in which it shines.

Le Macvin est une autre des spécialités du Château d'Arlay 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.

Macvin is another of Château d'Arlay's specialties, made by blending 30% of mature marc with freshly pressed grape must, followed by 6 to 12 months of barrel aging.

Servi très froid, une seule gorgée de cet apéritif onctueux de couleur jaune pâle rappelle la saveur des pêches et des poires, des figues, des raisins secs et d'un soupçon d'anis.

Served very cold, a single sip of this unctuous pale yellow aperitif fills the mouth with the flavour of peaches and pears, figs, raisins and even a hint of anise.

Un fin marc demeure cependant le choix du connaisseur par son arôme de noix des anciens vins du Jura, l'ardeur contenue par la distillation, le goût épicé complexe du vieux chêne.

But it is the underlying austerity of that fine old marc that wins the day: the walnut aroma of ancient Jura vines, the banked fires of the still, the complex spice of old oak.

[MacDonald 1990b; MacDonald 1990a]

Sur mon chemin, j'ai visité deux sites historiques.

Along the way, I paid heed to two historical sites.

Le premier est l'endroit où le capitaine James Cook a trouvé la mort, à l'abri d'une magnifique falaise de la baie Kealakekua .

The first was the place where Capt. James Cook was killed, in the lee of a magnificent cliff on Kealakekua Bay.

Il semble que Cook était plutôt fâché contre des indigènes qui lui avaient volé l'un de ses navires et qu'il leur avait vaillammant déclaré la guerre.

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

Ils lui ont courageusement résisté, avec tant de courage en fait qu'ils l'ont poignardé et assommé, avant de le noyer.

They resisted him manfully, so manfully, in fact, that they stabbed and clubbed him, before drowning him.

Après avoir exploré le parc, je me suis rendue à Hilo, la troisième ville en importance de l'État après Honolulu et Kailua-Kona.

After exploring the park, I journeyed into Hilo, which is the state's third largest city after Honolulu and Kailua-Kona.

Je m'y suis promenée avec George du Centre touristique, un acteur à ses heures qui a déjà joué le rôle d'un adjoint du gouverneur dans un épisode d'*Hawaï cinq zéro* dans lequel un fou essayait de dévaster la grande île en cachant des explosifs dans le cratère de Kilauea.

I drove around in the company of George from the visitors bureau, a sometime entertainer who once played a governor's aide in a *Hawaii Five-O* television episode that involved a madman who tries to devastate the Big Island by planting explosives in the Kilauea crater.

La ville de Hilo nous rappelle les années 1950 avec ses immeubles proprets du milieu du siècle et ses allées silencieuses.

Hilo puts one in a '50s frame of mind, with its tidy mid-century buildings and easygoing, sleepy ways.

On y voit des barbiers, des promenades et des gens souriants à profusion; ici, le fantôme de Norman Rockwell est en vacances perpétuelles.

There are a lot of barber shops, promenades and smiling citizens; here, it seems, the ghost of Norman Rockwell is on a permanent Hawaiian vacation.

Les gens y sont encore plus détendus qu'en Californie et rien ne l'illustre mieux que leurs vêtements qui, dans n'importe quel pays au monde, porteraient le nom de pyjamas.

The people are even more relaxed than Californians, and nothing illustrates that better than their clothes, which in any other land would be called pyjamas.

La plupart des femmes du plus de 40 ans ont un faible pour les *muumuus* fleures et la foule aux chemises à grandes fleurs du Ken's Pancake House à Hilo rappelle davantage les nombreuses serres d'orchidées de la ville qu'autre chose.

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.

[Dappen 1991a; Dappen 1991b]

«Manges tes carottes si tu veux avoir de bons yeux.» Vous avez certainement entendu cette expression lorsque vous étiez enfant.

"Eat your carrots, they're good for your eyes" is an expression you undoubtedly heard as a child and probably passed on to others.

Maintenant, cet adage pourrait bien en avoir engendré un autre: «Prends tes vitamines, elles aident à prévenir la cataracte.»

Now, there may be an addendum to the adage: "Take your vitamins, they prevent cataracts."

Des recherches menées récemment à l'université Western Ontario de London (UWO) par trois chercheurs ont révélé qu'un supplément quotidien de vitamines C et E réduit le risque de cataract sénile, celle qui se manifeste chez les gens âgés, dans un proportion d'au moins 50 pour cent.

Recent research at London's University of Western Ontario (UWO) concludes that taking daily supplements of vitamins C and E lowers the risk of senile cataracts — those which develop in the aged — by at least 50 per cent.

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.

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.

Les vitamines antioxydantes comme les vitamines C et E pourraient donc jouer un rôle important.

That's why anti-oxidant vitamins like C and E may be important.

«Il semble qu'elles aident à éliminer les radicaux libres», précise Dr John Trevithick, professeur de biochimie participant à l'étude de l'UWO.

"They appear to be instrumental in 'scavenging' free radicals," says Dr. John Trevithick, a professor of biochemistry and one of three researchers involved in the UWO study.

Selon lui, la vitamine E pourrait peut-être transformer les radicaux libres en substances inoffensives et la vitamine C pourrait lui insuffler la vigueur nécessaire pour en attaquer davantage.

According to Trevithick, vitamine E may be the workhorse that transforms free radicals into harmless substances. Vitamin C, meanwhile, may rejuvenate and reactivate vitamin E so it can scavenge more of the free radicals.

Pusiqu'il faudra attendu une dizaine d'années avant d'en connaître les resulats, que peut-on faire dans l'intervalle?

Attempts to fund such a study are underway, but the results of such work are nearly a decade away. So what should you do in the meantime? Pop pills? Wait until all the evidence is in?

Pour ce qui est de la dose quotidienne en question (400 unités de vitamine E et de 300 à 600 milligrammes de vitamine C), les spécialistes s'entendent pour dire que ces comprimés ne feraient pas de tort, même qu'ils pourraient aider.

At the daily levels being discussed (400 international units of vitamin E and 300 to 600 milligrams of vitamin C), experts agree that the pills will do no harm.

[Unknown 1991a; Unknown 1991b]

Nous désapprouvons les processus qui a mené à la conclusion de cet accord.

We were not in agreement with the way the Free Trade Agreement was signed.

La question n'a pas été suffisamment débattue au pays et trop peu de Canadiens comprenaient bien l'Accord.

We didn't think there was enough debate in Canada, nor did we feel that enough Canadians understood the agreement.

Il est si complexe qu'on peut difficilement prédire ses incidences sur les Dénés et autre peuples du Nord.

It is so complex that it's difficult to comprehend how it will ultimately affect the Dene Nation and the other peoples of the North.

J'ai peur que les ressources du Canada tombent aux mains des Américans sans que les gens puissent dire quoi que ce soit.

I'm quite concerned that Canada's resources might go to the Americans without our people having a say in the matter.

Par exemple, nous sommes presque obligés d'exporter aux États-Unis tout excédent de ressources comme le pétrole et le gaz, sans vraiment décider des quantités.

For example, if we have a surplus of resources like oil and gas we are virtually compelled to export them to the United States without much say as to the quantities.

Les peuples aborigènes se font dire qu'ils ne peuvent pas discuter de la constitution tant que le Québec n'aura pas réintégré la famille canadienne.

Aboriginal people are being told we are unable to discuss any kind of constitution building until Quebec is brought into the Canadian family.

Nous ne sommes donc pas en mesure de débattre nos difficultés associées à l'autonomie et à la souveraineté au sein du pays.

That prevents us from discussing our problems associated with self-government and sovereignty within Canada.

C'est tout le noeud du problème: il n'y a pas de processus.

And *that's* the problem; there is no process.

Nous espérions pouvoir commencer à résoudre un certain nombre de questions une fois que serait établi un processus.

We were hoping that once a process was spelled out, we could begin to work on these problem areas.

C'est en partie ce qu'exprimaient les Mohawks cet été à Oka.

That was part of the great frustration the Mohawks were expressing at Oka this past summer.

Les revendications territorials encore irrésolues seront une priorité puisque presque toutes les terres du Nord sont touchées.

Clearly, the main priority is to settle our outstanding land questions since virtually all land in the North is held under aboriginal title.

Nous sommes dans l'incertitude quant aux revendications de Dénés et des Métis.

Right now we're not sure where the Dene/Metis claim is going to go.

Nous avons présenté des modifications à l'entente avec le gouvernement fédéral et le Cabinet tranchera la question.

We are asking for changes in our agreement and the federal government has taken these changes before Cabinet for a decision.

L'instabilité régnera jusqu'à la mise en place de nouveaux arrangements.

Until a future arrangement is understood, there will be instability.

Bibliography

[Alvarado *et al.* 1990a] Sergio J. Alvarado, Michael G. Dyer, and Margot Flowers. Argument representation for editorial text. *Knowledge-Based Systems*, 3(2):87–107, June 1990.

[Alvarado *et al.* 1990b] Sergio J. Alvarado, Michael G. Dyer, and Margot Flowers. Argument comprehension and retrieval for editorial text. *Knowledge-Based Systems*, 3(3):139–162, September 1990.

[Beauparlant 1990a] Michel Beauparlant. L'école québécoise forme des bilingues fonctionnnels. *Langue et Société*, (33):28–29, hiver 1990.

[Beauparlant 1990b] Michel Beauparlant. Quebec schools produce functionally bilingual students. *Language and Society*, (33):28–29, Winter 1990. [translator unknown].

[BenHassine forthcoming] Nadia BenHassine. A formal approach to controlling style in language generation. Master's thesis, Department of Computer Science, University of Waterloo, forthcoming.

[Benton 1991] Peter M. Benton. The multilingual edge. *Byte*, 16(3):124–132, March 1991.

[Bishop *et al.* 1975] Yvonne M. M. Bishop, Stephen E. Fienberg, and Paul W. Holland. *Discrete Multivariate Analysis: Theory and Practice*. The MIT Press, 1975.

[Blake and Crewe 1978] Anthony Blake and Quentin Crewe. *Great Chefs of France*. Marshall Editions Ltd., London, 1978.

[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.

[Carbonell *et al.* 1981] Jaime G. Carbonell, Richard E. Cullingford, and Anatole V. Gershman. Steps toward knowledge-based machine translation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-3(4):376–392, 1981.

[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.

[Cherry and Vesterman 1981] L. L. Cherry and W. Vesterman. Writing tools — the STYLE and DICTION programs. Technical Report CSTR-91, AT&T, February 1981.

[Chipaux 1990a] Françoise Chipaux. Bsouss, the martyred village. *Manchester Guardian Weekly*, 143(17):13, October 28 1990.

[Chipaux 1990b] Françoise Chipaux. Bsouss, village libanais martyr. *Le Monde*, pages 1, 3, October 21/22 1990.

[Chomsky 1981] Noam Chomsky. *Lectures on Government and Binding.* Foris Publications, Dordrecht, Holland, 1981.

[Cluett 1976] Robert Cluett. *Prose Style and Critical Reading.* Teachers College Press, Columbia University, New York, 1976.

[Coltman 1973] Derek Coltman. *The Hundred Glories of French Cooking.* Doubleday Canada Ltd., Toronto, 1973.

[Conseil canadien de Développement social 1990a] Conseil canadien de Développement social. Les programmes sociaux du Canada en péril. *Les Programmes Sociaux du Canada en Péril*, pages 1–2, 7, 1990.

[Conseil canadien de Développement social 1990b] 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.

[Corbett 1971] Edward P. J. Corbett. *Classical Rhetoric for the Modern Student.* Oxford University Press, 2nd edition, 1971.

[Crystal and Davy 1969] David Crystal and Derek Davy. *Investigating English Style.* Longman Group Limited, 1969.

[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.

[Denny 1974] Norman Denny. *Friday or the Other Island.* Penguin Books, Harmondsworth, Middlesex, 1974.

[Dibben —] S. Dibben. Paris VII. [No additional information given].

[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 Language*, Austin, Texas, June 1990.

[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].

[Dobson 1990] Annette J. Dobson. *An Introduction to Generalized Linear Models.* Chapman and Hall, New York, 1990.

[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.

[Eglowstein 1991] Howard Eglowstein. Can a grammar and style checker improve your writing? *Byte*, 16(8):238–242, August 1991.

[Flaubert 1857] Gustave Flaubert. *Madame Bovary*. Michel Lévy, Paris, Garnier-Flammarion, 1966 edition, 1857.

[Flaubert 1957] Gustave Flaubert. *Madame Bovary*. Random House, Inc., Toronto, Modern Library edition, 1957. [translated by Francis Steegmuller].

[Fowler 1968] H. W. Fowler. *A Dictionary of Modern English Usage*. Oxford Univeristy Press, Oxford, second edition, 1968.

[Friends of the Earth 1986] Friends of the Earth. Groundwater pollution in New Brunswick. *infoetox*, 3(2):15–16, June 1986.

[Friends of the Earth 1987] Friends of the Earth. In the aftermath of Bhopal. *infoetox*, 4(1):1–2, March 1987.

[Greilsamer 1990a] Laurent Greilsamer. Simulacre. *Le Monde*, page 8, 21/22 October 1990.

[Greilsamer 1990b] Laurent Greilsamer. Will Vichy ever go on trial? *Manchester Guardian Weekly*, 143(17):13, 28 October 1990.

[Gresset and Richard 1975] Michel Gresset and Claude Richard. Partridge festival. In *Pourquoi ces nations en tumulte*. Gallimard, Paris, 1975.

[Guillemin-Flescher 1981] Jacqueline Guillemin-Flescher. *Syntaxe comparée du français et de l'anglais*. Ophrys, Paris, 1981.

[Heidorn *et al.* 1982] G.E. Heidorn, K. Jensen, L.A. Miller, R.J. Byrd, and M.S. Chodorow. The EPISTLE text-critiquing system. *IBM Systems Journal*, 21(3):305–326, 1982.

[Hergé 1947] Hergé. *L'Etoile mystérieuse*. «Les Aventures de Tintin». Casterman, Paris Tournai, 1947.

[Hergé 1948] Hergé. *Les 7 Boules de cristal*. «Les Aventures de Tintin». Casterman, Paris Tournai, 1948.

[Hovy 1988] Eduard H. Hovy. *Generating Natual Language Under Pragmatic Constraints*. Lawrence Erlbaum Associates, Inc., Hillsdale, New Jersey, 1988.

[Hovy 1990] Eduard H. Hovy. Pragmatics and natural language generation. *Artificial Intelligence*, 43:153–197, 1990.

[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.

[Kaplan and Bresnan 1982] Ronald M. Kaplan and Joan Bresnan. Lexical-functional grammar: A formal system for grammatical representation. In Joan Bresnan, editor, *The Mental Representation of Grammatical Relations*, chapter 4. MIT Press, 1982.

[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, December 1990.

[La Reynière 1971] La Reynière. *Cent merveilles de la cuisine français.* Seuil, Paris, 1971. [La Reynière is a pseudonym for Robert Courtine].

[Landsbergen 1987] Jan Landsbergen. Montague grammar and machine translation. In P. Whitelock, M.M. Wood, H.L. Somers, R. Johnson, and P. Bennett, editors, *Linguistic Theory and Computer Applications*, pages 113–148. Academic Press, London, 1987.

[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.

[Leermakers and Rous 1986] R. Leermakers and J. Rous. The translation method of Rosetta. *Computers and Translation*, 1(3):169–183, 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, June 1986.

[Les Ami(e)s de la Terre 1987] Les Ami(e)s de la Terre. Les répercussions de Bhopal. *infoetox*, 4(1):1–2, March 1987.

[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. Enquete stylistique sur la traduction automatique de documents techniques. *International Review of Applied Linguistics in Language Teaching*, 25(2):154–168, May 1987.

[Lonsdale-Cooper and Turner 1961] Leslie Lonsdale-Cooper and Michael Turner. *The Shooting Star.* Methuen & Co Ltd., London, 1961.

[Lonsdale-Cooper and Turner 1962] Leslie Lonsdale-Cooper and Michael Turner. *The Seven Crystal Balls.* Methuen Children's Books. Methuen & Co Ltd., London, 1962.

[Lytinen and Schank 1982] Steven L. Lytinen and Roger C. Schank. Representation and translation. *Text*, 2(1-3):83–111, 1982.

[Lytinen 1984] Steven L. Lytinen. *The Organization of Knowledge in a Multi-lingual, Integrated Parser.* PhD thesis, Department of Computer Science, Yale University, 1984. [Also available as Technical Report YALEU/CSD/RR #340].

[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.

[Madden 1991] Frank Madden. Correct grammar. *Computers and the Humanities*, 25:58–61, 1991.

[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.

[Mayoux 1979] Sophie Mayoux. *Les Grands Chefs.* Editions du Fanal, Paris, 1979.

[McKeown 1985] Kathleen R. McKeown. Discourse strategies for generating natural language text. *Artificial Intelligence*, 27:1–42, 1985.

[Michelin 1972] Michelin, 1972. [No additional information given].

[Milic 1982] Louis T. Milic. The annals of computing: Stylistics. *Computers and the Humanities*, 16:19–24, 1982.

[Murray 1990a] Nickolas Murray. Ending the endless cable rate hikes. *Canadian Consumer*, (2):50, 1990.

[Murray 1990b] Nickolas Murray. Les exagérations des câblodistributeurs! *Le Consommateur Canadien*, (2):50, 1990.

[Neijt 1986] A. Neijt. Esperanto as the focal point of machine translation. *Multilingua*, 5(1):9–13, 1986.

[Neuman 1991] Michael Neuman. Rightwriter 3.1. *Computers and the Humanities*, 25:55–58, 1991.

[Newmark 1988] Peter Newmark. *A Textbook of Translation.* Prentice Hall International (UK) Ltd, Hertfordshire, 1988.

[O'Connor 1946] Flannery O'Connor. Partridge festival. In *The Complete Stories.* Farrar, Strauss and Giroux, New York, 1946.

[Pierre 1991a] Gloria Pierre. Encore à propos d'enseignement et de recherche. *Affaires Universitaires*, 32(2):2, février 1991.

[Pierre 1991b] Gloria Pierre. More thoughts on teaching and research. *University Affairs*, 32(2):2, February 1991.

[Pierre 1991c] Gloria Pierre. Publier, bon. Et enseigner alors? *Affaires Universitaires*, 32(1):2, janvier 1991.

[Pierre 1991d] Gloria Pierre. What about teaching? and publishing? *University Affairs*, 32(1):2, January 1991.

[Pub. C.E.A. ——] Pub. C.E.A. [No additional information given].

[Pub. Najac ——] Pub. Najac. [No additional information given].

[Pub. PTT ——] Pub. PTT. [No additional information given].

[Pub. Tippex ——] Pub. Tippex. [No additional information given].

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

[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.

[Richardson and Braden-Harder 1988] Stephen D. Richardson and Lisa C. Braden-Harder. The experience of developing a large-scale natural langauge text processing system: Critique. In *Proceedings of the 2nd Conference on Applied Natural Language Processing*, Austin, February 1988.

[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].

[Sidner 1979] C.L. Sidner. *Towards a computational theory of definite anaphora comprehension in English discourse*. PhD thesis, MIT, Cambridge, MA, 1979.

[Slocum 1988] Jonathan Slocum. A survey of machine translation. In Jonathan Slocum, editor, *Machine Translation Systems*, pages 1–48. Cambridge University Press, Cambridge, 1988.

[S.N.C.F. 1971] S.N.C.F. *Les chemins de fer en France*. S.N.C.F., Paris, 1971.

[S.N.C.F. 1973] S.N.C.F. *The Railways of France*. S.N.C.F., Paris, 1973.

[Strunk and White 1979] William Strunk, Jr. and E. B. White. *The Elements of Style*. Macmillan Publishing Co. Inc., third edition, 1979.

[Tournier 1967] Michel Tournier. *Vendredi ou les limbes du Pacifique*. Gallimard, Paris, 1967.

[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, December 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.

[Valigra 1991] Lori Valigra. Japan shows progress in machine translation. *The Institute*, 15(5):1,8, May/June 1991.

[Vinay and Darbelnet 1958] J.P. Vinay and J. Darbelnet. *Stylistique comparée du français et de l'anglais*. Beauchemin ltée, Saint-Laurent, Québec, 1958.

[Wilks 1973] Yorick Wilks. An artificial intelligence approach to machine translation. In Roger C. Schank and Kenneth Mark Colby, editors, *Computer Models of Thought and Language*, pages 101–129. Freeman, San Francisco, 1973.