Current Research in Machine Translation: A Reply to Harry Somers

by
Pierre Isabelle
CITI
1575 Chomedey Blvd., Laval, Quebec, Canada H7V 2X2
isabelle@citi.doc.ca

1. Introduction

The gist of Somers' paper can be summarized as follows. The classical second-generation (G2) machine translation system architecture is so flawed with 'general or specific deficiencies' that it has not and will not be successful in producing practical MT systems. Most current research efforts are still corrupted by these erroneous principles. A small number of projects reject the G2 orthodoxy, which makes them better candidates for success.

Somers correctly characterizes the key feature of the G2 approach as 'modularity' along the following axes: 1) different linguistic levels such as morphology, syntax, and semantics are described separately ('stratificational' linguistic models); 2) source language, target language (and possibly, contrastive) descriptions appear in different modules: analysis, generation (and possibly transfer); and 3) linguistic descriptions are kept separate from processing algorithms. He then proceeds to argue that this modular design is responsible for a number of fundamental shortcomings.

2. Alleged deficiencies of G2 systems

Is Modularity Outdated?

The first purported deficiency is that G2 design reflects old-fashioned computational and linguistic techniques.

On the computational side, Somers sees the G2 architecture as strongly connected with the 'procedural algorithmic strictly-typed programming style'. I fail to see any such connection. In fact, just the opposite appears to be true. For example, logic programming turns out to facilitate the development of highly modular descriptions (Isabelle & Macklovitch [5]), which can give rise to *reversible* G2-type MT systems (Isabelle,Dymetman & Macklovitch [4], Dymetman, Isabelle & Perrault [3]) .

On the linguistic side, Somers associates stratificational models with the outdated transformational-generative (TG) model of grammar. I don't believe this connection to be a necessary one. Though the particular number and precise definition of the strata may be quite different from those of TG, many of the more recent linguistic models do rely in a fundamental way on some form of stratification. An obvious example is LFG [6], but even HPSG is nowadays described as a stratificational theory.

Stratification and Bottom-up Thinking

The second claimed deficiency is that stratificational linguistic models are somehow the source of a perverse bottom-up way of approaching MT. Namely, an approach in which the system designer tends to concentrate on the lower strata (morphology, syntax) until he/she runs out of ideas. Then he/she formulates some ad hoc transfer and generation schemes that try to make do with whatever the outcome of the first stage was. This is contrasted to a healthier approach in which one works one's way down from carefully defined intermediate representations. I fully agree with Somers that some top-down thinking is called for. However, this is not an argument against stratificational models per se, but a criticism of certain practices within the stratificational paradigm. The preferred top down approach actually presupposes the existence of some upper stratum to work down from.

Stratification, Compositionality and Literality

Another alleged deficiency of stratificational models is that, in view of their commitment to compositionality, they would seemingly encourage 'structure preserving translation as a first choice'. Somers' elliptical argument appears to involve at least three different issues:

1) Stratificational linguistic models (SLM's) are committed to compositional translation. Somers' point is unclear because, generally speaking, SLM's are models of natural languages, not translation models. GPSG, for instance, features a compositional syntax/semantics mapping, but it was never intended as a theory of translation. However, a few attempts have been made to extend some linguistic theories in the relevant direction. For example, Kaplan & al. [7] propose to build into LFG 'co-description' mechanisms which make it possible to define an LFG-based translation model featuring both: a) 'vertical' compositionality between different strata of the same language; and b) 'horizontal' compositionality between pairs of corresponding strata in two different languages. Somers does not make clear whether or not this is the sort of thing he has in mind.

2) Compositionality encourages literal translation. A translation model is compositional if the translation of any complex object is a function of the translation of its parts. This in itself does not preclude the production of non-literal translations. It would do so only if we stipulated that: a) the translation of any complex object must be a TRIVIAL function of the translation of its parts; and b) the objects so related belong to language-specific representation systems. No one would want to argue for such an extreme form of compositionality. On the other hand, if translation is to be modelled by some finite procedure, then some milder form of compositionality *must* be at play. Thus, the real questions are: (i) how complex is the language-to-language compositionality function? and (ii) how much language-independence is there in the objects that it relates? Somers admits that he 'does not have a ready solution here'.

3) For human translators, literal translation is always a last resort. In the eternal debate between 'literal' and 'free' translation (Larose [9]), Somers apparently choses to militate for the latter. The issues underlying this debate are exceedingly thorny and frustrating. Fortunately, particular theoretical postures on these issues appear largely orthogonal to automation issues. In my opinion, the responsibility of MT practitioners is not to side with either camp, but rather to aim for systems capable of producing the full range of translation alternatives. Still, it seems obvious that free translations will be even harder for machines to attain than literal ones and it does not seem very sensible to start with the harder problems. At any rate, Somers does not offer any real clue as to how to make computers less literal-minded.

In summary, then, I don't believe that Somers has really mounted a serious case against linguistic stratification and compositional translation.

Second-generation systems don't work

Somers' last blow against G2 systems is the observation that they do not work very much better than first-generation (G1) systems. It is obviously true that G2 systems aren't doing well on the marketplace. But then, no MT technology is doing really better on that count. Considering the full size of the translation market, MT is at best marginal. It is clearly false that G1 systems are selling well. Few translation services remain committed to any MT system for longer than a couple of years. The exceptions would seem to be: 1) those who benefited from a *custom-built* system (e.g. TAUM-METEO); and 2) those who acquired enough MT expertise to extensively *customize* some general-purpose MT system.

Because of their greater modularity, G2 systems are in principle more open to customization, but customizing them requires a high degree of expertise. My bet is that over the next ten or fifteen years, custom-built or customized G2 systems will make some modest gains in real-life applications.

3. The G2 atavism

Somers argues that much of current MT research is rather hopeless, for it remains committed to the G2 principles he has attempted to discredit.

According to him, the knowledge-based (KBMT) approach pursued at CMU focuses on problems like discourse pragmatics and the use of domain knowledge 'as viewed from a traditional viewpoint'. That is, KBMT endorses to some extent the G2 architecture (e.g. a stratificational linguistic model). I suspect that Messrs. Nirenburg and Carbonell would disagree with this view since they have tended to present KBMT as something radically different from G2 systems. However, in recent years KBMT practitioners have moved away from their previous radical anti-syntax position, narrowing the gap with G2-type systems. Somers does not altogether condemn KBMT: he simply (and rightly to my mind) points out that the step up from toy implementation to practical systems is a huge one, and that it is not clear that the results would warrant the investment.

Somers views as equally misguided another large section of current MT research: the one he characterizes as being in quest of 'better linguistic theories'. This characterization is meant to pick up a rather broad variety of MT efforts, including systems like EUROTRA (based on an eclectic linguistic theory) and ROSETTA (based on Montague grammar), a variety of systems based on such theories as categorial grammars, GPSG, LFG, or HPSG, and another group based on 'logic grammars'. His main criticism is that all of the fancy linguistic models inevitably crack under real-life conditions. He is unfortunately right: no existing theory is very good at accounting for real-life linguistic data, and many of the better existing systems are based on rather silly theories. Somers appears to imply that the search for better linguistic theories is irrelevant for MT. Again, I disagree. The better

G1 systems are not doing well at all, and a sizable proportion of their failures is a direct consequence of their inadequate linguistic models. Existing G2 systems sometimes do worse, only because their deeper linguistic analyses have so far been achieved at the expense of *breadth of coverage* and *robustness*. These two issues need to be addressed, but abandoning modularity is clearly not the way to do it.

Somers then mentions that the ‘sublanguage approach’ raises several theoretical issues that need to be examined. For example, we need a better definition of the term sublanguage and we need to better understand the relation of sublanguages to the ‘core grammar’. As an application strategy, Somers has no qualms with sublanguage-specific systems, and, of course, I fully agree with him.

4. Alternative avenues

In the final section of his paper, Somers examines a number of alternative avenues that he views as more promising because, in his view, they reject G2 orthodoxy.

Sublanguage Plus

One of the current UMIST projects concerns ‘a system with which individual sublanguage MT systems can be created, on the basis of a bilingual corpus of typical texts’. The system comprises an application-independant MT engine, just like in the G2 approach. In addition, it includes an interactive component meant to help a computational linguist develop specific grammars and dictionnaries, on the basis of statistical analyses of text corpora.

This project does not in any way appear to reject G2 principles. On the contrary, it deals with methods for industrializing the development of G2 sublanguage-specific systems. The practical value of large text corpora, concordances and frequency lists for developing MT systems has always been acknowledged by MT practitioners. But it may well be true that more sophiticated data extraction methods are called for. Still, a G2 system remains a G2 system no matter what the development methods are, and the ‘Sublanguage Plus’ approach endorses G2 principles rather than rejecting them.

Automatic Grammar Updating

A second 'alternative avenue' discussed by Somers is the development of MT systems that can learn from corrections made to their output by a human post-editor. I fully agree with Somers that the learning ability is of the utmost importance. However, the obvious problem with learning-oriented research in NLP is that it is pretty difficult to find out how we manage to learn before we have a better idea of just what it is we learn. In practice, one is forced to assume some linguistic (or translation) model which is known to be severely flawed. In the particular project mentioned by Somers (the PECOF project), the assumptions are precisely those that define G2 systems. Just as in the previous case, the focus of the project is not on some alternative to G2 principles, but rather on an alternative way of injecting knowledge into a G2 system.

Dialogue MT

The third 'alternative avenue', dialogue MT, actually goes off in two different directions. This is because the expression 'dialogue MT' is meant to be ambiguous between: 1) MT *by means of dialogues*, that is, using dialogues with the user to improve the performance of the system; and 2) MT *of dialogues*. It is not clear what is gained by the ambiguity, since these two dimensions appear to be orthogonal.

The standard term for the first sense is interactive translation, and the concept has been around for quite a while (cf. Martin Kay's MIND system and the ALPS Transactive MT system). Interaction with a user is just a particular way for a system to access the knowledge it needs to perform some task. In itself, the notion has little to do with the structure of the knowledge being acquired. Typical G2 systems, like any other type of MT system, can resort to interaction. In fact, as far as I can tell, the interactive systems mentioned by Somers do adhere to G2 principles.

As for the other sense of the tern 'dialogue MT', translation of dialogues, there has so far been relatively little work on the topic. It is certainly a worthwhile endeavor, for it is sure to bring forward a number of issues that are absent or unimportant in narrative text. But again, choosing to apply MT to dialogues, rather than to some other type of text, cannot be viewed as an alternative to G2 principles.

Corpus-based MT

The research avenues that we have thus far reviewed do not appear to be targeting genuine alternatives to G2 principles. However, Somers mentions one more trend, corpus-based MT (CBMT), which may at first glance look like a more serious contender, since it explicitly rejects the *rationalist* viewpoint that has pervaded most of the MT research of the last two decades, including most work on G2 systems. According to this now traditional approach, it is the linguist who is seen as the primary source of the linguistic knowledge to be injected into an MT system. In formulating his/her linguistic descriptions, the linguist freely appeals to his/her own intuitions about linguistic structure. Corpora, concordances and frequency lists are more or less optional tools.

By contrast, CBMT adopts an *empiricist* viewpoint. The primary source of linguistic knowledge resides in the linguistic data itself, that is, bilingual text corpora. The role of the linguist is deemphasized, or even abolished. Somers mentions two different varieties of CBMT research work.

The first one is the *probabilistic* approach of Brown & al. [1] at IBM. Under this approach, probabilistic techniques are used to automatically extract from text corpora the ‘paramaters’ needed to fill in the substance of some ‘language model’ and some ‘translation model’, which together constitute the core of an MT system. Since the focus of their research was knowledge extraction techniques, Brown & al. [1] used rather simplistic models to start with (i.e. an *n-gram* linguistic model). However, one can very well imagine applying a similar approach to the specification of more sophisticated linguistic models.

The second variety of CBMT mentioned by Somers is the *memory-based MT* of Sato & Nagao [10]. The authors examine different means of extracting translational equivalences from a database made up of annotated pre-translated texts. The annotations amount to an explicit specification of compositionality of the translation relationship between these texts. [1] Sato & Nagao then look at mechanisms that would make it possible to translate new texts on the basis of *analogies* with the pre-translated ones.

Somers views the two varieties of CBMT as somehow rejecting the G2 principles of modularity. I disagree. What distinguishes CBMT from more traditional approaches is its empiricist stance, not the architecture of the resulting linguistic descriptions. Nothing

1. More specifically, sentences of the SL and TL texts are represented by dependency trees. These trees and their subtrees are then paired by a *translate(SL_Node, TL_Node)* relation.

appears to preclude the application of empiricist techniques to the development of G2 systems. The IBM researchers now appear to be turning to a system architecture which is basically that of G2 transfer-based systems (Brown et al. [2]). As for the memory-based work of Sato & Nagao, it aims at automatically extracting from a bilingual corpora a set of translational equivalence statements. The resulting statements are meant to constitute no more and no less than the transfer component of a typical G2 system. Whether they are generated online during some translation task or precompiled is immaterial to the discussion.

To sum up, CBMT does not constitue an alternative to G2 principles. It can however constitute an alternative way of building G2 systems.

5. Conclusion

Somers' paper falls short of its declared aim of demonstrating that: 1) G2 principles of modularity are inadequate; 2) current knowledge-based and linguistics-based MT research is wrong in adhering to these modularity principles; and 3) clearcut alternatives to these principles have now emerged.

Finally, I would like to point out that Somers' review of current MT research contains some important omissions. For instance, it seems to me that a comprehensive review of the MT scene can no longer ignore some recent results in *multilingual text generation* (Kittredge [8]): in certain situations, generation clearly has an edge over MT.

6. Références

[1] Brown P., Cocke J., Della Pietra S., Della Pietra V., Jelinek F., Lafferty J., Mercer R., Roosin P., A Statistical Approach to Machine Translation, **Computational Linguistics**, 16 :2, 79-85, 1990.

[2] Brown P., Della Pietra S., Della Pietra V., Lafferty J., Mercer R. 1992. Analysis, Statistical Transfer and Synthesis in Machine Translation, **Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation** (TMI-92), Montreal.

[3] Dymetman M., Isabelle P. & Perrault F., A Symmetrical Approach to Parsing and Generation, in **Proceedings of the 13th International Conference on Computational Linguistics** (COLING-90), Helsinki, 1990.

[4] Isabelle P., Dymetman M. & Macklovitch E., CRITTER: a Translation System for Agricultural Market Reports, **Proceedings of the 12th International Conference on Computational Linguistics** (COLING-88), Budapest, 1988.

[5] Isabelle P., Macklovitch E., Transfer and Machine Translation Modularity, **Proceedings of the 11th International Conference on Computational Linguistics** (COLING-86), Bonn, 1986.

[6] Kaplan R. & Bresnan J., **Lexical Functional Grammar: A Formal System for Grammatical Representation**, in J. Bresnan (1982a), 1982, 173-281.

[7] Kaplan R., Netter K., Wedekind J., Zaenen A., Translation by Structural Correspondences, **Proceedings of the Fourth European ACL Conference**, Manchester, 1989.

[8] Kittredge R., Multilingual Text Generation as an Alternative to Machine Translation, **Proceedings of the 30th Annual Conference of the American Translators Association**, Learned Information, 1989, 465-69.

[9] Larose R., **Paramètres d'évaluation des traductions: théories contemporaines et approche textologique**, thèse de Ph.D., Université de Montréal, 1984.

[10] Sato S., Nagao M., Toward Memory-Based Translation, **Proceedings of COLING-90**, 247-252, 1990.