

Copyright Issues in Translation Memory Ownership

Ross Smith

Translation Manager
PricewaterhouseCoopers Spain
ross.smith@es.pwc.com

Introduction

In the last two decades terminological databases and translation memory (TM) software have become ubiquitous in the professional translation community, to such an extent that translating without these tools has now become almost unthinkable in most technical fields. Until recently, however, the question of who actually owned these tools was not considered relevant. Most users installed the software on their desktop computers and built their termbanks and TMs without ever considering the possibility that they could be extracted and sent elsewhere, or that the content of their databases might in fact belong wholly or partly to someone else. With the advent of high-speed data transmission over computer networks, however, these resources have been released from the confines of individual PCs and have begun circulating around the Internet, causing a major shift in the manner in which they are perceived and uncovering new commercial possibilities for their exploitation.

The first outlet for TMs as tradable products was set up in 2007 as a joint initiative between Multilingual Computing, Inc. and International Writers' Group, with the name of TM Marketplace (www.tmmarketplace.com). The creators of TM Marketplace use the term "TM assets" to refer to the products traded on this site. As a result, translation memory files can now be bought, sold and licensed as individual assets. Termbanks, already commercialised in the form of word-lists and specialised glossaries, are also affected by the Internet revolution because as well as being easy to transfer electronically, they are particularly vulnerable to illegal copying.

In this new scenario, a fundamental question has arisen: Who actually owns the assets? In other words, who holds the intellectual property rights to these resources and what legal protection is available to them? With the increasingly widespread use of this technology it is important to seek a solution to the ownership issue, since otherwise conflicts will inevitably arise. When an agency or corporate client instructs a translator to deliver the TM file created in the translation process together with the translation itself, are they acting legitimately as the legally authorised owners of the file? Or are they perhaps exploiting their relative position of dominance to coerce translators into handing over items of value which they would rather keep locked up on their hard discs? To be able to buy, sell, licence or assign these resources we

need a clear understanding of who owns what and the best starting point for achieving such an understanding is an analysis of intellectual property legislation in the European Union and the United States, as well as on an international scale.

It is advisable firstly to clarify the meaning of "intellectual property" and "copyright", as these terms tend to be confused by the lay user. According to the World Intellectual Property Organisation (WIPO), intellectual property can be split into two main areas: copyright plus related rights, and industrial property. These are generally understood to include rights relating to the following:

- Literary, artistic and scientific works (copyright)
- Performances of performing artists, phonograms, and broadcasts (related rights)
- Inventions in all fields of human endeavour, scientific discoveries, industrial designs, marks and commercial names, protection against unfair competition (industrial property)
- All other rights resulting from intellectual activity in the industrial, scientific, literary, and artistic fields

(source: WIPO publication No. 489)

It can therefore be appreciated that copyright is one area within the more comprehensive field of intellectual property. As intellectual creations, the content of databases and translation memories are covered by copyright (often called "authors' rights" in other European languages), not by industrial property rights.

Applicable legislation

The most important copyright laws which can be invoked with regard to the protection of authors' rights in the creation of termbanks and TMs, and which will be referred to below in the detailed analysis of these tools, are:

- International treaties: Berne Convention 1886 (last amended in 1979); TRIPS Agreement 1994; WIPO Copyright Treaty 1996
- US legislation: Copyright Act 1976; Digital Millennium Copyright Act 1998
- European legislation: EU Parliament and Council Directive 2001/29/EC on the harmonisation of certain aspects of copyright and related rights in the information society ("Copyright Directive"); EU Parliament and Council Directive 96/9/EC on the legal protection of databases ("Database Directive"); derived Member State laws.

On an international level, the precise provisions of the above treaties that concern collections of data and databases *per se* are as follows:

1. Berne Convention: Article 2(5). "Collections of literary and artistic works such as encyclopaedias and anthologies which, by reason of the selection and arrangement of their contents, constitute intellectual creations shall be protected as such." Article 2(1):

“The expression ‘literary and artistic works’ shall cover every production in the literary, scientific and artistic domain, whatever may be the mode or form of its expression.”

2. TRIPS Agreement: Article 10 (2). “Compilations of data or other material, whether in machine-readable or other form, which by reason of the selection or arrangement of their contents constitute intellectual creations shall be protected as such”.
3. WIPO Copyright Treaty: Article 5. “Compilations of Data (Databases) ... Compilations of data or other material, in any form, which by reason of the selection or arrangement of their contents constitute intellectual creations, are protected as such”.

On a European level, It is very significant for our present purpose that the European Union has passed legislation specifically governing the legal protection of databases as this provides a relatively clear framework for specifying who owns what with regard to terminological data banks. A number of specialists have suggested that this legislation should also apply to translation memories on the basis that they are a kind of database as well, as is discussed farther on. It should be noted that no specific law governing the protection of databases has yet been enacted in the United States, even though draft legislation on this area has existed for a number of years.

Copyright protection for termbanks

As mentioned above, the existence of specific legislation in Europe governing ownership rights to databases facilitates the task of identifying the owner of a given termbank. In this respect, Article 3.1. of the above-mentioned Directive 96/9/EC states: “databases which, by reason of the selection or arrangement of their contents, constitute the author's own intellectual creation shall be protected as such by copyright”. The essential justification for this protection is stated in the following terms in the preamble (Clause 7) of the Directive: “the making of databases requires the investment of considerable human, technical and financial resources while such databases can be copied or accessed at a fraction of the cost needed to design them independently.” What justifies the copyright protection, therefore, is the creative effort *and* the investment in time and money required to assemble the database. It is worth mentioning, in passing, that this protection exists irrespective of whether a database is formally registered or not. Database protection is also provided in the above-mentioned international treaties (TRIPS and WCT) and in US law, where the creativity/originality factor is seen as a key criterion for database copyright cover. What is different in EU legislation is that in addition to copyright it provides a second layer of cover called a “sui generis database right” where it is sufficient to demonstrate that the author has made a “substantial investment” in qualitative and quantitative terms when making the database to obtain legal protection, without having to provide evidence for a creative or original element. The “sui generis” right has a duration of fifteen years, which is considerably shorter than copyright, which in Europe lasts for the author's lifetime plus seventy years.

The EU Database Directive and the Member State laws based on it provide protection for the structuring of the database and the arrangement of the content but this protection does not apply to the content itself, as is made clear in Article 3.2.: “The copyright protection of databases provided for by this Directive shall not extend to their contents and shall be without prejudice to any rights subsisting in those contents themselves”. Again, this reflects the approach historically taken in the major international treaties. The content of databases, including termbanks, is often factual, non-original information which under traditional intellectual property doctrine cannot be copyrighted. In such cases, authors have nothing to worry about. Alternatively, the content may already be under copyright which will remain with the author of that content, and here authors need to be more wary.

One problematic area in this regard, which is particularly relevant in the case of termbanks, is the inclusion of quotes to illustrate or explain specific terms. While there are no legal stipulations which set precise limits on the size of quotes, copyright laws and jurisprudence demand that the name of the quoted work and its author be mentioned if possible. The question of the size of quotes and extracts brings into play the thorny issue of *fair use* or *fair practice*, which is one of the most controversial areas in the global intellectual property debate. The Berne Convention (Article 10) states the following in this connection: “It shall be permissible to make quotations from a work which has already been lawfully made available to the public, provided that their making is compatible with fair practice, and their extent does not exceed that justified by the purpose”. For its part, The European Union Copyright Directive (Art.5.3) allows “quotations for purposes such as criticism or review, provided that they relate to a work or other subject-matter which has already been lawfully made available to the public and that, unless this turns out to be impossible, the source, including the author’s name, is indicated, and that their use is in accordance with fair practice, and to the extent required by the specific purpose”. In theory, therefore, it is lawful to quote copyrighted material without having to obtain the author’s permission provided that the purpose is valid, the extent is reasonably short and the source is stated, but the vagueness of these provisions means that in practice the story is quite different. Few authors are prepared to risk a confrontation on the tenuous grounds of *fair use* with the battle-hardened lawyers of the world’s leading media corporations, who nowadays hold the rights to almost everything that is published in literary or musical form. The result is that unless an author is sure that a source is in the public domain, or unless he has specific authorisation to publish, the most advisable course is often considered to be to avoid quotes altogether.

As M. Johansson points out (2007), this is very much the case for termbank compilers. The decision to unify the three largest databases of the European Institutions into one termbank – IATE – brought to light numerous cases of quotes (“cut and pastes”) from various sources which lacked any reference to the original work or author. Some of these were from well-known copyrighted reference sources, such as encyclopaedias or dictionaries. Although the doctrine of *fair practice* might perhaps be invoked in such cases, Johansson insists that terminologists should be careful: “There is no coherent *fair use* principle applied universally; copyright law, while based on a number of international conventions, is still adopted nationally with national variations and therefore we would need to investigate the national law of the country where the

source is located." He concludes: "It may be very difficult to know what country a source is from, especially on the Internet" (Johansson 2007). There is a dual problem, therefore: on the one hand, there is no universal, consistent doctrine of *fair use* and therefore invoking it is always risky; on the other, even if *fair use* seems to be applicable, it is often problematic to determine (as one must) which national laws are involved. On this basis, his advice to termbank creators is to quote as little as possible and if the source cannot be given, not to quote at all.

Despite these exceptions, it must be stressed that the creators of termbanks enjoy considerable legal protection, especially in the European Union. Their ownership rights to the structure and arrangement of the termbank are assured and they will not usually need to be concerned about the terms they include since, logically, it is almost impossible to copyright individual words. In any event the terms they include will usually constitute universal, factual data which do not qualify for intellectual property cover. We must reiterate, however, that in cases where sources are known to be protected, e.g. explanatory quotes, a reference to the author and work being cited must always be included and if possible, their permission should also be requested.

Copyright protection for translation memories

We shall now analyse the lines of argument that may be taken into account when attempting to determine who holds the intellectual property rights to a given TM.

For translators who work using TM software the answer to the ownership question may seem simple: we make the memory using our effort and specialised knowledge, therefore the intellectual property rights to the TM belong to the translator. Yet the end-client who is paying for the translation and who supplied the source document may see things as being equally simple, but from the opposite side: the client created the source document and will buy the translation, so is it not logical that the relevant TM, which contains both, should also belong to that person? In some cases there will also be a third party involved, namely a translation agency or company, which may have generated the TM and which will therefore regard itself, not its own client or the freelancer it has contracted for the job, to be the lawful owner of the relevant intellectual property rights. The common practice here is for corporate LSPs to assign the TM and the rights to it directly to the customer, as added value in the translation service provided.

The number of potential owners of a translation memory can at least be restricted to three: the translator, the translation buyer (client), and any agency involved. Ownership could also be shared. Any claim by the software producer can of course be discarded, since the program supplied by the producer does not generate, but merely facilitates, the translation. TagEditor belongs to SDL-Trados, but the item made using that technology does not. If such were the case, all documents made using MS Word, for instance, would belong to Microsoft. Therefore, the party creating the translation, the party purchasing the translation and the party acting as intermediary between them, if any, are the candidates for ownership.

Owing to the complexity of this issue there is no consensus among the few legal experts who have ventured to express an opinion on it. The majority of those who have actually taken a stance, however, clearly advocate that, unless copyright is previously transferred under contract, translation memories belong to the translators who create them. This approach has been advocated in writing by J. Marcos (2001) and TMMarketplace (2005), as well as informally by PricewaterhouseCooper's IP department and by a government copyright lawyer (Smith 2007). However, it has been questioned by another specialist lawyer consulted for the purpose of this paper, who considers that ownership might be split between the translator and the source-text author. What is evident is that this complexity is not due to a lack of specific legislation but to the nature of translation memories themselves, and the manner in which information is segmented and stored by them. This can be illustrated in more detail by examining the arguments in favour of individual translators as holders of the intellectual property rights.

To qualify for intellectual property protection, a written work, among other things, has to be original and of a certain size. Commonplace expressions cannot be protected because by definition they are not original, while any "work" which is just one sentence long (setting aside advertising slogans and minimalist poems) will almost never be afforded copyright cover. These are the main arguments employed to advocate, from a legal viewpoint, that the translator who creates the TM also owns it. The kind of documents that are translated using TM systems are not usually composed of language that can be regarded as original, or attributable to a specific author or entity. For instance, no-one can claim intellectual property rights to phrases such as "profits have fallen by 12%" or "insert the card in the left-hand slot". At the same time, the TM program dissects the source document into segments which, at most, are a sentence long and therefore are disqualified from copyright protection due to their insignificant size. These arguments therefore rule out the source document author as owner of the TM in favour of the translator. The translator's claim to the intellectual property rights is further boosted by the traditional "sweat of one's brow" argument, whereby the translation memory should be regarded as belonging to the translator because it is the product of his or her own effort and specialised knowledge.

This position would also be solidly supported if translation memories could be equated to databases since in that case what would be protected would not be just the author's creative effort, but also the investment in time and money made to compile and store the most frequently used sentences and phrases. Most lawyers specialising in intellectual property who have studied this issue suggest that the protection provided for databases should also cover translation memories, since they are a kind of database. EU Directive 96/9/EC defines a database as follows:

"the term 'database' should be understood to include literary, artistic, musical or other collections of works or collections of other material such as texts, sound, images, numbers, facts, and data; it should cover collections of independent works, data or other materials which are systematically or methodically arranged and can be individually accessed".

Most translation memory programs employ database software and professionals unthinkingly tend to refer to "databases" when describing individual translation memories (the term is used constantly, for instance, in the TMMarketplace White Paper). A specific TM is a document mechanically broken down into individual sentences or phrases which are stored in electronic form with their other-language equivalents. A TM cannot be regarded as a "collection" *per se* though it is certainly "methodically arranged" as segments are stored in pairs in a manner that optimises the speed of retrieval. Entries can also be "individually accessed" when necessary for the purpose of correction or verification. It is relevant to note in any case that in the EU Directive there is no specific requirement that a database should actually be created for the purpose of retrieving individual items of information. In the last instance, of all the formats covered by intellectual property legislation, the data base is the one that comes closest to a translation memory file. As the copyright in the data base belongs to the person that created it, these arguments favour the position that the translator is the owner of the TM. It should be reiterated that the translator owns the structure, not the content, but as mentioned above, that content is unlikely to qualify for protection in any event.

However, other experts disagree with this line of reasoning and consider that the party supplying the source document has an equally solid claim to ownership. Specifically, while it is true that the source document is atomised within the translation memory and each segment of text cannot be regarded as an original creation, it could be argued that, nevertheless, the source text as a whole remains inside the relevant TM file and is merely being stored in a different form, from which it could be re-generated at any time. From this perspective, the source-text half of the translation memory continues to belong to the client since it has not undergone any significant transformation, while the translated half belongs to the translator. The ownership issue is further complicated by the fact that many translation memories contain original document segments from a number of different sources, and often contain target segments by various different translators as well. If the author of the source document is deemed to retain intellectual property rights to the document even after it has been segmented and stored in the TM file, then there will be as many owners to the source half of a given TM file as there are authors. The viewpoint of the translation purchaser (usually the same person as the source-document creator) should also be taken into account. Many translation service users and even some translators regard it as common sense to think that if you write a document, have it translated and then buy the translation, you own both the source and translated documents, in whichever format (paper, txt, html, xml, tmx, ttx, etc.) they may be stored.

We should note here that, although the notion that a translation memory contains a source-document half and a translated-document half may be useful for the arguments of those who propose that copyright remains with the original author or client, in fact this approach is misleading. TM software does not work like an alignment program, with the source text on one side, conveniently broken down into sentences, and the translated text neatly lined up on the other. Rather, a TM is an expandable collection of bilingual segments which have been transformed into something quite different from the source/target documents. TMs are bunches

of codes and tags which are only comprehensible if rebuilt using TM programs. If we open a TM document, or database, and bring segments onto the screen purely in order to visualise their content, we see a few identifiable words in each language, surrounded by lots of code which for most users is incomprehensible. The source and target documents have been substantially altered. In fact, there is no source or target document: what exists are numerous, relatively small and separate entries created by the translator using the translator's own judgement, effort and expertise and arranged or tagged in such a way that they can be easily retrieved by TM software.

Since these are the basic criteria for a medium to be regarded as a data base, it seems reasonable to advocate, once again, that TMs should be governed by data-base legislation. Accordingly, they are owned by whoever made them, i.e. the translators. This is applicable even when a TM is generated from existing aligned texts, since only a specialised translator is capable of examining each aligned segment and deciding if there is an acceptable semantic correspondence between source and target.

Conclusions

Looking at all the above claims and counterclaims we can see that there are strong legal arguments to support the view that translators own the intellectual property rights to the translation memories they generate. The fact that certain reputable specialists have taken this line must be regarded as significant and tips the balance in the translator's favour. With regard to termbanks, there is little doubt that the person creating the database holds the copyright to it since this is stipulated by law, subject to the limitations mentioned above with regard to content and bearing in mind the caution that should be exercised when including illustrative quotes. If translation memories are also deemed to be databases then this solid protection can be extended to include them as well and the translator's claim to TM ownership is further strengthened. The existence of "sui generis" database rights in the EU is a key issue for these purposes, in Europe at least.

In this respect, we should recall that under the employment legislation of many countries, workers automatically assign intellectual property rights in the works they create to their employers, so in the case of salaried employees who create termbanks or TMs, these rights will automatically pass to the organisations for which they work.

It should also be mentioned that, although the legal and practical issues are complex, LSPs and their clients have the opportunity to avoid most potential pitfalls by entering into a written agreement before the translation work commences. Within this context, Byrne (2007) suggests that purchasers should seek to conclude a prior agreement in which the translation work is defined as "work made for hire". Sections 101 and 201 of the U.S. Copyright Act define nine categories of service which can be treated as "work made for hire", and one of these is translation, provided that such work (in the case of a freelance) is specially ordered or commissioned. If these conditions are met, then: "The employer or other person for whom the

work was prepared is considered the author for the purposes of this title, and, unless the parties have expressly agreed otherwise in a written instrument signed by them, owns all of the rights comprised in the copyright” (Section 201.b U.S. Copyright Act 1976). In the U.S., therefore, this approach seems to provide solid protection for the commissioning party; however, it might not be desirable for the translator, who effectively signs away all his or her rights in this respect. The “work made for hire” doctrine does not exist in EU legislation or Member State laws derived from it. In any event, whether or not the translation is treated as “work for hire”, the legal specialists cited above agree that if a prior written agreement is concluded which specifies who owns any translation memory generated in the translation process, it is highly unlikely that subsequent legal problems over ownership will arise.

Sources:

Byrne, T. (2007). Copyright Protection for Translation Memories. Web seminar presentation by SDL Trados; November 2007/May 2008.

Davison, Mark (2007). The Legal Protection of Databases. Cambridge University Press.

Johansson, M. (2007). Copyright aspects of feeding terminology databases. FIT-Europe Technical Seminar of Copyright, Intellectual Property and Translation Tools, Barcelona 13 October 2007.

Marcos, J. (2001). Un enfocament jurídic de les memòries de traducció. Revista Tradumàtica October 2001. Autonomous University of Barcelona.

Saint-Robert, M.J. (2007). La protection des bases de données au regard du droit. FIT-Europe Technical Seminar of Copyright, Intellectual Property and Translation Tools, Barcelona 13 October 2007.

Smith, R. (2007). Intellectual Property and CAT: the PwC Experience. FIT-Europe Technical Seminar of Copyright, Intellectual Property and Translation Tools, Barcelona 13 October 2007.

World Intellectual Property Organisation. (2004). WIPO Intellectual Property Handbook: Policy, Law and Use. Geneva.

Zetsche, J. et al. (2005). TM Marketplace White Paper on sharing translation memory data aligned from third-party documents: legal considerations (www.tmmarketplace.com)