

You Must Remember This: The Copyright Conundrum of “Translation Memory” Databases

Francie Gow†

Abstract

Translation memory databases (compilations of texts linked with their translations) can be valuable resources in the process of translating subsequent texts. This article explores the circumstances under which such compilations might be considered sufficiently original to attract copyright protection that is independent of any copyright already subsisting in the underlying translations and source texts. Various characteristics of the tools and the translation industry in general make the analysis highly fact-specific; whether particular translation memory databases attract protection, and, if so, who can claim to be their “authors”, must be evaluated on a case-by-case basis. Any protection that is granted may be significantly restricted by the competing layers of copyright in the contents of the database. Ultimately, granting protection at the database level does little to promote the goals of copyright law. Reforms of the translation industry’s traditional business model are recommended to enable a greater pooling of linguistic resources among translation professionals without prejudice to their clients’ legitimate interests in protecting confidential information and benefiting economically from their property rights.

I. Introduction

Translators are increasingly using computer-assisted translation (CAT) tools both as research aids and to increase their productivity.¹ One such tool, in particular, has received considerable attention in the past decade: translation memory² (TM)³ software. Translation memory is a database tool⁴ for storing previously translated texts connected with their original texts, also known as source texts, so that the translator can quickly answer questions such as, “Have I translated something like this before, and if so, how?”

I will review the features of TM databases and consider the legal basis for granting them protection under the *Copyright Act*. I will also consider who might be entitled to the right, and the extent to which the scope of that right may be limited by pre-existing copyright in the stored texts as well as conflicting contract and confidentiality rights.

These compilations of archived translations contain extensive linguistic information that is potentially valuable both to translation providers and their clients. Because of the way the current industry model is structured, most of this information ends up locked away after its initial use. I will conclude by examining the copyright and confidentiality issues and assignment and

licensing practices that have created this problem, and recommending adjustments that could maximize value from the information while respecting the rights of all stakeholders: translators, translation agencies, and clients.

II. How Translation Memory Works

TM tools come in many forms, but they need to perform two essential functions to earn the name: the “alignment” of source and target texts, and “matching”, the comparison of two source texts for similarity at approximately the sentence level. Both of these tasks are basic for humans, but extraordinarily difficult for computers. What makes a French sentence “equivalent” to an English sentence for the purposes of alignment? What does it mean for two English sentences to be “similar” in a way that will be useful for a translator? How do we even come close to approximating these processes in an algorithm?

The leading TM developers,⁵ working independently, have come to very different solutions to these problems,⁶ and some of the differences in design may affect the copyright analysis. The principal difference lies in how the data is stored. In some of the tools, an individual cell in the database contains a single sentence or

†© 2007, F. Gow. The author holds an MA Translation from the University of Ottawa and works as a legal translator at the Translation Bureau of the Government of Canada. She will complete a B.C.L. and LL.B. at McGill University in December 2007. The author would like to thank Professor David Lametti of McGill University, Professor Lynne Bowker of the University of Ottawa, and Bruno Barrette and Joseph Reynaud of Stikeman Elliott LLP for their contributions to the development of the ideas in this paper.

sentence fragment, along with its translated equivalent. The text as a whole is not preserved.⁷ I will call these “sentence-type TMs”. In other cases, an individual cell in the database comprises the full source text along with the full target text (this is known as a “bitext”).⁸ I will call these “bitext-type TMs”.⁹

III. Translation Memory and Copyright

For years, translators and their clients have been asking “Who owns TMs?” Is it the translator who produced the translation, or the client who paid for it? Some researchers, such as Bowker,¹⁰ have alluded to this problem in passing, while others, such as Topping,¹¹ have explored it in the context of ethics. Although everyone recognizes this as a copyright issue, the hybrid features of translation memory databases defy easy analogy with more familiar objects of copyright protection.

Currently, the problem is avoided through contract. According to Paula Shannon of Lionbridge, a leading provider of translation and globalization services,

the current industry practice is for the client to take the rights by contract to any potential copyright in TM databases, regardless of where the copyright might actually originate. This is accomplished either through the broadest definition of work for hire, as in the U.S., or called out as a specific deliverable. In addition, most of the clients will identify this deliverable as constituting their intellectual property, subjecting it to additional restrictions.¹²

It is understandable that clients would want to protect their investment. Translators, however, are generally in favour of sharing TMs in order to maximize their productivity. A perusal of the postings on electronic discussion boards, such as *TranslatorsCafé.com*, reveals that many translators believe they should have ownership of the TM since they are the ones putting their time and effort into building it up. Meanwhile, agencies tend to feel that sharing TMs is not ethical because the clients would not appreciate having exactly the same content in their texts as their competition.¹³

Topping also offers some general advice, noting that “the laws which apply to translation database ownership are murky” and suggesting “if you are a translator, be honourable about which translation databases you share, and careful about how you apply databases you have not created”; if you are a translation agency, “create a sharing policy”; and if you are a client, “ask your vendors about their sharing policies”.¹⁴

A. Layers of Copyright Protection

The most obvious complicating factor is that TM databases are made up of texts that are themselves potentially subject to copyright protection. A TM easily falls within the definition of compilation in the *Copyright Act*,¹⁵ which means that an independent copyright in the

TM as a whole cannot override the copyright that may already exist in the individual works it contains.¹⁶ In other words, rights with respect to the TM may only be exercised to the extent that such exercise does not infringe the copyright in the individual texts.

This situation is not unfamiliar to copyright analysis; we need only think of more traditional composite works such as newspapers and film. More recently, it has become a central focus in studies of multimedia products. Multimedia products probably provide the most fruitful analogy to TM databases. Irini A. Stamatoudi describes their layers as follows:

The three essential layers of protection with regard to a multimedia product are: (1) the protection of the contents of a multimedia product, (2) the multimedia product itself (as a compilation of the works it includes, but not necessarily protection as such from the point of view of intellectual property), and (3) the protection of its technical base. Although this distinction of parts in a multimedia product is theoretically possible, in practice it is not always clear. [...] This is undoubtedly an area which stresses the need for flexibility as far as legal regulation of information technology is concerned. Moreover, any problems regarding the protection of any one of these three layers will inevitably have repercussions on the other two. All three layers have to co-exist in regulative harmony in order for the creation and proper functioning of the product not to be impeded.¹⁷

In the context of TMs, the first layer will almost always have to be considered. It is possible to have a TM made up entirely of texts that are in the public domain, either because the term of protection has expired, or because they were never subject to copyright protection. However, typical TMs contain at least some protected source texts. Texts need to be in digital format for use in a TM, and it may be prohibitively expensive to digitize texts whose copyright has expired. Also, older texts that have been digitized are often literary texts, which are not well suited to use in TMs because their contents are not likely to be repeated in future texts. Finally, by definition, public-domain source texts need to be linked to their translations, which may themselves still be protected.

The second layer is the principal focus of this paper, if we compare the TM to “the multimedia product itself”.¹⁸ Stamatoudi rightly points out that while we are necessarily concerned with protecting this layer, such protection does not necessarily have to come in the form of an intellectual property right. Whether or not the database is sufficiently original to qualify for its own copyright protection is only one part of the story; contract, confidentiality, and technological protection measures must all be considered as well.

The third layer is perhaps less problematic in the TM context than in the multimedia context. Unlike with some of the multimedia tools examined by Stamatoudi, it is relatively easy to separate the analysis of intellectual property protection for TM software itself from the analysis of the databases it is used to create. On the other hand, we cannot ignore the “technical base” altogether. As discussed in Part II above, its features differ

from one product to another in ways that could potentially affect the analysis of the first two layers, particularly the first.¹⁹

B. Copyright in Translation Memory Databases

i. Translation memory as compilation: database or collective work?

Copyright protection in Canada is solely statutory,²⁰ so we must look to the *Copyright Act* (the Act) to determine whether a given work is subject to protection. Section 5 of the Act states that subject to certain conditions, “copyright shall subsist in Canada, for the term hereinafter mentioned, in every original literary, dramatic, musical and artistic work.”²¹ If the first layer of source texts and translated texts is protected, it is because those texts fall easily into the category of literary works.

Characterizing the database itself so that it fits into section 5 is somewhat trickier. If we look to the definitions in section 2, we see that “‘literary work’ includes tables, computer programs, and *compilations of literary works*” (emphasis added).²² A compilation is defined as

- (a) a work resulting from the selection or arrangement of literary, dramatic, musical or artistic works or of parts thereof, or
- (b) a work resulting from the selection or arrangement of data;²³

Paragraph (a) of the compilation definition is meant to cover collective works and paragraph (b) is meant to cover databases.²⁴ Databases are not specifically defined in the Act, but “collective works” are defined as follows:

- (a) an encyclopaedia, dictionary, year book or similar work,
- (b) a newspaper, review, magazine or similar periodical, and
- (c) any work written in distinct parts by different authors, or in which works or parts of works of different authors are incorporated;²⁵

Paragraph (c) seems to be a perfect description of TMs. How, then, do we explain that a TM feels more like a database than a collective work? It would never occur to a translator or a client to refer to a TM as collective work, and the use of the term “database” in this context is now firmly entrenched in the industry.

The answer can be found by considering the purpose of the translation memory. The value of a TM to a translator is generally not in the ideas expressed in the texts stored therein. Of course, they may help the translator gain an understanding of the subject matter, which can be of some assistance, but a few minutes of Internet searching will often accomplish the same thing. The reason that the source and target texts are linked to one another is that they then are able to express a *relationship of linguistic equivalence*, which is what is valuable to the translator. Even a collection of the translated texts alone is of very little assistance. Although any given translation is supposed to be a foreign-language equivalent of

its source text, the translation standing alone is not a fixed expression of this equivalence.²⁶

One might be tempted to call a TM a database on the basis that each cell expresses the “fact” that fragment X of the source text is the linguistic equivalent of fragment Y of the target text.²⁷ However, the flaw in that analysis is that it does not sufficiently take into account the highly contextual nature of linguistic equivalence.

Take the following fragment of French source text: “Un, deux, trois, j’irai dans les bois”. Imagining that this fragment has occurred several times in a given database, we might plausibly find one or more of the following connected to it in various cells:

- (a) One, two, three, I am going into the forest
- (b) 1, 2, 3, I’m off to the woods
- (c) One, two, three, into the woods I flee
- (d) One, two, buckle my shoe
- (e) One, two, three, I am going into the words
- (f) One, two, three, I am going into the drink.

Is it a “fact” that each of these is linguistically equivalent to the French? The answer is probably yes for examples (a) through (d), depending on the purpose of the translation. If the goal is to produce a literal translation, many variations of (a) or (b) would do the job. We can see already that it would be a mistake to talk about “*the linguistic equivalent*” of a fragment of text, and this holds for the most mundane sentences as well as for nursery rhymes.

If the purpose is to preserve most of the literal meaning of the original text but with any necessary slides in connotation for the sake of preserving rhythm and rhyme (“I flee” is a semantic stretch), then (c) could be considered equivalent. Finally, the author may simply have been trying to convey a familiar counting rhyme, in which case (d) would be the best translation for a Western English-language audience.

Examples (e) and (f), which are meant to illustrate errors that might be made by an inattentive translator or by machine translation respectively (the former introducing a typographical error and the latter mistaking the noun “bois” for the verb),²⁸ are unlikely to be considered linguistically equivalent to the French in any context. But the translator looking at the TM is intelligent enough to know what was meant, so even these may be useful.²⁹ Errors can also be repaired by future users of the database, which would increase the number of true equivalents it contains.

Instead of characterizing the “facts” in a TM as “facts of linguistic equivalence” between the connected bits of source- and target-language text in each cell, a more nuanced approach would be to think of a TM as a compilation of facts that “a translator at some time considered this fragment of target text to be linguistically

equivalent to this fragment of source text in a particular context”.

This may be a somewhat less elegant description of a class of facts than “the telephone numbers of all dentists practising in St. John’s, Newfoundland in 2006”, but as the latter case shows, elegance is a relative standard, and it is not our first priority. The above description of the facts seems accurate, and it has the virtue of reconciling our conflicting intuitions about the nature of databases and the nature of translation. This is the information that is truly valuable to translators, even if they are merely seeking inspiration for a similar or identical fragment that must be translated in an entirely new context, and TMs are specifically designed to provide it.

ii. The standard of originality

Just because our TM can be classified as a work under the Act does not mean that it automatically qualifies for copyright protection. Section 5 specifies that subject to certain conditions, all *original* works will be protected for a term.

Although this paper focuses on the Canadian context in particular, it is important when looking at translation issues to look abroad, as well. It is not unusual for translators to work with clients in several different countries, or for large clients to deal with translators in more than one country. Originality is a touchstone requirement of copyright regimes,³⁰ which makes sense if the purpose of copyright is to promote the creation and dissemination of works of the intellect.³¹ It is clearly written into Canadian and U.S. copyright legislation,³² and although it does not appear as a stated requirement in the international *Berne Convention for the Protection of Literary and Artistic Works*,³³

[t]he records of various diplomatic conferences adopting and revising the Berne Convention reflect that the reason why Article 2(1) of the Convention does not state explicitly that works are intellectual creations is that that element of the notion of works was considered to be evident.³⁴

It is generally accepted that the work must at the very least “originate from the author”,³⁵ as opposed to being copied from another author, and that artistic merit and innovation constitute too high a standard.³⁶ There is some divergence, though, between the interpretations of originality among jurisdictions. The United Kingdom has traditionally adopted a low “sweat of the brow” threshold, whereby skill and labour combined with an absence of copying are sufficient to qualify a work for protection.³⁷ However, Daniel J. Gervais argues that the United Kingdom may currently be moving toward a more demanding threshold under the influence of the European Union.³⁸

The U.S. Supreme Court set a higher threshold in *Feist Publications v. Rural Telephone Service Company, Inc.*,³⁹ stating that “[o]riginal, as the term is used in copyright, means only that the work was independently created by the author (as opposed to copied from other

works), and that it possesses at least some minimal degree of creativity”.⁴⁰ The Court holds that this is still low, and that “the vast majority of compilations”⁴¹ will presumably pass the test, but that “[t]here remains a narrow category of works in which the creative spark is utterly lacking or so trivial as to be virtually nonexistent”.⁴²

The Supreme Court of Canada recently reinterpreted the test for originality in *CCH Canadian Limited v. Law Society of Upper Canada (CCH)*, coming down somewhere in between the two poles:

[A]n original work under the *Copyright Act* is one that originates from an author and is not copied from another work. [...] In addition, an original work must be the product of an author’s exercise of skill and judgment. The exercise of skill and judgment required to produce the work must not be so trivial that it could be characterized as a purely mechanical exercise. While creative works will by definition be “original” and covered by copyright, creativity is not required to make a work “original”.⁴³

Most of the debate about where the line actually lies arises when looking at factual compilations, i.e., databases, which “usually hover and sometimes inch past the threshold”.⁴⁴ Even with the high U.S. standard, it is possible for factual compilations to possess the requisite originality. In fact, most databases judicially considered in the United States since the decision in *Feist* have been found to be sufficiently original.⁴⁵

Before applying the standard of originality to TM databases, it is important to remember that we are not applying it to the expression of ideas in the texts contained in the database. Those texts are literary works subject to their own copyright protection at a different level. Compilations “are considered to be authored through a process of ‘selection and arrangement’”, as opposed to “the act of writing, drafting or composing” a work.⁴⁶ The “selection and arrangement” wording, which we saw above in the Act’s definition of a compilation,⁴⁷ affects the infringement analysis: instead of assessing the works for substantial similarity, “the focus of attention is on the author’s original selection and arrangement of elements, and nothing more”.⁴⁸

iii. Applying the standard of originality to translation memories

We have to look at how TMs are built to identify parts of the process that might involve the requisite level of skill and judgment. First, texts must be gathered to feed the database. Each text must be available in both the source and target languages, and both versions must be in electronic format. Texts could be selected by the client from among internal documents, selected by the translator or team of translators from among their own past translations, or obtained by either party from the public domain or from other sources under fair dealing provisions or with authorization from the copyright holders.

Next, the source texts have to be “aligned” with their corresponding target texts. Automatic alignment is a feature common to all translation memory software, but each tool handles the task somewhat differently. Alignment tools rarely give perfect results (although they are rapidly improving), so it is usually necessary for a human to edit the alignments to get optimal results from the database. As difficult as the task can be for a computer, alignment would qualify as a “purely mechanical exercise” for the purposes of our standard of originality. Any bilingual human being could perform the task easily for his or her language combination, and different people would make the same alignments, even without specialized language training.

At this point, the database is usable. However, its performance can still be improved considerably with post-editing. As we saw above, there may be mistakes in some of the cells. Typographical errors in the source text are problematic because they make it more difficult for the TM tool to identify a fragment of text as similar to the one currently being translated. Typographical errors or translation errors in the target text can also be problematic, since some TM tools have a feature whereby all perfect matches in the source text will be automatically replaced by their corresponding translations stored in the database. These errors get reproduced in the new text, either slowing down the revision process, or not getting caught at all. Finally, while some fragment pairs may be technically error-free, they may not be helpful, either because the same pair occurs over and over again, or because the solution appropriate in the context of the earlier translation is so unlikely ever to be appropriate again that its presence in the database will only distract the translator. An experienced translator will have developed a sense of which cells are likely to create “noise” during the translation process, and can increase the efficiency of the tool by deleting them from the database.

To meet a low “sweat of the brow” standard of originality, a TM database would simply have to be sufficiently large. However, it is not clear exactly how large the database would have to be to qualify for protection. It may be possible to create a database that fails to meet even this lowest standard, if one exercises no judgment in the selection of texts,⁴⁹ relies only on the automatic alignment tool, and does no post-editing.

If it is possible to build a TM that fails to meet even the lowest standard of originality, is it equally possible to build one that meets the highest standard? Here we must look to the U.S. standard, which requires a “creative spark”.⁵⁰ The Second Circuit elaborated on the meaning of creative selection or arrangement in *Matthew Bender & Co. v. West Publishing Co.*, reaffirming that “neither novelty nor invention is a requisite for copyright protection”,⁵¹ and adding the refinement that “when it comes to the selection or arrangement of information, creativity inheres in making non-obvious choices from among more than a few options”.⁵²

In the initial selection of texts to feed the database, technical concerns may limit one’s choices. In a case where there are not sufficient resources to convert texts into digital format, the translator and/or client will be limited to those texts for which both the source and target texts are available in electronic format. If only a small number of these exist, which may well be the case, it may be worthwhile to enter them all into the database, since even texts generally unrelated to the new text may provide useful examples of general language translation here and there. If significant numbers of texts are available, which may be the case for very large clients or for translation providers with large translation archives, a selection of texts will have to be carried out at the outset.

Assessing which texts are most likely to be relevant for future tasks does require making non-obvious choices. The task can require even more judgment or creativity when there are resources available to digitize texts, which usually increases the universe of possibilities considerably. Finally, identifying and locating relevant bilingual material from the public domain can require subject-matter expertise and advanced research skills.

The alignment process is too mechanical to contribute to a finding of original selection and arrangement under any standard that requires more than labour, although it can be argued that there is some skill involved in manipulating the alignment tool. However, there is no judgment, and certainly no creative choice.

The post-editing process may be relevant. Correction of typographical errors in the source or target texts will not suffice in Canada or in the United States to attract the protection of copyright.⁵³ The deletion of redundant cells from the database is unlikely to be considered sufficiently creative, either. However, it requires considerable skill and judgment to identify other kinds of noise that could be removed from the database to improve its performance. Such an exercise necessarily involves making non-obvious choices. However, making one or two such deletions is probably not enough on its own for the database to qualify for copyright protection. Paradoxically, even translators who have made extensive changes of this kind would have trouble demonstrating their exercises of creative judgment to a court, since they will necessarily have deleted the evidence!⁵⁴

Therefore, for a database to pass the relatively high U.S. standard, a party would have to demonstrate that non-obvious choices were made in the initial selection of texts. One might argue that the originality standard applies to the “selection *and* arrangement” and not to the “selection *or* arrangement” of a compilation. Requiring creativity in both seems excessive, and the Second Circuit did use “or” in the passage cited above,⁵⁵ but the interpretation may vary in different jurisdictions or over time.

If original arrangement is required in addition to original selection, meeting the standard will be more

difficult, but perhaps not impossible. Because TM tools generally scan the archived texts for matches in the order in which they are entered, a tool's performance could conceivably be enhanced by entering the texts in order of relevance. This would arguably constitute a creative arrangement of information in the database. The problem is that relevance can only be measured in relation to the new text that one is trying to translate, so the argument will only hold in those cases where the database has been built with a specific job in mind. This does occur, but the argument does not hold if the same database is used for future jobs, in which case the order of relevance may be different. Furthermore, any improvement in the TM's performance will be trivial unless the database is extremely large.

In conclusion, databases will fall along a wide spectrum of originality. Whether a given TM qualifies for protection will depend on whether the relevant jurisdiction applies a sweat-of-the-brow standard, or a more rigorous standard such as Canada's non-trivial skill and judgment standard, or the U.S. requirement of at least a minimal degree of creativity. It will also depend on how much judgment was required to select the texts for the database, and possibly whether creative selection is enough given the requirement for creative "selection and arrangement". Unless Canada and the United States end up with different interpretations of that last point, it seems likely that the vast majority of cases would lead to the same result regardless of which of the two standards is applied.

iv. Identifying the "author" of a database

All works that qualify for copyright protection have at least one author, a physical person from whom the work originates, even if that requirement is sometimes only implicit in the copyright legislation.⁵⁶ The Canadian *Copyright Act* states that "[s]ubject to this Act, the author of a work shall be the first owner of the copyright therein",⁵⁷ which makes it important to be able to identify him or her. There are some exceptions to the first-owner rule, but these apply to photographs, film, sound recordings, and communications signals⁵⁸ and are not relevant in the context of TM. In addition to allowing us to determine the first owner, knowing the identity of the author or authors is necessary to fix the term of copyright protection.⁵⁹

In the case of compilations, our analysis must focus on Stamatoudi's second layer of protection. Although the "facts" of the database are gathered by selecting and arranging literary works,⁶⁰ we are not interested in the authors of those underlying works, only in the people who exercise judgment in selecting and arranging them.

The issue of initial ownership has been hotly debated in the translation industry for years, with both clients and translation providers instinctively feeling that they have a legitimate claim.⁶¹ The preceding analysis shows that the initial owner will be whoever performs

the creative selection of texts at the outset. That might be either party, or both, depending on the facts of the case. The Act defines a work of joint authorship as "a work produced by the collaboration of two or more authors in which the contribution of one author is not distinct from the contribution of the other author or authors".⁶²

v. The role of translation industry contracting practices

The identity of the author is arguably a moot point in the translation industry context, since the copyright will almost invariably end up with the client. We already see this phenomenon with respect to the translations themselves: the party paying for the translation will either acquire the copyright in the translation automatically because the translator is an employee, or will have the copyright assigned to him or her in the service contract with an independent translation vendor.⁶³

Ownership of copyright in Canada is governed by section 13 of the Act, and it includes a specific provision covering "work made in the course of employment":

13. (3) Where the author of a work was in the employment of some other person under a contract of service or apprenticeship and the work was made in the course of his employment by that person, the person by whom the author was employed shall, in the absence of any agreement to the contrary, be the first owner of the copyright, but where the work is an article or other contribution to a newspaper, magazine or similar periodical, there shall, in the absence of any agreement to the contrary, be deemed to be reserved to the author a right to restrain the publication of the work, otherwise than as part of a newspaper, magazine or similar periodical.⁶⁴

This provision has two important implications: the first is that employers with in-house translation services will be the first holders of copyright in the works produced by their employees (whether these be translations or protectable TM databases), and the second is that translation service providers with multiple translators on staff will be the first owners of copyrightable material, and not the employed translators who actually "author" such materials.⁶⁵

Notably, subsection 13(3) does *not* cover the relationship between a client and a freelance translator or translation firm. In those cases, translation providers will be the first holders of copyright in any copyrightable work they produce for the client. If the client wants to use the work (beyond use that is permissible under copyright exceptions such as fair dealing), it will have to have the copyright assigned to it or get a licence from the holder:

13. (4) The owner of the copyright in any work may assign the right, either wholly or partially, and either generally or subject to limitations relating to territory, medium or sector of the market or other limitations relating to the scope of the assignment, and either for the whole term of the copyright or for any part thereof, and may grant any interest in the right by licence, but no assignment or grant is valid unless it is in writing signed by the owner of the right

in respect of which the assignment or grant is made, or by the owner’s duly authorized agent.⁶⁶

Assignments of copyright to the client in the translation contract are the norm for the translations themselves, generally at no cost above and beyond the rates paid to perform the work. A person generally commissions a translation because he or she needs to use it. For that very reason, the client has a strong argument that the holder has granted an implied licence even in cases where assignment is not mentioned in the contract.

The implied licence argument is not as strong in the case of protectable TM databases, since they are at least as useful to the translation provider for future work as they are to the client, if not more useful. However, clients generally have considerable bargaining power and can easily impose a clause asserting that they will own any intellectual property rights that might be created in the course of the work.⁶⁷ With respect to databases, this may be because they plan to continue using the TM themselves (for example, by providing it to their other translation providers to improve their speed and consistency), or because they are worried that some of the confidential information contained in the source texts in their database will be leaked to their competitors.⁶⁸

On the other hand, there may be cases where assignment is not dealt with in the contract. Small clients who need translation services infrequently or who only deal with a single provider may have no interest in “owning” a TM database. They may not even know that their translator is using a TM tool, or even what that is. They just want their translation. In those cases, the translation provider will retain any right in the database that may have arisen.

The above discussion assumes that the translation provider is the sole initial owner of the database. If it was the client who exercised the creative selection of texts, or participated in a joint creative selection, the right will arise there first, either alone or jointly with the translation provider.

But even if the translation provider is the sole initial copyright holder, and even if he or she retains a right in the translated texts by simply granting a non-exclusive licence to the client to use them, it is important to remember that the client holds the copyright in the source text. (Otherwise, the client would not be entitled to authorize the translation in the first place.) There is no reason to imagine that the copyright in the source text would ever flow to the translation provider. Any use it wants to make of its databases that contain source texts owned by its clients would therefore have to respect that underlying layer of independent copyright protection. This problem will be explored in more detail in Part III.C., below.

vi. Translation memory and the EU *sui generis* database right

Many TMs will not qualify for protection under copyright legislation because they lack originality in their selection and arrangement. However, they could still potentially qualify for protection under the *sui generis* database right created by Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases.⁶⁹

Databases need not be original to have economic value. Often, their value lies in the very exhaustiveness that prevents the selection process from being characterized as original.⁷⁰ The *sui generis* right was developed to prevent free-riding without extending copyright protection to mere facts. The Directive defines the object of protection as follows:

Member States shall provide for a right for the maker that there has been qualitatively and/or quantitatively a substantial investment in either the obtaining, verification or presentation of the contents to prevent extraction and/or re-utilization of the whole or of a substantial part, evaluated qualitatively and/or quantitatively, of the contents of that database.⁷¹

Note that in the Directive, database makers are only called authors if they qualify for regular copyright protection. Those who qualify only for the more limited *sui generis* protection (or no protection at all) are simply referred to as makers.

Some translation vendors and clients arguably make substantial investments in building their TM databases, so this may be an option for protection in certain jurisdictions (although this right does not yet exist in Canada or the United States). The term granted is shorter than for copyright (only 15 years), but because it can be renewed an unlimited number of times as the database is updated, it actually has the potential to be perpetual.⁷²

However, recent decisions related to sporting events have interpreted the substantial investment requirement narrowly, excluding any investments that the database maker has made in the creation of the data itself. In a case involving British football league fixtures, the Court of Justice of the European Communities held that

the expression “investment in ... the obtaining ... of the contents” of a database as defined in Art. 7(1) of the directive must be understood to refer to the resources used to seek out existing independent materials and collect them in the database. It does not cover the resources used for the creation of materials which make up the contents of a database.⁷³

In a similar case involving the British Horseracing Board, the Court of Appeal of England and Wales also held that the database in question was not covered by the *sui generis* right on the basis that most of the investment

had been in the creation of the data as opposed to its collection or verification.⁷⁴

If this continues to be the accepted interpretation, it will be an obstacle to TM database makers seeking protection under the *sui generis* database right. TM databases are generally not developed for their own sake, but are rather seen as a useful by-product of large investments in translation proper. The most significant part of the investment in any TM database will be in the creation of the translations that fill it. This is true both from the point of the translation vendor investing human resources and the client investing financial resources.

vii. What does copyright in a translation memory give the holder?

If one hits the jackpot and qualifies as an author of a compilation protectable under copyright legislation, what exactly is the prize? Less than you might imagine. The holder of copyright in a compilation of facts would be automatically entitled to the following rights under the Act:⁷⁵

3. (1) For the purposes of this Act, “copyright”, in relation to a work, means the sole right to produce or reproduce the work or any substantial part thereof in any material form whatever, to perform the work or any substantial part thereof in public or, if the work is unpublished, to publish the work or any substantial part thereof, and includes the sole right

[..]

(f) in the case of any literary, dramatic, musical or artistic work, to communicate the work to the public by telecommunication,

[..]

and to authorize any such acts.⁷⁶

The author of the work also gets non-assignable moral rights under section 14 of the Act (including the right to the integrity of the work and the right to be associated with the work), but this is not particularly meaningful in the context of translation memory. Anonymity is not unusual in the translation industry, even for translated texts,⁷⁷ so it seems especially unlikely that database compilers would fight for attribution. Even though there may be some exercise of non-obvious choice involved in selecting texts, it would be a stretch to call this an expression of the compiler’s personality. Secondly, the concept of integrity as it applies to novels or paintings has little meaning in the context of translation databases.

If we limit the analysis to the economic aspects of copyright, section 3 can be thought of as conferring both positive and negative rights. The positive rights include the holder’s exclusive right to produce or reproduce the whole or substantial parts of the work and the right to communicate the work to the public. The right to authorize may be seen as both a positive and a negative right, since the flip side of a right to authorize is a right to deny authorization. The copyright holder can therefore prevent acts of copying that are not covered by the fair dealing exceptions.⁷⁸

The distinction between positive and negative rights matters in this context because of the underlying layer of independently copyrightable material that makes up the database. Normally the data in a database is simply a series of unprotectable facts, but as we have already seen, translation memories are a unique hybrid of database and collective work with features of each. Because of the nature of the tool, the “facts” in the database that provide valuable linguistic information to the translator are inseparable from the literary works that are loaded into it, and it is necessary to take the copyright in these works into account when discussing what can legitimately be done with a copyright in a translation memory database.

The consequence of this is that the holders of copyright in a TM database will somehow have to deal with the copyright in all the underlying literary works before exercising their positive rights, since copying or communicating a TM database necessarily requires copying or communicating the literary works it contains. They will, however, be able to exercise their negative rights (including moral rights, if applicable) regardless of how many different players hold copyright in the texts.

The negative right to prevent others from copying is not as robust for databases as for other types of works:⁷⁹ because it is the selection and arrangement that are protected and not the underlying facts, one would have to copy most, if not all, of the database to be found to be infringing. However, this distinction makes little practical difference here, since potential copiers would have to copy the entire database to get any use out of it.

C. Non-Infringing Use of Translation Memory Databases

i. Strategies for overcoming the positive rights problem

There are four possible strategies for dealing with the positive rights problem: (a) acquire the rights in all the works in the database (both source texts and translations) in addition to the right in the database, (b) rely on the express or implicit consent of all the copyright holders before copying or communicating the work, (c) limit one’s uses of the underlying works to those covered by the fair dealing exceptions, or (d) fill the database only with materials that are in the public domain.

Clients are in the best position to overcome the positive rights problem, since only they have the possibility of acquiring both the database right and all the rights in the contents. They already have the rights in any protectable source texts, it is normal industry practice for the translation rights to be assigned to them in the translation contract, and they have the necessary bargaining power to insist on having any rights in the TM assigned to them as well, regardless of whether the translation provider would prefer to keep them.

Because translation providers will almost never hold the copyright in protected source texts, those who wish to continue using TMs developed in the course of a contract will have to rely on one of the other strategies. There will be no intellectual property limitations, for example, on the use they can make of databases that they have created exclusively from materials in the public domain. When some of the source texts are protected, they may be able to secure licences from the copyright holders. However, in many cases, this will be an uphill battle, especially if the holder has an interest in protecting the information contained in the source texts from competitors.

The extent to which they are able to rely on the fair dealing exception will depend on what use they plan to make of them. I will first review the scope of the fair dealing and fair use exceptions in Canada and the United States and then explore whether they might be relied upon by translators who wish to share⁸⁰ their TMs with others or continue using them internally to assist with work from other clients.

ii. The fair dealing/fair use exception

Not all unauthorized use of copyright-protected material constitutes infringement under copyright legislation. The Canadian *Copyright Act* provides a closed list of circumstances under which the unauthorized use of copyrighted work might be considered “fair”: research or private study, and, subject to certain conditions, criticism, review, or news reporting.⁸¹

Carys Craig writes that “[t]he fair dealing defence performs an integral function within the copyright system: it permits substantial uses of copyright-protected works, which would otherwise be infringing, in order to ensure that copyright does not defeat its own ends”.⁸² The purposes of copyright⁸³ should therefore be kept in mind when assessing whether a given use is fair.

Of course, it is possible to make infringing uses of works even within the Canadian statute’s enumerated list of activities. The test for whether such use will indeed be considered “fair”, and therefore non-infringing, was most recently developed by the Supreme Court of Canada in *CCH*: it involves balancing a variety of factors, namely, the purpose of the dealing, the character of the dealing, the amount of the dealing, alternatives to the dealing, the nature of the work, and the effect of the dealing on the work to determine whether the dealing is fair.⁸⁴

The corresponding provision in the U.S. copyright legislation⁸⁵ forgoes the closed list and provides an open-ended, contextual fair use exception into which any use considered “fair” could conceivably fall.⁸⁶ With the contextual factors elaborated in *CCH*, fair dealing has arguably been broadened in Canada to resemble the U.S. test in substance.

iii. Using translation memories internally

Before exploring whether internal use of translation memories might fall under the fair dealing provision, it will be useful to take a step back and look at the bigger picture of how translators use the texts themselves after a translation job has been delivered. Even before the advent of TM technology, it was not unusual for translators to keep copies of the source texts and their translations for future reference.⁸⁷ As I wrote in Part III.B.i, keeping an archive of the translations without their corresponding source texts is not particularly useful, and even if it were, the copyright in the translation is generally assigned to the client. In my experience, this is not something that is specifically negotiated with clients, which means that translators are keeping copies of the source and target texts without their express authorization. The first question we must answer is whether this practice is justifiable under copyright law.

The mere fact that it is a common business practice is not a defence on its own. In the words of David Vaver: “Just because a particular class of users has acted in a particular way for years without objection does not mean the usage is legal.”⁸⁸ However, customary usage can be a factor supporting an argument of implied licence.

There is a strong argument for implied licence in this case, if we look at the translation process. The client may deliver the source text in hard copy, electronically, or both, but the translation will almost invariably be created and delivered electronically, occasionally accompanied by a hard copy. If the source text is delivered electronically, the first thing that any professional translator will do is make a backup copy. He or she should also periodically back up the translation in progress, especially for larger documents. In the second case, there is no possibility of infringement if the contract stipulates that copyright in the translation remains with the translator until the work is delivered and paid for. In the first case, there is implied authorization to copy the source text; that it is simply part of the job that the translator has been commissioned to do.

Copyright law prohibits unauthorized copying, not access, so the only thing that would prevent a translator from being able to refer to those copies later would be a clause in the contract stipulating that any copies remaining in the translator’s possession be destroyed or returned to the client when the work is delivered. Otherwise, the translator is constrained only by confidentiality clauses and his or her professional code of ethics from showing unpublished or internal documents to third parties.

A client might argue that the translator’s implied licence to make copies of the source text for the purpose of delivering the final product includes an implied obli-

gation to use the work only once, or even to destroy it as soon as the job is done. This argument is not convincing with respect to clients who frequently employ the same translator. One of the main reasons for working with a single translation provider, apart from being satisfied with quality, is that it is the best way to ensure consistency across all of one's translated documentation. It is unreasonable to expect translators to be able to produce such consistency by relying on memory alone.

Even for clients who do not have a regular translation provider, the argument is weak. One of a translator's main selling points is experience. Although translators do gain automatic reflexes with time, requiring them to do less research for certain terms and sentence structures they encounter frequently, they still rely on research, and large archives are one of the best tools for producing the results desired by the client.

A client may indeed, if asked, agree that it wants any incidental copies destroyed after the contract, especially when the documents are sensitive. However, to argue that it expects translators to do that automatically, without being asked, is to argue that it expects translators to destroy the translations they have done for all other clients as well, or to negotiate each time for a licence to continue using them. Most clients likely do expect to benefit from the work a translator has done before.

A similar argument can be made for management consultants. Depending on the provisions in the contract, a consultant might either keep his or her copyright in the consulting report or assign it to the client. In the latter case, although the consultant technically can't copy the same report to deliver to a future client, he or she can certainly adapt the same ideas to the new client's situation and use the expression necessary to do that. Because consultants, like translators, are paid for their experience, it is reasonable to expect that they will keep a library of past projects to build on in the future.

The preceding analysis can equally be applied to the act of keeping a TM database created over the course of a contract. Some clients know that a TM is being created, while others may be unaware that the translator is using the technology. However, like making a backup copy, creating a new copy of the source text in a TM is simply a step in the process of delivering the speed and quality that the client is paying for.

What if the translator wants to build a TM from archived translations after they have been delivered? This requires making new copies and so could potentially infringe the client's copyright. It may not be possible to get a licence, in which case the translator would have to rely on fair dealing for any texts that are not in the public domain.

For the purposes of the Canadian fair dealing test, such a use would have to be characterized as research or private study in order to be caught by section 29 of the Act. In *CCH*, the Supreme Court of Canada held that

“[r]esearch must be given a large and liberal interpretation in order to ensure that users' rights are not unduly constrained. [...] research is not limited to non-commercial or private contexts”.⁸⁹ The Court also held that research conducted by lawyers for profit constituted research for the purposes of section 29. By analogy, linguistic research conducted by translators should also be covered. We can therefore proceed to the balancing stage of the test, looking at each of the factors listed by McLachlin C.J.⁹⁰

1. Purpose of the dealing

Applying “an objective assessment of the user/defendant's real purpose or motive in using the copyrighted work”⁹¹ leads us to commercial profit. TMs are created to increase translation speed, which means more money for translators getting paid by the word, and to increase translation consistency, which means translators can sell their services with a promise of higher quality products.

2. Character of the dealing

When creating a TM, a single electronic copy of the full text is made in the database.⁹² In bitext-type TMs, the text will remain whole, but in sentence-type TMs, the text is “destroyed”, leaving only a collection of distinct fragments, which may or may not in themselves attract copyright protection.⁹³ The texts copied into the TM are used in the same way they were before, as a reference archive, only more efficiently.

The Canadian government has stated that it is a public policy principle that the *Copyright Act* “be drafted, to the extent possible, in technologically neutral terms”.⁹⁴ This concern for technological neutrality is apparent when we look at the inclusion of section 80 of the Act to cover copying of music recordings for private use on other platforms, or the Supreme Court's recent decision in *Robertson*, in which it was decided that newspapers could validly reproduce freelancers' articles in CD-ROM format, since the CD-ROMs in question were found to preserve the essence of the newspaper as a collective work.⁹⁵

If it is indeed fair dealing for translators to keep reference archives of their previous work, as I argue, the principle of technological neutrality should allow them to update these archives into a more efficient format for the same purposes, even if this requires making incidental copies. Furthermore, this should be true regardless of whether the texts remain whole or are eventually fragmented in the database.

3. Amount of the dealing

Generally, the entire work is copied, but as the Supreme Court points out in *CCH*, “[i]t may be possible to deal fairly with a whole work”.⁹⁶

4. *Alternatives to the dealing*

To include a text in a TM, it is necessary to create a new copy. There is no reasonable alternative. Copying only part of the text is not as effective, since any part of a text may contain potentially valuable information for a translator when aligned with its target-language counterpart. The value of any given fragment of the text cannot be determined in advance, since it is dependent on its similarity to unknown future texts.

5. *Nature of the work*

The normal motivations for commissioning a translation are the desire to communicate the source text to a different audience or the desire to understand a foreign text oneself. In the latter case, the client probably does not hold the copyright, so the translation can only be commissioned for uses falling under fair dealing. In the former case, we may be dealing with texts that are published or are destined for publication, documents internal to an organization, or correspondence.

6. *Effect of the dealing on the work*

Copying the text into a TM will not have any effect on the work, since the translator is not making copies of anything to which he or she does not already have legitimate access, nor is he or she making copies to provide access to others.

7. *Conclusion*

Although it is difficult to establish that copies are made fairly when the purpose is for commercial gain, *CCH* shows that it is possible. A lawyer making copies for research purposes is not very different from a translator making copies for research purposes. The fact of copying the entire work, which can also be a sign of unfairness, is offset in this case by the lack of reasonable alternative and the complete lack of effect on the client’s market. All the factors considered together should easily lead to a finding of fair dealing.

While this is good news for translators, they still want to maintain good relations with their clients in order to continue getting their business. If clients feel that a translator cannot be trusted with their documentation, they will go elsewhere. However, as long as translators take adequate steps to protect the information, they have met their obligations. If a client wants to maintain tighter control over sensitive information, it can always take steps to include provisions for the destruction of copies in the contract. Translators are not in the best position to know what is sensitive and what is not, so it should not be their responsibility to suggest this. They should, however, treat *all* documents that they archive with or without authorization as though they are of the utmost secrecy.

A client who tries to rely on copyright purely as an extra weapon to protect confidential information is using it in a way that is in direct conflict with copyright’s

stated purpose, particularly in North America. Sunny Handa writes that one of the earliest purposes for copyright legislation was actually to provide an effective method of state censorship in 16th-century England.⁹⁷ However, we have since rejected this in favour of theoretical underpinnings that are diametrically opposed to the original purpose. He describes Canada’s social utility model for justifying copyright as a system whereby “authors are granted limited rights through a system of copyright protection in order to optimally encourage (i) the creation and (ii) the dissemination of their works, with goal of maximizing social utility”.⁹⁸ Social utility in this context is defined as knowledge and progress.⁹⁹ To allow clients to protect their sensitive information through copyright instead of more appropriate mechanisms like contract would be to turn back the clock to a 16th-century justification that runs counter to our society’s needs.

Finally, even if keeping copies for internal use and copying them to create translation memories constitutes fair dealing, care will have to be taken in defining what constitutes “internal” use. Solo freelancers and small partnerships are easy cases. For larger translation firms, internal use probably covers all employees. However, sending the database to a subcontractor may not be considered acceptable. Bringing subcontractors on-site and providing them with access to the memories might be acceptable from a purely copyright perspective, although care would have to be taken not to breach confidentiality.

iv. **Copying translation memories for other clients or translators**

Because of the potential of carefully stocked TM databases for increasing translation speed and improving translation quality, they have value in the translation marketplace. Translation providers may therefore be tempted to sell their memories, sell licences for their use, or pool their memories with those of their colleagues.

While they are free to do so when the TM is stocked only with public-domain content, they would not be able to make copies of protected texts for this purpose, since it would fall outside of the closed list of acceptable dealings in subsections 29 to 29.2. (Remember our assumption that someone else holds the copyright in at least the source text.)

There may still be a chance that an application of the more open-ended contextual test in the U.S. legislation would lead to a finding of fair use, although this seems unlikely if the complete texts are copied and communicated to others for direct profit. On the other hand, the fact that the translator is generally not competing in the same market might influence the final outcome. A factor that could further complicate this issue is the fact that a collection of unilingual texts (called a “unilingual corpus”) of a particular type or on a particular topic may be a valuable source of information, linguistic or other-

wise, in its own right. Since a bitext-type TM also happens to be two unilingual corpora “glued” together, so to speak, it is possible for a TM to be used in a way that competes more directly with the original authors in the markets of their source texts, even if that is not the translator’s intent at the time of sale.

However, even without such competition, there may be an unreasonable negative effect on the client if sensitive source texts end up in the hands of competitors. If the documents have not already been made publicly available by the client, the action is likely to constitute a breach of the translator’s obligation of confidentiality.

The above arguments are most clearly applicable to bitext-type TMs, but it is worth re-examining the question looking specifically at sentence-type TMs. Although a full copy of a text is made when entering it into the database, it is subsequently “chopped up” into smaller fragments. These fragments are usually at the sentence level, although they may also consist of headings or list items.¹⁰⁰

It is likely that the client’s copyright in the whole text does not subsist in these fragments. There might be some exceptions to this, although David Vaver clearly thinks not: “The occasional hyperbole to the contrary — that the taking of even a single sentence from the likes of a Dickens or a Shakespeare may infringe — is simply nonsense.”¹⁰¹ However, Vaver was writing about taking a single sentence for use in a new context. A sentence-type TM creates a wholly different scenario, since all the fragments of the whole text are in there somewhere, and although the user can only refer to one at a time, copying the TM still means copying the whole text (minus a few redundant or otherwise useless fragments that may subsequently have been deleted from the database).

Further complicating the matter is the fact the fragments are stored in their original order, and it may be possible to reconstruct the text by looking at the fragments in the alignment tool.¹⁰² One company avoids this problem by using scrambling algorithms,¹⁰³ although this is only viable after any misalignments have been corrected.

Finally, even if the client cannot claim copyright in any single sentence or fragment, it is possible for a single sentence to contain confidential information: “And the CARAMILK secret is” Translators could arguably remove all confidential sentences from the database before sharing it with others, but what a client considers sensitive will not necessarily be so obvious to an outsider in every case. Translators have an incentive to make the database more valuable by including as much linguistic information as possible and so may not always err on the side of caution. Clients have no incentive to spend resources for the benefit of the translator by going through the texts and flagging what they consider truly confidential. They are most likely to say that *everything* is confidential, leaving the translator no further ahead.

All of these factors taken together seem to indicate that this second use of translation memories is impermissible under *any* circumstances without the authorization of the holder of copyright in the texts.

IV. Promoting Translation Memory Reuse

A. Why Translators Want To Reuse Translation Memory

While the above result will come as no surprise to translators, they will nonetheless find it frustrating. The global demand for translation is increasing sharply in the information age without a corresponding increase in the number of translation professionals.¹⁰⁴ Increasing the number of bitexts in circulation would ease some of the growing pressures on the industry by allowing the existing professionals to meet the demand more quickly. Suzanne Topping has described the need for such pooling, while recognizing the significant obstacles to making it a fully legal and ethical reality.¹⁰⁵

Vaver argues that lawyers should not be able to make infringement claims against their colleagues in the profession for using their forms (see the text accompanying note 106, below). Even if they technically have copyright in something original they have created, it goes against the spirit of the profession and would not be in the public interest to enforce it. In support of his argument, he cites sociologist R. Greenwood:

The ethics governing colleague relationships demand behaviour that is cooperative, equalitarian, and supportive. Members of a profession share technical knowledge with each other. Any advance theory and practice made by one professional is quickly disseminated to colleagues through the professional associations. The proprietary and quasi-secretive attitudes toward discovery and invention prevalent in the industrial and commercial world are out of place in the professional.¹⁰⁶

This analysis seems equally accurate when applied to the translation profession,¹⁰⁷ as does Vaver’s following observation:

The second lawyer, if he is doing his job, does not simply copy the form; he decides whether and how far it suits his client’s purposes and tailors it accordingly. This involves a separate exercise of professional skill and judgment.¹⁰⁸

Translators can best serve all of their clients if they have extensive “precedents” to work with.¹⁰⁹ Every job is a new context, however, and the professional must consider even identical passages carefully before inserting them into a new text.

A central concept in understanding the conflicting motivations of clients and translators, alluded to in Greenwood’s passage above, is the fact that the same text has value to both parties, but for completely different reasons. Clients value the text for the ideas it contains; to them, the expression is usually little more than a vehicle for those ideas. Translators value the text for the expres-

sions it contains, particularly when they are connected to linguistically equivalent expressions in another language; to them, *the ideas are usually little more than a vehicle for those expressions*. Ironically, the less valuable a passage is to the client, i.e., because it is a banal statement lacking novelty, the *more* valuable it will probably be to a translator, who prizes phrases and structures that are likely to reappear in a variety of contexts.

It is true that expressions, as opposed to ideas, are the very things that copyright protects.¹¹⁰ However, if a new client happens to say the same thing as the old client in a new text (not uncommon outside of literature, especially at the sentence level), it would be nonsensical to claim that the translator must now find a different way to translate it just because the client holds a copyright in the first translation. If that were true, clients could also try to enjoin other translators from using “their” particular translation of the same phrase. Although there may potentially be dozens of ways to translate even a short phrase,¹¹¹ there may only be a few “best” ways, and maybe only one or two good ways in a given context. It would be contrary to public policy to fence these off.

To summarize, translators recognize their clients’ ownership of the source texts and even the target texts for which they have assigned the rights. It is understandable that clients would interpret the impulse to continue to use, and especially to sell, those texts as an attempt to free-ride, particularly when clients have spent significant resources developing multilingual terminology for emerging markets. Translators, on the other hand, are interested in a “layer” of the text that they instinctively, and most likely correctly, feel cannot be owned by anyone. On that logic, what can justifiably prevent them from gathering mere “facts” together into an original database and using it for their own purposes? Surely not copyright. But it is difficult to separate the “layer” of text that is valuable to the translator from the “layer” that is valuable to the client. For reasons explained above, this is not fatal to unauthorized internal use. However, the combined effects of the client’s copyright and the translators’ confidentiality obligations do justifiably prevent the unauthorized sharing of translation memories with others.

B. A Brief Economic Analysis of Translation Memory Reuse

Translation memories are a valuable economic resource, and because they are designed to be used by translators, it is translators who can make the most efficient use of that resource. Even clients who recognize the reuse potential instead of simply wanting to protect information must hand over the databases to translators, in-house or otherwise, in order to get any value from them. However, the division of rights in different layers of the memory among different parties seems to prevent efficient use at the outset. If the desired outcome is that translators end up with access to the TM database

without constraints on copying, is there a way of reallocating rights to achieve this?

The Coase theorem states that “when transaction costs are zero, an efficient use of resources results from private bargaining, regardless of the legal assignment of property rights”.¹¹² Because transaction costs are usually higher than zero, a useful corollary to the Coase theorem is that “when transaction costs are high enough to prevent bargaining, the efficient use of resources will depend on how property rights are assigned”.¹¹³ This realization led Cooter and Ulen to formalize a principle they call the normative Coase theorem: “Structure the law so as to remove the impediments to private agreements.”¹¹⁴

Because the rights in the various layers arise in two different places, a transfer of some kind will be necessary to “herd” them all together. At first glance, it might seem that transaction costs for such a transfer are indeed near zero. The parties do not need to spend resources locating one another, since they are already in a business relationship based on a translation contract. All that is required for the transaction to occur is the addition of a clause or two to a contract already under negotiation. But if this is the case, how can we explain the fact that the intellectual property is still ending up with the “wrong” party?

The flaw in this analysis is that it fails to take into account the double aspect of the value of the text. As was discussed in the previous section, the texts contained in the TM database are valuable to translators as a mass of examples of language use, particularly when linked to examples of how they have been translated in the past. But each source text is also valuable for the ideas it expresses, and each translation for the expression of those same ideas in a form that can reach a different audience. The clients are in a better position to make efficient use of that aspect of the text, creating a stronger force directing toward them those rights that originate with the translator.

Since there are two definitions of “efficient use” at play, we cannot necessarily rely on the modified Coase theorem to justify reassigning the right in the source text to the translator. Such a solution would rightly conflict with our intuition. However, translators are not necessarily worse off as a result of this “flow”. It just so happens that the rights are travelling in the necessary direction, at least initially.

Although we ultimately want TMs to end up back in the hands of translators, the TMs must go “through” the client first. Pooling all the rights in the hands of a single party is a necessary prerequisite for later reuse of a database by translators who were not involved in its creation. Once the client has all the rights, restrictions imposed by competing layers of the database disappear, and the client can copy and distribute freely.

The real problem in the current system is not that private bargaining results in the rights flowing in the wrong direction. Rather, another factor comes into play to prevent the now “unshackled” database from then

finding its way into the hands of other translators, the most desirable destination from an efficiency point of view. This factor is the client's lack of incentive to filter out confidential information and release the remaining linguistic information into the broader translation marketplace. If we want to intervene to promote efficiency, it is this factor that requires correction.

C. An Efficient Solution

Jost Zetzsche is the founder of TM Marketplace, a limited liability company whose *raison-d'être* is to make just such a correction. Zetzsche writes that some clients already recognize the potential value of internal reuse of TMs, but he adds that they "could astronomically multiply the value of these assets as the only party with legal standing to share these TM assets with other parties".¹¹⁵

One difficulty with convincing translation clients to participate as *vendors* in the market for TMs is that they are not in the business of selling language products and services. They buy language products and services in order to help them sell *other* products and services. They would have to dedicate resources to cleaning up the databases to rid them of confidential information and start working outside of their usual client network. Somebody has to make this worth their while.

TM Marketplace promises clients an instant network of potential licensees, technical assistance with making their TMs fit for outside commercial use, and advice on setting appropriate licensing fees. There is little risk to the client, who only pays transaction fees to TM Marketplace for successful transactions.¹¹⁶ This lowering of transaction costs to both the client and the end users of the TM database should allow private bargaining to do its job and get the databases into the right hands. Clients may never have an incentive to sell the databases outright, but translators only need a licence to make efficient use of them. This scenario represents one effective solution to a complex problem.

V. Conclusion: the Policy Perspective

With Part IV rounding out the picture of how TM databases currently are or could be used in the translation industry, we should take one final look at the TM ownership puzzle, this time from a policy perspective. The preceding discussion was an attempt to answer the question, "Who, if anybody, should own TM?" What follows is an attempt to answer a slightly different question: "Should *anybody* own TM?"

Under the current regime, nobody will be granted copyright in a non-original TM database, although various parties may own its contents. Which databases are original must be determined on a case-by-case basis, looking in particular at the judgment exercised during the selection of the texts. It is possible to consider some databases to result from original selection and arrange-

ment. However, we should only extend copyright protection to these if doing so advances the purposes of the copyright regime.

Vaver challenges our modern tendency to claim "that anything that might be of use to somebody is potentially valuable and should be turned into a commodity; and if commodified, it almost goes without saying that it should be protected".¹¹⁷ He adds: "The idea that some creative work needed no protection, because it would occur anyway, was not [always] as heretical as copyright campaigners would make it seem today."¹¹⁸ Desirable creative work will "occur anyway" if free-riding can be sufficiently prevented through mechanisms already in place (e.g. contract, tort, or technological protection measures) or if there is sufficient incentive to create *despite* the possibility of free-riding.

TMs are databases of unprotectable facts inextricably connected to protectable texts, so the copyright protection that already exists in the content layer of the database already does most of the work. The facts cannot be copied because they also happen to be texts. Protection is also available through contract and technological limitations to access.

Would there be adequate protection for original selections and arrangements of texts and translations drawn exclusively from the public domain? These could be valuable, and their creation and dissemination should be encouraged. But translation memories are not created for their own sake; they are created because they will be useful for a specific job or series of jobs. This means that even a TM filled with public-domain texts will be created anyway, regardless of the reduced protection. Under the circumstances, granting copyright protection would only make the database harder to disseminate in the short term, and so is not justifiable from a policy perspective.¹¹⁹

There seems to be no good reason to grant copyright in the database as a compilation. While a TM may not have an "owner" in the legal sense, its *de facto* owner will be the one who owns the contents, or controls the licensing. If the industry can continue to operate effectively without a compilation right, it should not be granted.

It is time to stop asking, "Who owns TMs?" Instead, we should look for ways to make the most efficient use of our language resources while respecting the legitimate interests of all stakeholders, including the clients, the translators, and the general public. First, I would recommend that clients learn to profit from their potentially valuable linguistic assets by licensing their TMs to translators who can use them. Second, I would recommend that translators who have created TM databases for clients be allowed to continue using them internally under the fair dealing provisions, provided that they do everything necessary to meet their non-disclosure obligations. It is possible for everyone to benefit.

Appendix: Illustrations of TM Tools

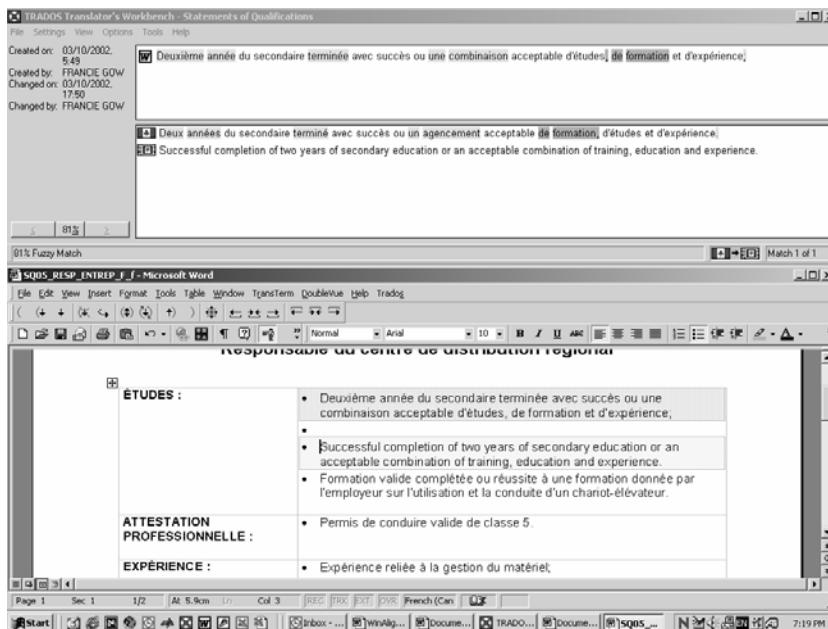


Figure 1: In SDL-Trados, an example of a sentence-type TM tool, each cell consists of a pair of corresponding sentences or sentence fragments. The second window at the top of the screen illustrates the content of a single cell. The top window shows the sentence from the new text to be translated that is being compared to the contents of the database. The full new text is shown in its word-processing program in the bottom half of the screen.

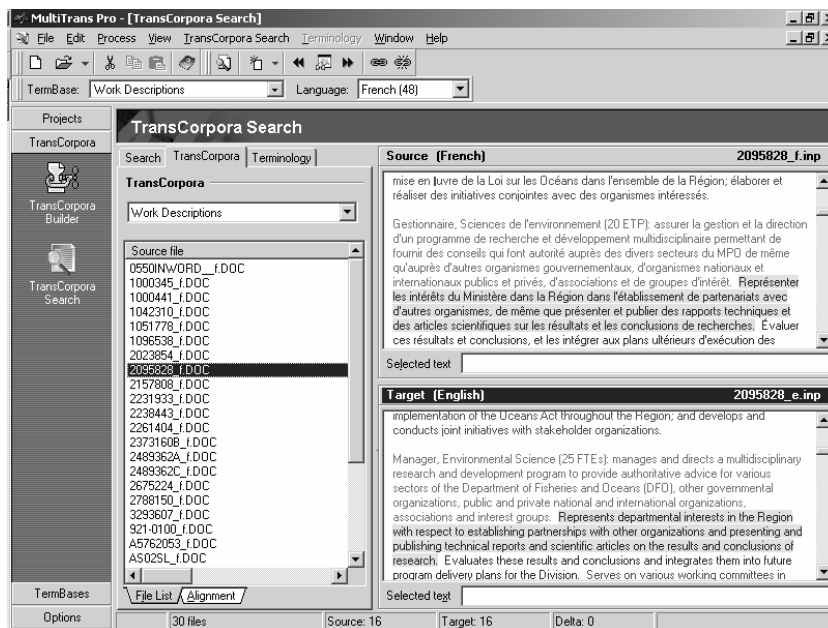


Figure 2: In MultiTrans, an example of a bitext-type TM tool, a single “cell” of the database consists of a complete bitext, as illustrated in the two linked windows on the right side of the screen. The source text is shown in the top window and the target text in the bottom.

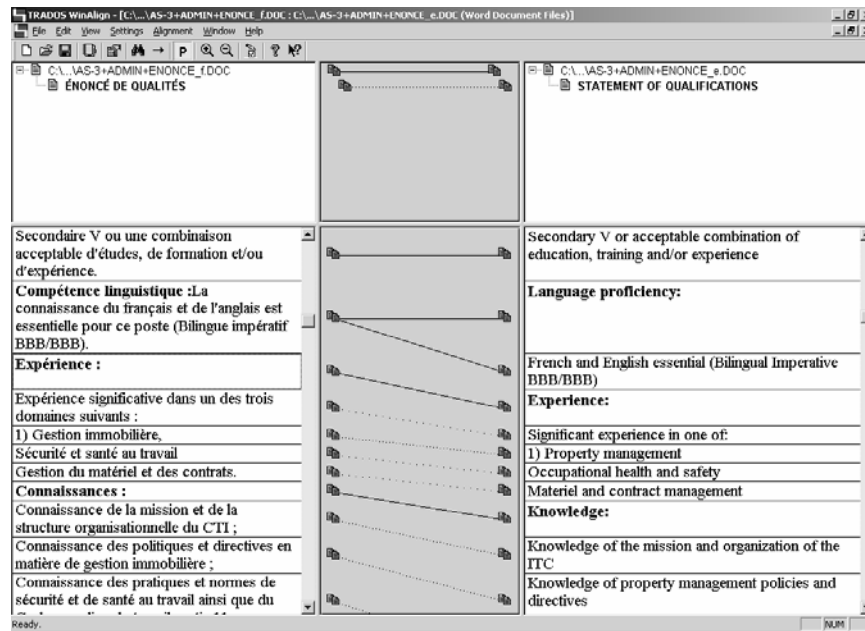


Figure 3: WinAlign, the alignment tool that comes with SDL-Trados, generates alignments between source and target sentences automatically, but misalignments can be corrected manually.

Notes:

- ¹ My statements regarding the translation industry are based on a combination of external sources, my professional experience as a legal translator and translation technologist, primarily within the Translation Bureau of the Government of Canada, and my graduate research at the School of Translation and Interpretation at the University of Ottawa.
- ² For a detailed comparative study of various types of translation memory tools, see Francie Gow, *Metrics for Evaluating Translation Memory Software* (M.A. Thesis, University of Ottawa School of Translation and Interpretation, 2003), online: <<http://www.chandos.ca/thesis.html>>.
- ³ “TM” is the standard abbreviation for “translation memory” in the translation industry. One difficulty with writing about translation memory in an intellectual property context is that “TM” is also the standard abbreviation for “trademark”. In this paper, “TM” will never be used to designate the latter concept.
- ⁴ The term “translation memory” may be used to describe either the type of software or the actual database produced by the tool. References to “a translation memory” should be interpreted to mean “a translation memory database”.
- ⁵ The leading developers include Atril (Déjà Vu), MultiCorpora (MultiTrans), SDL International (SDL-Trados), and STAR Group (Transit).
- ⁶ See, e.g., Tim Hallett, “Transit and TRADOS: Converging functions, diverging approaches” (2006) 5:4 *Localisation Focus* 9; Nancy A. Locke & Marc-Olivier Giguère, Review of MultiTrans 3.0 (2002) No. 51 Vol. 13 Issue 7 *MultiLingual Computing and Technology*, online: <<http://new.multilingual.com>>; and Catherine Arrouart & Claude Bédard, “Éloge du bitexte” (2001) 73 *Circuit* 30.
- ⁷ In some cases, if the sentences are stored in their original order, it may be possible to reconstruct a full text. The implications of this will be discussed in Part III.C.iv, below.
- ⁸ It is important to note that although the bitext constitutes a single cell, the alignment tool will create additional links between the two texts at the sentence or paragraph level *within* the cell. When the tool identifies a fragment of source text in your database that is similar to what is currently being translated, the program should be able to do more than simply indicate that the equivalent sought is “somewhere” in the translated text. The translator wants to be taken straight to that equivalent without spending time searching.
- ⁹ See the Appendix below for illustrations of the two types of tools.
- ¹⁰ Lynne Bowker, *Computer-Aided Translation Technology: A Practical Introduction* (Ottawa: University of Ottawa Press, 2002).
- ¹¹ Suzanne Topping, “Sharing Translation Database Information” (2000) No. 33 Vol. 11 Issue 5 *Multilingual Computing & Technology*, online: <<http://new.multilingual.com>>.
- ¹² Interview of Paula Shannon, Chief Sales Officer and Senior Vice President, Lionbridge Technologies Inc. (14 December 2006).
- ¹³ To read a thread on the Translator’s Café online forum that deals with TM ownership issues, see <<http://www.translatorscafe.com/cafe/MegaBBS/thread-view.asp?threadid=4881&start=61>> [Consulted July 30, 2007].
- ¹⁴ *Supra* note 11.
- ¹⁵ *Copyright Act*, R.S.C. 1985, c. C-42 [the Act], s. 2.
- ¹⁶ *Ibid.*, s. 2.1(2).
- ¹⁷ Irimi A. Stamatoudi, *Copyright and Multimedia Products: A Comparative Analysis* (Cambridge: Cambridge University Press, 2002) at 27 [footnote omitted].
- ¹⁸ *Ibid.*
- ¹⁹ The potential effects of these differences will be discussed in Part III.C.iv, below.
- ²⁰ *Compo Co. v. Blue Crest Music Inc.*, [1980] 1 S.C.R. 357 at 373.
- ²¹ *Supra*, note 15, s. 5.
- ²² *Ibid.*, s. 2, [emphasis added].
- ²³ *Ibid.*
- ²⁴ Madeleine Lamothe-Samson, “Les conditions d’existence du droit d’auteur : n’oublions pas l’auteur et sa créativité!” (2003) 15 C.P.I. 619 at 643 [Lamothe-Samson].
- ²⁵ *Supra* note 15, s. 2 [emphasis added].
- ²⁶ For persuasive evidence of this, try “back-translating” even the most formulaic document. It is easy to end up surprisingly far from the original source text, *even if you recently did the translation yourself*.
- ²⁷ It makes little difference for these purposes whether the fragment is a sentence, a paragraph, or even the whole text, along with its translated counterpart in each case.
- ²⁸ Modern machine translation programs are actually unlikely to make that particular mistake, since they are programmed to recognize cues such as articles before a word indicating that it is probably a noun. The example is simply meant to illustrate that they often make inappropriate and

unexpected selections in cases of ambiguity that a translator would never make.

- ²⁹ In reality, it would rarely make sense to store pure machine translations in a TM, since they tend to be unreliable unless they have been revised by a human. Such a database might be useful for researchers studying the performance of machine translation, but not as a source of inspiration for humans doing subsequent translations. The rest of the paper will be based on the assumption that only translations done by humans are stored in the databases (or, at worst, machine translations that have been revised by humans).
- ³⁰ See, e.g., *supra* note 24 at 641; David Vaver, *Copyright Law* (Toronto: Irwin Law, 2000) at 57 [Vaver, *Copyright*].
- ³¹ See *Théberge v. Galerie d'art du Petit Champlain inc.*, [2002] 2 S.C.R. 336 at paras. 30-31 [Théberge]; U.S. Const. art. 1, §8, cl. 8; Sunny Handa, *Copyright Law in Canada* (Markham, ON: Butterworths, 2002) at 28.
- ³² *Copyright Act* 17 U.S.C. §102.
- ³³ *Berne Convention for the Protection of Literary and Artistic Works*, 9 September 1886, as last revised at Paris on 24 July 1971 and amended in 1979, S. Treaty Doc. No. 99-27 (1986), online: <www.wipo.int/treaties/en/ip/berne> [Berne Convention].
- ³⁴ Daniel J. Gervais, “Feist Goes Global: A Comparative Analysis of the Notion of Originality in Copyright Law” (2002) 49 *Journal of the Copyright Society of the USA* 949 at 972.
- ³⁵ *University of London Press, Limited v. University Tutorial Press Limited*, [1916] 2 Ch. 601 at 609 [London Press].
- ³⁶ Lamothe-Samson, *supra* note 24 at 641-42.
- ³⁷ See *Walter v. Lane*, [1900] A.C. 539; *London Press, supra* note 35; *Ladbroke (Football), Ltd. v. William Hill (Football) Ltd.*, [1964] 1 All E.R. 465 at 469 (H.L.).
- ³⁸ *Supra* note 34 at 958.
- ³⁹ *Feist Publications v. Rural Telephone Service Co., Inc.*, 499 U.S. 340 (1991) [Feist].
- ⁴⁰ *Ibid.* at 345.
- ⁴¹ *Ibid.* at 359.
- ⁴² *Ibid.*
- ⁴³ *CCH Canadian Limited v. Law Society of Upper Canada*, [2004] 1 S.C.R. 339 at para. 25 [CCH].
- ⁴⁴ Gervais, *supra* note 34 at 955.
- ⁴⁵ *Ibid.*, see text accompanying note 41.
- ⁴⁶ Teresa Scassa, “Distinguishing Functional Literary Works from Compilations: Issues in Originality and Infringement Analysis” (2006) 19 I.P.J. 253 at 258.
- ⁴⁷ *Supra* note 15, s. 2.
- ⁴⁸ Scassa, *supra* note 46 at 259.
- ⁴⁹ There are two ways to build a TM database. One is to archive a block of previous translations, and the other is to build the database “on the fly”. Since the new translation generally gets incorporated into the database immediately for future reuse, in some cases sentence by sentence while you are translating, it is possible to start translating with an empty database and allow it to build up by itself with every new translation. Even with an initially empty database, the tool can be useful right away if you are translating a text with a high degree of internal repetition. The on-the-fly method clearly requires no judgment in the selection process, but even the archiving method can be devoid of judgment if you take the approach of archiving everything you have indiscriminately. See Bowker, *supra* note 10 at 107-10.
- ⁵⁰ *Feist, supra* note 39 at 359.
- ⁵¹ *Matthew Bender & Company, Inc. v. West Publishing Co.*, 158 F.3d 674 (2nd Cir. 1998) at 677 [Matthew Bender].
- ⁵² *Ibid.* at 682.
- ⁵³ *CCH, supra* note 43 at para. 35; *Grove Press, Inc. v. Collectors Publication, Inc.*, 264 F. Supp. 603 at 605 (C.D. Cal. 1967) as cited in *Matthew Bender, ibid.* at 688, note 13.
- ⁵⁴ Theoretically, software could be designed to track these deletions, but implementation costs would outweigh any potential benefits, particularly given the remoteness of the possibility of any given translation provider becoming involved in judicial proceedings in this regard.
- ⁵⁵ See text accompanying note 52.
- ⁵⁶ Lamothe-Samson, *supra* note 24 at 636-637.
- ⁵⁷ *Supra* note 15, s. 13.
- ⁵⁸ Lamothe-Samson, *supra* note 24 at 638-641.
- ⁵⁹ *Supra* note 15, ss. 6-12.
- ⁶⁰ See Part III.B.i, above, for more on this topic.
- ⁶¹ See, e.g., Antonio S. Valderrábanos, “Who Owns What? Some Insights on TM Ownership”, online: TransRef — The Translation Reference Center <www.transref.org/u-articles/TMOwnership1.asp> and Translator’s Café, *supra* note 13.
- ⁶² *Supra* note 15, s. 2.
- ⁶³ Betty Cohen, “Bien connaître ses droits” (2003) 82 *Circuit* 10 at 10.
- ⁶⁴ *Supra* note 15.
- ⁶⁵ The United States has a similar provision in its copyright legislation, but there the term “work made for hire” is used. *Supra* note 32, §201(b).
- ⁶⁶ *Supra* note 15.
- ⁶⁷ In *Robertson v. Thomson Corp.*, 2006 SCC 43, [2006] 2 S.C.R. 363 [Robertson], which deals with assignments of copyright in freelance articles for the purpose of reprinting them in electronic databases, there is also a situation of unequal bargaining power. Giuseppina d’Agostino recommends that the balance be redressed either through legislative reform with regard to copyright contract issues, or at least through the application of equitable doctrine by the courts, so that freelancers are not left without any rights at all. See Giuseppina D’Agostino, “Canada’s Robertson Ruling: Any Practical Significance for Copyright Treatment of Freelance Authors?” (2007) CLPE Research Paper 5/2007, vol. 03, no. 02. A similar equitable argument can be made for assignments of copyright in translation memory databases, so that the stronger bargaining power of the clients does not leave them without any possibility of being compensated for work in creating the database above and beyond the translation work itself.
- ⁶⁸ Jost Zetzsche, “Translation Memories: The Discovery of Assets” (2005) No. 72 Vol. 16 Issue 4 *MultiLingual Computing & Technology*, online: <<http://new.multilingual.com/>>.
- ⁶⁹ EC Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases [1996] OJ L77/20, online: <<http://eur-lex.europa.eu/>> [the Directive].
- ⁷⁰ Howard P. Knopf, “The Database Dilemma in Canada: Is ‘Ultra’ Copyright Required?” (1999) 48 U.N.B.L.J. 163 at 172.
- ⁷¹ *Supra* note 69, art. 7(1).
- ⁷² David Lametti, “Coming to Terms with Copyright” in Michael Geist, ed., *In the Public Interest: The Future of Canadian Copyright Law* (Toronto: Irwin Law, 2005) 480 at 512.
- ⁷³ *Fixtures Marketing Ltd. v. Oy Veikkaus Ab*, C-46/02, [2005] E.C.D.R. 2 at para. 49.
- ⁷⁴ *The British Horseracing Board Ltd. v. William Hill Organization Ltd.*, [2005] EWCA Civ 863 at para. 35.
- ⁷⁵ I have omitted enumerated rights that relate to other kinds of works.
- ⁷⁶ *Supra* note 15, s. 3.
- ⁷⁷ Cohen, *supra* note 63 at 11.
- ⁷⁸ Further discussion of fair dealing under the *Copyright Act* will be found in Part III.C., below.
- ⁷⁹ *Feist, supra* note 39 at 359.
- ⁸⁰ “Sharing” should be understood here to mean “giving access”, whether gratuitously or for sale or barter.
- ⁸¹ *Supra* note 15, ss. 29-29.2.
- ⁸² Carys Craig, “The Changing Face of Fair Dealing in Canadian Copyright Law: A Proposal for Legislative Reform” in Geist, ed., *supra* note 72, 437 at 437.
- ⁸³ See text accompanying note 31.
- ⁸⁴ *Supra* note 43 at para. 53.
- ⁸⁵ *Supra* note 32, §107.
- ⁸⁶ Craig, *supra* note 82 at 440.
- ⁸⁷ Bowker, *supra* note 10 at 93-94.
- ⁸⁸ Vaver, *Copyright, supra* note 30 at 172.
- ⁸⁹ *Supra* note 43 at para. 51.
- ⁹⁰ *Ibid.* at para. 53.
- ⁹¹ *Ibid.* at para. 54.

⁹² See Part II above for a brief overview of how translation memory works.

⁹³ This possibility will be discussed in Part III.C.iv., below.

⁹⁴ Industry Canada and Canadian Heritage — Copyright Policy Branch — Archives, *Government Statement on Proposals for Copyright Reform* (March 2005), online: Canadian Heritage <<http://pch.gc.ca/progs/ac-ca/pda-cpb/index-e.cfm>>.

⁹⁵ *Robertson*, *supra* note 67.

⁹⁶ *Supra* note 43 at para. 56.

⁹⁷ Handa, *supra* note 31 at 28-29.

⁹⁸ *Ibid.* at 120.

⁹⁹ *Ibid.* at 135.

¹⁰⁰ See Figure 3 in the Appendix for an illustration.

¹⁰¹ Vaver, *Copyright*, *supra* note 30 at 144 [footnote omitted].

¹⁰² This problem is clearly illustrated in Figure 3 in the Appendix.

¹⁰³ TM Marketplace: Translation Memory Owner FAQ: FAQ#1 online: TM Marketplace, <<http://www.tmmarketplace.com/ownerFAQ.htm>>. TM Marketplace is discussed in Part VI.C., below.

¹⁰⁴ Canadian Translation Industry Sectoral Committee, *Survey of the Canadian Translation Industry: Human Resources and Export Development Strategy, Final Report of the Canadian Translation Industry Sectoral Committee* (30 September 1999) at 80, online: <<http://uottawa.ca/associations/csict/princi-e.htm>>.

¹⁰⁵ *Supra* note 11.

¹⁰⁶ R. Greenwood, "Attributes of a Profession" (1957) 2-3 *Social Work* 44-55, reprinted in H.M. Vollmer & D.L. Mills, eds., *Professionalization* (Englewood Cliffs, NJ: Prentice-Hall, 1966) 10 at 15, cited in David Vaver, "Copyright in Legal Documents" (1993) 31 *Osgoode Hall L.J.* 661 at 676, note 57 [Vaver, "Legal Documents"].

¹⁰⁷ This comparison should not be taken as a criticism of translation clients. "Proprietary and quasi-secretive attitudes" can be necessary and appropriate in industrial and commercial contexts. Conflicts arise when

actions taken in one context are judged by norms more suitable to another.

¹⁰⁸ Vaver, "Legal Documents", *supra* note 106 at 676.

¹⁰⁹ The codes of ethics of provincial associations of translators and interpreters often contain a rule promoting cooperation among members of the profession. The Association of Translators and Interpreters of Ontario Code of Ethics, for example, provides that, "Members shall foster the development of their profession, by sharing their knowledge with colleagues in a spirit of mutual assistance". Association of Translators and Interpreters of Ontario, By-laws of ATIO, Appendix 1, online: <http://www.atio.on.ca/info/ByLaws/Code_Ethics.pdf>.

¹¹⁰ Lamothe-Samson, *supra* note 24 at 632.

¹¹¹ See the discussion of "Un, deux, trois, j'irai dans les bois" at Part III.B.i., above.

¹¹² As formulated by Robert Cooter & Thomas Ulen, *Law and Economics*, 2d ed. (Reading, MA: Addison-Wesley Longman, 1997) at 82. The Coase Theorem was first published by Ronald H. Coase in his landmark article "The Problem of Social Cost" (1960) 3 *J.L. & Econ.* 1.

¹¹³ Cooter & Ulen, *ibid.*

¹¹⁴ *Ibid.* at 89.

¹¹⁵ *Supra* note 68.

¹¹⁶ FAQ #9, *supra* note 103.

¹¹⁷ Vaver, "Legal Documents", *supra* note 106 at 682.

¹¹⁸ *Ibid.* at 682.

¹¹⁹ Elizabeth F. Judge makes the same argument with respect to Crown copyright: "This incentive system is difficult to square with Crown copyright" and "In the Crown copyright context, copyright could be as likely to keep information from circulating as to provide an incentive to publish. Crown copyright could, in theory, be used to censor materials, delay access, or to chill discussion". Elizabeth F. Judge, "Crown Copyright and Copyright Reform in Canada" in Geist, ed., *supra* note 72, 550 at 571, 572.