

Proposal for a Foreign Language Drafting Aid

Guy Lapalme, Elliott Macklovitch

Laboratoire RALI, Université de Montréal

Montréal, Canada

E-mail: lapalme@iro.umontreal.ca, macklovi@iro.umontreal.ca

Abstract

We examine a range of existing authoring aids (understood in the widest sense), from full-scale NLG to a new type of interactive MT. All of these depend on an independently available meaning representation, either a language-neutral formal model or a parallel text in another language. When no such meaning representation is available, however, current authoring aids can offer little more than word completion, as in augmentative communication systems or SMS. We consider the particular situation of a domain specialist who is called upon to draft a text in a second or a foreign language, and propose ways in which a new authoring tool might be able to assist him/her by suggesting multi-word completions, which should be more beneficial.

1. Introduction

The natural language generation (NLG) process is often separated into two phases: *What to say?* which corresponds to the building of a semantic representation, and *How to say it?* which chooses the words and sentences to convey the meaning encoded in that semantic representation. Although these phases have been defined primarily for engineering the modules of automatic text generators, they also have cognitive motivation and, informally at least, seem to correspond to the way humans plan and produce their communications.

Automated aids designed for human authors can also benefit from this division of the writing task. In certain situations where the intended meaning of a projected text is independently available, automated aids can be developed that provide the author with considerable assistance in generating the desired text. We will consider two such contexts in some detail. The first is where a non-linguistic, fully explicit, formal meaning representation is independently available. In this case, current NLG techniques allow for the automatic generation of certain types of texts directly from the data, and even in multiple languages. The other context is the case of translation, where the source text – a little like in painting by numbers – provides the semantic outline of the desired target text and the tones of the message to be transmitted. The problem with using a natural language text as a meaning representation is that it is rarely wholly unambiguous: not only do most source texts generally under-determine the meaning required to achieve a high-quality translation, but they often allow for multiple, equally acceptable translations. This remains the principal obstacle to fully automatic, high-quality machine translation (MT) today, and, despite impressive recent progress in this field, it is not likely to change for some time. Hence, where publication quality is a *sine qua non*, the only viable alternative is to retain the services of a human, either as a post-editor of the raw MT output, or as a translator who will hopefully have access to automated support tools. The former approach (post-editing) is not always cost-effective and humans often dislike it; the latter is generally referred to as machine-aided human translation (MAHT). In this paper, we will report on a

novel form of MAHT that proposes multi-word completions of the target text, which the translator may accept or reject in whole or in part.

However, where *no* meaning representation (linguistic or otherwise) is independently available – and this corresponds to the vast majority of text drafting situations today – current authoring aids can offer little more than simple word completion. The central question we want to raise in this paper is whether it would be possible to improve on this, and if so how.

2. Text Generation

Canadian weather bulletins provide an interesting example of the case where a fully specified, non-linguistic knowledge base is available, and this for a number of reasons. First, the text of these simple bulletins, which is derived from the numerical content of huge meteorological databases, is highly stereotypical in form. For many years, these short texts were manually drafted by expert meteorologists and then translated into English or French by human translators. Beginning in 1977, the MÉTÉO system (Chandioux & Guéraud, 1981) succeeded in largely removing the human translators from the loop,¹ thereby providing MT with one of its first major success stories; but the meteorologists were still required to draft the source texts that were automatically translated by MÉTÉO. More recently, however, NLG techniques have been developed that allow for the parallel generation, directly from the numerical databases, of a large subset of these weather bulletins in both English and French, thereby dispensing with the need for the manual drafting of the text in one language and its machine translation into the other. Originally developed over a decade ago (c.f. Goldberg et al., 1994), this technology has only recently come into full daily use at Environment Canada, and still not for all types of public weather forecasts.

It turns out, however, that the particular constellation of factors that make weather bulletins such an ideal application for both MT and NLG – the rigid and highly restricted sublanguage in which they are drafted; the large volume and steady rate at which they are published; and

¹ Some translators were still required to revise the system's output.

the independent existence of a formal database from which they are derived – is relatively rare, or at least in this combination. In most other cases where an underlying knowledge base exists independently of the text generation process, the type of texts required are rarely as entirely predictable as weather bulletins are. Hence, even in such sophisticated NLG systems as MDA (Lapalme et al., 2003), the interactive composition of a text describing a pharmaceutical product involves a fair amount of word (or concept) selection from menus that list permissible completions or extensions, according to a grammar and a pre-established semantic hierarchy. However, once the author's intentions have been fully specified, ideally in a language-independent manner, the system may then be able to generate parallel versions of the desired text in multiple languages. But the situation here is quite different from that of weather bulletins, if only because there isn't a constantly evolving demand for this kind of pharmaceutical documentation.²

Hartley & Paris (1997) argue that in certain high-tech industries where large volumes of multilingual documentation are required, a trend has begun to emerge which eschews machine translation in favour of multilingual NLG. They also point to a concomitant blurring of the distinction between a translator and a technical writer, whereby a major part of what they call the *authoring* task now involves “the specification of the conceptual model (or message) underlying task-oriented instructions from which the documentation itself can be written.” (p.115) Hartley & Paris convincingly demonstrate the advantages that this approach offers over sequential translation from a source text, although they admit that these advantages crucially hinge on the availability of an underlying knowledge base. “If such a knowledge base already exists or can be constructed automatically, then multilingual generation is a viable option for producing multilingual documents.” (p.117) And indeed, there has been much exciting work in NLG over the last decade that is devoted to precisely this problem, i.e. finding ways to assist domain experts who do not have the skills or the time to master formal knowledge representation systems; see in particular the various implementations of the *what you see is what you mean* approach, e.g. (Power & Scott, 1998). To call this kind of work *authoring*, however, seems somewhat abusive, or at least at variance with the most widespread definition of this verb, which (according to Encarta's online dictionary) means, “to write or be responsible for the final form of a book, report, or other **text**” (our emphasis). Be that as it may, in the great majority of text drafting situations, a fully specified knowledge base is not available and the construction of one is simply not a viable option; hence, full-scale NLG is just not applicable in these situations.

3. Predictive Text Technology

But even when a full semantic model is not available, it may still be possible to help a person draft a text by guessing what she intends to write based on what she has

already keyed in. The idea of automatically extending, or completing the tokens (if not the full text) that a user is typing has a long history that goes back to the earliest UNIX systems, which included both command-name and file-name completion. The success of the operation here clearly depends on there being a relatively small number of items for the completion program to select among, exploiting a very simple heuristic: use the list of available command names when typing at the start of the line and the list of available filenames otherwise. This principle also applies to interactive development environments (e.g. Eclipse and Visual Studio) or XML editors (e.g. oXygen and XML Spy), which suggest correct completions given the context of the piece of text the user is drafting. The semantic model in this case is more complex because it involves the parsing of both the program text the user is writing, as well as previous program parts. This is possible because both the text and its environment are defined in the context of artificial languages which have a rigid grammar and a well-defined semantics.

The widespread popularity of SMS (short message services) on mobile phones constitutes another, extremely successful application of predictive text technology. Owing to the telephone's limited keypad, every sequence of key presses that is meant to convey a word is necessarily ambiguous; and here, the great leap forward came with the addition of a monolingual dictionary and statistical language models, which together allow many of the words in the message to be automatically disambiguated. In fact, today's SMS systems can often complete a word before the user has finished typing it, and they will reorder the completions in the suggestion box with each new key the user enters. Some systems now propose multiple word expressions and others can even acquire the user's preferred vocabulary from past messages or files.

Predictive text technology is also used to assist people with various types of learning disabilities or those who, for whatever reason, feel intimidated or lack the requisite language skills to compose a written text (even in their native tongue). For example, the Canadian company Quillsoft has developed the WordQ program, a plug-in that inserts a small pop-up in a word processing or text editing window in which completions are suggested for the current word; and after each word, or at the end of the sentence, the software uses speech synthesis to read the completed unit back to the user. For children with impaired writing skills whose spoken vocabulary far exceeds their written vocabulary, this kind of automated word completion followed by synthesized read-back helps to reinforce links between the two forms of language. At the other end of the disability spectrum, there are augmentative communication systems designed for people with severe motor handicaps, for whom every keystroke demands considerable physical effort. For example, the Sibylle system (Wandmacher et al., 2007) offers two types of assistance to its users: an on-screen keyboard, which is dynamically reconfigured so as to minimize the effort required to select the next letter of a word; and another completion window in which predictions of the following word are proposed on the basis of those that have preceded. Because spelling and other input errors are common in this particular context, Sibylle's word completion system is not based on a standard frequential dictionary, like the other word

² For any given region in Canada, a new public weather bulletin is published four times a day, every day of the year. Documentation on a pharmaceutical product will normally be published once and will last for the life of the product.

completion systems discussed so far, but rather on an n-gram model that operates on the level of characters; and a similar model is used on the sentence level to calculate the word predictions.

4. Machine translation

What distinguishes the process of translation from other acts of writing is that the translator works simultaneously with not one but two texts: a source text in one language and a target text that he is drafting in a second language. A central part of the translator's job is to ensure the semantic equivalence of these two texts. However, the availability of the source text makes it possible to envisage novel kinds of authoring aids that are not possible in other, monolingual contexts.

As we previously noted, fully automatic machine translation is still not capable of systematically producing publication-quality translations. In fact, until recently, MT output was not generally considered revisable by most translators, but only suited for gisting purposes. Machine-aided human translation (MAHT) eliminates the MT component, leaving the translation to the human translator, while attempting to provide her with reliable tools that will hopefully increase productivity. However, surprisingly few types of aids have been proposed for MAHT thus far: basically, tools for terminology management and repetitions processors (more commonly known as translation memories). The problem with the former is that its impact on productivity is minimal; the problem with the latter is that the great majority of texts do not contain a sufficient level of full-sentence repetition to make them cost-efficient.

For some years now, the RALI laboratory has been developing the TransType system (Foster et al., 1997; Langlais et al., 2004), a novel approach to translation automation that aims to strike a compromise between MAHT and fully automatic MT. In contrast to MAHT, it retains an MT component; it thus has a greater potential to increase translator productivity than current MAHT tools. Unlike fully automatic MT, however, it doesn't allow the system to operate with complete autonomy, but instead seeks to conjugate the power of the machine with the discernment of the human translator, in a way that allows the two to productively collaborate in the text drafting process.

TransType represents a new kind of interactive MT in which the focus of the interaction is on the target text, unlike classic interactive systems, where the aim of the human intervention is to disambiguate the source text; c.f. (Kay, 1973). In TransType, a statistical MT system is embedded within an interactive editing environment. The MT system's contributions to the target text take the form of predicted extensions to the current linguistic unit, based upon its automatic translation of the source text and a target language model. (See Figure 1 at the end of this paper for a snapshot of a TransType session.) TransType is intended to be used by a professional translator whose contribution to this joint undertaking is to serve as guarantor of the quality of the target text being drafted; as such, she must evaluate the system's proposed extensions and either accept or reject them, in whole or in part. To reject a proposed completion, the translator simply continues typing the target text she has in mind. However, TransType can immediately take advantage of this

additional, reliable information on the desired target in order to propose an alternative (and hopefully better) completion, which the user may again accept or reject; and so on, until a fully acceptable translation has been composed. Obviously, the more accurate the system's predicted completions, the less the user will have to intervene, which in principle should translate into increased productivity gains. For more on the results of extensive user trials with TransType, see (Macklovitch, 2006).

A major advantage of this interactive-predictive approach is that it actively involves the human translator in the text generation process, doing that for which she is most qualified, i.e. translating, as opposed to performing linguistic analyses, or post-editing (i.e. correcting) raw MT output. TransType thus avoids one of the principal objections that translators have long harboured against MT, to wit, that it places them in a role of subservience to a machine whose comprehension of natural language is considerably less than theirs.

5. Drafting in a Foreign Language

Let us now return to the situation where no meaning representation is available and where, as we have seen, current authoring aids can offer little more than simple word completion. Needless to say, for people with serious learning disabilities or physical handicaps, there is absolutely nothing wrong with word completion; on the contrary, it can appreciably facilitate text entry for them. However, in the absence of such obstacles, word completion can be rather frustrating for an "ordinary" person who is attempting to draft a text. Simply put, the time and effort required to evaluate the list of proposed completions usually outweigh the benefit of not having to type a few characters. This at least is the case for most moderately literate monolinguals.³

But what about people who are called upon to draft texts in a second or a foreign language? In the domain of scientific research, for example, English has become the lingua franca for international conferences and scientific publications and, as a consequence, many researchers whose mother tongue is not English are obliged to write technical articles in that language. It is our contention that authors in this situation might be more tolerant of word completion than native English speakers. At the very least, it could help them avoid certain errors of spelling and grammar during the text drafting process.

What sorts of errors do authors in this situation typically commit? (And again, we're focusing here on people who may have considerable competence in the foreign language, although this will tend to be more passive than active competence, i.e. it derives principally from reading texts in their domain of technical specialization.) An Anglophone drafting in French, for example, will often make errors of gender agreement, either selecting the incorrect article for a masculine or feminine noun, or

³ Contrary to what is suggested in the CFP for this workshop, ordinary text composition does not strike us as an undertaking that is generally fraught with peril. Writing is certainly a complex task, but educated adults are not often overwhelmed by these complexities, and when they do lose their orientation, they usually manage to recover it. Which is not to say that the texts they author are always stellar or models of clarity.

failing to properly inflect a noun modifier. Such errors arise, of course, because nouns are not marked for gender in English. Another major difference between the two languages is the high degree of inflection in the French verbal system; this too will result in Anglophones committing not just more errors in their French texts, but also different types of errors than those found in the texts of Francophones. On the other hand, a Francophone drafting in English will frequently have trouble with the null article, which English uses to denote generic reference (among other things). Since French does not generally allow nouns with no article, Francophones will tend to insert one where no English speaker ever would, e.g. *'she loves the classical music'*. Both Anglophones and Francophones will often err in their selection of governed prepositions, since these can be wholly unpredictable and do not directly translate from one language to the other. So, for example, no Anglophone would ever use the preposition 'of' to introduce the complement of the verb 'depend', although one could well find this in a text drafted by a Francophone; and conversely, no self-respecting Francophone would ever write *'dépendre sur'* (unless under the pernicious influence of an English calque). Note that some of these errors might be detected *a posteriori* by a sophisticated monolingual grammar checker.⁴ What we're suggesting here, however, is that it might be more convenient for people drafting in a foreign language to have a specialized foreign language drafting tool that would hopefully avoid the generation of such errors in the first place.

Such a system would naturally be more helpful if it were able to go beyond simple word completion and propose multi-word or even full-sentence completions, as the TransType system does. In fact, a series of extended user trials with TransType led us to fundamentally modify our approach to text completion. Instead of systematically predicting a fixed amount of text after each character the user types, later versions of the system attempted to make predictions that would maximize the expected benefit to the user in each context; c.f. (Foster et al., 2002). This expected benefit was estimated from two components: a statistical translation model which calculated the probability that a predicted completion would be correct or incorrect, and a user model which factored in (among other things) the time required to read a more or less lengthy prediction. In addition, later versions of TransType allowed the user to adjust its predictions to her preferences and taste, e.g. by setting a confidence threshold, or fixing a minimal length for a completion, or a maximum number of alternate completions. Moreover, if on a given text the system's predictions were wholly unsatisfactory (for whatever reason), the user could always turn off the prediction engine and type the translation on her own. We suggest that this same kind of flexibility should be retained for monolingual authoring aids that attempt to suggest completions that extend beyond the current word.

Of course, the crucial question that remains is this: In the absence of an underlying meaning representation or a parallel text in another language, what would be the

source of such multi-word completions? How likely is it that general-purpose language models trained on large, publicly available corpora (of journalese or parliamentary debates, for example) will be able to propose extended completions that match an author's intentions? Not very likely, in our view. Our experience with TransType suggests that the predictions would be too diffuse and would require too many interactions before they manage to zero in on the sentence the author has in mind. However, the situation that we evoked above, i.e. of a person drafting a specialized text in a foreign language, is slightly different and would not require that the predictor rely exclusively on general-purpose texts for its training material. Here, ideally, we would prefer to train the system's language model on a corpus composed of native speakers' texts taken from the same specialized domain that the author is drafting in. Moreover, our foreign language drafting tool could perhaps take a cue from some of the simpler word completion systems described above and allow the user to explicitly select certain multi-word expressions in this corpus so that they are added to the system's dictionary. In the domain of computational linguistics, for example, a non-English speaking user might flag such multi-word terms as 'cross-language information retrieval' or 'statistical machine translation', thereby allowing the system to propose their completion after only a few characters.

Another potential advantage of this interactive-predictive approach is that it could *learn* from its interactions with the user, i.e. from the completions the user has accepted and those that have been rejected, in order to improve its underlying model. Most word completion systems adapt their predictions to new input from the user, to ensure their compatibility with the input prefix. However, what we're proposing here is not simply adapting the predictions of a fixed model, but rather updating the parameters of that model on the fly so that it would actually improve with use. We have done some initial experiments in this vein (Nepveu, 2004) using a cached language model with the TransType system, and the gains, though modest, were encouraging.

In short, this kind of interactive text prediction would seem to offer interesting ways of harnessing the power of language models for the purposes of foreign language drafting. Would the additional restrictions that come with this application – i.e. highly targeted training material and a more tolerant author drafting in a second language – suffice to make an interactive-predictive text completion system beneficial for such a user? In particular, would the specialized language models allow for accurate multi-word or phrasal predictions? It is difficult to say without actually implementing the system and then carrying out tests. However, we find the idea sufficiently intriguing to want to undertake the development of such a system. Anyone care to join us?

6. References

- Chandioux, J. & Guéraud, M-F. (1981). MÉTÉO : un système à l'épreuve du temps. *META*, 26(1), pp. 18--22.
- Goldberg, E., Driedgar, N., Kittredge, R. (1994). Using natural-language processing to produce weather forecasts. *IEEE Expert*, 9(2), pp. 45--53.

⁴ Both Antidote and the grammar checker in Word, for example, are quite good at detecting errors of agreement in French, although their performance is less impressive on governed prepositions.

- Foster, G., Isabelle, P., Plamondon, P. (1997). Target-Text Mediated Interactive Machine Translation. *Machine Translation*, 12(1-2), pp. 175--194.
- Foster, G., Langlais, P., Lapalme, G. (2002). User-Friendly Text Prediction for Translators. In Proceedings of *EMNLP 2002*, Philadelphia, PA., pp. 148--155.
- Hartley, A. & Paris, C. (1997). Multilingual Document Production: From Support for Translating to Support for Authoring. *Machine Translation*, 12(1-2), pp.109--128.
- Kay, M. (1973). The MIND System. In R. Rustin (ed.) *Natural Language Processing*, New York, N.Y.: Algorithmics Press, pp. 155--188.
- Langlais, P., Lapalme, G., Loranger, M. (2004). TransType: Development-Evaluation Cycles to Boost Translator's Productivity. *Machine Translation* (Special Issue on Embedded Machine Translation Systems), 17 (2), pp. 77--98.
- Lapalme, G., Brun, C., Dymetman, M. (2003). MDA-XML : une expérience de rédaction contrôlée multilingue basée sur XML. In Proceedings of *TALN 2003*, Batz-sur-Mer, France, pp. 379--384.
- Macklovitch, E. (2006). TransType2: The Last Word. In Proceedings of *LREC-2006*, Genoa, Italy pp. 167--172.
- Nepveu, L. (2004). Adaptation de modèles de traduction dans le cadre du projet TransType. Mémoire de maîtrise en Informatique, Université de Montréal.
- Power, R. & Scott, D. (1998). Multilingual authoring using feedback texts. In Proceedings of *COLING-ACL 1998*, Montreal, Canada, pp. 1053--1059.
- Reiter, E. & Dale, R. (1997). Building applied natural language generation systems. *Natural Language Engineering*, 3(1), pp. 57--87.
- Wandmacher, T., Béchet, N., Barhoumi, Z., Poirier, F., Antoine, J-Y. (2007). Système Sibylle d'aide à la communication pour personnes handicapés. In Proceedings of *TALN 2007*, Toulouse, France, pp.539--548.

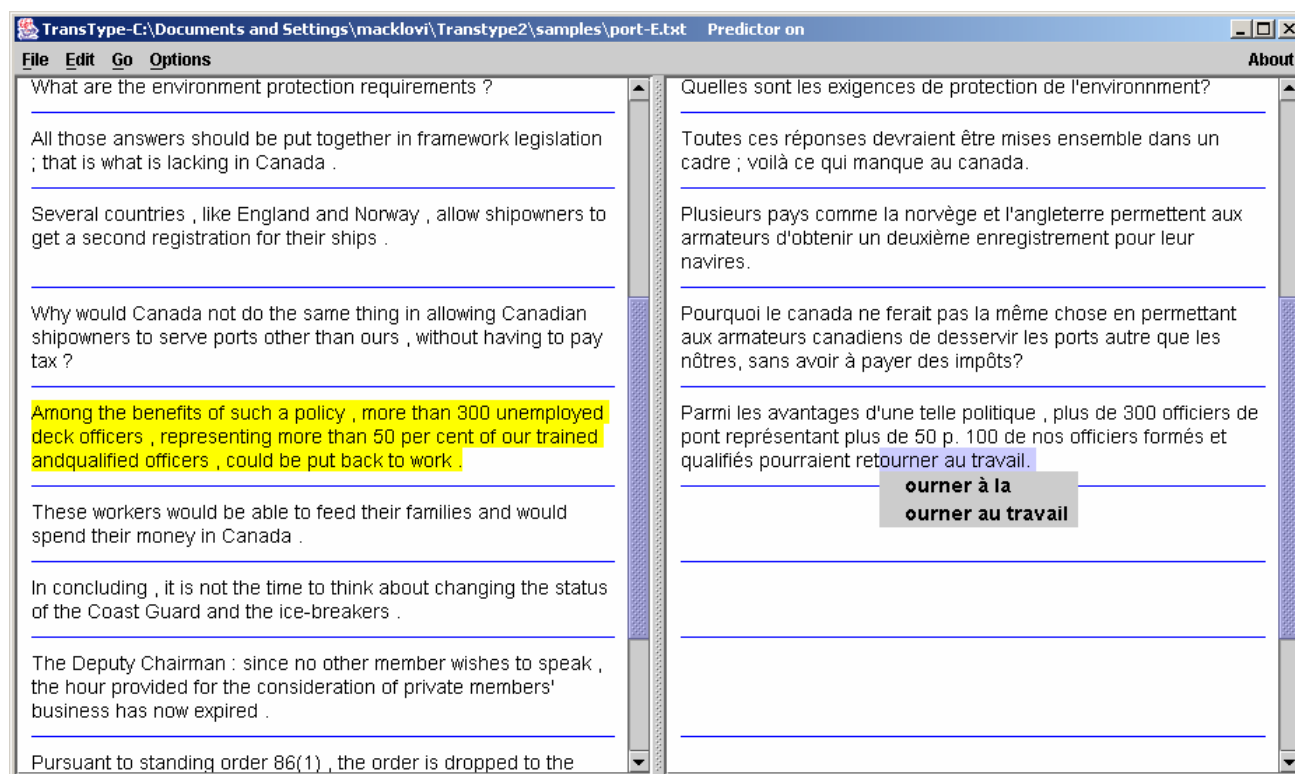


Figure 1 : Snapshot of TransType session

The English source text appears in the left pane; on the right is the French translation currently being drafted. To assist the human translator, TransType proposes completions to the target translation, generated on the fly, using a statistical translation model combined with target language model. Here, three completions are suggested: to accept the top one, the translator need only hit the Enter key; the others can be inserted into the text, in whole or in part, by pointing and clicking with the mouse. Other versions of TransType are capable of proposing longer predictions, up to full-sentence length.