TransSearch: What are translators looking for?

Elliott Macklovitch, Guy Lapalme, Fabrizio Gotti
Laboratoire RALI
Université de Montréal
Montreal, Canada
{macklovi,lapalme,gottif}@iro.umontreal.ca

Abstract

Notwithstanding machine translation's impressive progress over the last decade, many translators remain convinced that the output of even the best MT systems is not sufficient to facilitate the production of publication-quality texts. To increase their productivity they turn instead to translator support tools. We examine the use of one such tool: *TransSearch*, an online bilingual concordancer. From the millions of requests stored in the system's logs over a 6-year period, we extracted and analyzed the most frequently submitted queries, in an effort to characterize the kinds of problems for which translators turn to this system for help. What we discover, somewhat surprisingly, is that our system seems particularly well-suited to help translate highly polysemous adverbials and prepositional phrases.

1 Introduction

TransSearch (henceforth TS) is a Web-based, bilingual concordancer that allows its users to query large databases of past translations in order to find ready-made solutions to a host of translation problems. The system was first developed in the early 1990s at the CITI, an Industry Canada research centre, as one illustration of the practical applications that derive from the then-novel concept of translation analysis; see (Isabelle et al., 1993). Re-christened TSrali and now administered by Terminotix, a private sector partner of the RALI, the service now boasts about two thousand regular users, the majority of whom are Canadian translators working between English and French; but there are also a fair number of regular users, both freelancers and translation services, outside of Canada. Information on the current translation databases which TS subscribers have access to is given in Table 1 below. More detailed information is available on the TSrali website.[1]

Corpus	words (M)
Canadian Hansard (1986-2007)	273
Canadian court rulings (1986-2007)	92
Canadian Senate (1996-2007)	26
International Labour Org (Eng-Fr)	44
International Labour Org (Eng-Sp)	37
International Labour Org (Fr-Sp)	36
TOTAL	508

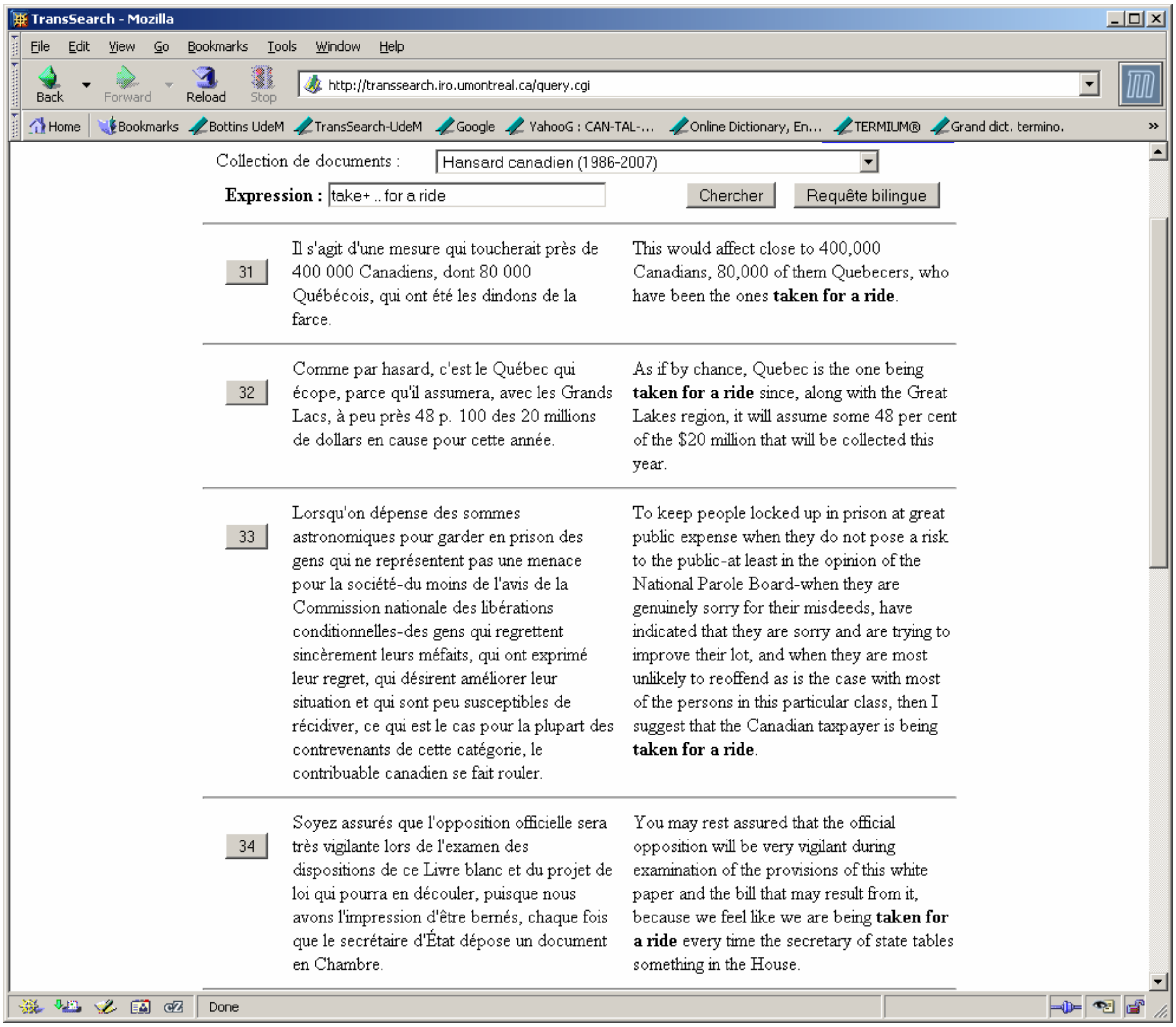
Table 1: Size of current translation databases

Figure 1 on the next page provides a snapshot of a typical user session with TS. As mentioned above, the system is accessed over the Internet; the databases reside on a centralized server and users submit their queries using a Web browser. TSrali processes thousands of such queries every day; Table 2 shows the number of queries submitted each month between October 2006 and September 2007.

Month	# Queries	Month	# Queries
Oct-06	206,986	Apr-07	180,647
Nov-06	218,729	May-07	210,472
Dec-06	138,701	Jun-07	189,191
Jan-07	167,501	Jul-07	130,726
Feb-07	196,542	Aug-07	143,521
Mar-07	204,286	Sep-07	137,336

Table 2: Queries per month

[1] http://www.tsrali.com/


TransSearch - Mozilla
File Edit View Go Bookmarks Tools Window Help
Back Forward Reload Stop
http://transsearch.iro.umontreal.ca/query.cgi
Home Bookmarks Bottins UdeM TransSearch-UdeM Google YahooG : CAN-TAL-... Online Dictionary, En... TERMIUM® Grand dict. termino.
Chercher
Requête bilingue
Done

Collection de documents : Hansard canadien (1986-2007)

Expression : take+ .. for a ride

31	Il s'agit d'une mesure qui toucherait près de 400 000 Canadiens, dont 80 000 Québécois, qui ont été les dindons de la farce.	This would affect close to 400,000 Canadians, 80,000 of them Quebecers, who have been the ones **taken for a ride**.
32	Comme par hasard, c'est le Québec qui écope, parce qu'il assumera, avec les Grands Lacs, à peu près 48 p. 100 des 20 millions de dollars en cause pour cette année.	As if by chance, Quebec is the one being **taken for a ride** since, along with the Great Lakes region, it will assume some 48 per cent of the $20 million that will be collected this year.
33	Lorsqu'on dépense des sommes astronomiques pour garder en prison des gens qui ne représentent pas une menace pour la société-du moins de l'avis de la Commission nationale des libérations conditionnelles-des gens qui regrettent sincèrement leurs méfaits, qui ont exprimé leur regret, qui désirent améliorer leur situation et qui sont peu susceptibles de récidiver, ce qui est le cas pour la plupart des contrevenants de cette catégorie, le contribuable canadien se fait rouler.	To keep people locked up in prison at great public expense when they do not pose a risk to the public-at least in the opinion of the National Parole Board-when they are genuinely sorry for their misdeeds, have indicated that they are sorry and are trying to improve their lot, and when they are most unlikely to reoffend as is the case with most of the persons in this particular class, then I suggest that the Canadian taxpayer is being **taken for a ride**.
34	Soyez assurés que l'opposition officielle sera très vigilante lors de l'examen des dispositions de ce Livre blanc et du projet de loi qui pourra en découler, puisque nous avons l'impression d'être bernés, chaque fois que le secrétaire d'État dépose un document en Chambre.	You may rest assured that the official opposition will be very vigilant during examination of the provisions of this white paper and the bill that may result from it, because we feel like we are being **taken for a ride** every time the secretary of state tables something in the House.

Figure 1 : Screenshot of a typical *TransSearch* session

The user has queried the system for translations of the figurative expression "to take for a ride". The '+' sign appended to "take" in the Expression field expands to all inflected forms of the verb, and the ellipsis allows for anything to appear between the verb and the preposition phrase, e.g. a direct object. The system responds to the query by displaying all the matches it finds in the Hansard database, each in its sentential context; and alongside each result it also displays the translation of that sentence, in which the user can find solutions that other translators have come up with for this particular translation problem.

All of these queries are recorded in the TS logfiles, along with details on the source of the query, the date and time it was submitted, how many matches were found, how the results were displayed, etc. The TS logfiles literally contain millions of such records, and they constitute a goldmine of information on how translators actually operate.

In this paper, we pursue a line of research first undertaken by (Simard and Macklovitch, 2005). There, the authors analyzed the TS logfiles in order to verify certain hypotheses about the syntactic nature of source-language translation units; in particular, they were interested in the extent to which the queries submitted to TS correspond to well-defined syntactic chunks. In this paper, we will also examine the TS logfiles, but with a view to determining the principal types of translation problems for which users turn to TS, as evidenced by the most frequent queries submitted to the system over the last six years.

2 The Most Frequent Queries

The TS logfiles that were analysed for this research contain information on over 7.2 million queries submitted by users of the system over a six-year period. In order to process this large volume of data, we first sorted the queries by length, grouping all one-word queries together, all two-word queries, etc. Basic statistics on these length-based groupings are given in Table 3 below.

Single-word queries	955,282	13.2%
Two-word queries	2,862,013	39.6%
Three-word queries	1,996,643	27.7%
Four-word queries	935,351	13.0%
Five-word queries	311,130	4.3%
Queries of six+ words	159,749	2.2%
Total	7,220,175	100.0%

Table 3: Frequency of queries in terms of their length

As Table 3 shows, most of the queries submitted to TS – nearly 40%, in fact – are made up of two words, followed by three-word queries, and only then by single-word queries, which are only slightly more frequent than four-word queries. Now with an interactive concordancer like TS, it is the users who decide when they want to query the system; moreover, they are free to formulate their particular translation problem as a query to TS anyway they like. A priori, there is no obvious reason why so many of the queries they submit should be constituted of just two words. On the other hand, there *is* good reason why such a small percentage of the submitted queries are made up of five words or more. (Macklovitch, Simard & Langlais, 2000) demonstrate that a clear correlation exists between the length of the queries submitted to TS and the system's non-response rate. Indeed, what they found was that the system returned no results at all for 65% of five-word queries, 70% of six-word queries, 78% of seven-word queries, and so on, until, when the queries reached fourteen words, all of them came up empty; and this, despite the fact that users were querying a database of about fifty million words. Although the TS databases have grown considerably since that study was conducted, there is little reason to believe that this pattern in the non-response rate has changed significantly. Furthermore, experienced users of TS have probably intuited this fact about the system; they know that if they submit a query that is over six words long, the chances that the system will come up with a useful result are slim, and so they tend to avoid these longer queries. Still, this doesn't explain the particular distribution of the shorter queries that we see reflected in Table 3. Why, for example, do the two- and three-words queries so clearly outnumber single word queries. We will return to this intriguing question later in the paper.

Within each length-based class, we then extracted from the logfiles the actual queries that were submitted most often. Tables 4-7 on the next page list the top twenty of these for two-, three-, one- and four-word queries respectively.

3 Analysis

Perhaps the first thing to notice about the queries in these tables is that all eighty of them are in English. This, despite the fact that TS offers translated collections in English, French and Spanish (in all six combinations). Furthermore, TS is a fully bidirectional system; i.e. users can submit queries to the same parallel corpus in either the source or the target language, and don't even have to specify the language of their query. The fact that eighty of the most frequent queries found in the TS log over the last six years are all in English tells us something important about the demographics of the translation market in Canada: namely, that translation in this country is preponderantly from English to French. Of course, this comes as no surprise, but rather confirms what has been empirically demonstrated in other surveys.[2] Given the predominance of English to French translation in Canada, it is only natural that the great majority of the translation problems which French-speaking translators submit to TS derive from their English source texts.[3]

[2] See, for example, the 1999 report by the Canadian Translation Industry Sectoral Committee.

[3] The most frequent French query encountered in the TS logs is "par ailleurs", which occurs 349 times.

Query	Freq.	Query	Freq.
as such	1195	as of	533
over time	1046	most importantly	530
consistent with	743	more importantly	511
in turn	708	where appropriate	499
hard work	680	as per	493
along with	669	due diligence	486
subject to	655	build on	483
based on	649	in particular	478
as required	609	focus on	472
as appropriate	587	where possible	461

Table 4: The twenty most frequent two-word queries submitted to *TransSearch*

Query	Freq.	Query	Freq.
as a result	1131	at this point	611
in terms of	994	a number of	572
at this time	913	in the future	564
in conjunction with	887	in light of	544
in support of	817	look forward to	539
with respect to	777	in accordance with	514
as part of	772	in consultation with	514
in line with	722	as a whole	504
in keeping with	716	in response to	490
make a difference	707	course of action	474

Table 5: The twenty most frequent three-word queries submitted to *TransSearch*

Query	Freq.	Query	Freq.
ultimately	851	overall	431
accordingly	619	insight	427
leverage	587	challenging	422
specifically	549	overarching	417
similarly	549	typically	409
alternatively	512	essentially	405
consistently	497	meaningful	386
hopefully	458	corporate	383
whereby	454	successful	383
historically	432	momentum	377

Table 6: The twenty most frequent one-word queries submitted to *TransSearch*

Query	Freq.	Query	Freq.
as a result of	882	in the near future	396
at the same time	840	on the basis of	374
from time to time	560	for the benefit of	367
out of the blue	508	in the face of	366
for the most part	473	in the first place	365
with a view to	453	on an ongoing basis	333
in a timely manner	444	At the same time	327
as it relates to	440	with this in mind	326
on the other hand	419	in the event that	314
in an effort to	404	on the basis that	312

Table 7: The twenty most frequent four-word queries submitted to *TransSearch*

More surprising than the predominance of English among the most frequent queries submitted to TS is the rarity of bona fide terms. Among the eighty queries listed in these four tables, there are only a few that qualify as true terms: "due diligence", certainly, and perhaps "course of action".[4] Now obviously, it would be a mistake to interpret this to mean that the texts which TS users have to translate contain few technical terms. Rather, what we can say is that the system's users do not generally look to TS to help them translate technical terminology, but in all likelihood turn to other, more specialized resources. (We will return to this point below.)

Notice as well that among the most frequent queries listed in Tables 4 – 7 there are almost no figurative expressions; perhaps the one exception being "out of the blue" found in Table 7. This too came as something as a surprise to us; for in the many demonstrations of TS that we have given over the years, we have frequently used such figurative expressions as "tourner autour du pot" or "take for a ride" – the example given in Figure 1 above – to illustrate the richness of the data lying dormant in past translations and the remarkable resourcefulness shown by the translators in transposing them into another language. We have also argued that such figurative expressions, which are not always listed in standard lexical resources like dictionaries or term banks, play a far more important role in the use of language than is generally recognized. Be that as it may, what the data in Tables 4 – 7 show us is that working translators (as opposed to computational linguists) do not often submit such expressions to TS.

So what are the types of problems for which translators do regularly turn to *TransSearch*? Judging from the most frequent queries listed in these tables, what emerges most strikingly is the high proportion of prepositional phrases. By our count, no fewer than forty-five of the eighty queries begin with a preposition or include a preposition as their headword. Indeed, this is the case for all twenty of the four-word queries listed in Table 7 and for sixteen of the twenty three-word queries listed in Table 5. A few of these can be considered compound prepositions, e.g. "along with", "as of", "as per"; but many are more complex prepositional groups[5] that generally conform to the pattern Prep1 + Noun + Prep2, e.g. "in support of", "as a result of", "in response to", "with a view to". Furthermore, most of these complex prepositions are frozen, both syntactically – i.e. one cannot generally insert an adjective before the noun, or modify the type of determiner if one appears – and semantically as well; i.e. they must be translated as a unit and not compositionally. To illustrate these properties with one example, consider "in light of", which will not admit a determiner or any modifiers before the noun; and if we look up its French equivalent in TS, the most frequent translation seems to be "compte tenu de".

The complex prepositions that we have just discussed are all transitive, i.e. they require a noun phrase complement. Another important subclass among the prepositional groups that appear in Tables 4, 5 and 7 are *in*transitive; i.e. they do not require a complement but can stand alone as an adjunct or sentence adverbial. This is the case, for example, of "as such", "at this time", "for the most part" and "on the other hand". Interestingly, eleven of the most frequent single-word queries are also adverbs, as are "most/more importantly" that are grouped with the two-word queries. In addition, we also find four instances of reduced subordinate clauses in Table 4, which can also be seen to function as sentence adverbials: "as required", "as appropriate", "where appropriate" and "where possible". Hence, in answer to the question posed above – For what types of translation problems do translators consult TS? – we can propose the following preliminary response: **syntactically**, many of the most frequent queries submitted to the system are made up of complex, often idiomatic prepositions and/or adverbials. Which then invites the following follow-up question: What is it about these precise syntactic constructions that translators find problematic? We will try to suggest a plausible answer to this in the next section.

Another type of syntactic construction that appears with some frequency in the Tables is made up of a predicate (either a verb or an adjective) that governs a particular preposition, e.g. "consistent with", "subject to" or "focus on". The translation problem here is that one cannot simply translate the preposition independently of the predicate;

[4] We hesitate to include the nouns "leverage" and "momentum" that appear in Table 6 because, in the Hansard at least, both are almost always used in a figurative sense and not as technical terms.

[5] This being the terminology of (Quirk et al. 1972), p. 300 ff.

rather, it is the translation of the *target* predicate which determines the form of the governed preposition. So, for example, the most frequent French translation of “consistent with” is “conforme à”; and while “focus on” admits numerous French translations – e.g. “se consacrer à”, “se concentrer sur”, “viser” – in each of these, it is the French verb which determines the proper form of the French preposition (if indeed it takes a preposition). One could characterize the problem of governed prepositions in terms of translational **compositionality**; although the non-compositionality displayed here is only partial, in contrast to such fully idiomatic phrasal verbs as “look forward to” and (less obviously) “make a difference”.

Finally, consider the five adjectives that appear in Table 6: “challenging”, “overarching”, “meaningful”, “corporate” and “successful”. None of these (perhaps with the exception of “overarching”) seems terribly difficult to translate, and so their presence among the most frequently submitted queries is somewhat puzzling. In an effort to understand what is going on here, we submitted these adjectives to TS ourselves and examined the first twenty results for each. What we found was that TS proposes an impressive variety of French equivalents for these English adjectives. Take the example of “meaningful”: among the first twenty TS results, we found no less than fifteen different French translations! And yet to a native English speaker, the adjective in question is not exceedingly polysemous; if anything, “meaningful” tends to be somewhat vague, so that its precise meaning in a particular sentence – and hence, the most appropriate translation – is partially determined by the particular noun it modifies.

4 Interpretation

To summarize our analysis of the eighty most frequent queries in the TS logfiles, what we found, on the one hand, was that there were surprisingly few bona fide terms or figurative expressions, while, on the other, there was a unexpectedly high proportion of adverbs and prepositional groups, many of which function as adverbials. We also found a fair number of non-compositional expressions, such as governed prepositions and idiomatic phrasal verbs; but these are the kinds of translation problems that we fully expected to find in the TS logfiles. It is the former phenomenon that is more intriguing and which calls for an explanation.

The first, fairly obvious point to make is that TS co-exists alongside a panoply of other resources to which translators have ready access. Chief among these, in Canada at least, are two large-scale terminology banks, *Termium* and *Le Grand dictionnaire terminologique*, that are now accessible over the Internet.[6] In both cases, their numerous records are classified by technical domain and generally include detailed definitions, citations, usage notes, etc. All of this is in stark contrast to a concordancer like TS, which only provides its users with access to large volumes of raw, unannotated text. What is more, the collections currently offered in TS only cover a few technical domains, most notably court rulings and labour relations; otherwise, the majority of the texts found in TS are parliamentary debates. Hence, for translation problems involving technical terminology, Canadian translators would be well-advised to consult one of these large term banks, rather than TS. And this, presumably, is what they do; which would account for the rarity of bona fide terms among the most frequent queries submitted to TS.

In addition to these bona fide term banks, translators also have access to an impressive number of bilingual dictionaries, although for the moment not many of the standard bilingual reference works, for English and French at least, are freely available online. Still, there is nothing to prevent translators from consulting such bilingual dictionaries as the Robert-Collins or the Hachette-Oxford in printed form, even if most of them now draft their texts on a computer. This is not the place to conduct a detailed comparison of the advantages and disadvantages of TS as opposed to such printed dictionaries, but suffice it to say that TS fares somewhat better here than it did as source for translations of technical terminology. If, for example, we look at the adjectives that appear in Table 6, what we find in consulting the Robert-Collins is, first, that there is no entry for “overarching”, while for “meaningful”, it offers only three French equivalents. And the same is generally true for the remaining adjectives in Table 6: not only does the printed word-

[6] The URL for *Termium* is http://www.termiumplus.gc.ca/ and that of *Le Grand dictionnaire terminiologique* is http://www.granddictionnaire.com . The latter is freely available to the public; for the former, there is a monthly subscription fee.

book have far fewer equivalents to propose, but the indications by which it distinguishes them are more cryptic and less explicit than the full sentential context which TS provides. What is more, an interactive concordancer like TS makes it easy for the user to query the exact adjective + noun combination that he or she is grappling with in the text under translation; and again, the system's databases are so large that there is more than a reasonable chance that TS will come up with useful suggestions. Hence, for cases like these vague-ish adjectives, which take on a different semantic shade according to the noun they modify, there would seem to be good reason for users to consult TS over a standard bilingual dictionary.[7]

Let us now return to the question of the prepositional groups which, as we saw above, are so surprisingly numerous among the eighty most frequent queries that we extracted from the TS log. As we suggested above, part of the explanation for this preponderance of prepositional groups in the logfiles may derive from the fact that it is much easier to locate equivalents for phrasal expressions in an online concordancer like TS than it is in a printed dictionary. Consider, for example, such phrases as "with respect to" and "in the face of". In most printed bilingual dictionaries, one cannot look these up as such, but first must look up the head noun and then hunt for the phrase in question among a plethora of expressions that are listed, seemingly pell-mell, at the end of the entry. The Robert-Collins proposes four equivalents for "with respect to" and two for "in the face of"; and in both cases no indications are provided which would help a user select among them. Looking up these same phrases in TS, on the other hand, is easy and direct, and one finds many more suggested equivalents, each displayed in its full sentential context.

No doubt, this ease of consultation (which is common to all electronic reference works, compared to their printed counterpart) is one important factor which encourages translators to submit such phrases to TS. But this cannot be the whole story. For how is it, one may ask, that TS users, most of whom are professional translators, feel the need to look up such phrases as "at this time" (submitted 913 times), "as a result of" (882 times), or "most importantly" (530 times) in the first place? On the face of it, none of these appears particularly difficult to translate; and the same could be said for the majority of the frequent queries listed in Tables 4 - 7. What exactly is going on here?

A definitive answer to this question would no doubt require interviewing (or at least closely observing) many TS users as they work with the system, something we haven't been able to do. Nevertheless, we want to suggest a tentative, hopefully plausible explanation, for which we shall borrow and extend a metaphor first introduced by (Langé et al., 1997) in a stimulating article entitled "Bricks and Skeletons: Some Ideas for the Near Future of MAHT." Here is how the authors first introduce the notion of translation "bricks", which they also refer to later as "building blocks":

> "…we see here one possibility for future systems: To offer an extended TM capability that would deal not only with sentences, but also with the elementary 'bricks' that sentences are made of, including terms, phrases or clauses." (p. 42)

Our immediate concern here is not with the authors' proposal for a more powerful translation memory (TM) that can perform sub-sentential matches, although this is certainly an interesting idea.[8] Instead, we want to focus on the notion that the translation of a sentence can be viewed as being made up of certain bricks (or blocks), which correspond to technical terms and perhaps larger phrases, and which ideally can be retrieved from a TM or a term bank. Suppose that in translating a given source sentence, a translator has all such elements in hand. What is left for him to do? Pursuing the analogy of Langé et al., we would suggest that the translator still has to *assemble* these bricks into a coherent sentence; and to do this, he needs to relate those blocks, or constituents, one to another, i.e. set them in a certain order and place them in grammatical relations that reflect their semantic roles. Moreover, once this target sentence has been assembled, the translator also has to relate it, both logically and in terms of the larger discourse of the text, to the sentences that precede and follow it. In languages like English and French, the

[7] Counterbalancing TS' richness, on the other hand, is the fact that users have to sift through and evaluate all the examples it makes available; whereas a printed bilingual dictionary, which needs to be concise for reasons of publishing costs, attempts to condense and synthesize this information.

[8] And one which we ourselves have worked on; c.f. (Macklovitch and Russell, 2000).

mortar with which the bricks of a translation are assembled into a final translation – or, if you prefer, the hinges[9] linking the blocks together – are largely made up of function words, typically prepositional phrases and other types of adverbials. Now as we have seen, these are just the kinds of queries that users appear to submit most frequently to TS. In sharp contrast to technical terms, an important property of these prepositional and adverbial phrases is that they generally admit multiple, often numerous translations, with the most appropriate target equivalent "depending on the context", as translators like to say.[10] If we can judge from the TS logfiles, and in particular from the most frequent queries submitted to this system, translators apparently require assistance in coming up with a suitable translation for just these kinds of phrases, i.e. those that furnish the mortar that allows them to fix the building blocks in place. Given the enormous size of its databases, the quality of the translations that they contain and the ease which the system allows these to be queried, TS seems to respond quite adequately to this need.

5 Conclusion

At the outset of this paper, we described TS as a system that allows its users to query large databases of past translations in order to find ready-made solutions to a host of *translation problems*. A priori, one might be tempted to think that translation problems are all of the same type: "I don't know how to translate this word or this expression." What our analysis of the TS logfiles has shown, somewhat surprisingly, is that this is not the case; on the contrary, there are different types of translation problems for which different types of resources may be best suited. Term banks, bilingual dictionaries and perhaps translation memories may be the tools best suited to help a translator obtain the appropriate target equivalent for the building blocks of his translation. But once he has these, he still has to assemble them into a coherent target text; and for this, he needs a quick and easy way to obtain multiple equivalents for the prepositional phrases and adverbials that serve to link those blocks together in just the right way. And for this type of translation problem, which is no less serious than the other, a bilingual concordancer like *TransSearch* appears to be particularly well suited.

References

Canadian Translation Industry Sectoral Committee (CTISC) (1999): Survey of the Canadian Translation Industry: Human Resources and Export Development Strategy. Available online at: http://www.uottawa.ca/associations/csict/princi-e.htm

Pierre Isabelle, Marc Dymetman, George Foster, Jean-Marc Jutras, Elliott Macklovitch, François Perrault, XiaoPo Ren and Michel Simard. Translation Analysis and Translation Automation. Proceedings of the Fifth International Conference on Theoretical and Methodological Issues in Machine Translation, Kyoto, Japan, 1993.

Jean-Marc Langé, Éric Gaussier, and Béatrice Daille. Bricks and Skeletons : Some Ideas for the Near Future of MAHT. *Machine Translation*, special issue on New Tools for Human Translators, eds. Pierre Isabelle and Ken Church, vol. 12, nos. 1-2, p. 39-51, 1997.

Elliott Macklovitch, Michel Simard and Philippe Langlais. TransSearch: A Free Translation Memory on the World Wide Web. *Second International Conference On Language Resources and Evaluation* (LREC), vol. 3, p. 1201-1208, Athens Greece, June 2000.

Elliott Macklovitch and Graham Russell. What's been Forgotten in Translation Memory. *Proceedings of the fourth conference of the Association for Machine Translation in the Americas* (AMTA), Cuernavaca Mexico, p. 137-146, October 2000.

Randolph Quirk, Sidney Greenbaum, Geoffrey Leech and Jan Svartvik. *A Grammar of Contemporary English*. Longman Group Limited, London, 1972.

Michel Simard and Elliott Macklovitch. Studying the Human Translation Process through the TransSearch Log-Files. In *Knowledge Collection from Volunteer Contributors: Papers from the 2005 Spring Symposium*, ed. Timothy Chklovski, Pedro Domingos, Henry Lieberman, Rada Mihalcea, and Push Singh, 70-77. Technical Report SS-05-03. American Association for Artificial Intelligence, Menlo Park, California, March 2005.

[9] The French term for 'hinge' is 'charnière', which is sometimes used to refer to prepositions and adverbials.

[10] Presumably, what translators refer to by this oft-repeated mantra are a host of more or less imponderable factors, ranging from the flow of the target text to the avoidance of awkward repetitions, and including the translator's personal preferences.