

Accounting for Style in Machine Translation

Chrysanne DiMarco

Department of Computer Science
University of Waterloo
Waterloo, Ontario, Canada N2L 3G1
(cdimarco@watdragon.waterloo.edu)

Graeme Hirst

Department of Computer Science
University of Toronto
Toronto, Ontario, Canada M5S 1A4
(gh@ai.toronto.edu)

1 An advocacy of style in machine translation

A significant part of the meaning of any text lies in the author's style. Different choices of words and syntactic structure convey different nuances in meaning, which must be carried through in any translation if it is to be considered faithful. Up to now, machine translation systems have been unable to do this. Subtleties of style are simply lost to current MT systems.

The goal of the present research is to develop a method to provide MT systems with the ability to understand and preserve the intent of an author's stylistic characteristics. Unilingual natural language understanding systems could also benefit from an appreciation of these aspects of meaning. However, in translation, style plays an additional role, for here one must also deal with the *generation* of appropriate target-language style.

Consideration of style in translation involves two complementary, but sometimes conflicting, aims:

- The translation must preserve, as much as possible, the author's stylistic intent—the information conveyed through the manner of presentation.
- But it must have a style that is appropriate and natural to the target language.

The study of *comparative stylistics* is, in fact, guided by the recognition that languages differ in their stylistic approaches: each has its own characteristic stylistic preferences. The stylistic differences between French and English are exemplified by the predominance of the pronominal verb in French. This contrast allows us to recognize the greater preference of English for the passive voice:

- (1) (a) Le jambon se mange froid.
(b) Ham is eaten cold.

Such preferences exist at the lexical, syntactic, and semantic levels, but reflect differences in the two languages that can be grouped in terms of more-general stylistic qualities. French words are generally situated at a higher level of *abstraction* than that of the corresponding English words, which tend to be more concrete (Vinay and Darbelnet 1958, 59). French aims for *precision* while English is more tolerant of *vagueness*. (Duron 1963, 109).

So, a French source text may be abstract and very precise in style, but the translated English text should be looser and less abstract, while still retaining the author's stylistic intent. Translators use this kind of knowledge about comparative stylistics as they clean up raw MT output, dealing with various kinds of stylistic complexities.

2 Comparative stylistics of English and French

Let us now consider in more detail some of the stylistic differences between English and French. All our examples will be taken from a textbook for translators: J.P. Vinay and J. Darbelnet, *Stylistique comparée du français et de l'anglais*, (1958).

At the *lexical level* we have situations such as the following, shown in (2), in which the two languages differ in the use of the premodifying pronouns. English is more concrete in its use of a personal pronoun, *your*, while French uses an abstract impersonal construction, *du*.

- (2) (a) Reçu du client.
(b) This is your receipt.

Not only individual words but whole phrases can express different nuances. The English preference for concreteness is illustrated below by its tendency to mark the stages of a process while French uses a single word:

- (3) (a) à mesure que les kilomètres s'allongeaient derrière eux
(b) as they covered mile after mile

One must also consider different preferences in *syntactic structure* between the source and target languages. The structure can reflect the essential nature of a language, as in (4) below. In French, adverbial phrases or clauses are placed by preference at the head of a sentence, especially if they have a causal sense—the cause precedes the effect. English, on the other hand, presents information in order of importance in the text.

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

In addition to lexical or syntactic re-organization in translation, there may be a complete re-statement, a change in *semantic structure*:

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

In the example, English is more concrete and personal in its use of a deictic (*this*), while French is correspondingly more abstract and impersonal in its use of nouns (*un pareil état d'esprit*).

To deal with style in machine translation, we must deal with all the types of lexical, syntactic, and semantic problems. We must recognize stylistic effects and formally represent knowledge about stylistics. This may seem a rather idealistic goal for, after all, basic problems of syntax and semantics still remain in MT. However, to ultimately win wider public acceptance, MT will have to be less obvious; a machine translation should not sound like a translation.

3 What exactly is style?

Having accepted in general terms that style should be a concern of MT systems, we must clarify just what style is in this context before we can begin to think about how it can be accounted for in MT.

For machine translation, where we expect to deal with large amounts of similar types of text, the analysis of *group style* is of more interest than the idiosyncratic style of any one writer. By group styles, we refer to a characteristic of text that, although possibly produced by one individual, shares the stylistic standards of a body of writers. Group style can be subdivided into two major types, each associated with a different view of stylistics: *literary style* and *utilitarian style*.

Utilitarian group styles are distinguished from the literary type by their association with a genre of text that has a particular function or purpose, such as medical textbooks or newspaper articles. In such styles, the writer accommodates her language to what readers expect in the specific, restricted, situation. The fact that utilitarian group styles *are* more restrictive than literary styles, as they are characteristic of a particular genre, suggests that the problem of codifying utilitarian style will be more tractable. For this reason, we have focused on utilitarian texts in general and newspaper writing in particular, using as our main source of examples English translations in *The Manchester Guardian Weekly* of articles from *Le Monde*.

So far, this view of style will help us with our aim of preserving an author's stylistic intent. If, for example, the author is a journalist, then dealing with utilitarian group style will mean that we will attempt to understand and codify the lexical, syntactic, and semantic choices that are characteristic of a newspaper's manner of presentation.

But our other goal, style appropriate to the target language, requires us to go further. We must account for the fact that different languages tend to favour particular stylistic characteristics—different *stylistic goals*. If we are to properly translate from one language to another, then we must adopt a more completely *goal-directed* definition, so that we can correlate lexical choice, syntactic structure, and semantic

organization with such stylistic goals.

Thus, we propose a view of style for MT that is group-based, utilitarian, and, most significantly, goal-directed to allow us to achieve a codification of style that will be both precise and expressive.

From a review of previous work in stylistics, we conclude that a foundation exists for a codification of stylistics for MT purposes:

- The need for a *vocabulary* of stylistic terms is an accepted idea. However, the nature and structure of such a vocabulary, which must describe more than basic syntax and be amenable to systematic construction, has not been addressed. This will be our first task (section 4).
- Similarly, while large bodies of stylistic rules already exist, they have not been organized into the formal structure necessary for a computational approach. We will develop a structured method for representing stylistic rules by constructing formal *stylistic grammars* (section 5).
- The feasibility of the goal-directed analysis of style necessary to MT is supported by previous research, but an actual goal-based codification has not been attempted. In the grammars that we will develop, we will incorporate a *goal-directed* knowledge of stylistics.

In this codification, our methods are intended to be language-independent. So far, we have adapted our work to English and French (DiMarco 1990), focusing on syntactic, rather than lexical or semantic, aspects of style.

4 A vocabulary of style

4.1 Stylistic goals

To create a vocabulary for talking about stylistics, we will construct a single lexicon to describe both English and French stylistics. It will be constructed from terms that are associated with group-based, utilitarian, goal-directed stylistics. Let us start by considering the kinds of stylistic goals to be dealt with in the lexicon. *Comparative*, or language-dependent, stylistic goals are, by definition, characteristic of a particular language. As a consequence, in the same situation, two languages may aim for opposing

goals. (For example, French tends to be a more static language than English, as there is a predominance of the noun over the verb.) Hence, language-dependent goals are necessarily part of the study of style in translation.

We have chosen six representative goals (suggested by Vinay and Darbelnet (1958)) that represent the opposite ends of three dimensions:

- abstraction and concreteness;
- staticness and dynamism;
- clarity and obscurity.

4.2 Abstract elements of style

As the foundation of a vocabulary of style, we propose a set of stylistic terms that make abstract stylistic features explicit that are only implicit in existing terminology. These stylistic terms are based on effects of *concord* and *discord*, which we define as follows:

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

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

These notions are used to define a set of fifteen abstract elements, examples of which are given below. Full details are given by DiMarco (1990).

4.2.1 Sample abstract elements

A common type of abstract element is the *monoschematic*, the simplest stylistic sentence type:

Monoschematic: A stylistic constituent with a single dominant component and no accompanying subordinate or coordinate components.

Monoschematic sentences, such as the following, are simple uninterrupted clauses:

- (6) He was born in 1598 in the small town of Fuente de Cantos in Badajoz province.¹

The most common abstract element in the texts that we have studied is the *centroschematic*:

¹ *Manchester Guardian Weekly*, 14 February 1988, 15.

Centroschematic: A stylistic constituent 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 the rest of the sentence. Such structures are built with subordination and coordination. In the first example below, the main clause is supported by a complex, but subordinate and coordinate, relative clause structure:

- (7) Neither these devices nor the cramped viewing rooms which are too narrow and whose ceilings are much too low for the big altarpieces manage to spoil the works.²

Two subordinate clauses, the first adverbial, the second relative, accompany but do not overshadow the main clause in the following sentence:

- (8) Not all that long ago, the famous collector Charles de Bestegui, when unable to get hold of certain paintings, was quite prepared to adorn the walls of his Venetian palace with copies, which happily rubbed shoulders with his numerous genuine canvases.³

Finally, the complex but imitative postmodification in the next example incorporates a substantial amount of information without weakening the main clause:

- (9) Silvia, a commanding woman in her 50s. a shrew falsely mellowed by religion, promptly organised prayer sessions on the lines of Tupperware meetings.⁴

As we have seen, centroschematic sentences can be quite complex, containing subordinate and coordinate constructions, but, in all cases, a dominant component, usually the main clause, provides the organizing core.

We can also observe stylistic effects related to changes in the concordant and discordant organization of components over the entire length of a sentence:

² *Manchester Guardian Weekly*, 14 February 1988, 15.

³ *Manchester Guardian Weekly*, 7 February 1988, 14.

⁴ Adapted from the *Manchester Guardian Weekly*, 7 February 1988, 16.

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 sentence-initial position. However, there is a terminal concord as the subsequent main clause contains no incongruities and restores the sense of harmony:

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

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 see an example of a dissolution in the Biblical text below. A strong initial concord, created by imitative clauses, is set against a terminal discord, created by a syntactic inversion:

- (11) 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.⁶

The organization of a sentence can be interrupted or perturbed by a *heteropoisal* ("different weights") component:

Heteropoise: A stylistic constituent with an offsetting component that contributes either to an overall stylistic balance or imbalance.

In the following heteropoisal example, a participle clause interrupts the main clause, yet does not disturb the flow of the sentence:

- (12) The measure, which brought no protests, was decided, according to a company spokesman, because of the dangers to which these employees are exposed in travelling to particularly exposed sub-tropical countries.⁷

⁵ Quirk *et al.* (1985, 652).

⁶ Matthew 7:27, Revised Standard Version (1952).

⁷ *Manchester Guardian Weekly*, 21 February 1988, 14.

On the other hand, an interruption can perturb the balance of a sentence. In the following example, an initial offset, an appositive noun phrase, provides a contrast to the main clause:

- (13) The heir to a fortune, her friend did not care about passing examinations.⁸

4.3 Primitive elements of style

Having defined a set of abstract stylistic elements, we must next consider how to use them in practical stylistic analysis. Because they are "abstract" elements, the level of definition is still too general to apply to the interpretation of an arbitrary sentence. These *stylistic* elements must be related to more basic *syntactic* elements, whose stylistic characteristics are more specific and concrete. We have catalogued existing stylistic terms and we observe that there are two views of stylistic analysis: *connective* (an ordering on a single level of sentence structure) and *hierarchic* (an ordering across multiple levels). We will use this notion of two views to guide the definition of more-primitive stylistic elements that will provide a precise syntactic basis to the vocabulary, but also allow a mapping to the abstract elements.

These *primitive stylistic elements*, although intended to be more low-level than the abstract elements, must still be associated with particular stylistic effects. But what should be used as the basis for the definition of "stylistic effect"? We define the stylistic effect of an individual component in terms of its contribution to both the connective ordering and the hierarchic ordering:

- For the connective view, a component acquires a characteristic stylistic quality from its inherent cohesive effect, its degree of bonding with the rest of the sentence.
- For the hierarchic view, stylistic effect correlates with the inherent degree of subordination or superordination of a component.

To define primitive stylistic elements, we introduce the terms *conjunct*, *antijunct*, *subjunct*, and *superjunct* and use superscripts on the

⁸ Quirk *et al.* (1985, 1314).

terms to indicate the degree of integration or subordination. In the connective view, we classify syntactic components as either *conjunct*⁴ (extremely integrating), *conjunct*³ (strongly integrating), *conjunct*² (integrating), *conjunct*¹ (less integrating, but still having a connective effect), or *conjunct*⁰ (neutral). Similarly, we use the terms *antijunct*⁰ through *antijunct*⁴ to indicate increasingly disconnective effects; *antijunct*⁰ is equivalent to *conjunct*⁰.

There is a complementary vocabulary of primitive elements for the hierarchic view. The stylistic effects of syntactic components are correlated with the degree of subordination or superordination; the classifications are analogous to the connective: *subjunct*⁰ through *subjunct*⁴ (increasingly subordinate) and *superjunct*⁰ through *superjunct*⁴ (increasingly superordinate), where *subjunct*⁰ and *superjunct*⁰ are the same.

The following sentences provide an illustration of how the primitive elements are used in stylistic analysis. Distinct stylistic effects are associated with the different characteristics of the marked components. In the first example, the postmodifying *conjunct*⁰ verbless clause has a *neutral* primitive-element classification for it lacks the cohesive effect of a subject-verb "block".

- (14) Most people consider these books rather valuable.

In the next example, the postmodifying *conjunct*¹ non-finite clause is mildly connective due to the presence of a verbal component.

- (15) These books, being rather valuable, are expensive.

In the last example, the postmodifying *conjunct*² relative clause is strongly connective because of the presence of a subject-verb block.

- (16) These books, which are rather valuable, are expensive.

For both the connective and hierarchic views, we have classified all syntactic components into a particular category of primitive element, as in the examples above. The complete classification is given by DiMarco (1990). With the development of a stylistic vocabulary of abstract elements and primitive shapes, the basic

tools needed to construct complementary stylistic grammars for English and French are now available.

5 The syntactic stylistic grammars

Next, we will define a method for using the stylistic vocabulary to systematically build up stylistic grammars for English and French. To construct the stylistic grammar, we will use a methodology whose steps are as follows:

- the classification of primitive stylistic elements of the language in question;
- the correlation of the stylistic effects of these elements with the abstract stylistic elements; and
- the correlation of patterns of these abstract elements with specific stylistic goals.

This methodology produces a stylistic grammar with three levels of rules: primitive elements, abstract elements, and stylistic goals. Together, these levels form a language-independent schema for a goal-directed grammar for style in translation.

5.1 Grammar of primitive elements

The English and French stylistic grammars are both constructed using the same vocabulary and methodology described earlier. The grammars are, however, distinct at the base-level grammar of primitive elements: the French grammar has a *functional* basis. A functional grammar, unlike traditional grammars, provides not just a system of rules for judging whether a sentence is grammatically correct, but a means of understanding what the writer is trying to express and why he has used a particular sentence construction, even at the lowest level of syntactic detail.

At this level of the grammar, we assign to each syntactic component a degree of *conjunctness*, in the connective view of sentence structure, and *subconjunctness*, in the hierarchic view.⁹ For example:

conjunct² postmodification →

⁹ The rules are given in phrase-structure notation. After the arrow, each alternative is listed on a separate line, with the occasional interpolation of an example.

wh-relative clause

the man who founded the religious association

that-relative clause

the laboratory that will not insist on a qualified pharmacist being present at the point of distribution

conjunct² adverbial clause

Our homely evening fire, when the fiery heat of the desert daylight is done, is pleasant.

Furthermore, we define components to be either concordant or discordant as in the following rule. If postmodification is moderately or strongly conjunct, then it is considered concordant, for there are no disconnective constructions to disturb the solidarity of the noun-postmodifiers block.

concordant postmodification →

conjunct³ postmodification

conjunct² postmodification

conjunct¹ postmodification

The underlying syntactic terms are quite different for English and French. The following rule for French contrasts with the one shown above for English.

conjunct² postmodification →

relative clause

Le géranium est une plante qui vit longtemps.

epithet nominal group

Elle porte un costume tailleur.

juxtaposed apposition

Les arbres squelettes ressemblent à des vieillards grotesques.

epithet with no intermediary

Un chapeau doublé de feutre.

We see here that the French grammar, unlike its English counterpart, uses functional syntactic terms, such as *epithet* and *apposition*. However, in both English and French grammars, the syntactic terms are correlated with the same vocabulary of stylistic terms, such as *conjunct² postmodification*, so that at the next level of the grammar, abstract elements, the two grammars can be integrated.

5.2 Grammar of abstract elements

The grammar of abstract elements correlates the stylistic effects of the primitive elements with the abstract elements. At this level, the stylistic grammar rules are identical for English and French.

initial concord →

concordant major + (clause)*

*John went to get his book, so he said.
Le géranium est une plante qui vit
longtemps, mais la rosé est éphémère.*

concordant clause + major + (clause)*

*// we can judge from the canvases on
the walls of the gallery, those who are
responsible are myopic.
Si la maman chante, elle éveillera
l'enfant.*

centroschematic →

concordant sentence

*It is unthinkable to move the whole of
the works which are the only Zurbarán
paintings in the palace where they were
hung in the year when the artist died.
Notre professeur a puni sévèrement
un élève qui avait fraudé.*

heteropoise →

sentence with heteropoisal noun phrase

*Lives, a great number of them children,
were lost.
Notre professeur, homme doux et bon,
a puni sévèrement un élève qui avait
fraudé.*

In the following rule for connective *resolution*, an inherently concordant, monoschematic, component is embedded at the end of a more complex and more relatively discordant sentence. Thus, there is a conflict between concordant and discordant components that is resolved at the end of the sentence.

resolution →

initial discord + final concord

*In its energy, its lyrics, its advocacy of
frustrated joys, rock is a symphony.
Pâle comme un fantôme dont tout le
monde a peur, il marchait.*

5.3 Grammar of stylistic goals

At the top level of the grammar, the rules are again identical for English and French. Patterns of the abstract elements are correlated with specific stylistic goals, as in the following examples. We define *clarity* to be associated with simple sentences (*monoschematic*) or sentences that have a dominant concordant component (*centroschematic, resolution*).

clarity →

monoschematic

centroschematic

resolution

We define *concreteness* to be associated with sentences that express an effect of immediacy by emphasizing a particular component, which may be highlighted because it is discordant (*dissolution*) or because it is in a prominent position (*heteropoise*).

concreteness →

heteropoise

dissolution

We associate *staticness* with standard simple structures.

staticness →

monoschematic

Thus, the application of the same stylistic vocabulary and development methodology allows the complete integration of the English and French stylistic grammars at the higher levels of description.

5.4 Implementation

The English and French grammars, from which the above rules are taken, have been implemented in a single computational system, *STYLISTIQUE* (DiMarco 1990). *STYLISTIQUE* is a syntactic stylistic parser that interprets the English and French grammars and produces an analysis at the three levels of stylistic goals, abstract elements, and primitive elements.

6 STYLISTIQUE and machine translation

Our goal has been to construct the foundations of a system that could preserve style in machine translation. So far, we have developed English and French stylistic grammars that can be interpreted by a single parser. The next step must be to define a mapping between these grammars.

The following pairs of sentences demonstrate stylistic differences between English and French that we can now recognize and name using the stylistic grammars. The first group demonstrates cases where we want to preserve the same stylistic effect from French to English. Recall example (4) of section 2, showing different preferences in syntactic ordering between English and French. In a machine translation system incorporating STYLISTIQUE, the change in structure in this example follows from the recognition and preservation of an *initial concord* between French and English. This is because, in French, the initial adverbial clause is more concordant than in English. When we translate into English, we must begin with the main clause itself to reproduce the effect of concord.

We also see initial concord preserved in the next pair of sentences. The French begins with a *concordant* non-finite clause, while the English begins with a strong *concordant* word, *because*, which tends not to be used in the initial position in French.

- (17) (a) Ma première lettre ayant pu s'égarer, je me permets de vous écrire de nouveau. (*lit.*, My first letter having perhaps gone astray, I am writing you again.)
(b) Because my first letter may have been lost, I am writing you again.

Our final example of the preservation of an initial concord occurs in the next pair of sentences: a *concordant* cleft structure in French is replaced by a strong *concordant* initial adverb in English:

- (18) (a) Il y eut un moment où il faillit se faire prendre. (*lit.*, There was a moment when he just missed being captured.)
(b) Once he was almost captured.

In all these examples, we observe that different situations can be recognized as initial con-

cords that must be preserved, without the necessity of separate heuristic rules.

The second group of examples demonstrates cases where we have to choose different stylistic effects for French and English, each effect being characteristic of the particular language. In the first pair of sentences, we can recognize the difference between a *static* French construction and a more characteristically *dynamic* English sentence and choose to translate accordingly:

- (19) (a) Les gens ont applaudi sur le passage des troupes. (*lit.*, People cheered along the passage of the troops.)
(b) People cheered as the troops marched by.

In the next pair of sentences, STYLISTIQUE'S grammars can recognize the French tendency to *abstractness* in a minimally subjunct noun while English, more *concrete*, uses a more subordinated, subjunct, deictic premodification:

- (20) (a) Toute partie de la carte que nous ne voyons pas à l'instant même n'existe pas pour nous. (*lit.*, Any part of the map that we do not see at the exact moment does not exist for us.)
(b) All that part of the map that we do not see before us is a blank.

Another instance of the French tendency to *abstractness* can be recognized in the following example, in which French uses a minimally subjunct noun while English uses a more subordinated, more strongly subjunct noun phrase:

- (21) (a) Ils passèrent dans un rapide mouvement, (*lit.*, They passed by in a quick motion.)
(b) Quick-moving feet pattered by.

Occasionally, it is French that will be more *concrete* in its use of a more strongly heteropausal relative clause, while English, more *abstract*, uses a less obtrusive participle clause without a subject:

- (22) (a) Là aussi flotte une brume légère qui estompe les durs contours des idées et fond les couleurs de la passion. (*lit.*, There, also, drifts a light mist that softens the hard contours of ideas and melts the colours of passion.)
(b) There, too, is a haze rubbing away the

hard edges of ideas, softening and blending the hues of passion.

We might even use STYLISTIQUE's grammars to recognize more complex situations in translation, such as the next group of examples. In the next pair of sentences, an *initial concord* in French is replaced by an *initial discord*, an initial heteropose in English, in the form of a participle clause that lacks a subject:

- (23) (a) Les organisateurs ont pu s'assurer le concours de professeurs des différentes écoles. (*lit.*, The organizers were able to secure the help of teachers from various schools.)
(b) Participating in the program are teachers from the various schools.

In the following pair of sentences, a French *hierarchic resolution*, the completion of a complex sentence by a strongly subordinated, subjunct noun phrase, is replaced by an English *hierarchic dissolution*, the completion of a complex sentence by a minimally subjunct noun phrase:

- (24) (a) C'est aux tribunaux d'enfants qu'il incombe de diriger les sujets appropriés vers une institution chargée du redressement des garçons particulièrement difficiles. (*lit.*, It is the juvenile courts that it is incumbent upon to direct the appropriate subjects towards an institution charged with the responsibility of the straightening-out of particularly difficult boys.)
(b) The proper bodies to direct suitable boys into an organization intended to reclaim the exceptionally tough are the juvenile courts.

Finally, an initial French *centroschematic* construction is replaced by a final English *monoschematic* structure:

- (25) (a) D'après les chiffres qui ont été fournis, la révolte a coûté la vie à 3.000 civils. (*lit.*, According to the figures which had been furnished, the rebellion cost the lives of 3,000 civilians.)
(b) The rebellion cost the lives of 3,000 civilians, a survey showed.

7 Conclusion

With the STYLISTIQUE grammars and parser, we now have the machinery to recognize and name stylistic differences between English and French style. What is needed next is the representation of *comparative stylistics*, the knowledge of how to map between the source-language internal stylistics to the target-language internal stylistics, *i.e.*, from the source-language stylistic grammar to the target-language stylistic grammar. The solution to this problem will necessarily involve the generation and editing of natural language text with specific stylistic constraints.

STYLISTIQUE could form the basis of a system that preserves or modifies style in translation. It could be integrated with a complete MT system or, alternatively, it could act as a post-editor to appropriately improve the output of a basic MT system. The incorporation of stylistic analysis into MT systems could significantly reduce the current reliance on human post-editing and improve the quality of MT output.

Acknowledgements

We gratefully acknowledge the contributions of Robert Cluett, Brian Fitch, Eduard Hovy, and Henry Schogt to the development of this work. Nathalie Japkowicz, Jean-François Lamy, Yves Lespérance, Hector Levesque, and Julie Payette helped with the French translations. The work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada.

References

- [DiMarco 1990] DiMarco, Chrysanne. *Computational stylistics for natural language translation*. Doctoral dissertation, published as technical report CSRI-239, Department of Computer Science, University of Toronto.
- [Duron 1963] Duron, Jacques. *Langue française, langue humaine*. Paris, Larousse.
- [Quirk et al 1985] Quirk, Randolph; Greenbaum, Sidney; Leech, Geoffrey; and Svartvik, Jan. *A comprehensive grammar of the English language*. Longman Group Limited.
- [Vinay and Darbelnet 1958] Vinay, J.P., and Darbelnet, J. *Stylistique comparée du français et de l'anglais*. Paris, Didier.